

**République Algérienne Démocratique et Populaire**  
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique  
Université Djilali Bounaama Khemis Miliana



**Faculté des Sciences et de la Technologie**  
**Département de Technologie**

Mémoire du Projet de Fin d'Etudes  
Pour l'obtention du diplôme de

**Master**

En

« Télécommunications »

*Option :*

« Systèmes de Télécommunications »

**Titre :**

**Séparation aveugle des mélanges des signaux de parole par la méthode  
ACI**

**Réalisé par :**

- MAZOUNI Mohammed
- CHEKIRINE Abdelmalek

**Encadré par :**

Mme BOUNIF Aouda

Année Universitaire : 2018/2019

# *Dédicaces*

*Je dédie ce mémoire à :*

*Mes chers parents, qui m'ont soutenu, encouragé pour que je  
puisse mener à bien mes études, et qui attendu ce jour avec  
impatience.*

*A mes frères.*

*A mes amies de l'étude et mes camarades.*

*A tous ceux qui ont contribué à la réalisation de ce travail.*

*Mohammed*

# *Dédicaces*

*Je dédie ce mémoire*

*A mes chers parents ma mère et mon père*

*Pour leur patience, leur soutien et leur*

*Encouragements*

*A mes frères.*

*A mes amies et mes camarades.*

*Tous ceux qui sont présents dans mon cœur et absents entre ces  
lignes.*

*Sans oublier toute les professeurs que ce soit du Primaire, du  
moyen, du secondaire ou de L'enseignement supérieur.*

*Abdelmalek*

## Remerciements

*Nous remercions tout d'abord le grand Dieu pour l'achèvement de ce mémoire.*

*Nous exprimons nos gratitude à Monsieur le président de jury d'avoir accepté examiné ce mémoire.*

*Nous remercions Messieurs les membres de jury, d'avoir accepté de prendre part à ce jury ainsi que pour l'intérêt qu'ils l'ont portés à ce travail.*

*Nous remercions Mme BOUNIF notre encadreur, pour ses conseils et suggestions avisés qui nous aidés à mener à bien ce travail, et d'avoir rapporté à ce mémoire ces remarques et conseils.*

## Résumé

Au cours des dernières décennies, beaucoup d'attention a été accordée à la séparation aveugle de sources. Dans plusieurs situations, il est souhaitable de récupérer toutes les sources à partir des mélanges enregistrés, ou au moins de séparer une source particulière. Ce mémoire se focalise sur l'étude de la séparation aveugle de sources dans le cas des mélanges instantanés des sources de signaux de parole.

Nous avons mis en œuvre une technique de séparation appelée ACI, appliquée à des mélanges de sources de parole.

Pour valider nos résultats, nous avons évalué la qualité des sources estimées après la séparation des critères objectifs et subjectifs tels que : SDR, SAR, SIR et MOS. Finalement, nous avons conclu que la méthode de séparation aveugle par ACI donne une bonne estimation des mélanges instantanés.

**Mots clés :** la séparation aveugle de sources, mélange instantané, ACI, SDR, SIR, SAR, MOS

## Abstract

In recent decades, much attention has been paid to the blind separation of sources. In several situations, it is desirable to recover all sources from the recorded mixtures, or at least to separate a particular source. This memory focuses on the study of the blind source separation in the case of instantaneous mixtures of speech signal sources.

We implemented a separation technique called ICA, applied to mixtures of speech sources. To validate our results, we tested the quality of the estimated sources after the separation by objective and subjective criteria such as SDR, SAR, SIR and MOS. Finally, we concluded that the blind separation method by ICA gives a good estimate of instantaneous mixtures.

**Keywords:** blind source separation, instant mixing, ICA, SDR, SIR, SAR, MOS

## Liste des abréviations

**ACI** : Analyse en Composantes Indépendantes

**SAS** : Séparation Aveugle de Sources

**IAM** : Identification Aveugle de Mélanges

**BMI** : Blind Mixture Identification

**PCA** : Analyse en Composantes Principales

**SNR** : Signal to Noise Ratio (Rapport signal sur bruit)

**MOS** : Mean Opinion Score.

**SAR** : Rapport Signal sur Artefacts

**SDR** : Rapport Signal sur Distorsion

**SIR** : Rapport Signal sur Interférence

## Liste des figures

<b>Fig.1.1</b> : Vue schématique de l'appareil phonatoire .....	4
<b>Fig.1.2</b> : Audiogramme de la lettre 'A' (son voisé) .....	5
<b>Fig.1.3</b> : Audiogramme de lettre 'B' (son non voisé) .....	6
<b>Fig.1.4</b> : Schéma de production de parole.....	6
<b>Fig.1.5</b> : Etapes de la transformation d'un son en signal électrique.....	7
<b>Fig. I.6</b> : Représentation de la fréquence fondamentale.....	8
<b>Fig. I.7</b> : Evolution du pitch dans la phrase "les techniques de traitement numérique de la ..... parole.....	9
<b>Fig. 1.8</b> : Exemples de son voisé (haut) et non voisé (bas).....	11
<b>Fig. 1.9</b> : Audiogramme de signaux de parole.....	11
<b>Fig.1.10</b> : Architecture d'un système automatique de reconnaissance de la parole.....	12
<b>Fig.1.11</b> : Représentation temporelle d'un signal parole.....	14
<b>Fig.1.12</b> : Représentation fréquentielle (module) d'un signal de parole.....	15
<b>Fig.1.13</b> : Représentation d'un mélange des signaux de parole.....	18
<b>Fig.2.1</b> : Configuration générale de la séparation aveugle de sources.....	21
<b>Fig.2.2</b> : Diagramme de Vann de l'information mutuelle et l'entropie conditionnelle pour une classe X et une variable Y.....	27
<b>Fig.2.3</b> : Illustration d'un processus de séparation de deux sources à partir de deux observations.....	32
<b>Fig. 3.1</b> : Distribution de probabilité de la source $s_1(t)$ .....	37
<b>Fig. 3.2</b> : Distribution de probabilité de la source $s_2(t)$ .....	38
<b>Fig. 3.3</b> : Distribution de probabilité de la source $s_3(t)$ .....	38
<b>Fig. 3.4</b> : L'audiogramme des signaux sources $s_2(t)$ et $s_3(t)$ . .....	39
<b>Fig. 3.5</b> : L'audiogramme des mélanges X1 et X2. ....	40
<b>Fig. 3.6</b> : L'audiogramme des trois signaux sources $s_1(t)$ , $s_2(t)$ et $s_3(t)$ .....	40
<b>Fig. 3.7</b> : L'audiogramme des trois mélanges X1, X2 et X3.....	41
<b>Fig. 3.8</b> : L'audiogramme des deux sources indépendantes estimées.....	42
<b>Fig. 3.9</b> : Evolution de la valeur de l'entropie.....	45
<b>Fig. 3.10</b> : Variation du gradient de la valeur de l'entropie.....	46
<b>Fig. 3.11</b> : L'audiogramme des trois sources indépendantes estimées $\tilde{s}_1(t)$ , $\tilde{s}_2(t)$ et $\tilde{s}_3(t)$ ....	46
<b>Fig. 3.12</b> : Variation de l'entropie avec $\eta=0.9$ , N-itération=500) .....	48

<b>Fig. 3.13</b> : Variation du gradient de l'entropie avec $\eta=0.9$ , N-itération=500) .....	49
<b>Fig. 3.14</b> : Variation de l'entropie avec $\eta=0.1$ , N-itération=500) .....	49
<b>Fig. 3.15</b> : Variation du gradient de l'entropie avec $\eta=0.1$ , N-itération=500) .....	50
<b>Fig. 3.16</b> : Variation de l'entropie avec $\eta=0.25$ , N-itération=100.....	50
<b>Fig. 3.17</b> : Variation du gradient de l'entropie avec $\eta=0.25$ , N-itération=100.....	51



## Liste des tableaux

<b>Tableau 2.1</b> : l'échelle MOS.....	33
<b>Tableau.3.1</b> : Résultat du test MOS pour le premier cas.....	42
<b>Tableau.3.2</b> : L'évaluation de la qualité de la source estimée $\tilde{s}_2(t)$ en fonction du pas du gradient. ....	43
<b>Tableau.3.3</b> : L'évaluation de la qualité de la source estimée $\tilde{s}_3(t)$ en fonction du pas du gradient. ....	43
<b>Tableau.3.4</b> : L'évaluation de la qualité de la source estimée $\tilde{s}_2(t)$ en fonction de nombre d'itérations.....	44
<b>Tableau.3.5</b> : L'évaluation de la qualité de la source estimée $\tilde{s}_3(t)$ en fonction de nombre d'itérations.....	44
<b>Tableau. 3.6</b> : Corrélation entre signaux estimés et originaux.....	44
<b>Tableau.3.7</b> : Résultat du test du MOS pour le deuxième mélange.....	47
<b>Tableau.3.8</b> : L'évaluation de la qualité d'estimation des sources $s_1(t)$ , $s_2(t)$ et $s_3(t)$ du deuxième mélange avec <code>nbre_iter = 25</code> .....	47
<b>Tableau.3.9</b> : L'évaluation de la qualité d'estimation des sources $s_1(t)$ , $s_2(t)$ et $s_3(t)$ du deuxième mélange <code>nbre_iter = 100</code> .....	47

# Table de Matière

**Remerciement**

**Résumé**

**Introduction générale.....1**

## **Chapitre 1 : Signal de parole**

1.1. Introduction.....3

1.2. Production de la parole.....3

1.2.1. Appareil phonatoire humain.....3

1.2.2. Mécanisme de production de la parole.....4

1.2.3. Modélisation de l'appareil phonatoire.....6

1.3. Représentation d'un signal de parole.....7

1.4. Caractéristiques du signal de parole.....8

1.4.1. Fréquence fondamentale.....8

1.4.2. L'énergie.....9

1.4.3. Le spectre.....9

1.5. Audiogramme.....10

1.6. Utilisation des présentations temporelle et fréquentielle en traitement de la parole.....12

1.6.1. Reconnaissance de la parole photo.....12

1.6.2. Analyse acoustique.....12

1.6.3. Codage.....13

1.6.4. Traitement et modification du signal.....13

1.6.5. Synthèse de la parole.....13

1.7. Classification des signaux .....13

1.7.1. Représentation temporelle de signal de parole..... 14

1.7.2. La représentation fréquentielle d'un signal de parole.....15

1.8. La variabilité du signal de parole.....15

1.9. Stationnarité.....16

1.10. Cotexte de mélange des signaux de parole.....17

1.11. Conclusion.....18

## **Chapitre 2 : Séparation aveugle de sources par l'Analyse en Composantes**

### **Indépendantes**

2.1. Introduction.....19

2.2. Technique de séparation à l'aveugle .....	19
2.3. Mélanges et séparation aveugle des sources .....	21
2.4. Analyse en composantes indépendantes (ACI) .....	22
2.4.1. Définition .....	22
2.4.2. Hypothèses fondamentales de l'ACI .....	23
2.4.3. Formulation mathématique de l'ACI.....	23
2.4.4. Mesure de l'indépendance.....	25
2.4.5. Différentes approches de l'ACI.....	25
2.4.5.1. L'information mutuelle .....	25
2.4.5.2. Entropie d'une suite de variables .....	26
2.4.5.3. La néguentropie .....	28
2.4.5.4. Le kurtosis .....	29
2.5 Application d'un algorithme ACI.....	30
2.5.1. Centrage des données.....	30
2.5.2. Détermination du nombre de sources.....	30
2.5.3 Réduction de dimensions.....	31
2.5.4. Blanchiment.....	31
2.6 Application de l'ACI à des mélanges des signaux de parole .....	31
2.6.1. Mélanges instantanés.....	31
2.6.2. Mélanges convolutifs.....	32
2.7. Les critères pour évaluation la qualité de la parole.....	33
2.7.1. Critères subjectifs.....	33
2.7.2. Critères Objectifs.....	33
2.8. Conclusion.....	35

### **Chapitre 3 : Résultats de simulation**

3 .1. Introduction.....	36
3.2. Environnement du travail.....	36
3.2.1. Machine du travail.....	36
3.2.2. Logiciel Matlab.....	37
3.3. Base de données.....	37
3.4. Résultats et discussion .....	39
3.4.1 Le mélange des signaux.....	39

3.4.2 Séparation des signaux mélangés.....	41
3.4.2.1 Séparation du premier mélange.....	41
3.4.2.2 Séparation du deuxième mélange.....	46
3.4.3. Discussion de la qualité d'estimation.....	48
3.4.3.1. Le pas d'adaptation.....	48
3.4.3.2. Le nombre d'itération.....	50
3.5. Conclusion.....	51

**Conclusion générale .....**

**Bibliographie**

# **Introduction générale**

## **INTRODUCTION GENERALE**

Le traitement de la parole est devenu très important pour le développement des outils de communications modernes. Il se situe entre le traitement du langage (c'est-à-dire du traitement de données symboliques) et le traitement du signal numérique. Cet axe de recherche a connu, depuis les années soixante, une évolution remarquable liée au développement des moyens et des techniques de télécommunications.

L'être humain est capable, grâce à son système auditif, de localiser une des sources d'un mélange de signaux provenant de son environnement.

Plusieurs travaux ont été menés dans ce domaine avec la terminologie de séparation aveugle des sources.

Depuis plusieurs années, la séparation de sources est un sujet de recherche particulièrement actif qui s'explique par les nombreuses applications envisageables. On peut notamment citer : les télécommunications, l'acoustique, la sismologie, le domaine biomédical, séparation d'extraits musicaux stéréophoniques, le rehaussement de la parole pour la téléphonie mobile, la localisation et la poursuite de cibles en radar et sonar, la séparation de locuteurs (problème dit de "cocktail party"), la détection et la séparation dans les systèmes de communication à accès multiple, etc.

Ce mémoire est organisé comme suit :

Le premier chapitre intitulé « Signal de parole » comporte des notions fondamentales sur le signal de parole, sa production et son modèle général. Nous donnerons aussi un aperçu sur le spectre du signal de parole, ses applications ainsi que le contexte du mélange des signaux de parole.

La séparation aveugle de sources par l'analyse en composantes indépendantes sera abordée dans le chapitre deux. La première section de ce dernier comportera les techniques utilisées dans la séparation des mélanges illustrée par des schémas, et aussi les approches utilisées en tant qu'ACI.

La troisième section décrira les notions des mélanges rencontrés dans le monde réel de l'audio et les problèmes rencontrés lors de leur séparation.

La section suivante illustrera la technique ACI à travers sa définition, ses hypothèses fondamentales, sa formulation mathématique. Nous verrons aussi les critères permettant la mesure de l'indépendance entre les signaux ainsi que les prétraitements nécessaires pour un algorithme ACI.

Le chapitre 3 présentera les résultats expérimentaux de la séparation d'un mélange de signaux de parole avec deux ou plusieurs sources par l'ACI. Cette partie sera suivie par une évaluation de la qualité de séparation à l'aide des critères tels que : SDR, SAR, SIR et le MOS.

Enfin, nous terminerons notre travail par une conclusion générale qui résume les points abordés.

# **Chapitre 1**

## **Signal de parole**



## 1.1. Introduction

La parole est sans doute le moyen de communication le plus simple et le plus efficace chez les humains. Depuis le début de la recherche dans le domaine du traitement du signal, les chercheurs ont toujours eu une attention particulière pour le signal de parole.

Grâce aux développements dans les domaines de l'informatique, de la linguistique, de la télécommunications, ..., le rêve de communiquer avec des machines est devenu de plus en plus réalisable.

Le développement des outils mathématiques et les moyens techniques, électroniques et informatiques, a permis de classer le traitement de la parole comme une composante fondamentale des sciences de l'ingénieur et une discipline à part pour les traités de signaux.

Ainsi, nous présentons dans ce chapitre des notions fondamentales sur le signal de parole, sa production et son modèle générale. Nous donnons aussi un aperçu sur le spectre du signal de parole, ses applications ainsi que sa représentation temporelle et même que fréquentielle ensuite nous avons expliqué sur les caractéristiques de signal de parole à l'aide par des graphes et nous donnons aussi la reconnaissance de la parole, à la fin on conclut le chapitre.

## 1.2. Production de la parole

Il n'y a pas un organe spécifique pour produire la parole, donc le système respiratoire contribue en particulier avec une partie du système digestif.

### 1.2.1. Appareil phonatoire humain

Le système de production des sons de l'être humain est complexe. La figure 1.1 montre les différents organes de l'appareil phonatoire humain. Les sons de la parole peuvent être classés en trois catégories : les sons voisés, les sons non voisés et les plosives.

## Les sons voisés

Les cordes vocales vibrent de façon quasi-périodique. Le signal de la parole est alors quasi-périodique et est caractérisé par sa fréquence fondamentale, typiquement, la période fondamentale des différents sons voisés varie entre 2 et 20 ms.

## Les sons non voisés

Les cordes vocales ne vibrent pas, l'air passe à haute vitesse entre les cordes vocales, le signal produit est équivalent à un bruit blanc.

## Les plosives

Ces sons sont obtenus lorsqu'on libère soudainement l'air comprimé par fermeture totale du conduit vocal [2].

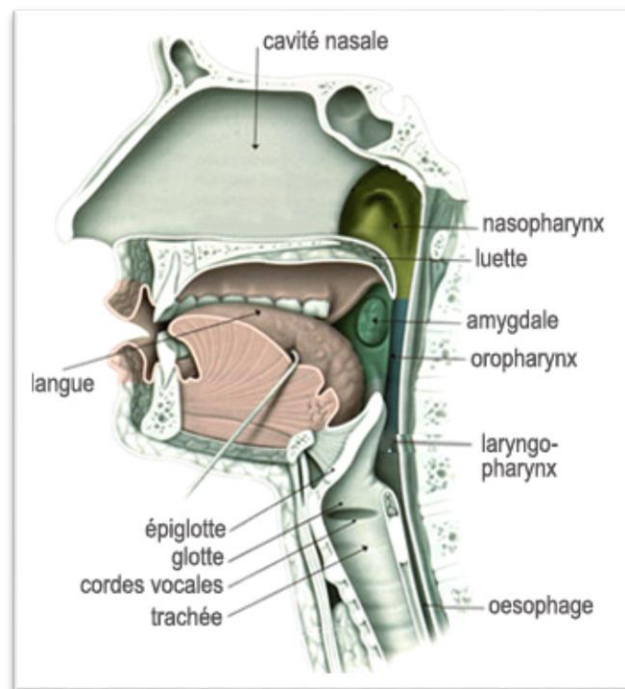


Fig.1.1 : Vue schématique de l'appareil phonatoire [3].

### 1.2.2. Mécanisme de production de la parole

- **Formation du flux d'air** : En particulier les poumons, mis en action par le diaphragme qui se contracte ce qui chasse l'air des poumons et c'est cet air dont les variations de pression au niveau du pharynx vont créer la parole.

- **Phonation** : consiste à créer du son au niveau des cordes vocales.

L'air expulsé des poumons arrive au niveau des cordes vocales au repos (fermées : phase d'apnée). La pression élevée exercée par l'air peut faire vibrer les cordes vocales ce qui donne naissance à un signal périodique. Dans le cas contraire, l'air passe sous forme d'un souffle qui sera considéré comme un bruit.

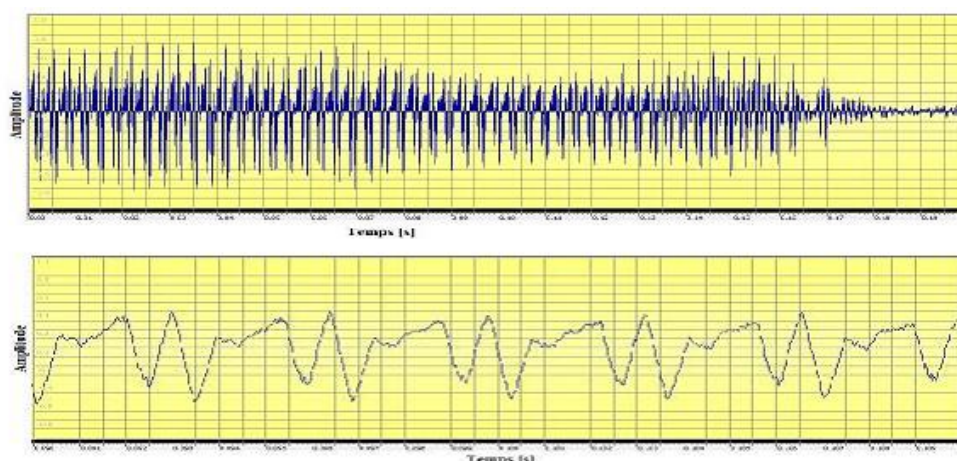
Donc, la parole est produite lorsque l'air, poussé hors des poumons, passe par les cordes vocales et le conduit vocal pour produire un son.

L'air circule dans le conduit vocal, ce qui engendre une vibration acoustique semblable au claquement des mains. Ce son est assez complexe, constitué d'une fréquence fondamentale (qui correspond à la fréquence moyenne du mouvement des cordes vocales), et de ses harmoniques.

En fait, l'appareil phonatoire contient deux sources sonores :

- Source quasi-périodique : cordes vocales en vibration
- Source de bruit : écoulements turbulents produits par une obstruction du passage de l'air dans le conduit vocal.

Les modes de production des sons diffèrent grandement, ce qui permet une grande variété de phonèmes. On peut diviser ces phonèmes en deux classes : voisés et non voisés. Les voyelles sont des exemples de sons voisés, alors que les fricatives sont des exemples de sons non voisés. Les figure 1.2 et 1.3 montre dans le domaine temporel une voyelle (/A/) et une voyelle (/B/). On remarque que le signal correspondant à une voyelle est périodique [10].



**Fig.1.2** : Audiogramme de la lettre 'A' (son voisé) [2].

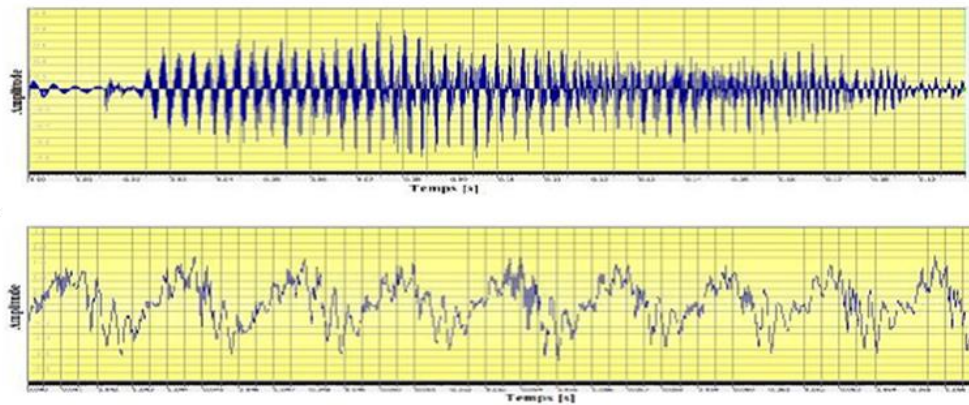


Fig.1.3 : Audiogramme de lettre ‘B’ (son non voisé) [2].

### 1.2.3. Modélisation de l’appareil phonatoire

La parole est une succession des sons dits voisés caractérisés par la vibration des cordes vocales et des sons non voisés (qui ne font pas intervenir les cordes vocales).

Le signal voisé est un signal pseudopériodique présentant des zones fréquentielles plus ou moins importantes. Les maxima de l’enveloppe spectrale sont appelés des formants.

D’après son fonctionnement, l’appareil phonatoire peut être modélisé par un circuit résonateur excité par une source d’énergie comme le montre la figure 1.4 [9].

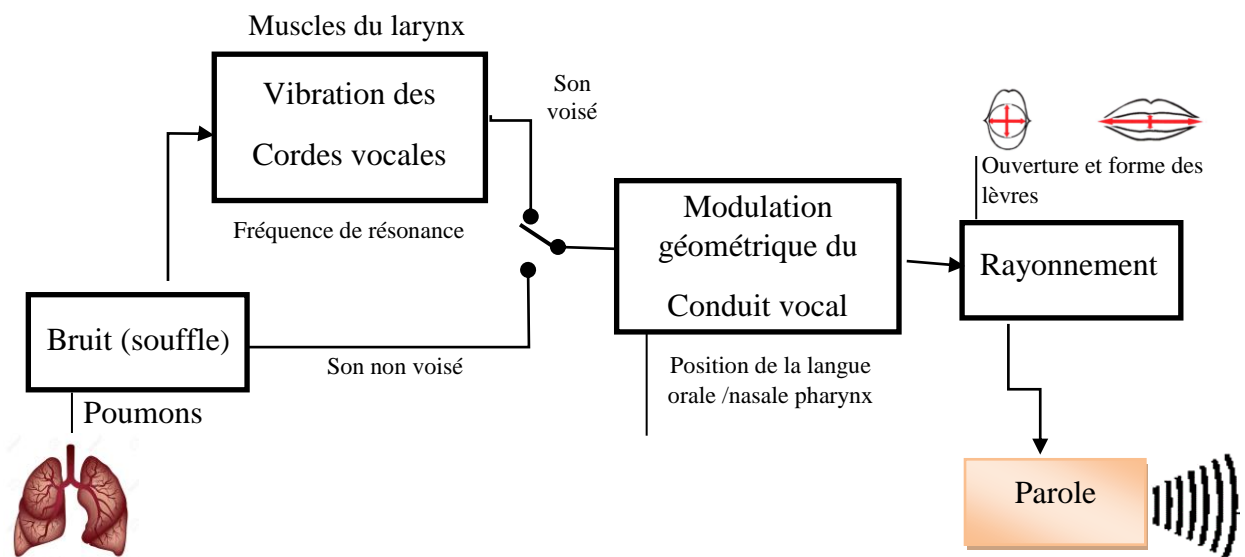


Fig.1.4 : Schéma de production de parole.

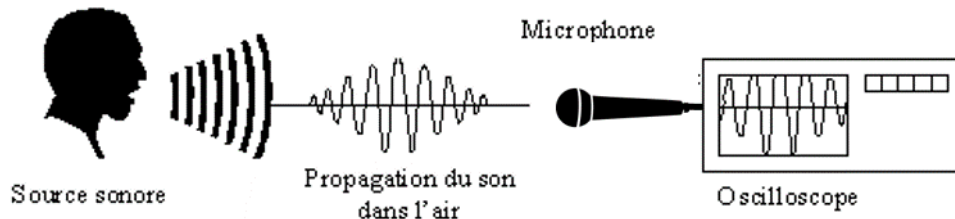
Le spectre du son émis par les cordes vocales est modulé par les propriétés résonantes du corps résonateur et par la position des lèvres. Les différentes modifications possibles dans le corps résonateur pour le traitement du son permettent d'obtenir un panel de sonorités extrêmement important.

### 1.3. Représentation d'un signal de parole

Un microphone (parfait) placé dans un champ acoustique délivre une tension électrique  $v(t)$  ou un courant électrique  $i(t)$  dont l'évolution temporelle suit fidèlement celle de la pression acoustique à cet endroit. Par définition de signal audio est la fonction  $v(t)$  ou  $i(t)$ .

Le signal audio peut être mesuré, enregistré, numérisé (ou digitalisé), traité, transformé, transmis (via un canal hertzien ou un câble) et finalement reproduit pour l'audition par un haut-parleur.

Le signal audio peut être multicanal, lorsqu'il résulte de l'acquisition simultanée de la pression par plusieurs microphones. Un signal audio stéréophonique est un exemple de signal multicanal (2 canaux : gauche et droit) [16].



**Fig.1.5** : Etapes de la transformation d'un son en signal électrique

Le signal audio peut être créé par d'autres moyens par une acquisition microphonique (synthétisé par des circuits électroniques ou par un ordinateur : exemple des signaux musicaux produits par des synthétiseurs). Dans ce cas, tout signal synthétisé comportant des composantes fréquentielles dans la bande 20 Hz – 20 kHz peut être qualifié de signal audio.

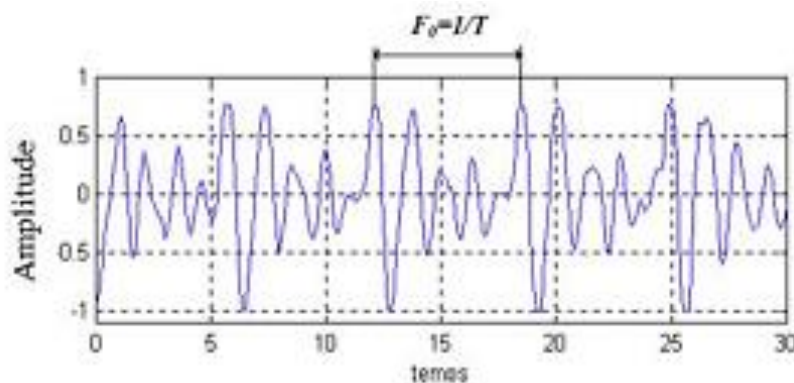
## 1.4. Caractéristiques du signal de parole

Le signal de parole est un vecteur acoustique porteur d'informations d'une grande complexité, variabilité et redondance. Les caractéristiques de ce signal sont appelées traits acoustiques. Chaque trait acoustique a une signification sur le plan perceptuel. Ces traits acoustiques sont : la fréquence fondamentale, le spectre fréquentiel et l'énergie.

### 1.4.1. Fréquence fondamentale

La fréquence fondamentale est par définition l'inverse de la période de vibration des cordes vocales.

Ses variations définissent le pitch qui constitue la perception de la hauteur, ou les sons s'ordonnent de grave à aigu. Seuls les sons voisés engendrent une sensation de hauteur tonale bien définie.



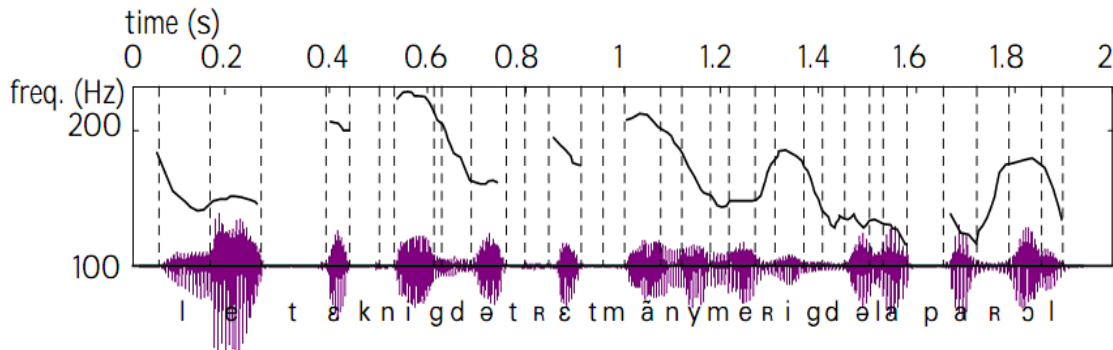
**Fig.I.6** : Représentation de la fréquence fondamentale [2].

L'extraction de la fréquence fondamentale n'est pas une tâche facile pour les trois raisons suivantes :

- La périodicité de vibration des cordes vocales n'est pas nécessairement parfaite.
- Il est difficile de séparer la fréquence fondamentale des effets du trait vocal.
- La plage de dynamique de la fréquence fondamentale est très grande Une analyse d'un signal de parole n'est pas complète tant qu'on n'a pas mesuré l'évolution temporelle de la fréquence fondamentale.

La figure I.7 donne l'évolution temporelle de la fréquence fondamentale de la phrase " les techniques de traitement de la parole ". On constate qu'à l'intérieur des zones voisées la fréquence fondamentale évolue lentement dans le temps. Elle s'étend approximativement de

70 à 250 Hz chez les hommes, de 150 à 400 Hz chez les femmes, et de 200 à 600 Hz chez les enfants.



**Fig.I.7** : Evolution du pitch dans la phrase "les techniques de traitement numérique de la parole" [16].

Dans la figure précédente la fréquence est donnée sur une échelle linéaire et les sons non-voisés sont associés à une fréquence nulle [2].

### 1.4.2. L'énergie

Elle est représentée par l'intensité du son qui est liée à la pression de l'air en amont du larynx. L'amplitude du signal de la parole varie au cours du temps selon le type de son, et son énergie dans une trame est donnée par :

$$E = \sum_{n=0}^{N-1} s^2(n) \quad (\text{Eq.1.1})$$

Avec : N la taille de la trame.

### 1.4.3. Le spectre

L'enveloppe spectrale ou spectre représente l'intensité de la voix en fonction de la fréquence, elle est généralement obtenue par une analyse de Fourier à court terme. La quasi-stationnarité du signal de parole permet de mettre en œuvre des méthodes efficaces d'analyse et de modélisation utilisées pour le traitement à court terme du signal vocal sur des fenêtres de durée généralement comprise entre 20ms et 30ms appelées trames, avec un recouvrement entre ces fenêtres qui assure la continuité temporelle des caractéristiques de l'analyse [3].

## 1.5. Audiogramme

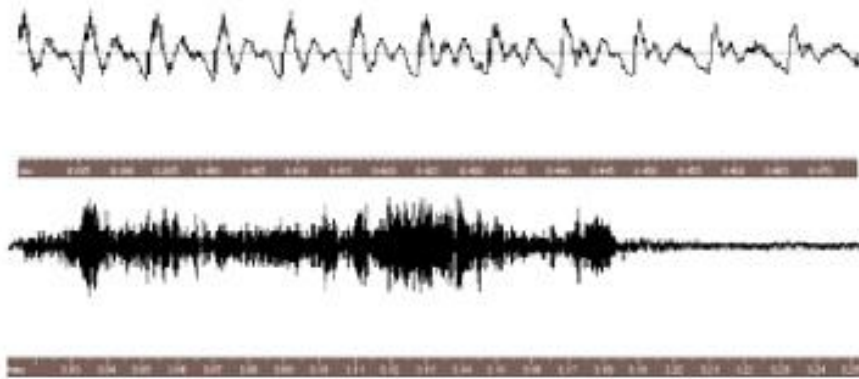
L'échantillonnage transforme le signal à temps continu  $x(t)$  en signal à temps discret  $x(k T_e)$  défini aux instants d'échantillonnage, multiples entiers de la période d'échantillonnage  $T_e$  ; celle-ci est elle-même l'inverse de la fréquence d'échantillonnage  $F_e$ . En ce qui concerne le signal vocal, le choix de  $F_e$  résulte d'un compromis. Son spectre peut s'étendre jusque 12 kHz. Il faut donc en principe choisir une fréquence  $F_e$  égale à 24 kHz au moins pour satisfaire raisonnablement au théorème de Shannon. Cependant, le coût d'un traitement numérique, filtrage, transmission, ou simplement enregistrement peut être réduit d'une façon notable si l'on accepte une limitation du spectre par un filtrage préalable. C'est le rôle du filtre de garde, dont la fréquence de coupure  $F_c$  est choisie en fonction de la fréquence d'échantillonnage retenue. Pour la téléphonie, on estime que le signal garde une qualité suffisante lorsque son spectre est limité à 3400 Hz et l'on choisit  $F_e = 8000$  Hz.

Pour les techniques d'analyse, de synthèse ou de reconnaissance de la parole, la fréquence peut varier de 6000 à 16000 Hz.

Par contre pour le signal audio (parole et musique), on exige une bonne représentation du signal jusque 20 kHz et l'on utilise des fréquences d'échantillonnage de 44.1 ou 48 kHz. Pour les applications multimédia, les fréquences sous-multiples de 44.1 kHz sont de plus en plus utilisées : 22.5 kHz, 11.25 kHz.

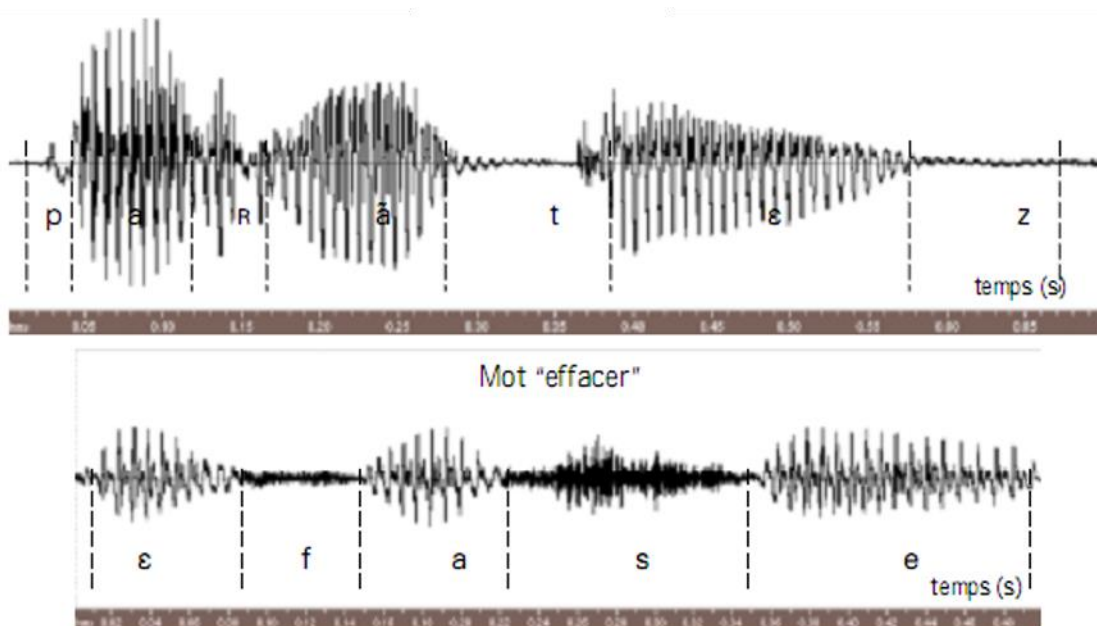
Parmi l'ensemble des valeurs possibles pour les échantillons  $x(k T_e)$  la quantification ne retient qu'un nombre fini  $2^b$  de valeurs ( $b$  étant le nombre de bits de la quantification), espacées du pas de quantification  $q$ . Le signal numérique résultant est noté  $x(k)$ . La quantification produit une erreur de quantification qui normalement se comporte comme un bruit blanc. Le pas de quantification est donc imposé par le rapport signal à bruit à garantir. Si le pas de quantification est constant, ce rapport est fonction de l'amplitude du signal ; les signaux de faible amplitude sont dès lors mal représentés. Aussi adopte-t-on pour la transmission téléphonique une loi de quantification logarithmique et chaque échantillon est représenté sur 8 bits (256 valeurs). Par contre, la quantification du signal musical exige en principe une quantification linéaire sur 16 bits (65536 valeurs).





**Fig.1.8** : Exemples de son voisé (haut) et non voisé (bas).

La figure I.8 donne une représentation plus fine de tranches de signaux voisés et non voisés. L'évolution temporelle ne fournit cependant pas directement les traits acoustiques du signal. Il est nécessaire, pour les obtenir, de mener à bien un ensemble de calculs ad-hoc.



**Fig.1.9** : Audiogramme de signaux de parole.

La figure I.9 représente l'évolution temporelle, ou audiogramme, du signal vocal pour les mots 'parenthèse' et 'effacer'. On y constate une alternance de zones assez périodiques et de zones bruitées, soient zones voisées et non voisées [16].

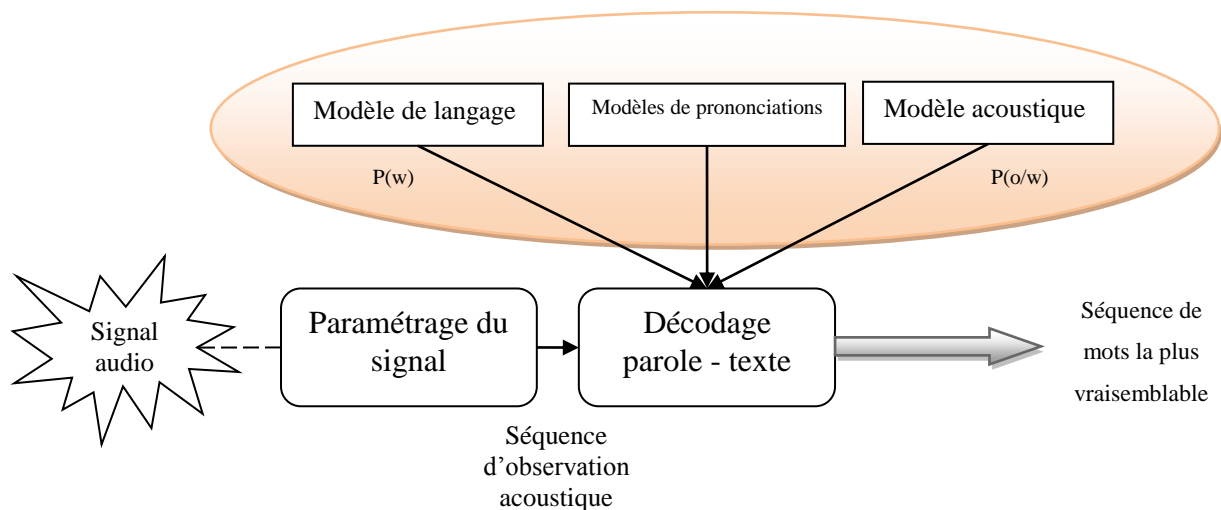
## 1.6. Utilisation des présentations temporelle et fréquentielle en traitement de la parole

Les représentations temps fréquence du signal de parole est utile dans plusieurs domaines d'utilisation tels que :

### 1.6.1. Reconnaissance de la parole

Une représentation temps-fréquence du signal intervient pour paramétrer le signal acoustique. Dans la reconnaissance par des systèmes experts en lecture de spectrogrammes, la quantité de données utilisées pour la para métrisation acoustique est faible.

Les vecteurs (qui représentent une trame) d'analyse sont constitués généralement de 8 à 20 coefficients, comportant des informations spectrales et d'autres informations (énergie totale de la trame par exemple). Environ une centaine de vecteurs par seconde semble nécessaire et suffisant [5].



**Fig.1.10** : Architecture d'un système automatique de reconnaissance de la parole [15].

### 1.6.2. Analyse acoustique

Le problème est ici de mesurer des paramètres ou des indices acoustiques. La qualité de la mesure (résolution, biais, etc..) et la lisibilité des paramètres ou des indices acoustiques sur

les coefficients ou les paramètres d'analyse sont fondamentaux, sans que le débit ou le temps de calcul soient obligatoirement des données critiques. L'analyse par synthèse est largement répandue, ce qui implique des représentations capables de reconstituer le signal.

### **1.6.3. Codage**

Les méthodes temps-fréquence peuvent contribuer à réduire le débit du signal. Pour évaluer ces méthodes, il faut rapporter la qualité de la parole codée au coût du codage, en termes de débit, de puissance de calcul, de résistance au bruit et aux erreurs de nombreuses revues partielles sur le codage existent.

### **1.6.4. Traitement et modification du signal**

Certaines applications impliquent de traiter ou de modifier le signal acoustique capté. Citons par exemple : la modification de la fréquence de voisement, des formants, des durées segmentales, la conversion de la voix d'un locuteur en la voix d'un autre locuteur, le rehaussement de parole noyée dans du bruit, la séparation de plusieurs sources, la compensation des particularités de la prise de son, etc.

Pour modifier un paramètre acoustique (fréquence fondamentale par exemple), il faut utiliser des méthodes qui en font un paramètre explicite. De même, des contraintes perceptives sont souvent incluses dans ce type de méthode (comme la courbe de sélectivité fréquentielle de l'oreille, par exemple).

### **1.6.5. Synthèse de la parole**

Afin de générer automatiquement à partir d'un texte un signal de parole, des méthodes de codage, de modification et de description acoustique de la parole sont nécessaires. Ce problème emprunte donc en partie aux trois points précédents.

Comme pour la reconnaissance, la représentation du signal n'est qu'un des éléments d'un système complet de synthèse, qui contient aussi des composantes phonétiques, phonologiques, lexicales, syntaxiques, voire sémantiques et pragmatiques [5].

## **1.7. Classification des signaux**

Pour faciliter l'étude des signaux, différents modes de classification peuvent être envisagés :

- Représentation temporelle des signaux.
- Représentation spectrale.
- Caractéristique morphologique (signal continu ou discret).

### 1.7.1. Représentation temporelle de signal de parole

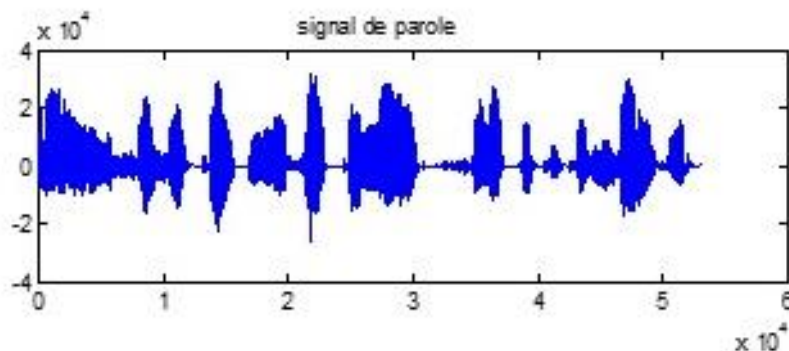
La première classification, basée sur l'évolution du signal en fonction du temps, fait apparaître deux types fondamentaux :

- Les signaux certains (ou déterministes) dont l'évolution en fonction du temps peut être parfaitement décrite par un modèle mathématique. Ces signaux proviennent de phénomènes pour lesquels on connaît les lois physiques correspondantes et les conditions initiales, permettant ainsi de prévoir le résultat.
- Les signaux aléatoires (ou probabilistes) dont le comportement temporel est imprévisible et pour la description desquels il faut se contenter d'observations statistiques.

En ce qui concerne les signaux aléatoires, ils sont dits stationnaires lorsque leur valeur moyenne est indépendante du temps, c'est-à-dire que les résultats de leur analyse statistique restent les mêmes quel que soit le moment où l'on en observe une partie déterminée.

De plus ces signaux aléatoires stationnaires sont ergodiques s'il est identique de faire une moyenne statistique à un instant donné sur différents essais ou de faire une moyenne temporelle suffisamment longue sur un seul de ces essais.

Le signal de la parole fait partie de la deuxième catégorie des deux signaux cités précédemment. La figure 1.11 illustre les variations temporelles d'un tel signal de parole [6].



**Fig.1.11** : Représentation temporelle d'un signal de parole

### 1.7.2. La représentation fréquentielle d'un signal de parole

Un signal parlé ou musical est plus complexe, puisque son allure varie constamment au cours du temps. Il contient des fréquences graves, moyennes et aiguës. Son spectre s'étend de 20 Hz à 20 kHz et varie en permanence entre ces deux fréquences extrêmes.

Le signal vidéo est encore plus complexe et son spectre s'étend du continu à quelques mégahertz. Le spectre d'un signal nous renseigne donc sur les différentes composantes fréquentielles qu'il contient.

Le spectre d'un signal est la représentation en fonction de la fréquence des amplitudes des différentes composantes présentes dans le signal [7].

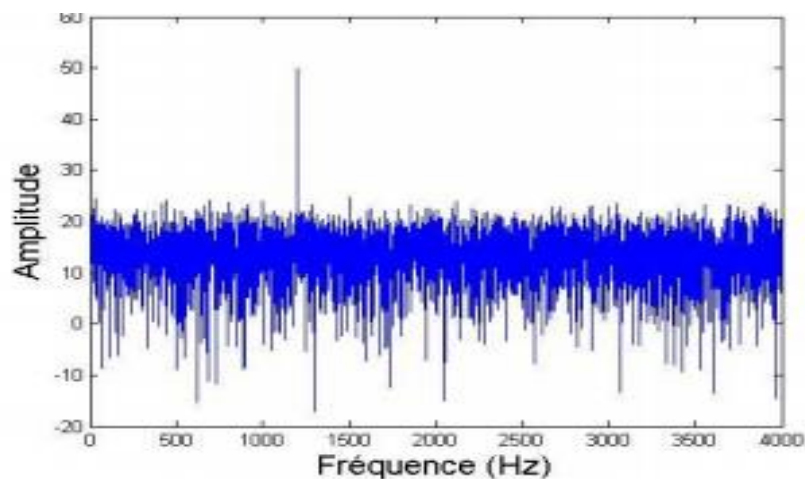


Fig.1.12 : Représentation fréquentielle (module) d'un signal de parole.

### 1.8. La variabilité du signal de la parole

La variabilité d'une personne à une autre (variabilité interlocuteurs) démontre les différences du signal de parole en fonction du locuteur. Cette variabilité, utile pour différencier les locuteurs, et également mélangée à d'autres types de variabilité : variabilité intra-locuteur, variabilité due aux conditions d'enregistrement et de transmission du signal de parole (bruit ambiant, microphone utilisé, lignes de transmission) et variabilité due au contenu linguistique.

Variabilités interlocuteurs proviennent des différences physiologiques (différences dimensionnelles du conduit vocal, fréquences d'oscillations des cordes vocales) et de différences de style de prononciation (p.ex. accent, niveau social). Certaines de ces

différences qui influencent la représentation de chaque locuteur, nous permettent de les séparer.

Variabilités intra-locuteur font que la voix dépend de l'état physique et émotionnel d'un individu. La voix humaine varie avec le temps ou les conditions physiologiques et psychologiques du locuteur. Cependant, ces variations intra-locuteur ne sont pas identiques pour tous les humains. En effet, hormis les variations lentes de la voix dues au vieillissement, certains phénomènes extérieurs tels que l'état de santé d'une personne ont une influence variable sur sa voix.

Une dégradation croissante des performances a été observée au fur et à mesure que le temps qui sépare la session d'enregistrement de la session de test augmente. De plus, le comportement des locuteurs se modifie lorsque ceux-ci s'habituent au système. Les modèles des locuteurs doivent donc être régulièrement mis à jour avec les nouvelles données d'exploitation du système. Les altérations de la voix dues à l'état physique (fatigue, rhume) ou émotionnel (stress), lorsqu'ils sont importants, peuvent mettre aussi en échec l'efficacité de certains systèmes [2].

## 1.9. Stationnarité

La stationnarité est une propriété particulièrement importante pour l'analyse des signaux aléatoires.

Les moments, qui dépendent dans le cas général de l'instant considéré, deviennent des quantités indépendantes du temps dans le cas stationnaire. La fonction de covariance  $Cov_X(t_1, t_2)$  devient quant-à-elle une quantité dépendant uniquement de l'écart entre  $t_1$  et  $t_2$ . Toutes les variables  $X(t_i, \omega)$  possèdent ainsi la même loi à un instant. Par suite :

$$E\{X(t_1)^n\} = E\{X(t_1 - \tau)^n\} = \dots = m_X^{(n)} \quad (\text{Eq.1.2})$$

On en déduit donc que tous les moments sont indépendants du temps.

À deux instants, la distribution conjointe ne dépend que de l'écart entre les deux instants et non des instants eux-mêmes :

$$P_{X(t_1), X(t_2)} = P_{X(t_1 - t_2), X(0)} \quad (\text{Eq.1.3})$$

On en déduit donc que :

$$\begin{aligned} Cov_x(t_1, t_2) &= E\{X(t_1)X(t_2)^*\} = E\{X(t_1 - t_2)X(0)^*\} \\ &= E\{X(0)X(t_2 - t_1)^*\} = E\{X(t)X(t - \tau)^*\} \end{aligned} \quad (\text{Eq.1.4})$$

Avec  $t$  quel conque et  $\tau = t_1 - t_2$ . La covariance devient une fonction de corrélation, qui ne dépend que de l'écart de temps. On pose :

$$R_x(\tau) = E\{X(t)X(t - \tau)^*\} \quad (\text{Eq.1.5})$$

On peut vérifier la stationnarité en calculant tous les moments, à tous les ordres. Ceci n'est pas forcément utilisable, et on se contentera souvent d'étudier une stationnarité au sens faible (par opposition à la stationnarité stricte), en définissant une stationnarité à l'ordre 1 « le moment d'ordre 1 » est indépendant du temps et fonction de covariance invariante par translation dans le temps.

Le signal à traiter est, dans la plupart des cas, mesuré par des capteurs et issu de notre environnement physique réel : ondes sonores, vibrations sismiques, images visuelles, etc. Les microphones par exemple captent les ondes sonores se propageant dans l'air.

En pratique, malheureusement, les signaux obtenus grâce à ces capteurs sont souvent des mélanges de plusieurs contributions issues de signaux originaux appelés sources. Dans le cas le plus général, ces contributions sont des signaux obtenus par filtrage non linéaire des sources physiques. Dans d'autres cas plus simples, elles peuvent être des versions pondérées et/ou décalées temporellement des sources.

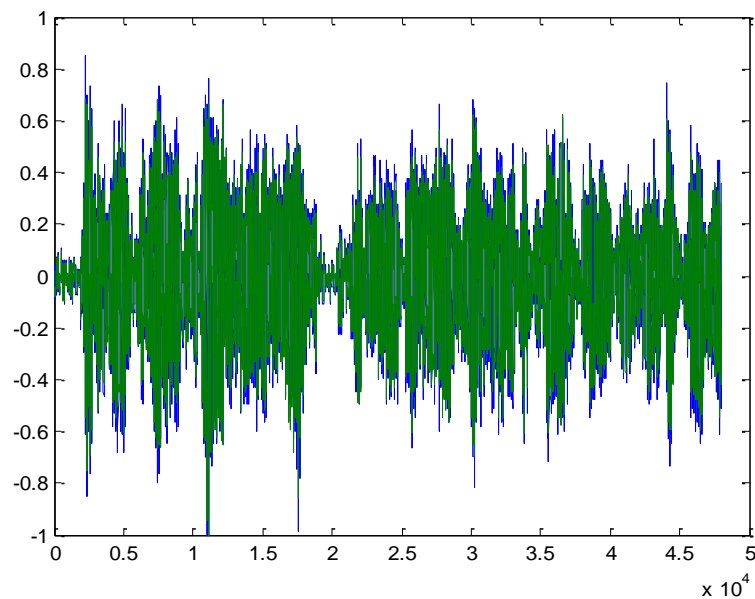
Le prochain chapitre sera consacré à l'étude de la séparation aveugle de sources qui consiste à estimer les signaux sources (ou éventuellement leurs contributions) à partir des mélanges obtenus sur chaque capteur [8].

## 1.10. Contexte de mélange des signaux de parole

La nature du mélange représente les hypothèses que l'on fait sur l'environnement Acoustique dans lequel l'enregistrement est effectué. En particulier, les enregistrements qui ont lieu dans un milieu naturel sont affectés par de multiples échos qui ont pour conséquence

d'introduire des directions multiples pour chacune des sources. Si l'enregistrement est effectué dans une salle anéchoïque, où l'on peut supposer qu'il n'y a plus d'écho, par contre les distances d'une source à chaque microphone étant différentes, les signaux des sources n'arrivent pas au même instant aux différents microphones. On définit trois hypothèses sur le mélange qui correspondent aux trois scénarios précédents :

Mélange instantané, mélange convolutif et mélange convolutif anéchoïque [1].



**Fig.1.13** : Représentation d'un mélange de signaux de parole.

## 1.11. Conclusion

Dans ce chapitre nous avons essayé d'écrire le signal de la parole qui est complexe dans sa description et son traitement. En premier lieu, nous avons discuté le mécanisme de sa production en montrant le fonctionnement de l'appareil phonatoire humain. Une modélisation de ce dernier a été exposée à l'aide d'un schéma.

Ensuite, nous avons donné les caractéristiques de ce signal suivi par les représentations temporelles et fréquentielles de quelques types de phonèmes.

Enfin, la formulation mathématique nécessaire pour le traitement de ce type de signaux est abordée.



# **Chapitre 2**

Séparation aveugle de  
sources par l'Analyse  
en Composantes  
Indépendantes

## 2.1. Introduction

La séparation aveugle de sources est un problème ancien qui date des années 80 avec des méthodes comme l'Analyse en Composantes Indépendantes (ACI). Ces méthodes s'appuient souvent sur l'utilisation des cumulants d'ordre quatre ou sur la maximisation d'une fonction d'entropie.

Dans ce chapitre, nous verrons comment séparer le mélange de certaines sources à l'aide de la méthode ACI. Dans un premier temps, nous aborderons les techniques utilisées dans la séparation des mélanges et nous essaierons d'illustrer par des dessins, puis nous évoquerons certaines des techniques utilisées en tant qu'ACI. La section 3 illustre des notions sur les mélanges rencontrés dans le monde réel de l'audio et les problèmes rencontrés lors de leur séparation.

La section suivante sera consacrée à l'étude de la technique ACI, en commençant par la définition, les hypothèses fondamentales sur lesquelles repose cette dernière seront énumérées. Ensuite, nous présenterons la formulation mathématique de l'ACI qui nous permet de la mieux comprendre.

Nous parlerons aussi des critères utilisés pour mesurer l'indépendance entre les signaux en présentant quelques équations. Ceci nous mène à évoquer les différentes approches utilisées pour élaborer une variété d'algorithmes d'ACI.

La cinquième section explique les prétraitements et la procédure d'application d'un algorithme ACI. Enfin, la séparation des mélanges des signaux de paroles par ACI sera discutée.

À la fin, le chapitre sera clôturé par une conclusion.

## 2.2. Technique de séparation à l'aveugle

Les techniques du traitement du signal trouvent des applications dans pratiquement tous les domaines de la technologie, particulièrement en audio. Une bonne partie des signaux audio sont des mélanges aux quels contribuent plusieurs sources.

L'étude des mélanges audio a de nombreuses applications dont la plus populaire est le problème du « cocktail party ». Ce nom provient de la situation rencontrée par exemple lors d'un joyeux déjeuner familial.

La séparation aveugle de sources(SAS) est une discipline plus au moins jeune dont les premiers travaux sont attribués à Héroult, Jutten et Ans au milieu des années 1980.

La séparation de sources consiste à restaurer les signaux sources inconnus  $S_j$  à partir de  $K$  observations  $X_i$ . Ces dernières sont des mélanges des signaux sources et proviennent généralement des capteurs, tels que des microphones, des antennes radiofréquences, des capteurs CCD...

La relation liant les sources et les observations est donnée par :

$$X = F.S \quad (\text{Eq.2.1})$$

Où :

$$S = [s_1 \dots s_l]^T \quad (\text{Eq.2.2})$$

Et

$$X = [x_1 \dots x_k]^T \quad (\text{Eq.2.3})$$

Sont deux vecteurs colonnes représentant respectivement l'ensemble des signaux sources et l'ensemble des signaux d'observations, et  $F$  est une matrice du mélange.

L'originalité de la SAS réside dans le fait qu'on se place dans un contexte aveugle, c'est-à-dire que l'opérateur de mélange ainsi que les signaux sources sont inconnus.

L'objectif de la SAS est de construire un système capable de fournir des sorties  $y_j$  égales aux signaux sources  $S_j$ .

La reconstruction des signaux sources est réalisée grâce au modèle :

$$Y = G.X \quad (\text{Eq.2.4})$$

Où ;

$$Y = [y_1, \dots, y_l]^T \quad (\text{Eq.2.5})$$

Est un vecteur colonne, représentant l'ensemble des signaux de sortie et  $G$  est l'opérateur inverse de  $F$ .

On peut considérer ici deux problèmes qui peuvent être différents et qui sont :

- Le problème de l'estimation du modèle de mélange, que l'on appelle problème d'Identification Aveugle de Mélanges (IAM) (Blind Mixture Identification (BMI)).
- Le problème de restitution des sources consistant à proposer une estimation des sources sans proposer une estimation de la structure de mélange.

Ces deux problèmes coïncident lorsque le mélange est inversible. La configuration générale de la séparation aveugle de sources est illustrée par la figure (Fig. 2.1).

Les mélanges sont reçus aux niveaux des capteurs qui peuvent être des microphones dans le cas des signaux audio et des caméras pour l'image, ces derniers représentent les observations qui sont les seuls signaux connus par le modèle de séparation [4].

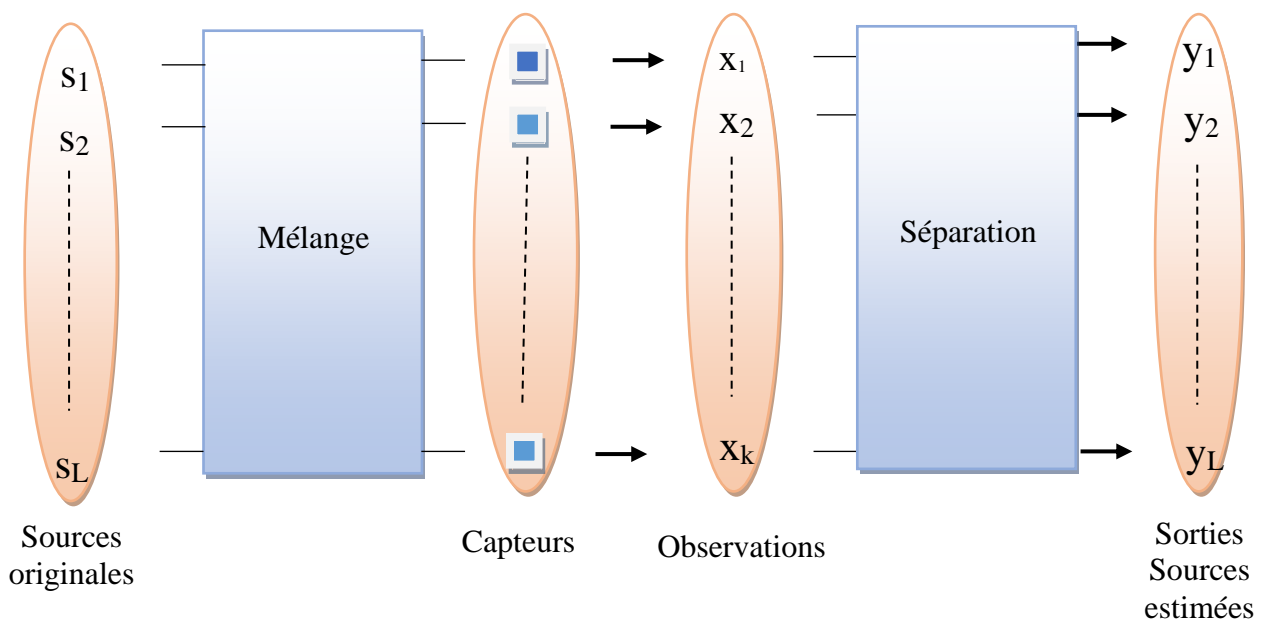


Fig. 2.1 : Configuration générale de la séparation aveugle de sources.

### 2.3. Mélanges et séparation aveugle des sources

La SAS permet de nombreuses applications telles que la biologie, le génie biomédical, la chimie et l'astrophysique d'une manière générale mais elle possède plusieurs degrés de difficulté. Donc, il y a plusieurs paramètres à prendre en considération. Plus particulièrement, les deux points

théoriques qui sont le type de mélange et le nombre  $K$  d'observations par rapport au nombre  $L$  des signaux sources.

Concernant le nombre  $K$  d'observations par rapport au nombre  $L$  de signaux sources, on conçoit que le cas surdéterminé ( $K = L$  et  $K > L$ ), est plus simple à résoudre que le cas sous déterminé ( $K < L$ ). Ce dernier cas ne peut être résolu qu'à l'aide d'importantes informations à priori.

Pour le type de mélange, on trouve dans la littérature plusieurs types. Une première catégorie, largement étudiée, concerne les mélanges « linéaires », c'est-à-dire le cas où les observations sont des mélanges linéaires des signaux sources. Dans ce cas on distingue trois sous-catégories :

- **Mélanges linéaires instantanés** : dans lesquels les valeurs des observations à un instant (lorsqu'il s'agit de signaux temporels) s'écrivent comme des combinaisons linéaires des valeurs des signaux sources au même instant. Dans ce cas, le mélange est décrit par le modèle de l'équation (2.1).

On trouve ce genre de mélange dans les applications où les temps de propagation sont négligeables.

- **Mélanges linéaires à atténuations et retards** : dans lesquels les temps de propagation, ou les décalages spatiaux, sont pris en considération.
- **Mélanges linéaires convolutifs** : dans lesquels les capteurs fournissent des superpositions de versions filtrées des signaux sources reçus. Ce type de mélange est le plus général des trois présentés [4].

## 2.4. Analyse en composantes indépendantes (ACI)

### 2.4.1. Définition

L'analyse en composantes indépendantes (ACI) est une technique qui a été initialement développée pour résoudre le problème de séparation aveugle de sources (SAS). En supposant l'hypothèse d'indépendance des sources, cette méthode utilise les statistiques de ces dernières. Elle exploite aussi la diversité spatiale et fréquentielle ce qui la rend multidimensionnelle (multi-capteur).

### 2.4.2. Hypothèses fondamentales de l'ACI

L'ACI repose sur trois hypothèses fondamentales qui sont :

1. Les sources  $S_j$  sont supposées statistiquement indépendantes. C'est grâce à cette hypothèse que l'on obtient les estimations des sources, appelées composantes « indépendantes » et notées  $Y_j$ .
2. Parmi les  $p$  sources, au plus une peut avoir une distribution gaussienne (car ses cumulants d'ordre supérieur à deux sont nuls).
3. On suppose que la matrice de mélange  $A$  est carrée et inversible et son inverse est noté  $W$ , On peut alors exprimer les sources en fonction des mélanges de sources selon le modèle d'ACI inverse tel que :

$$S = W.X \quad (\text{Eq.2.6})$$

On remarque que les sources sont une combinaison linéaire des mélanges de sources. Il suffit de trouver la bonne matrice de séparation  $W$  qui va nous permettre de retrouver les sources [11].

### 2.4.3. Formulation mathématique de l'ACI

Dans le contexte de la séparation aveugle de sources, l'ACI permet d'extraire les sources à partir de mélanges de ces sources. Dans le modèle d'ACI, on considère les sources et les mélanges de sources comme étant des variables aléatoires. Le modèle s'écrit sous la forme suivante :

$$\begin{bmatrix} x_1 \\ \dots \\ x_m \end{bmatrix} = \begin{bmatrix} a_{11} & \dots & a_{1n} \\ \dots & \dots & \dots \\ a_{n1} & \dots & a_{mn} \end{bmatrix} \begin{bmatrix} S_1 \\ \dots \\ S_n \end{bmatrix} \quad \text{ou } x = A.s$$

Où ;

- $X = (X_1, \dots, X_p)^T$  : un vecteur aléatoire observable  $p \times 1$  à valeurs continues des  $p$  mélanges de sources,
- $A = (a_{lk})$  : une matrice de mélange inconnue non aléatoire de dimension  $p \times p$ .

- $S = (S_1, \dots, S_p)^T$  : un vecteur aléatoire non observable  $p \times 1$ , à valeurs continues, des  $p$  sources que l'on souhaite retrouver.

Le problème semble impossible à résoudre puisque  $A$  et  $S$  sont tous les deux inconnues. Afin de pouvoir identifier les sources  $S_j$ , on peut les exprimer en fonction des mélanges de sources selon le modèle d'ACI inverse tel que (Eq.2.6).

On remarque que les sources forment une combinaison linéaire des mélanges de sources. Il suffit donc de trouver la bonne matrice de séparation  $W$  qui va nous permettre de retrouver les sources. Les différents algorithmes élaborés et appliqués ont démontré ses capacités à séparer des sources. Ces applications consistent donc à générer quelques signaux, puis les mélanger artificiellement avant d'utiliser l'algorithme pour retrouver avec succès les signaux originaux.

En pratique, le modèle d'ACI s'écrit comme :

$$X^T = A.S^T \quad (\text{Eq.2.7})$$

Où ;  $S$  est une matrice ( $n \times p$ ), qui contient les réalisations des sources. Il y a deux objectifs principaux à l'ACI. Le premier objectif est d'estimer la densité de probabilité  $p_{S_j}$ , de chaque source  $S_j$  à partir de la matrice d'observations  $X$  seulement. Le deuxième objectif est de retrouver les réalisations  $[S_1, \dots, S_n]$  comprises dans la matrice  $S$  c'est-à-dire à les estimer à partir de la matrice d'observations  $X$ .

Chacun des objectifs nécessite l'estimation de la matrice de séparation notée  $\hat{W}$ . L'estimation de la densité  $p_{S_j}$ , de chaque source  $S_j$ , se fait directement si on connaît la densité  $P_X$  des mélanges de sources ou bien si cette dernière a été estimée à partir des données. Les sources  $S_j$ , seront retrouvées en appliquant la formule ci-dessous:

$$S = W.X \quad (\text{Eq.2.8})$$

L'équation (2.8) représente le modèle d'ACI inverse. L'estimation de  $S$  est dénotée par la matrice  $Y$  contenant les « composantes indépendantes » telle que :

$$Y^T = \tilde{W}.X^T \quad (\text{Eq.2.9})$$

Chacune des approches d'analyse en composantes indépendantes se base sur des critères d'indépendance différents afin d'estimer la matrice  $W$  pour obtenir ensuite, des composantes  $(y_1, \dots, y_p)$ , réalisations des variables aléatoires  $Y_1, \dots, Y_p$  les plus indépendantes possible au sens de ce critère [11].

#### 2.4.4. Mesure de l'indépendance

Deux variables aléatoires  $X$  et  $Y$  sont dites indépendantes si :

$$p(X = x \text{ et } Y = y) = p(X = x).p(Y = y) \quad (\text{Eq.2.10})$$

Avec  $p(\cdot)$  dénote la densité de probabilité d'une variable aléatoire. Cette définition ne peut pas être utilisée directement. Généralement, l'indépendance est exprimée par :

$$E(XY) - E(X)E(Y) = 0 \quad (\text{Eq.2.11})$$

Où ;  $E(\cdot)$  dénote l'opérateur d'espérance mathématique. En plus, les variables indépendantes sont toujours décorrélées (covariance nulle) mais l'inverse n'est pas toujours vrai. Ceci permet de simplifier la recherche des composantes indépendantes mais elle n'est pas très intéressante en pratique car il existe plusieurs possibilités de décorréliser des variables [1].

#### 2.4.5. Différentes approches de l'ACI

Une approche est définie en fonction du critère utilisé pour mesurer l'indépendance. L'indépendance entre les signaux peut être mesurée par plusieurs critères dont les plus fameux sont :

##### 2.4.5.1. L'information mutuelle

L'information mutuelle entre des variables aléatoires est toujours positive et vaut 0 si et seulement si les variables sont indépendantes. On peut donc utiliser l'information mutuelle comme critère de dépendance afin de retrouver des composantes  $Y_1, \dots, Y_p$  les plus indépendantes possible.

Supposons que les mélanges  $X$  de sources ont été blanchis préalablement de façon à ce que :



$$COV(X) = I \quad (\text{Eq.2.12})$$

Où ; I représente la matrice identité.

L'entropie d'une transformation linéaire est décrite par l'équation suivante :

$$H(Y) = H(X) + \log |detM| \quad (\text{Eq.2.13})$$

Où ; X : le mélange ; Y : le signal estimer ; M : nombre de source.

A l'aide de l'équation (2.13), on peut exprimer l'information mutuelle des composantes indépendantes comme suit :

$$I(Y_1, Y_2, \dots, Y_p) = \sum_{i=1}^p H(Y_i) - H(Y) = \sum_{i=1}^p H(Y_i) - H(X) - \log|detW| \quad (\text{Eq.2.14})$$

Le blanchiment résout le problème d'ACI à une matrice orthogonale près. Puisque les mélanges de sources ont été blanchis préalablement, ceci implique que la matrice W est orthogonale. Ou, la valeur absolue du déterminant d'une matrice orthogonale vaut toujours 1. Le dernier terme à droite de l'équation vaut donc 0. De plus, l'entropie des mélanges de sources blanchies H(x) est une constante.

On obtient alors :

$$I(Y_1, Y_2, \dots, Y_p) = \sum_{i=1}^p H(Y_i) - \text{constante} \quad (\text{Eq.2.15})$$

Puisque l'information mutuelle est toujours positive, on n'a qu'à minimiser le terme  $\sum_{i=1}^p H(Y_i)$  afin d'obtenir des composantes les plus indépendantes possible [11].

#### 2.4.5.2. Entropie d'une suite de variables

Si X et Y sont deux variables aléatoires, on peut définir l'entropie du couple X, Y par :

$$S = - \sum p_{ij} \log p_{ij} \quad (\text{Eq.2.16})$$

Avec :  $p_{ij}$  : La proportion de particules

Ceci peut être généralisé à un nombre quelconque de variables.

Il résulte de la définition que si X et Y sont deux variables aléatoires indépendantes, on a :

$$S(X, Y) = S(X) + S(Y) \quad (\text{Eq.2.17})$$

L'information conjointe des variables indépendantes, est égale à la somme des informations individuelles.

En fait on a toujours

$$S(X, Y) \leq S(X) + S(Y) \quad (\text{Eq.2.18})$$

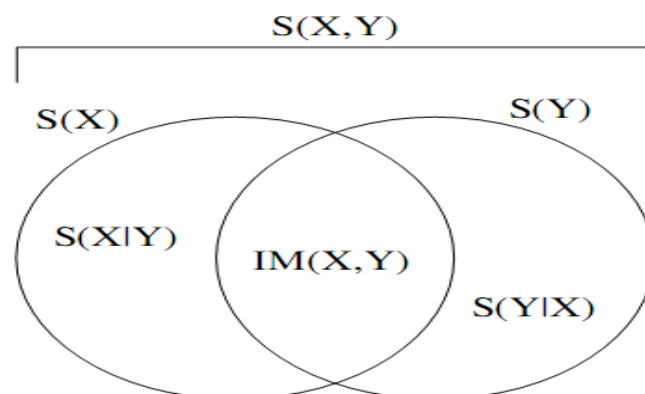
Avec l'égalité si et seulement si X et Y sont indépendantes. Ceci peut servir par exemple à définir une information partielle

$$S(X|Y) = S(X, Y) - S(Y) \quad (\text{Eq.2.19})$$

$S(X|Y)$ , c'est la quantité d'information réellement apportée par X si on connaît déjà Y. Ainsi, une information mutuelle est définie par :

$$IM(X, Y) = S(X, Y) - (S(X|Y) + S(Y|X)) \quad (\text{Eq.2.20})$$

C'est la quantité d'information présente dans X et dans Y.



**Fig. 2.2:** Diagramme de Vann de l'information mutuelle et l'entropie conditionnelle pour une classe X et une variable Y

Par exemple, pour le cas d'un texte en une langue, chaque lettre dépend des précédentes. Le signal  $X_t$  émis au temps  $t$ , c'est une variable aléatoire à valeurs dans un certain alphabet. Ces variables ne sont pas forcément indépendantes. L'entropie de la suite infinie  $X = (X_1, \dots, X_n, \dots)$  est définie par :

$$S(X) = \lim_{n \rightarrow \infty} \frac{1}{n} S(X_1, \dots, X_n) \quad (\text{Eq.2.21})$$

si cette limite existe, où  $S(X_1, \dots, X_n)$  est l'entropie de la distribution jointe du vecteur  $(X_1, \dots, X_n)$ .

Si les lettres successives du message sont toutes indépendantes et identiquement distribuées (la loi de  $X_n$  est égale à la loi de  $X_1$  pour tout  $n$ ), alors on a bien sûr :

$$S(X_1, \dots, X_n, \dots) = \lim_{n \rightarrow \infty} \frac{1}{n} S(X_1, \dots, X_n) = S(X_1) \quad (\text{Eq.2.22})$$

Autrement dit, dans le cas d'une source qui émet des lettres successives et indépendantes toujours avec la même loi, la quantité d'information moyenne pour chaque nouvelle lettre, est égale à l'entropie de la loi [12].

### 2.4.5.3. La néguentropie

L'entropie utilisée pour bâtir une mesure de non gaussianité est appelée la néguentropie. Elle vaut 0 pour une variable gaussienne et elle est toujours positive. Elle est définie comme suit :

$$J(Y) = H(Y_{gauss}) - H(Y) \quad (\text{Eq.2.23})$$

Où ;  $Y_{gauss}$  est une variable gaussienne ayant la même matrice de covariance que  $Y$ . Son entropie peut être évaluée par :

$$H(Y_{gauss}) = \frac{1}{2} \log |\det \Sigma| + \frac{n}{2} [1 + \log(2\pi)] \quad (\text{Eq.2.24})$$

L'information mutuelle de l'équation (2.15) peut être exprimée par :

$$I(Y_1, Y_2, \dots, Y_p) = \sum_{j=1}^p [H(Y_{j,gauss}) - J(Y_j)] - \text{constante} \quad (\text{Eq.2.25})$$

L'entropie  $H(Y_{j,gauss})$  est constante. On obtient alors :

$$I(Y_1, Y_2, \dots, Y_p) = \text{constante} - \sum_{j=1}^p J(Y_j) \quad (\text{Eq.2.26})$$

Ceci montre la relation entre l'information mutuelle et la négentropie. On constate que la minimisation de l'information mutuelle (trouver la transformation linéaire et inversible  $W$ ) est la même chose que maximiser la négentropie. L'équation (2.26), nous montre bien que l'ACI par la minimisation de l'information mutuelle est équivalente à la maximisation de la somme de mesures de non gaussianité des composantes indépendantes. On en conclut que la maximisation de la non gaussianité mène à trouver des composantes « indépendantes ».

Par contre, en pratique, il y a une différence importante entre les deux méthodes de résolution du problème d'ACI. La négentropie, ainsi que les autres mesures de non gaussianité, nous permet de trouver des composantes « indépendantes » l'une après l'autre, puisque nous trouvons le maximum de non gaussianité dans une seule direction  $w$  à la fois. Ceci n'est pas possible avec l'information mutuelle [11].

#### 2.4.5.4. Le kurtosis

Etant donné une variable aléatoire réelle  $X$  d'espérance  $\mu$  et d'écart-type  $\sigma$ , on définit son kurtosis comme :

$$\beta_2 = E\left[\left(\frac{X-\mu}{\sigma}\right)^4\right] \quad (\text{Eq.2.27})$$

Pour trouver l'inverse de la matrice de mélange afin d'identifier les sources, on procède source par source. On cherche un vecteur  $w$  tel que :

$$S = W^T \cdot X \quad (\text{Eq.2.28})$$

Ait le kurtosis le plus différent de 0 possible.

- On part d'un vecteur  $W$  pris au hasard.

- On calcule la direction où le kurtosis croit le plus vite s'il est positif ou décroît le plus vite s'il est négatif. Par des techniques de type « gradient », on peut trouver un optimum local. On a alors identifié une source.

## 2.5. Application d'un algorithme ACI

Avant d'appliquer l'algorithme ACI sur les données, il est généralement très utile de faire un prétraitement qui rend l'estimation du problème de l'ACI plus simple et mieux conditionnée. Parmi les techniques de prétraitement on cite :

### 2.5.1. Centrage des données

Le prétraitement le plus élémentaire et nécessaire est de centrer  $X$ , c'est-à-dire lui soustraire sa moyenne qui est définie par :

$$m_X = E\{X\} \quad (\text{Eq.2.29})$$

Où ;  $E\{\cdot\}$  dénote l'opérateur espérance mathématique.

Ceci rend la variable  $X$  de moyenne nulle et de même pour  $S$ . Ce prétraitement sert uniquement à simplifier les algorithmes ACI. Après l'estimation de la matrice de mélange  $A$  avec les données centrées, nous pouvons compléter l'estimation en ajoutant la moyenne du vecteur  $s$ . Cette valeur est donnée par :

$$m_s = A^{-1} \cdot m_X \quad (\text{Eq.2.30})$$

Où ;  $m_X$  est la valeur moyenne qui a été soustraite dans le prétraitement [1].

### 2.5.2. Détermination du nombre de sources

Dans une plusieurs expérience, le nombre de sources peut être inconnu , la détermination de nombre de sources est un problème plus important dans les mélanges ,mais si on parle sur les mélanges instantanés sans bruit surdéterminés, le nombre de sources peut être estimé à partir de la matrice de covariance des observations , le nombre de sources est égale le nombre de valeur propre non nulle de cette matrice de covariance, Cependant, dans d'autres expérience, la détermination du nombre de sources n'est pas facile, le nombre de sources est estimé avec le système de mélange [13].

### 2.5.3 Réduction de dimensions

Ce prétraitement est une technique qui consiste à réduire la dimension des données multidimensionnelles. En séparation de sources, ceci permet de limiter le nombre d'observations et d'éliminer la redondance dans ces dernières, afin de réduire le coût de calcul. L'analyse en composantes principales (PCA) est souvent utilisée pour accomplir cette tâche. Ce prétraitement est avantageux dans la réduction du bruit.

### 2.5.4. Blanchiment

Un autre prétraitement utile à l'algorithme ACI est de blanchir les variables observées. Après centrage des données, nous devons transformer linéairement le vecteur  $x$  observé de sorte que nous obtenons un nouveau vecteur  $\tilde{x}$  qui est blanc, c'est-à-dire ses composantes sont décorrélées et leurs variances sont égales à l'unité. En d'autres termes, la matrice de covariance de  $\tilde{x}$  est égal à la matrice identité :

$$E\{\tilde{x}\tilde{x}^T\} = I \quad (\text{Eq.2.31})$$

## 2.6. Application de l'ACI à des mélanges des signaux de parole

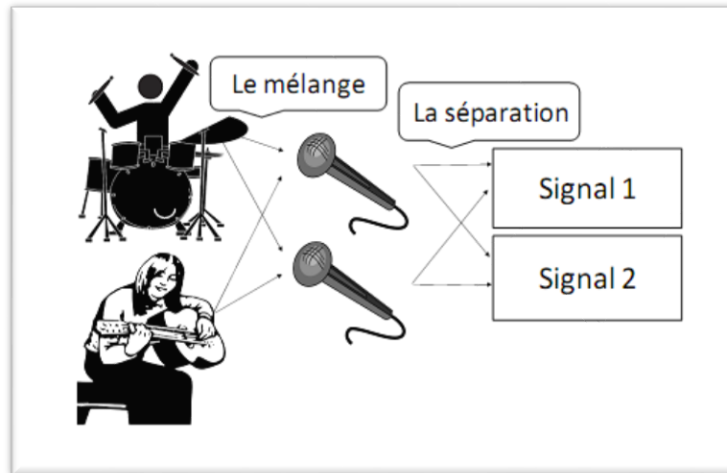
La séparation de sources a été étudiée initialement pour les mélanges instantanés. Une attention considérable a visé les applications potentielles dans le traitement de la parole, bien que le modèle convolutif semble plus adapté à cette application.

Une première application est la séparation de signaux de parole, telle que présentée dans le problème de « la soirée cocktail ».

### 2.6.1. Mélanges instantanés

Malheureusement, ce modèle d'ACI présenté (mélange linéaire instantané) ne permet pas de résoudre le problème de séparation du mélange illustré dans la figure 2.3, parce que les signaux ont tendance à être convolés et surtout parce qu'il faut prendre en compte les délais temporels entre chaque microphone et les sources.

De plus, dans un contexte réel, nous connaissons mal le modèle de mélange des voix, ce qui rend la séparation d'enregistrements réels difficile.



**Fig.2.3** : Illustration d'un processus de séparation de deux sources à partir de deux observations.

### 2.6.2. Mélanges convolutifs

Le problème le plus général de séparation de sources pour les mélanges convolutifs a été envisagé au début des années 1990. La différence entre le modèle convolutif et le modèle instantané est que les valeurs retardées des signaux de source contribuent à la sortie à un instant donné. Plus précisément, un signal d'observation  $x_m \in \mathbb{R}^T$  est décrit par un mélange:

$$x_m(t) = \sum_{n=1}^N s_{mn}^{img}(t) \quad (\text{Eq.2.32})$$

Où :

$$s_{mn}^{img}(t) = \sum_{l=1}^K a_{mn}(l) s_n(t-l) \quad (\text{Eq.2.33})$$

Dans notre travail, on se limitera aux mélanges linéaires invariants dans le temps. Les mélanges linéaires invariants dans le temps peuvent se modéliser à l'aide d'une opération de convolution. Cependant, on s'intéressera seulement au cas particulier des mélanges instantanés [13].

## 2.7. Les critères d'évaluation de la qualité de la parole

Les mesures de performances d'un système de séparation de sources se divise en deux type qui sont : les mesures subjectives d'écoute avec un casque audio de qualité, et les mesures mathématiques objectives.

### 2.7.1. Critères subjectifs

MOS (Note d'opinion moyenne) : c'est un critère utilisé pour déterminer la qualité des signaux estimés après le démixage, ce critère dépend de l'audition de ces derniers par plusieurs personnes. Chaque personne donne une note entre 1 et 5 selon la qualité subjective. Le tableau 2.1 représente les notes et les niveaux de critère MOS correspondants résultant d'un sondage fait par des experts. La moyenne des notes fournies par cette population constitue le MOS.

Note	Qualité de la parole	Niveaux de dégradation
5	Excellente	Imperceptible
4	bonne	perceptible mais non gênant
3	Moyenne	légèrement gênant
2	Mauvaise	Gênant
1	Très mauvaise	très gênant

Tableau 2.1 : l'échelle MOS [17]

### 2.7.2. Critères Objectifs

Ce type de critères permet d'évaluer la qualité des signaux par des moyens plus faciles à mettre en œuvre que les tests d'écoute. Récemment, Vincent et al. ont proposé un ensemble de critères de mesure de performances de séparation de sources. Dans le cas d'un mélange, le principe de ces dernières est le suivant :

On considère que l'estimée  $\hat{s}_i(t)$  d'un signal  $s_i(t)$  peut-être décomposée selon la somme suivante :

$$\hat{s}_i(t) = s_i^{cible}(t) + e_i^{interf}(t) + e_i^{art}(t) \quad (\text{Eq.2.34})$$



Où ;

- $s_i^{cible}(t)$  : est une dégradation acceptable du signal source  $s_i(t)$ .
- $e_i^{interf}(t)$  : est une dégradation due aux autres signaux sources considérés alors comme des interférences.
- $e_i^{art}(t)$  : sont les artefacts dûs au processus de séparation tels que le bruit musical.

Les mesures de performances introduites sont alors :

- Signal-to-Distorsion Ratio (SDR).
- Signal-to-Interference Ratio (SIR).
- Signal-to-Artefacts Ratio (SAR).

Ils sont définis respectivement par :

$$SDR = 10 \log_{10} \frac{\|s_i^{cible}\|^2}{\|e_i^{interf} + e_i^{art}\|^2} \quad (\text{Eq.2.35})$$

$$SIR = 10 \log_{10} \frac{\|s_i^{cible}\|^2}{\|e_i^{interf}\|^2} \quad (\text{Eq.2.36})$$

$$SAR = 10 \log_{10} \frac{\|s_i^{cible} + e_i^{interf}\|^2}{\|e_i^{art}\|^2} \quad (\text{Eq.2.37})$$

Il y a un autre type des mesures objectives de la qualité de la séparation, noté SNR (Ratio Signal to Noise) est un terme utilisé pour renseigner sur la qualité d'un signal estimé [14].

Ce rapport est exprimé en décibels par l'équation suivante :

$$SNR = 20. \log\left(\frac{U_{Utile}}{U_{bruit}}\right) \quad (\text{Eq.2.38})$$

## **2.8. Conclusion**

Dans ce chapitre nous avons essayé d'expliquer la séparation aveugle par ACI qui est une technique très prometteuse dans le domaine de la séparation des signaux. En premier lieu, nous avons présenté la technique de séparation à l'aveugle à l'aide des schémas et des équations mathématiques.

Ensuite, les critères variés pour créer les différents algorithmes ACI qui permettent de retrouver les sources indépendantes noyées dans les mélanges, ont été présentés.

Puisque notre travail s'intéresse au cas des mélanges instantanés, une formulation mathématique de l'ACI de ce modèle a été donnée.

Enfin, le problème de séparation des mélanges des signaux de parole a été discuté.

# **Chapitre 3**

## Résultats de simulation

### 3.1. Introduction

Dans ce chapitre nous allons présenter les résultats trouvés par la simulation, en utilisant le logiciel Matlab. Ce travail comporte trois parties principales et chacune s'exécute en plusieurs étapes.

Dans la première partie, nous allons s'entraîner à lire, à entendre et à afficher des signaux de paroles en Matlab.

La deuxième partie consiste à mélanger les signaux de paroles pour obtenir deux ou trois mélanges différents. Nous pourrions aussi traiter des mélanges disponibles à partir d'autres sources. Ensuite, nous allons écouter et représenter graphiquement ces mélanges.

Dans la troisième partie nous entamerons la procédure de séparation des signaux mélangés en utilisant la méthode ACI. Pour évaluer les performances de cette technique de séparation, nous allons utiliser des critères connus en traitement du signal de parole tel que le SDR, SAR, SIR et le MOS.

Enfin, nous terminerons ce chapitre par une conclusion qui décrit et discute le déroulement des différentes expériences.

### 3.2. Environnement du travail

Dans cette section, nous définirons logiciel utilisés dans notre travail. Pour réaliser ce projet.

#### 3.2.1. Logiciel Matlab

Entre 1985 et 1990, plusieurs logiciels interactifs de calcul scientifique sont apparus sur le marché dont MATLAB est l'un d'entre eux. Sa simplicité fait de lui un outil de choix pour la mise au point des algorithmes scientifiques.

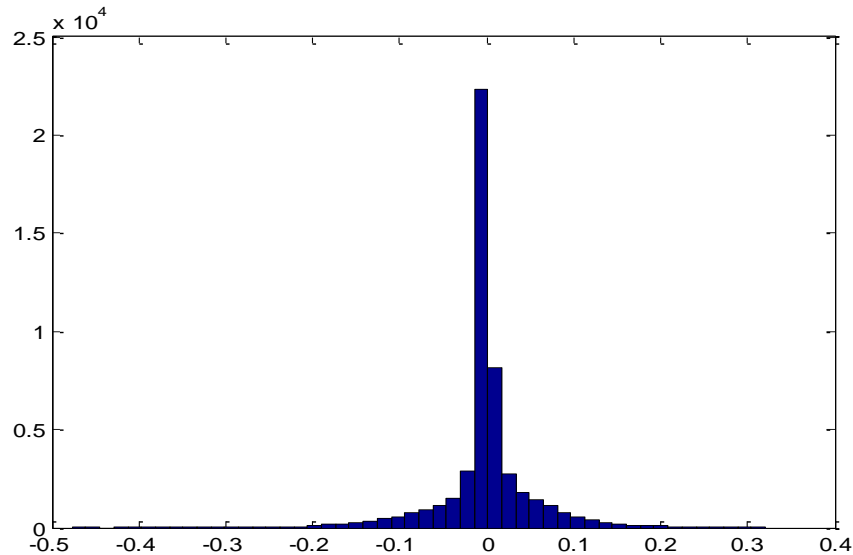
Grâce à sa puissance dans le calcul matriciel et ses nombreuses bibliothèques « Toolboxes » de fonctions préexistantes, Matlab simplifie et rend plus fiable la résolution des problèmes par l'utilisateur.

### 3.3. Base de données

Pour toutes les simulations qui suivent nous utiliserons deux signaux de paroles et un signal de bruit (extraits du corpus français des signaux audio) dont les caractéristiques sont les suivantes :

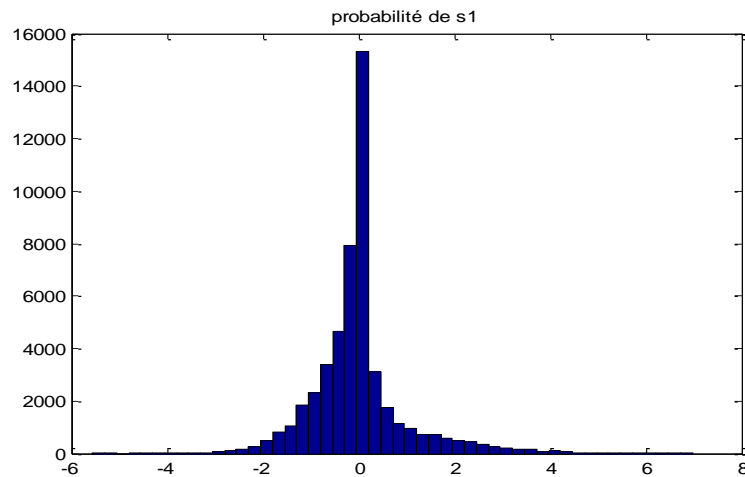
- une phrase en français prononcée par un locuteur femme, enregistrée sous le format « .wav » et échantillonnée à 8khz : « La bise et le soleil se sont disputés... ». Ce signal a une durée de 30 s, disponible dans une base de données TIMIT et représente la source  $s_1(t)$ .

La figure 3.1 représente la distribution de la loi de probabilité de la source  $s_1(t)$  et qui montre la non-gaussianité de cette dernière.



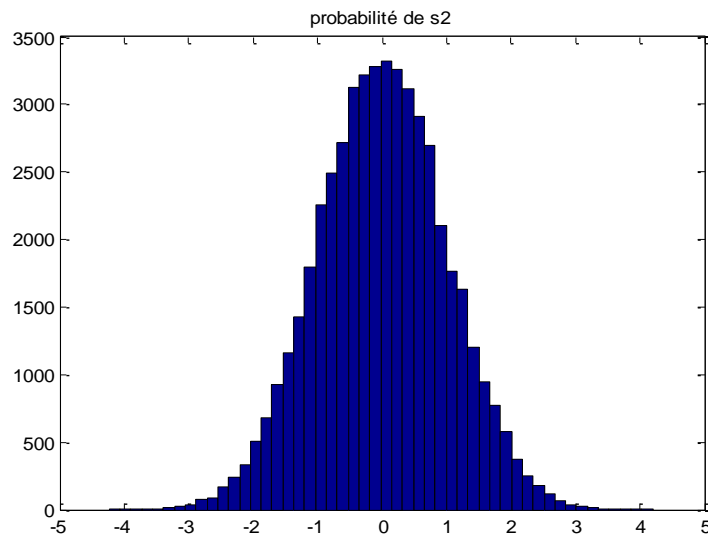
**Fig. 3.1** : Distribution de probabilité de la source  $s_1(t)$

- une phrase en français prononcée par un locuteur homme, enregistrée sous le format « .sig » et échantillonnée à 16kHz : « Un loup s'est jeté immédiatement sur la petite chèvre..... ». Ce signal a une durée de 4 s, disponible dans une base de données TIMIT et représente la source  $s_2(t)$ . Cette dernière présente aussi une distribution de probabilité non gaussienne comme le montre la figure 3.2.



**Fig. 3.2 :** Distribution de probabilité de la source  $s_2(t)$

- un bruit blanc gaussien : Ce signal représente la source  $s_3(t)$ , la figure 3.3 illustre sa distribution de probabilité qui est gaussienne.



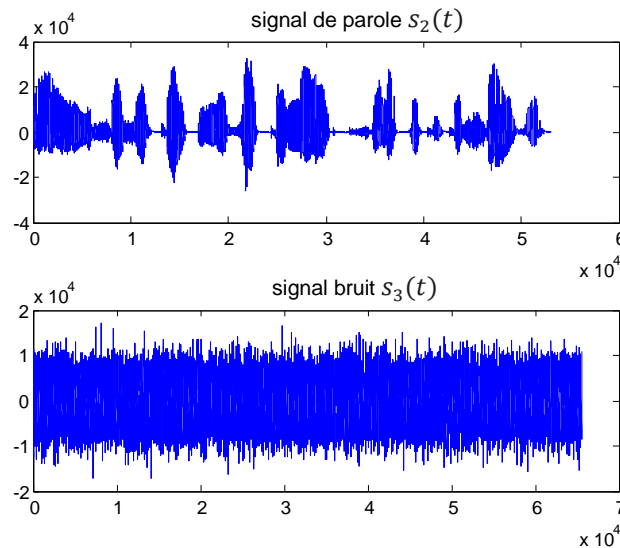
**Fig. 3.3 :** Distribution de probabilité de la source  $s_3(t)$

### 3.4. Résultats et discussion

#### 3.4.1 Le mélange des signaux

Avant d'entamer la procédure de séparation, il faut posséder des mélanges qui vont être les signaux d'entrée de ce système. Nous allons réaliser deux expériences avec des mélanges différents.

Pour le premier cas, les sources  $s_2(t)$  et  $s_3(t)$  sont mélangées par une matrice A appelée matrice de mélange et obtenue aléatoirement avec les instructions `rand(2,2)` ou `randn(2,2)`. La figure 3.4 affiche les signaux sources.



**Fig. 3.4 :** L'audiogramme des signaux sources  $s_2(t)$  et  $s_3(t)$ .

Pendant cette étape, nous avons pu lire, écouter et afficher des signaux audio. Nous avons mélangé instantanément un signal de parole et un bruit. Les signaux de parole sont purs et très clairs.

Après avoir écouté le premier mélange, nous avons remarqué que la voix humaine n'était pas claire à cause du bruit. Pour le deuxième mélange, nous n'avons entendu que du bruit car la parole a été noyée dans le bruit.

La figure 3.5 illustre les deux mélanges (X1 et X2) comme suit :

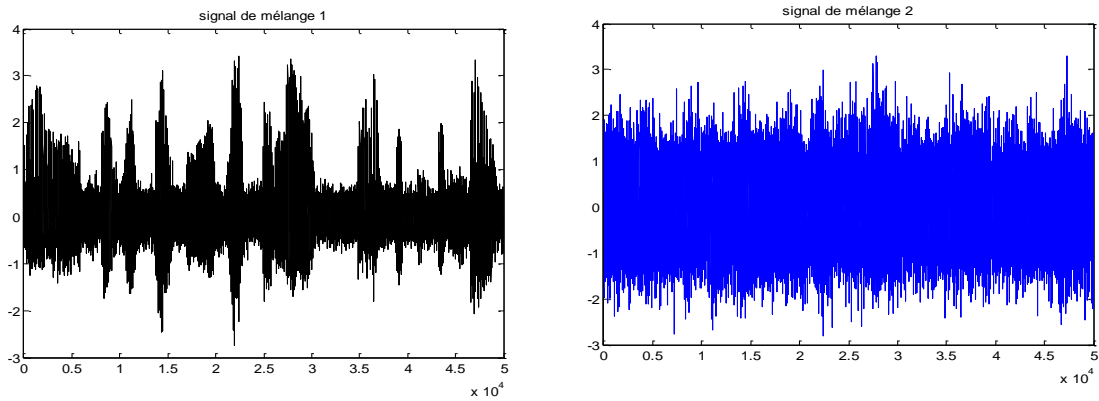


Fig. 3.5 : L'audiogramme des mélanges X1 et X2.

Le deuxième cas consiste à mélanger les trois sources  $s_1(t)$ ,  $s_2(t)$  et  $s_3(t)$  avec une matrice de mélange  $A=\text{rand}(3,3)$ . Les trois sources sont illustrées par la figure 3.6 ci-dessous :

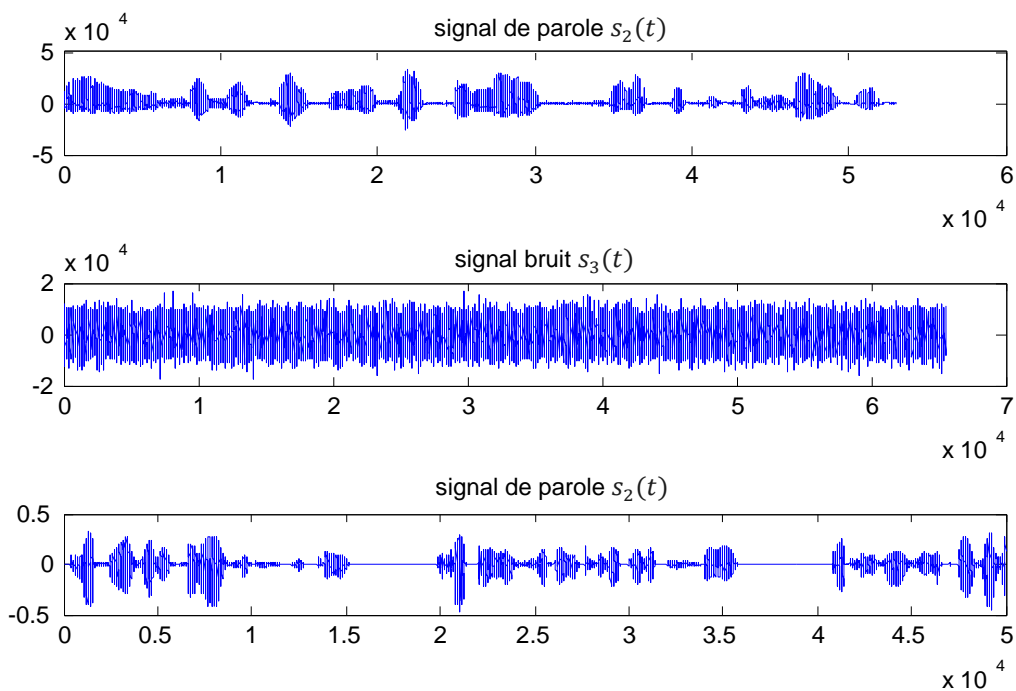
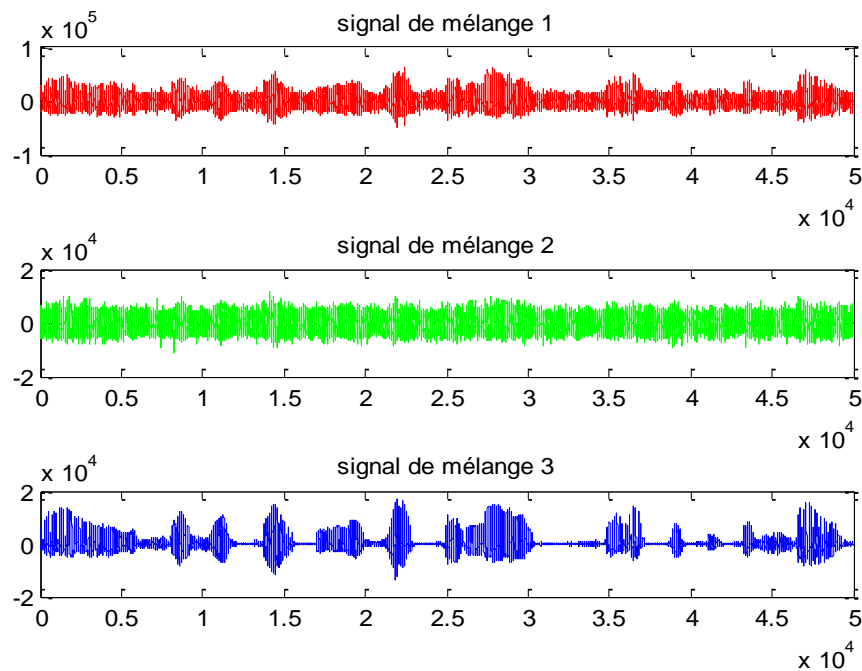


Fig. 3.6 : L'audiogramme des trois signaux sources  $s_1(t)$ ,  $s_2(t)$  et  $s_3(t)$



Les mélanges obtenus sont affichés sur la figure 3.7 suivante :



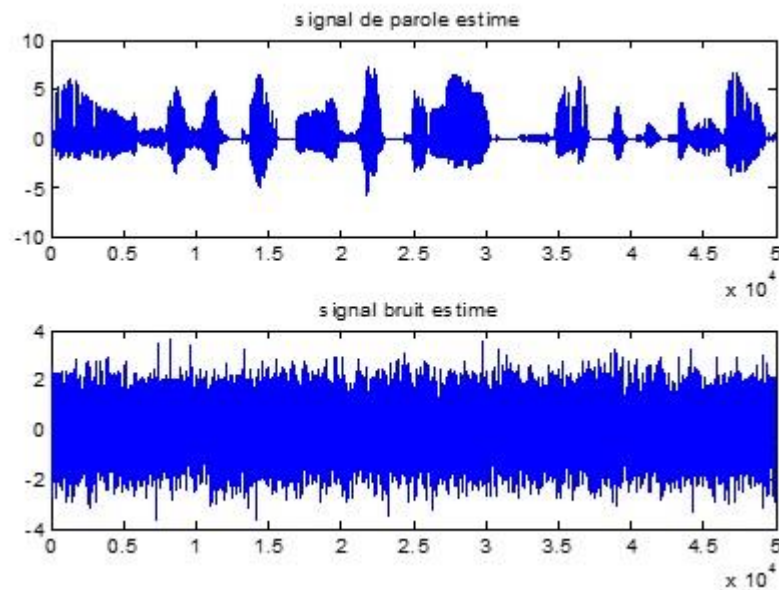
**Fig. 3.7** : L'audiogramme des trois mélanges X1, X2 et X3

### 3.4.2 Séparation des signaux mélangés

Lors de cette expérimentation nous allons appliquer deux versions d'algorithmes ACI c.-à-d. deux programmes Matlab pour séparer les deux types de mélanges définis dans la section précédente.

#### 3.4.2.1 Séparation du premier mélange

Le premier programme utilise comme entrée deux sources, avec l'algorithme ACI à base de l'entropie. Les sources indépendantes estimées sont données sur la figure 3.8 ci-dessous :



**Fig. 3.8** : L'audiogramme des deux sources indépendantes estimées

La figure 3.8 montre les deux signaux estimés dans le domaine temporel. Après avoir écouté ces derniers, nous avons remarqué que les deux voix étaient bien séparées au point à ne pas les distinguer des sources originales.

Pour pouvoir évaluer la qualité de séparation, nous avons mentionné, les critères objectifs et subjectifs accompagnant cette expérience, dans les tableaux 3.1, 3.2 et 3.3.

Les auditeurs	La note du MOS	Valeur moyenne
01	4	4.05
02	4.5	
03	4.75	
04	4	
05	3	

**Tableau.3.1** : Résultat du test MOS pour le premier cas

Le tableau 3.1, illustre les résultats du test MOS réalisé par un groupe de cinq auditeurs (normalement doivent être des experts) chacun attribue une note variant de 1 à 5 en écoutant le signal estimé. Pour ce cas la valeur moyenne du MOS est égale 4.05 qui est une qualité « Excellente ».

Valeur du pas du gradient	SDR(dB)	SIR(dB)	SAR(dB)
0.25	44.7974	44.7974	284.5705
0.9	13.7777	13.7777	289.4460
0.1	21.1708	21.1708	290.8724

**Tableau.3.2:** L'évaluation de la qualité de la source estimée  $\tilde{s}_2(t)$  en fonction du pas du gradient.

La matrice de dé-mélange  $W$  est retrouvée par un algorithme itératif à base de gradient utilisant un pas pour la mise à jour de cette dernière.

Valeur du pas du gradient	SDR(dB)	SIR(dB)	SAR(dB)
0.25	66.1041	66.1041	289.6071
0.9	10.2450	10.2450	287.0704
0.1	15.1843	15.1843	284.4705

**Tableau.3.3:** L'évaluation de la qualité de la source estimée  $\tilde{s}_3(t)$  en fonction du pas du gradient.

Les tableaux 3.2 et 3.3 ci-dessus, expriment l'évaluation de la qualité des deux sources estimées par des critères objectifs SDR, SIR, SAR déjà définis dans le chapitre précédent. Dans ces expériences, nous avons simulé les trois grandeurs pour trois pas différents : 0.25, 0.9 et 0.1.

D'après ces tableaux, nous remarquons que la qualité d'estimation est bonne pour le plus faible pas qui nécessite une convergence lente de l'algorithme vers la solution désirée.

Un grand pas accélère la convergence mais dégrade la qualité.

Dans une autre expérience, nous avons changé le nombre d'itérations et nous avons obtenu les résultats donnés dans tableaux 3.4 et 3.5 suivants :

Nombre d'itérations	SDR(dB)	SIR(dB)	SAR(dB)
100	41.5875	41.5875	282.1049
900	44.7768	44.7768	284.2722
10000	44.7768	44.7768	283.4658

**Tableau.3.4** : L'évaluation de la qualité de la source estimée  $\tilde{s}_2(t)$  en fonction de nombre d'itérations

Nombre d'itérations	SDR(dB)	SIR(dB)	SAR(dB)
100	35.0437	35.0437	293.6325
900	66.3650	66.3650	293.4370
10000	66.3650	66.3650	297.1308

**Tableau.3.5** : L'évaluation de la qualité de la source estimée  $\tilde{s}_3(t)$  en fonction de nombre d'itérations

D'après ces tableaux, nous observons que la qualité d'estimation des sources indépendantes augmente avec le nombre d'itération. Les SDR et SIR dépassent 40 dB ce qui traduit une bonne qualité de séparation.

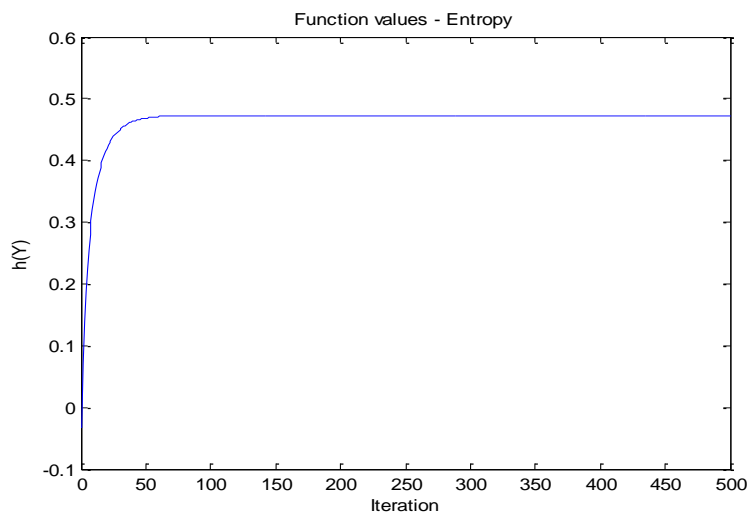
corrélation initiale (début de programme)		Corrélation finale (fin du programme)	
$Y_1$	$Y_2$	$Y_1$	$Y_2$
0.6663	0.1318	1.0000	0.0056
0.7417	0.9905	0.0002	1.0000

**Tableau. 3.6** : Corrélations entre signaux estimés et originaux

Nous avons aussi simulé la corrélation entre les signaux estimés et les sources originales pour voir s'ils sont les mêmes ou non. Le tableau 3.6 affiche les résultats trouvés.

Quand la corrélation est maximale et égale à un, nous disons que le signal estimé est le même que le signal original. Si ce n'est pas le cas la séparation ne s'est pas faite.

D'après le Tableau 3.6, nous remarquons que la corrélation initiale est différente de 1 c.-à-d., il n'y a pas de relation entre le signal estimé et le signal source au lancement du programme. A la fin, le résultat devient égal à 1, ce qui prouve que les sources sont estimées à partir de leurs mélanges.



**Fig. 3.9:** Evolution de la valeur de l'entropie

L'entropie est aussi un indice de la séparation et on aura l'indépendance lorsqu'elle est maximale. La figure 3.9, montre son évolution pour un pas d'adaptation égal à 0.25 et 500 itérations. Nous remarquons que la valeur de l'entropie se stabilise à 0.48 en donnant des signaux paroles estimés de qualité comparable à celle des signaux originaux.

La figure 3.10, montre La variation de son gradient, nous voyons clairement que lorsque l'entropie est maximale sa dérivée s'annule.

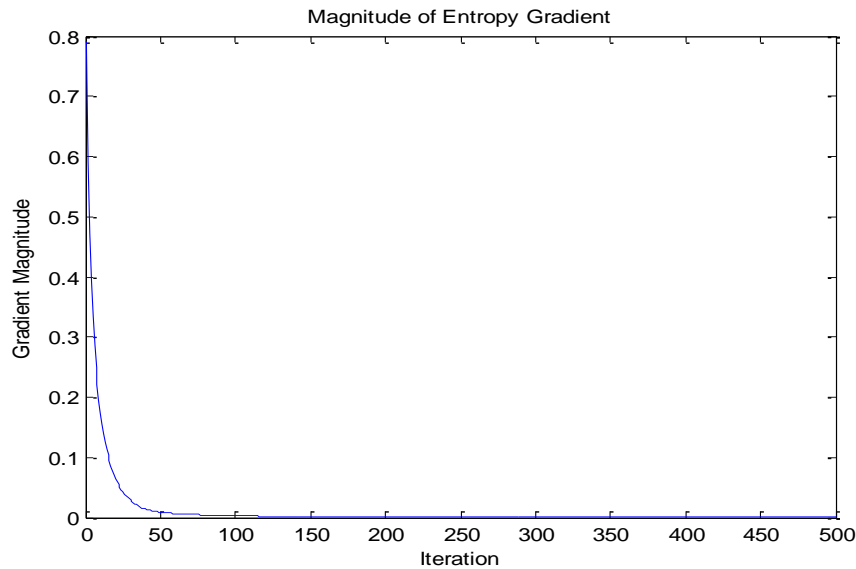


Fig. 3.10: Variation du gradient de la valeur de l'entropie.

### 3.4.2.2 Séparation du deuxième mélange

Cette expérience consiste à séparer un mélange obtenu en utilisant trois sources. Le deuxième programme utilise un algorithme ACI avec une autre approche infomax. Les sources indépendantes estimées sont données sur la figure 3.11 ci-dessous :

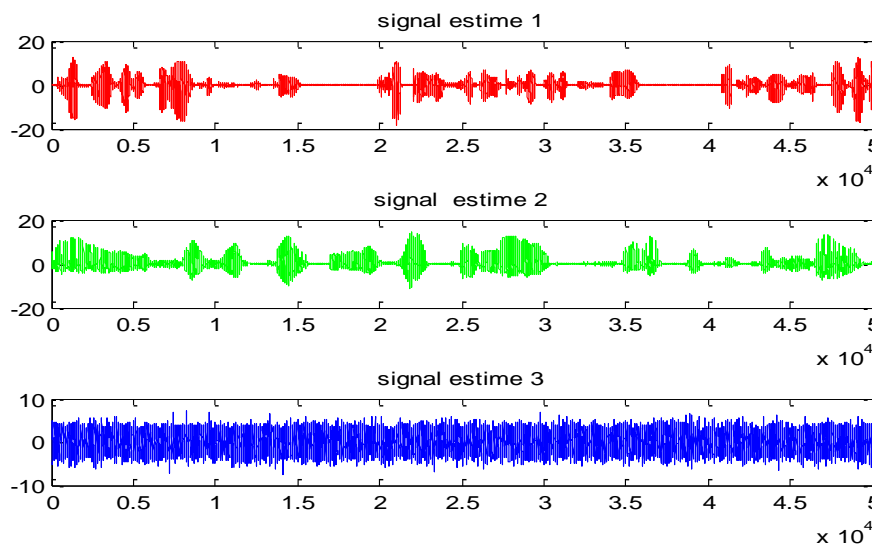


Fig. 3.11 : L'audiogramme des trois sources indépendantes estimées  $\tilde{s}_1(t)$ ,  $\tilde{s}_2(t)$  et  $\tilde{s}_3(t)$

La figure 3.11 montre respectivement les signaux estimés dans le domaine temporel. Après avoir écouté, nous avons remarqué que les trois voix étaient bien séparées.

De la même façon que la première expérience, nous avons simulé les critères objectifs et subjectifs pour ce cas et les résultats sont résumés dans les tableaux 3.7, 3.8 et 3.9 qui suivent :

Les auditeurs	La note du MOS	Valeur moyenne
01	4	4.15
02	4	
03	4.5	
04	3.75	
05	4.5	

**Tableau.3.7 :** Résultat du test du MOS pour le deuxième mélange

Le tableau 3.7 illustre les notes attribuées à la qualité du signal écouté respectivement par cinq auditeurs. La valeur moyenne du MOS vaut 3.9 ce qui reflète une qualité de séparation « Excellente ».

Les sources	SDR(dB)	SIR(dB)	SAR(dB)
S <sub>1</sub>	31.4975	31.4975	247.7636
S <sub>2</sub>	30.7483	30.7483	251.4205
S <sub>3</sub>	33.1266	33.1266	216.6631

**Tableau.3.8:** L'évaluation de la qualité d'estimation des sources  $s_1(t)$ ,  $s_2(t)$  et  $s_3(t)$  du deuxième mélange avec `nbre_iter = 25`

Les sources	SDR(dB)	SIR(dB)	SAR(dB)
S <sub>1</sub>	29.1367	29.1367	229.1618
S <sub>2</sub>	41.6922	41.6922	265.3394
S <sub>3</sub>	43.4652	43.4652	267.4173

**Tableau.3.9:** L'évaluation de la qualité d'estimation des sources  $s_1(t)$ ,  $s_2(t)$  et  $s_3(t)$  du deuxième mélange `nbre_iter = 100`

D'après les résultats mentionnés dans les tableaux 3.8 et 3.9 nous concluons que pour une meilleure estimation, il faut prendre un grand nombre d'itération si le pas du gradient est fixé. D'après les résultats des tests MOS, nous pouvons maintenant dire que nous avons une bonne estimation. Ceci implique que les signaux estimés sont presque identiques aux sources originales.

Pour affirmer une bonne qualité, il faut que les SNR dépassent les 40 dB.

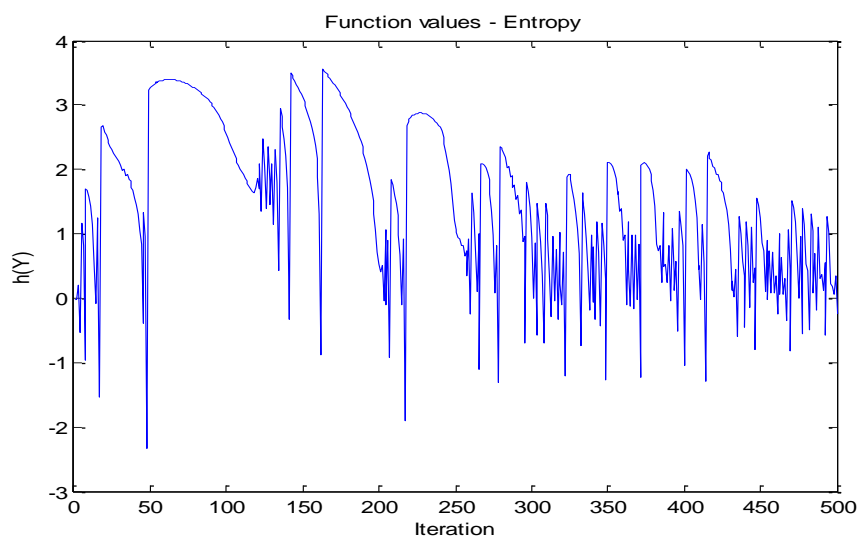
**Comparaison entre les deux approches :** Dans notre expérience nous avons utilisé deux approches différents, entropie et infomax, après les résultats obtenu nous avons remarqué que l'approche de l'entropie mieux que l'autre approche, parce que quand nous avons utilisé beaucoup de nombre d'itérations, l'approche de l'entropie très rapide par rapport de l'autre approche.

### 3.4.3. Discussion de la qualité d'estimation

La qualité d'estimation est liée à plusieurs paramètres, dans cette expérience nous allons jouer sur deux d'entre eux qui sont le pas d'adaptation et le nombre d'itération.

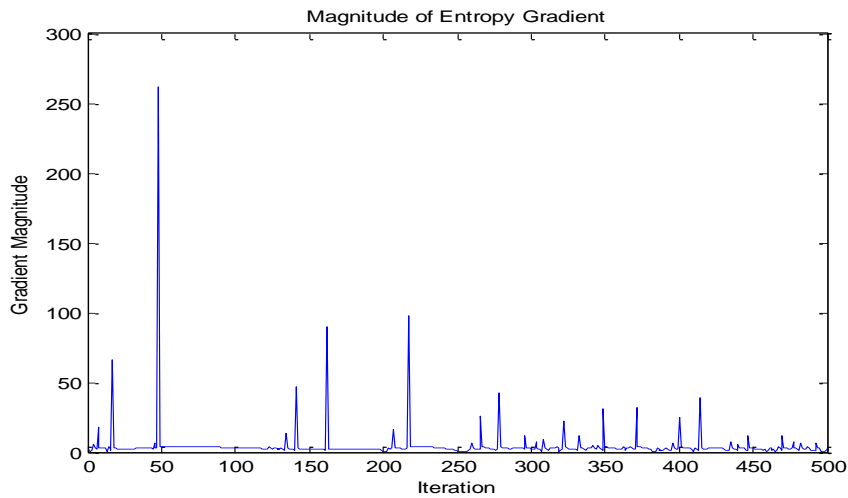
#### 3.4.3.1. Le pas d'adaptation

Dans cette expérience, nous avons attribué au pas d'adaptation la valeur 0.9, avec un nombre d'itération égal à 500. La variation de l'entropie et son gradient sont représentées sur les figure 3.12 et 3.13.



**Fig. 3.12 :** Variation de l'entropie avec  $\eta=0.9$ , N-itération=500)

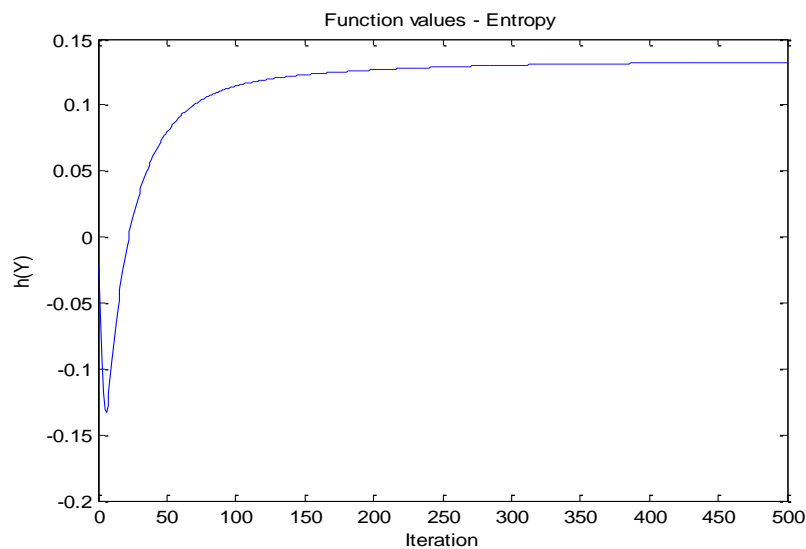




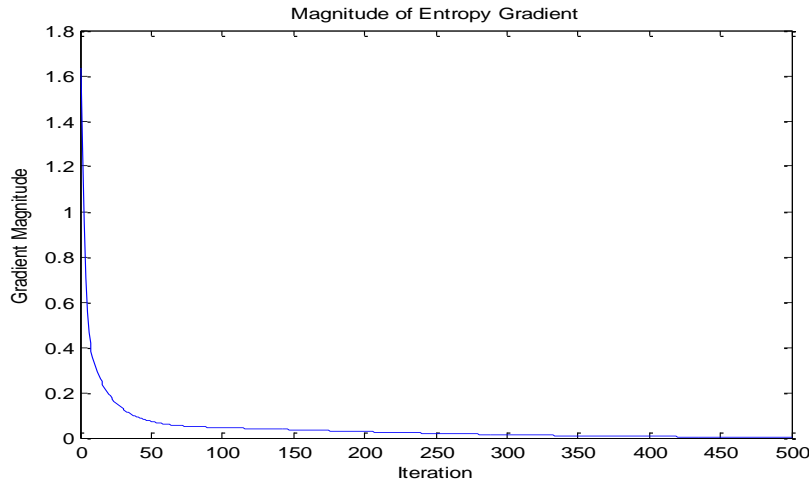
**Fig. 3.13 :** Variation du gradient de l'entropie avec  $\eta=0.9$ , N-itération=500)

Pour cette expérience, nous remarquons que l'entropie et son gradient ne se stabilisent pas. Ceci a conduit à des signaux estimés de mauvaise qualité.

Dans une deuxième expérience, nous avons changé la valeur du pas en la fixant à 0.1 tout en gardant le nombre d'itération le même (500) et nous avons obtenu les figures 3.14 et 3.15.



**Fig. 3.14 :** Variation de l'entropie avec  $\eta=0.1$ , N-itération=500)



**Fig. 3.15 :** Variation du gradient de l'entropie avec  $\eta=0.1$ , N-itération=500

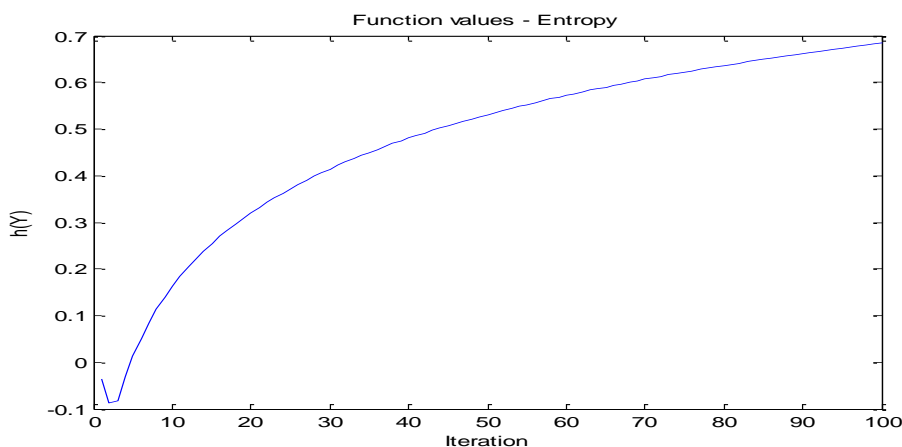
Nous remarquons que l'entropie et son gradient ont tendance à se converger vers une valeur fixe ce qui a conduit une qualité d'estimation moyenne.

### 3.4.3.2. Le nombre d'itération

Maintenant, nous agissons sur le nombre d'itération. Les figures 3.16 et 3.16 illustrent la variation de l'entropie et son gradient pour un pas d'adaptation égal à 0.25 et un nombre d'itérations égal à 100.

Nous observons que l'entropie et son gradient ne convergent pas rapidement ce qui donne une mauvaise estimation.

Pour une meilleure estimation des signaux il faut que la valeur de l'entropie prenne une valeur maximale et son gradient s'annule pour le pas et le nombre d'itérations choisis.



**Fig. 3.16 :** Variation de l'entropie avec  $\eta=0.25$ , N-itération=100

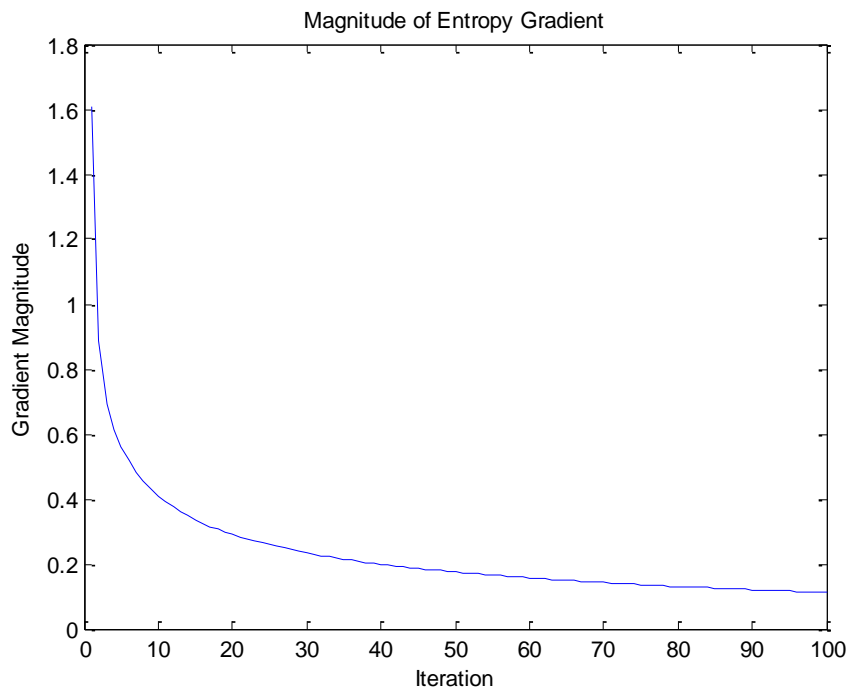


Fig. 3.17 : Variation du gradient de l'entropie avec  $\eta=0.25$ , N-itération =100

### 3.5. Conclusion

Dans ce chapitre nous avons implémenté l'algorithme ACI destiné à estimer les mélanges des signaux audio. Les résultats obtenus par la séparation des signaux indépendants contenus dans les mélanges ont montré que L'ACI se prêle robuste pour résoudre ce problème.

Tout d'abord, nous avons cherché des signaux de parole à partir d'une base de données. Premièrement, pour les lire, les afficher et les mélanger à l'aide d'une matrice de mélange, en utilisant le logiciel Matlab. Ensuite, nous avons séparé ses mélanges par la méthode ACI avec deux approches différentes : par maximisation de l'entropie et infomax.

Nous avons simulé la corrélation pour savoir s'il existait une relation entre le signal estimé et le signal original. L'entropie est l'une des techniques permettant la mesure de la non-gaussienneté de source. Pour cette raison, nous avons également simulé l'entropie et son gradient pour voir la convergence de l'algorithme ACI.

La qualité de la séparation a été évaluée par les critères MOS, SDR, SIR et SAR. Nous pouvons dire que cette dernière est meilleure si nous respectons les conditions et les hypothèses que nécessite l'ACI pour séparer. Par exemple, au plus, une des sources peut être gaussienne.

Nous avons appris qu'ils existent plusieurs algorithmes ACI parce qu'il y en a plusieurs méthodes pour mesurer l'indépendance entre les signaux.

# **Conclusion générale**

## **CONCLUSION GENERALE**

Dans ce mémoire, nous avons abordé la problématique de la séparation aveugle de sources (SAS). En particulier, nous nous sommes intéressés au problème de la séparation de mélanges linéaires instantanés de sources indépendantes.

La séparation aveugle de sources (SAS) profite de plusieurs techniques pour résoudre le problème, dans notre mémoire nous avons travaillé avec une technique appelée ACI. Cette méthode exploite la diversité spatiale et fréquentielle, ce qui la rend multidimensionnelle (multi-capteur) pour résoudre le problème de séparation aveugle de sources (SAS).

A travers le premier chapitre, nous avons noté que le traitement de la parole est une composante fondamentale des sciences de l'ingénieur. Elle présente une singularité qui la différencie fondamentalement des autres composantes du traitement de l'information qui tient sans aucun doute au rôle fascinant que joue le cerveau humain à la fois dans la production et dans la compréhension de la parole de façon pratiquement instantanée.

Le traitement de la parole doit faire face à de nombreux problèmes, entre autres le problème de mélange de signaux.

La formulation de l'ACI dans le deuxième chapitre, nous a permis de comprendre les principes des approches utilisées dans la mesure de la non gaussianité, ce qui élabore les différents algorithmes ACI pour séparer les mélanges des signaux. Au plus une des sources peut être gaussienne, c'était l'hypothèse fondamentale d'existence d'une indépendance entre les sources.

Dans le troisième chapitre, nous avons implémenté deux algorithmes ACI en MATLAB qui ont permis de séparer des signaux des paroles contenues dans des mélanges.

Les performances de ces différents algorithmes ont été évaluées par des critères objectifs et subjectifs tels que MOS, SIR, SDR et SAR.

Grace aux résultats obtenus au chapitre 3, nous concluons que l'ACI est une technique statistique très robuste pour estimer des signaux noyés dans des mélanges.

# **Bibliographie**



## *Bibliographie*

- [1] Yasmina ben Abderrahmane, « Séparation aveugle de signaux de parole utilisant les statistiques d'ordre supérieur et la décomposition, en sous-espaces », thèse de doctorat, institut national de la recherche scientifique - INRS, Université du Québec.
- [2] Amiar Samra, Ferhoune Nessrine, « Dé-bruitage du signal de parole », diplôme de master, université Larbi ben m'hidi de oum el bouaghi , 2012.
- [3] Aziza yassamine, « Modélisation AR et ARMA de la Parole pour une Vérification Robuste du Locuteur dans un Milieu Bruité en Mode Dépendant du Texte », diplôme de magister, Département d'Electronique, Université Ferhat Abbas – Sétif.
- [4] Moussa Sofiane Karoui, « Méthodes de séparation aveugle de sources et application à la télédétection spatiale. Traitement du signal et de l'image », Université Paul Sabatier-Toulouse III, 2012. Français.
- [5] Christophe d'alessandro, Christian demars, « Représentations temps-fréquence du signal de parole », LIMSI-CNRS, BP 133, 91403 Orsay Cedex, France.
- [6] Francis Cottet, « Aide-mémoire et traitement du signal », Nouvelle présentation 2005, éditions Dunod, Paris, 2000.
- [7] Jean-philippe muller, « Analyse spectrale », version juillet 2002.
- [8] J.-F. BERCHER, « Signaux aléatoire », École Supérieure d'Ingénieurs en Électrotechnique et Électronique, Novembre2001.
- [9] Abdelouahed Sara, « Analyse spectro-temporelle du signal vocal en vue de dépistage et du suivie des dysphonies chronique d'origine », université abou beker belkaid ,tlemcen,juin 2012.
- [10] Lucie BAILLY, « Interaction entre cordes vocales et bandes ventriculaires en phonation : Exploration in-vivo, modélisation physique, validation in-vitro », école doctorale de l'université du Maine, Le Mans, France, juin2009.
-



- [11] Marc-Olivier Billette, « Analyse en composantes indépendantes avec une matrice de mélange éparses », Mémoire présenté à la Faculté des études supérieures en vue de l'obtention du grade de maître ès sciences (M.Sc.) en Statistique, Université de Montréal, Canada, Juin 2013
- [12] Yann Ollivier, « Aspects de l'entropie en mathématiques », École normale supérieure de Lyon, avril 2002.
- [13] Fangchen FENG, « Séparation aveugle de source : de l'instantané au convolutif », Thèse de doctorat de l'Université Paris-Saclay préparée à Université Paris-Sud, septembre 2017
- [14] Mathieu Parvaix, « Séparation de sources audio informée par tatouage pour mélanges linéaires instantanés stationnaires », Sciences de l'ingénieur [physics], Institut National Polytechnique de Grenoble-INPG, 2010. Français.
- [15] Luiza Orosanu, « Reconnaissance de la parole pour l'aide à la communication pour les sourds et malentendants », thèse de doctorat, l'Université de Lorraine, Décembre 2015.
- [16] Thierry. Dutoit, « Introduction au Traitement Automatique de la Parole », Faculté Polytechnique de Mons, octobre 2000
- [17] BOUASLI, Salim NOUMERI Ahmed, « Compression et codage de la parole par la Transformée KLT », Université Djilali Bounaama Khemis Miliana, 2016.
-