UNIVERSITÉ DU QUÉBEC

UNIVERSITÉ DU QUÉBEC

INSTITUT **N**ATIONAL DE LA **R**ECHERCHE **S**CIENTIFIQUE – **INRS**

CENTRE **É**NERGIE, **M**ATÉRIAUX ET **T**ÉLÉCOMMUNICATIONS, **EMT**

SEPARATION AVEUGLE DE SIGNAUX DE PAROLE UTILISANT LES STATISTIQUES D'ORDRE SUPERIEUR ET LA DECOMPOSITION EN SOUS-ESPACES

THÈSE DE DOCTORAT

INSTITUT **N**ATIONAL DE LA **R**ECHERCHE **S**CIENTIFIQUE – **INRS**

CENTRE **É**NERGIE, **M**ATÉRIAUX ET **T**ÉLÉCOMMUNICATIONS, **EMT**

Yasmina BENABDERRAHMANE

UNIVERSITÉ DU QUÉBEC

INSTITUT **N**ATIONAL DE LA **R**ECHERCHE **S**CIENTIFIQUE – **INRS**

CENTRE **É**NERGIE, **M**ATÉRIAUX ET **T**ÉLÉCOMMUNICATIONS, **EMT**

SEPARATION AVEUGLE DE SIGNAUX DE PAROLE UTILISANT LES STATISTIQUES D'ORDRE SUPERIEUR ET LA DECOMPOSITION EN SOUS-ESPACES

Par :
Yasmina BENABDERRAHMANE

Thèse présentée pour l'obtention
du grade de Philosophiae Doctor (Ph.D)
en Télécommunications

a été dûment accepté par le jury d'examen constitué de :

Examinateur interne et président de jury :		Sofiène Affes
Examinateurs externes	:	Martin Bouchard, Université d'Ottawa
		Pierre Dumouchel, Centre de recherche informatique de Montréal
Directeur de recherche	:	Douglas O'Shaughnessy
Codirecteurs de recherche	:	Sid-Ahmed Selouani. Université de Moncton

© *droits réservés à Yasmina Benabderrahmane, 2011*

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 du moins de séparer une source particulière. Cette thèse est focalisée sur l'étude de la séparation aveugle de sources dans le cas de mélanges instantanés et convolutifs ; en particulier, des sources de signaux de parole.

La première partie de ce rapport de thèse est consacrée aux mélanges instantanés. Nous présentons les résultats d'une étude comparative sur les méthodes de séparation aveugle de signaux de parole. Nous avons développé une méthode utilisant les moments d'ordre supérieur : HOS (*High Order Statistics*), exploitant l'hypothèse de l'indépendance des signaux. La deuxième approche que nous avons développée est basée sur la méthode de l'analyse en composantes principales orientées (OPCA : *Oriented Principal Components Analysis*). C'est une méthode utilisant les statistiques du second ordre et qui représente une extension de l'analyse en composantes principales (PCA) visant à maximiser le rapport de puissance d'une paire de signaux. L'avantage de cette méthode par rapport aux autres techniques de second ordre est l'absence de l'étape du pré-blanchiment. L'évaluation des techniques développées a été effectuée en les comparant avec une approche de référence connue : Fast-ICA (*Fast-Independant Component Analysis*).

Dans la deuxième partie de cette thèse, nous nous placerons dans une situation plus réaliste et donc plus complexe où nous considérerons un mélange convolutif. Les expériences ont été réalisées en considérant deux types de mélange : un mélange convolutif suivant le modèle HRTF (*Head Related Transfert Function*) et un mélange convolutif anéchoïque. Les deux mélanges sont enregistrés dans des milieux non réverbérants. Nous avons appliqué une approche que nous avons appelée FOPCA. L'idée était d'étendre l'algorithme OPCA aux mélanges convolutifs en l'appliquant dans le domaine fréquentiel. Cette méthode présente l'avantage de ne pas produire

le problème de permutation de fréquences. L'évaluation a été réalisée en faisant la comparaison des résultats avec deux algorithmes de référence connus : *Degenerate Unmixing Evaluation Technique* (DUET) and *Convolutive Fast Independent Component Analysis* (C-FICA). Le cas sous-déterminé a été aussi considéré.

Toutes les techniques ont été objectivement comparées en utilisant cinq métriques différentes : le domaine temporel, le spectrogramme, le coefficient de corrélation, le rapport signal-à-interférence : SIR, ainsi que le PESQ (*Perceptual Evaluation Signal Quality*). Pour des fins de comparaison subjective, nous avons utilisé une méthode basée sur la note d'opinion moyenne : le MOS. Les résultats des expériences ont été réalisés en utilisant les bases de données TIMIT, Noizeus ainsi que d'autres enregistrements plus longs des signaux de parole.

Mots clés : Séparation Aveugle de Sources, signaux de parole, Statistiques du Second-Ordre, Statistiques d'Ordre Supérieur, mélanges instantanés, mélanges convolutifs.

Yasmina Benabderrahmane Douglas O'Shaughnessy

Dédicaces

« À la mémoire de mes parents, mon très cher père Mohamed et ma très chère mère Lahouaria : Paix soit en leur âme, et que le paradis soit leur dernière demeure. Je tiens à leur exprimer tout mon profond **respect et mon infinie gratitude »**

À mes trois frères Larbi, Salim et Nour-Eddine que j'adore, leurs épouses, à tous mes neveux et nièces, à mon époux Saïd et enfin à mes deux filles chéries Sarah et Farida qui sont la lumière de ma vie.

Citation

«La théorie, c'est quand on sait tout et que rien ne fonctionne. La pratique, c'est quand tout fonctionne et que personne ne sait pourquoi. Ici, nous avons réuni théorie et pratique : Rien ne fonctionne... et personne ne sait pourquoi!»

Albert Einstein

Remerciements

En premier lieu, je tiens à remercier tout particulièrement mes deux professeurs Douglas O'Shaughnessy et Sid Ahmed Selouani de m'avoir acceptée au sein de leur équipe. Leurs qualités scientifiques et humaines ainsi que leur expérience dans le domaine de la recherche ont été de précieux atouts durant ces quatre années de thèse. Leurs compétences, leurs conseils et leurs idées ont été essentiels dans l'aboutissement de ce travail. Je leur en témoigne toute ma reconnaissance.

En outre, Je tiens à exprimer toute ma gratitude en mon nom et aux noms de l'ensemble des étudiants de l'INRS-EMT à Mr Douglas O'Shaughnessy, comme directeur des programmes pour sa gentillesse, son respect pour les étudiants son habilité à aplanir nos soucis et trouver toujours solution à tout problème.

Je tiens à remercier les membres du jury, Messieurs, Martin Bouchard de l'Université d'Ottawa et Pierre Dumouchel du Centre de recherche informatique de Montréal, pour avoir accepté d'évaluer ce travail, ainsi que Mr Sofiane Affès pour m'avoir fait l'honneur de présider ce jury.

Mes remerciements vont également à Mr Jacob Benesty de m'avoir permis d'utiliser la salle d'enregistrement disponible au laboratoire INRS-Télécom, et à Mehrez Souden, d'avoir pris le temps de m'expliquer le fonctionnement des différents appareils. J'exprime par ailleurs toute ma sympathie à tous les enseignants et les étudiants chercheurs que j'ai côtoyés à l'INRS-Télécommunications et qui ont participé à la bonne ambiance de travail et d'enthousiasme dans nos diverses collaborations, particulièrement : D. Hammou, A. Djaiz pour son support moral, A. M. Habib, Imèn, Marcelo, Francisco, M. Djamah, M. Bensalah, J. P. E. Aquije qui a partcipé à l'élaboration de l'interface ainsi qu'à L. Abed pour son support technique pour le logiciel Latex et toutes mes excuses à tous ceux que je n'ai pas pu nommer.

Je veux également remercier Emilia Moldovan, pour ses précieux conseils et suggestions concernant la vie professionnelle et personnelle et pour le temps que nous

avons passé ensemble à rigoler de tout et de rien et surtout pour son support moral pendant les moments difficiles par lesquels je suis passée.

Je remercie également tous mes amis et plus particulièrement : Krimo mon quatrième frère et sa chère épouse Lynda, les premiers que j'ai rencontrés à mon arrivée ici au Canada, Lenda qui m'appelait souvent et m'encourageait, en me répétant sans cesse que j'étais capable d'aller jusqu'au bout, Zoubida, Leila, Caroline, Habib, El Hadi B. et El Hadi T., Lamia et mon cousin Sid Ahmed, sans oublier mes amies d'Oran, Fatima et Khédidja.

Je veux également remercier Nathalie Aguiar, au delà de son professionnalisme, elle a installé au sein du centre INRS-EMT, un environnement positif et un climat chaleureux et fort agréable, toujours disponible avec le sourire.

Mes remerciements également à Sylvain Fauvel pour son support informatique ainsi qu'à Hélène Sabourin pour son professionnalisme.

Enfin, il me serait impossible de terminer ces remerciements sans adresser une pensée chaleureuse à mes très chers parents auxquels j'exprime toute ma reconnaissance et leurs dédie cette thèse. A mes trois frères : Larbi, Salim et Nour-Eddine qui m'ont toujours encouragée à aller toujours plus loin dans mes études avec une très grande fierté.

Et un simple merci à mon époux Saïd, et mes filles adorées Sarah et Farida qui m'ont soutenu tout au long de mes études pour répondre aux attentes des chercheurs et ont été ma source de motivation continuelle.

Table des matières

1 Introduction générale **1**
- 1.1 Généralité et contexte . . . 1
- 1.2 Objectifs et méthodologie . . . 3
- 1.3 Contributions . . . 5

2 La Séparation de Sources : Problématique et état de l'art **7**
- 2.1 Problématique . . . 7
- 2.2 État de l'art et recherches dans le domaine de la séparation de sources 9
  - 2.2.1 Application au traitement de la parole . . . 13
- 2.3 Principe . . . 15
  - 2.3.1 Nature du mélange . . . 16
  - 2.3.2 Mélange et modèle linéaire instantané . . . 16
  - 2.3.3 Détermination du mélange . . . 22
  - 2.3.4 Niveau de connaissance a priori sur les sources et le mélange . 22
- 2.4 Hypothèses nécessaires . . . 22
- 2.5 Principe de la séparation idéale . . . 24
- 2.6 Choix entre domaine fréquentiel et domaine temporel . . . 24
- 2.7 Ambiguïté de séparation . . . 26
  - 2.7.1 Ambiguïté de permutation . . . 26
  - 2.7.2 Ambiguïté d'échelle . . . 26
  - 2.7.3 Ambiguïté de filtrage . . . 27
  - 2.7.4 Ambiguïté complète . . . 27
- 2.8 Évaluation de la séparation de sources . . . 27
  - 2.8.1 Critères de comparaison . . . 28
  - 2.8.2 Critères subjectifs de qualité . . . 28
  - 2.8.3 Critères objectifs de qualité . . . 29
- 2.9 Conclusion . . . 30

3 Méthodes de décomposition en sous-espaces appliquées à la séparation de sources **32**
- 3.1 Introduction . . . 32
- 3.2 Analyse en composantes principales : PCA . . . 33
  - 3.2.1 Recherche des composantes principales . . . 34
  - 3.2.2 Réalisation . . . 35

3.3 Analyse en composantes principales orientée (OPCA) et la séparation aveugle de sources 38
3.3.1 Définition 38
3.3.2 Algorithme OPCA de K. I. Diamantaras et Th. Papadimitriou 39
3.4 Analyse en Composantes Indépendantes : ICA 40
3.4.1 Définition 42
3.4.2 Indépendance 42
3.4.3 Principales approches du problème ICA 42
3.5 Algorithme ICA 49
3.5.1 Prétraitement pour l'ICA 50
3.6 Principe de base de l'algorithme Fast-ICA 52
3.6.1 L'algorithme C-FICA 53
3.7 Quelques différences et similarités entre PCA et ICA 54
3.8 Comparaison entre PCA et OPCA 54
3.9 Conclusion 55

4 Séparation aveugle de sources pour les mélanges instantanés 57
4.1 Introduction 57
4.2 Principe de la séparation aveugle de sources 58
4.2.1 Critère de séparation 58
4.2.2 Modèle initial 59
4.2.3 Architecture du réseau 59
4.3 Séparation aveugle de deux sources utilisant les statistiques d'ordre supérieur 60
4.3.1 Règle d'adaptation 62
4.3.2 Estimation des pas d'adaptation 64
4.4 Séparation aveugle de deux sources utilisant l'analyse en composantes principales orientée (OPCA) 65
4.5 Expériences et résultats 70
4.5.1 Méthode Fast-ICA 70
4.5.2 Méthode utilisant les statistiques d'ordre supérieur : HOS . . . 72
4.5.3 La méthode OPCA 74
4.6 Critères d'évaluation de la qualité du signal de parole 74
4.6.1 Le PESQ 75
4.6.2 Le MOS : Note d'opinion moyenne 76
4.6.3 Le rapport signal-interférence : SIR 77
4.7 Rehaussement d'un signal de parole en présence d'un bruit avec la méthode OPCA 79
4.7.1 Bruits enregistrés à partir de la base de données NOISEX : bruit blanc 79
4.7.2 Signaux de parole bruités enregistrés à partir de la base de données : NOIZEUS 81
4.8 Conclusion 83

5 Séparation aveugle de sources pour les mélanges convolutifs 86
5.1 Introduction 86
5.2 Problématique et éléments classiques de séparation de sources dans le cas convolutif 87
5.2.1 Formulation du problème de séparation de sources avec mélange convolutif 87
5.2.2 Relation avec le système de mélange instantané 88
5.3 Différents modèles de mélanges convolutifs 89
5.3.1 Modèle de filtres « Head Related Transfer Function » : HRTF 90
5.3.2 Modèle de mélange anéchoïque 91
5.4 Modèles de séparation dans le domaine fréquentiel 92
5.5 Méthodes de référence de séparation aveugle de sources avec mélange convolutif 93
5.5.1 La méthode C-FICA 93
5.5.2 La méthode DUET 95
5.5.3 Séparation Aveugle de Mélanges Convolutifs par la méthode FOPCA 97
5.6 Expériences et résultats 100
5.6.1 Cas d'un mélange de deux signaux de parole 100
5.6.2 Cas de la séparation aveugle de sources sous-déterminée 123
5.7 Séparation aveugle de sources en milieu de réverbération : enregistrement réel 127
5.8 Discussion 128
5.9 Conclusion 131

6 Conclusion générale 132

A Rappels sur le kurtosis 151

B Algorithmes de séparation aveugle de sources :SOBI, JADE, COM1, COM2 165
B.1 L'algorithme SOBI 165
B.2 L'algorithme COM2 166
B.3 L'algorithme COM1 167
B.4 L'algorithme JADE 167

C Formulaire de l'Analyse en Composantes Principales 168

D Décomposition et maximisation de la variance 170

E Méthode par Déflation 172

F Influence du blanchiment sur les performances de la séparation 173
F.1 Blanchiment spatial d'ordre deux 173
F.2 Détermination de la matrice de blanchiment 173
F.3 Influence du blanchiment sur les performances de la séparation 175

Table des figures

2.1 Schéma synoptique de l'algorithme de séparation aveugle de sources dans le domaine Temps-Fréquence. 13
2.2 Problème multiple concernant le mélange convolutif. 18
2.3 Principe de la séparation de sources mélangées convolutivement. . . . 24

3.1 Analyse en composantes principales : Nuage de données à deux dimensions [94]. 33
3.2 Plan de projection [95]. 35
3.3 Problème du "*cocktail party*". 41

4.1 Système de séparation de sources. 58
4.2 Architecture du réseau de la séparation de sources de Herault et Jutten. 60
4.3 Schéma synoptique d'une séparation de deux sources 61
4.4 Schéma synoptique de la séparation aveugle de source utilisant la méthode OPCA appliquée à des signaux aléatoires avec une modulation multi-niveaux en amplitude : PAM [119]. 66
4.5 Signaux de parole originaux, signal de parole de l'homme (en haut) et signal de parole de la femme (en bas). 71
4.6 Signaux mélangés instantanément. 71
4.7 Signaux séparés avec la méthode Fast-ICA. 71
4.8 Spectrogramme des signaux de parole originaux. 72
4.9 Spectrogramme des signaux mélangés instantanément. 72
4.10 Spectrogramme des signaux estimés par la méthode Fast-ICA. 72
4.11 Signaux de parole estimés par la méthode HOS. 73
4.12 Spectrogramme des signaux estimés par la méthode HOS. 73
4.13 Évolution des coefficients des mélanges c_{ij}. 73
4.14 Signaux de parole estimés par la méthode OPCA. 74
4.15 Spectrogramme des signaux estimés par la méthode OPCA. 74
4.16 Schéma synoptique du calcul de la mesure du PESQ [123]. 75
4.17 Signaux originaux : parole et bruit blanc. 79
4.18 Signaux de parole et bruit blanc mélangés. 79
4.19 Signaux estimés par la méthode OPCA : parole et bruit blanc. 79
4.20 Spectrogrammes des signaux originaux : parole et bruit blanc. 80
4.21 Spectrogrammes des signaux mélangés. 80

4.22 Spectrogrammes des signaux estimés par méthode OPCA : parole et bruit blanc. 80
4.23 Signal de parole original. 81
4.24 Signaux de parole et du train mélangés avec des rapports signal-à-bruit de 0 et 5 dB. 81
4.25 Signaux de parole et du bruit du train de banlieue estimés par la méthode OPCA. 82
4.26 Spectrogramme du signal de parole original. 82
4.27 Spectrogramme des signaux de parole et bruit du train mélangés avec un RSB de 0 et 5 dB. 82
4.28 Spectrogramme des signaux de parole et du train estimés par la méthode OPCA. 83

5.1 Modèle de mélange convolutif [125]. 88
5.2 Schéma de réalisation d'un seul mélange linéaire convolutif de sources [126]. 88
5.3 Schéma de principe de la reproduction des sources au niveau des oreilles du mannequin. 90
5.4 Représentation des signaux sources et des signaux au niveau des oreilles [129]. 91
5.5 Modèle de mélange anéchoïque [125]. 92
5.6 Système de séparation de sources dans le domaine fréquentiel. 92
5.7 Schéma synoptique de la BSS utilisant la méthode FOPCA dans le domaine fréquentiel pour les mélanges convolutifs. 98
5.8 Le modèle convolutif (HRTF) : positions des sources définies par les angles 30 et - 80 degrés par rapport à la tête d'un mannequin. 101
5.9 Signaux de parole originaux de la base de données TIMIT. 102
5.10 Signaux mélangés par convolution avec le modèle HRTF. 103
5.11 Signaux estimés par la méthode FOPCA. 103
5.12 Spectrogrammes des signaux de parole originaux 103
5.13 Spectrogrammes des signaux de parole mélangés convolutivement avec le modèle HRTF. 104
5.14 Spectrogrammes des signaux de parole estimés par la méthode FOPCA. 104
5.15 Signaux originaux : signal "speech$_1$" et signal "speech$_2$ ". 104
5.16 Signaux mélangés par convolution avec le modèle HRTF. 105
5.17 Signaux estimés par la méthode FOPCA pour les signalux "speech$_1$" et "speech$_2$". 105
5.18 Spectrogramme des signaux originaux, "speech$_1$" et "speech$_2$". 105
5.19 Spectrogramme des signaux "speech$_1$" et "speech$_2$" mélangés par convolution avec le modèle HRTF. 106
5.20 Spectrogramme des signaux estimés par la méthode FOPCA pour "speech$_1$" et "speech$_2$". 106
5.21 Signaux de parole originaux : signal original d'un locuteur s_2 et signal original d'une locutrice s_1. 106
5.22 Signaux s_1 et s_2 mélangés convolutivement utilisant le modèle HRTF. 107
5.23 Signaux estimés $\hat{s}_1$ et $\hat{s}_2$ par la méthode FOPCA. 107

5.24 Spectrogramme des signaux de parole originaux s_1 et s_2. 107
5.25 Spectrogramme des signaux s_1 et s_2 mélangés par convolution utilisant le modèle HRTF 108
5.26 Spectrogramme des signaux $\hat{s}_1$ et $\hat{s}_2$ estimés par la méthode FOPCA . 108
5.27 Signaux de TIMIT estimés par la méthode C-FICA. 109
5.28 Spectrogrammes des signaux de parole de TIMIT estimés par la méthode C-FICA. 109
5.29 Histogramme représentant les signaux mélangés suivant le modèle HRTF.110
5.30 Signaux estimés par la méthode DUET. 111
5.31 Spectrogrammes des signaux de parole estimés par la méthode DUET. 111
5.32 Signaux mélangés par convolution suivant le modèle anéchoïque. . . . 111
5.33 Spectrogramme des signaux mélangés par convolution suivant le modèle anéchoïque. 112
5.34 Signaux de TIMIT estimés par la méthode FOPCA dans le domaine temporel. 112
5.35 Spectrogramme des signaux de TIMIT estimés par la méthode FOPCA. 112
5.36 Signaux de TIMIT estimés par la méthode C-FICA. 113
5.37 Spectrogramme des signaux de TIMIT estimés par la méthode C-FICA. 113
5.38 Histogramme dans le cas d'un mélange convolutif anéchoïque. 113
5.39 Signaux de TIMIT estimés par la méthode DUET dans le domaine temporel. 114
5.40 Spectrogramme des Signaux estimés par la méthode DUET. 114
5.41 Comparaison des rapports SIR avant et après séparation par la méthode FOPCA (Signal de parole prononcé par l'homme : TIMIT). 118
5.42 Comparaison des rapports SIR avant et après séparation par la méthode FOPCA (Signal de parole prononcé par la femme : TIMIT). 119
5.43 Comparaison des rapports SIR avant et après séparation par la méthode DUET (Signal de parole prononcé par l'homme : TIMIT). 119
5.44 Comparaison des rapports SIR avant et après séparation par la méthode DUET (Signal de parole prononcé par la femme : TIMIT). 119
5.45 Comparaison des rapports SIR avant et après séparation par la méthode C-FICA (Signal de parole prononcé par l'homme : TIMIT). 120
5.46 Comparaison des rapports SIR avant et après séparation par la méthode C-FICA (Signal de parole prononcé par la femme : TIMIT). 120
5.47 Comparaison des rapports SIR avant et après séparation par la méthode FOPCA (Signal de parole prononcé par l'homme s_1 et enregistré par D. Nion et al.). 120
5.48 Comparaison des rapports SIR avant et après séparation par la méthode FOPCA (Signal de parole prononcé par la femme s_2 et enregistré par D. Nion et al.). 121
5.49 Salle d'enregistrement en milieu réverbérant : quatre hauts parleurs et quatre microphones. 127
5.50 Salle d'enregistrement en milieu réverbérant : deux microphones dans les oreilles du Mannequin. 128

E.1 Déflation classique. Extraction de 3 signaux $\{s_p\}_{p=1}^{3}$ ayant typiquement (mais pas forcément) la même modulation. La déflation conventionnelle estime les signaux d'entrée un par un. 172

F.1 Schéma général de la séparation de sources incluant l'étape de blanchiment. 175

Liste des tableaux

4.1 Comparaison du PESQ des méthodes Fast-ICA, HOS et OPCA. . . . 76
4.2 Comparaison du MOS pour les méthodes Fast-ICA, HOS et OPCA. . 77
4.3 Comparaison du SIR des méthodes Fast-ICA, HOS et OPCA (calcul du SIR fourni par la "toolbox C-FICA v1-1"). 78
4.4 Comparaison du SIR des méthodes Fast-ICA, HOS et OPCA (calcul du SIR fourni par J. Thomas et al. [139]). 78
4.5 Résultats du SIR pour différents rapports signal-à-bruit. 83

5.1 Comparaison du PESQ pour les méthodes C-FICA, DUET et FOPCA 114
5.2 Comparaison du SIR des méthodes C-FICA, DUET et FOPCA (calcul du SIR fourni par J. Thomas et al. [139]). 115
5.3 Comparaison du Gain entre le SIR de sortie et d'entrée pour les méthodes C-FICA, DUET et FOPCA (TIMIT). 116
5.4 Comparaison du SIR des méthodes C-FICA, DUET et FOPCA (calcul du SIR fourni par la "toolbox C-FICA v1-1"). 116
5.5 Comparaison du gain entre le SIR de sortie et d'entrée pour les méthodes C-FICA, DUET et FOPCA 116
5.6 Comparaison des SIR des méthodes C-FICA, DUET et FOPCA avec différents angles pour les réponses impulsionnelles des filtres du modèle HRTF. 117
5.7 Comparaison du SIR pour les méthodes C-FICA, DUET et FOPCA pour les signaux " speech1 " et " speech2 ". 117
5.8 Résultats du SIRst pour les signaux de parole recueillis par D. Nion et al. pour la méthode FOPCA. 118
5.9 Coefficient de corrélation entre les signaux originaux et estimés pour les deux types de mélange : HRTF et Anéchoïque. 122
5.10 Comparaison des MOS pour les méthodes C-FICA, DUET et FOPCA. 123
5.11 Comparaison des SIRst pour les approches C-FICA, DUET et FOPCA avec différents angles pour les réponses impulsionnelles des filtres du modèle HRTF : cas sous-déterminé. 125
5.12 Comparaison des SIR pour les approches C-FICA, DUET et FOPCA utilisant le modèle de mélange anéchoïque : cas sous-déterminé (trois signaux de parole et deux microphones). 126

5.13 Comparaison du PESQ pour les méthodes C-FICA, DUET et FOPCA utilisant le modèle anéchoïque : cas sous-déterminé (trois signaux de parole et deux microphones). 126
5.14 Comparaison des SIR avant et après traitement avec la méthode FOPCA. 128

Liste de Sigles et abréviations

- A : matrice de mélange instantané composée de coefficients scalaires réels a_{ij}.
- BSS : séparation Aveugle de sources (Blind Sources Separation).
- C : matrice de covariance.
- D : matrice diagonale avec les éléments $diag(x_1, \ldots, x_M)$: matrice diagonale dont les termes diagonaux sont les x_i.
- $E\{y(n)\}$: espérance du signal $y(n)$.
- Filtre AR : filtre Auto Régressif, aussi appelé Auto-Regressive en anglais.
- Filtre ARMA : Filtre Auto Regressive, Moving Average.
- H : entropie.
- h : coefficient d'un filtre.
- I : matrice identité.
- ICA : Analyse en Composantes Indépendantes (Independent Component Analysis).
- i.i.d. : indépendant et identiquement distribué.
- $kurtx(n)$: kurtosis du signal x à l'instant n.
- M : nombre d'observations disponibles.
- $\overline{M}$: matrice centrée.
- $\tilde{M}$: matrice centrée réduite.
- MA : Moving Average.
- $n(t)$: bruit.
- N : nombre de sources présentes dans les observations.
- OPCA : Analyse en Composantes Principales Orientées (Oriented Principal Component Analysis).
- P : matrice de permutation contenant seulement un « 1 » dans chaque colonne et chaque ligne.
- PCA : Analyse en Composantes Principales (Principal Component Analysis).
- $p_i(s_i(t))$: densité de probabilité.
- RSB : Rapport signal à bruit.
- $R_x(n)$: *matrice d'auto-corrélation du vecteur x à l'instant n.*
- $r_{xy}(n)$: vecteur d'intercorrélation des signaux x et y à l'instant n.
- $s(t)$ resp. $s(n)$ vecteur colonne composé de signaux sources $s_i(t)$ resp. $s_i(n)$.
- $\hat{s}(t)$ vecteur colonne composé de signaux sources estimés $\hat{s}_i(t)$.
- SIR : Rapport signal à interférences (Signal to Interference Ratio).
- $S_x(t, f)$: densité de spectre de puissance du signal x(t).
- T : nombre d'échantillons de chaque observation.
- U^T : *matrice transposée.*
- W : matrice de séparation.
- $x(t)$ resp. $x(n)$: vecteur colonne composé de signaux d'observation $x_i(t)$ resp. $x_i(n)$.

- *X(f)* : transformée de Fourier du vecteur *x(n).*
- *X(t, f) : Transformée de Fourier à court terme.*
- *X(Z)* : transformée en Z du vecteur *x(n).*
- λ *:* valeurs propres d'une matrice.
- $\rho_{xy}(n)$: coefficient de corrélation.
- σ_{xy} : écart type.
- * : opérateur de convolution.
- μ : pas d'adaptation.
- ∇ : gradient.
- $\Pi p_i(s_i(t))$: produit des densités de probabilité.

Chapitre 1

Introduction générale

1.1 Généralité et contexte

Le traitement de la parole représente un élément très important dans les sciences de l'ingénieur. Il se situe entre le traitement du langage (c'est-à-dire du traitement de données symboliques) et le traitement du signal numérique. Cette discipline scientifique a connu depuis les années 60 une expansion fulgurante, liée au développement des moyens et des techniques de télécommunications. Comme la parole représente un vecteur d'information dans notre société humaine, elle occupe une place privilégiée, d'où l'importance particulière du traitement de la parole.

L'extraordinaire singularité de cette science, qui la différencie fondamentalement des autres composantes du traitement de l'information, 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 et à l'étendue des fonctions qu'il met, inconsciemment, en œuvre pour y parvenir de façon pratiquement instantanée.

Le traitement de la parole fait partie de la discipline du traitement de signal. Le signal est le support de l'information et le moyen de communication entre les hommes et les machines. Il est, dans la plupart des cas, mesuré par des capteurs et issu de notre environnement physique réel : ondes sonores, vibrations sismiques, images visuelles, etc.

Le traitement de signal connaît un grand développement qui continue jusqu'à l'heure actuelle. Tous les outils mathématiques, algorithmiques calculatoires utilisés ont conduit à une révolution technologique dans plusieurs domaines d'applications : reconnaissance de la parole, rehaussement d'images, télécommunication,...

Le traitement de signal a pour vocation d'extraire l'information véhiculée par le signal. Ceci s'appuie sur la connaissance soit du signal, soit du système, soit des deux. Les systèmes de traitement audio voient en leur entrée, selon le cas, un ou plusieurs si-

gnaux. Cependant, l'environnement externe impose à ces derniers, une transformation au cours de leur propagation telle qu'une superposition avec d'autres signaux de l'environnement. Un cas typique de tels systèmes se trouve dans les organes auditifs de l'être humain. En effet, l'être humain est capable de focaliser, dans le mélange provenant de son environnement, sur une des sources de signal qu'il reçoit à son oreille.

Dans cet objectif, plusieurs travaux ont été menés dans ce domaine et plus précisément avec la terminologie de séparation aveugle de sources : "*Blind Source Separation : BSS*" [1], [2], [3], [4].

Depuis plusieurs années, la séparation de sources est un sujet de recherche particulièrement actif. Cet engouement peut s'expliquer par le large spectre d'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.

La grande majorité des technologies modernes, dans la plupart des disciplines, utilisent des capteurs pour mesurer des grandeurs utiles pour l'étude ou le contrôle d'un phénomène physique. Les microphones par exemple captent les ondes sonores se propageant dans l'air, les antennes radio mesurent l'évolution d'un champ électromagnétique, les détecteurs photosensibles numérisent l'image obtenue par un système optique, enfin les électrodes placées sur le crâne d'un patient permettent d'enregistrer le champ électrique associé à son activité cérébrale.

En pratique, malheureusement, les signaux obtenus grâce à ces capteurs sont souvent des mélanges de plusieurs contributions issues de signaux originels 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.

La séparation aveugle de sources consiste à estimer les signaux sources (ou éventuellement leurs contributions) à partir des mélanges obtenus sur chaque capteur. Le corps humain possède un organe réalisant ce type d'opération : le cerveau. Celui-ci peut en effet se focaliser sur un phénomène extérieur particulier détecté par les sens. On peut ainsi écouter une personne s'exprimant dans une salle bruitée où d'autres personnes parlent en même temps (un hall de conférences par exemple). La zone du cerveau dédiée à la vision peut aussi fixer notre attention sur une zone précise du champ visuel.

La séparation de sources, en tant que discipline scientifique est relativement récente et s'est inspirée, à son commencement, du cerveau humain et de son organisation en réseaux neuronaux [5]. Elle s'est aussi basée sur des résultats obtenus en déconvolution

aveugle, en adaptant un critère particulier, le kurtosis ou auto-cumulant d'ordre quatre qui est aujourd'hui encore l'un des plus utilisés. La séparation aveugle de sources a ensuite connu durant deux décennies des évolutions constantes, et a trouvé un vaste champ d'applications, ce qui en fait actuellement un domaine de recherche majeur en traitement du signal. On utilise les techniques de séparation aveugle de sources dans des domaines aussi divers que le traitement de sources audio, les télécommunications multicanaux, l'analyse de signaux biomédicaux, l'imagerie astrophysique multi-spectrale, le traitement de données géophysiques, la détection et la localisation radar et bien d'autres [6], [7], [8], [9], [10].

Comme le modèle du signal de la séparation aveugle de sources est sensiblement le même que celui d'un système de transmission MIMO (*Multiple-Input Multiple-Output*), son application aux systèmes MIMO est très prometteuse. En effet, le canal MIMO et le traitement de réception dans un système MIMO sont équivalents, respectivement, aux mélange et matrice de séparation dans la séparation de sources.

Depuis deux décades ce problème a suscité un grand intérêt et a engendré de nombreuses contributions. Malgré des avancés considérables dans ce domaine, l'application dans la pratique de cette technique est loin d'être satisfaisante et beaucoup de travail reste à effectuer pour améliorer celle-ci.

1.2 Objectifs et méthodologie

Cette thèse est focalisée sur l'étude de la séparation aveugle de sources dans le cas de mélanges instantanés et convolutifs ; en particulier, des sources de signaux de parole. Le problème de séparation de sources consiste à recouvrer des signaux utiles provenant de différentes sources, qui ont été mélangées c.-à-d. : filtrées et additionnées en se propageant vers un ensemble de capteurs. Ce problème se rencontre dans plusieurs applications.

Nous montrerons qu'il est possible d'effectuer la séparation de signaux de parole mélangés instantanément en utilisant les statistiques d'ordre deux et les statistiques d'ordre supérieur à deux. Pour les signaux mélangés convolutivement, nous utiliserons seulement les statistiques d'ordre deux. Notre méthodologie rend possible la séparation des mélanges convolutifs dans le domaine de Fourier, fréquence par fréquence. L'approche aveugle de la séparation, à laquelle nous nous intéressons, offre en outre l'avantage de ne requérir que des hypothèses faibles sur le mélange : en dehors de sa structure générale, souvent supposée linéaire, aucun paramètre n'est supposé connu a priori. La suite de ce manuscrit sera organisée comme suit :

- Dans le deuxième chapitre, nous exposerons la problématique et l'état de l'art

sur les méthodes de séparation de sources, en présentant les principales idées utilisées pour résoudre ce problème ; nous évoquerons les méthodes basées sur les statistiques d'ordre deux et les statistiques d'ordre supérieur à deux. Nous parlerons aussi des algorithmes de séparation les plus connus.

• Dans le troisième chapitre, nous présenterons le problème de séparation de sources en expliquant son principe, en donnant son modèle mathématique et en citant quelques exemples d'application de la séparation de sources. Nous détaillerons tous les calculs mathématiques des méthodes utilisant les statistiques du second ordre : l'analyse en composantes principales (PCA) et l'analyse en composantes principales orientées (OPCA) et nous présenterons aussi tout le détail des calculs mathématiques des méthodes utilisant les statistiques d'ordre supérieur : L'analyse en composantes indépendantes (ICA).

• Dans le quatrième chapitre, nous présenterons notre contribution à la séparation de sources. Nous traiterons le cas des mélanges instantanés sans bruit additif, où nous décrirons nos approches et verrons comment nous obtenons la séparation des signaux de parole. Nos approches sont : HOS (High Order Statistics), méthode utilisant les statistiques d'ordre supérieur, et OPCA, méthode utilisant les statistiques du second ordre. Nous nous focaliserons sur l'évaluation de ces méthodes en les comparant avec une approche de référence connue : Fast-ICA.

• Dans le cinquième chapitre, nous nous placerons dans une situation plus complexe où nous considérerons le mélange convolutif. Deux types de mélange convolutifs non réverbérants sont présentés : un mélange convolutif suivant le modèle HRTF (Head Related Transfert Function), un mélange convolutif anéchoïque. Les deux mélanges sont enregistrés dans des milieux non réverbérants. Ces deux mélanges sont utilisés pour toutes les expériences. Pour accomplir la séparation dans ce contexte, nous appliquerons la méthode FOPCA et nous ferons son évaluation en faisant la comparaison des résultats avec deux algorithmes de référence connus : C-FICA et DUET. Cette méthode a l'avantage de ne pas produire de permutation de fréquences, qui est habituellement présente dans les méthodes fréquentielles. Nous avons aussi considéré le cas sous-déterminé : trois signaux de parole et deux microphones. À partir de la modélisation mathématique de cette situation, la vraisemblance des paramètres de ce modèle est formulée et un algorithme itératif est utilisé pour son estimation. Nous évaluerons les résultats des différentes expériences en utilisant des mesures de performance.

• Nous terminerons ce manuscrit par une conclusion générale dans laquelle sont ré-

sumés les travaux présentés dans cette thèse et sont abordées les perspectives qu'ouvrent ces derniers pour de futurs travaux de recherches prometteurs.

1.3 Contributions

Les principales contributions à la recherche de ce travail de thèse portent sur la séparation de signaux de parole autant dans les systèmes de mélanges instantanés que dans les systèmes de mélanges convolutifs. Nous avons effectué des tests des méthodes sur deux types de mélange convolutifs différents. Ces travaux nous ont permis de soumettre un article au journal : *EURASIP Journal on Audio, Speech, and Music Processing*. L'article est intitulé : "Blind Speech Separation Using Frequency Oriented PCA Method for Convolutive Mixtures". Nous avons aussi participé à la conférence *Interspeech* qui a eu lieu au Japon en 2010 et nous participerons aussi à *Interspeech 2011* qui aura lieu à Florence (Italie). C'est une conférence de renommée concernant le traitement de la parole. Cette participation nous a value d'être classés parmi les cinq meilleurs articles.

Une contribution importante à souligner est l'écriture d'un chapitre, comme co-auteure dont le titre est : "Convolutive Blind Separation Speech Mixtures using Auditory based Subband Model", dans un recueil qui regroupe les meilleurs articles en : « Electrical Engineering and Applied Computing », édité par *Springer* et à paraître le 29 juin 2011.

Une autre contribution importante concerne nos travaux dans le cas d'un mélange instantané. Nous avons fait une étude comparative entre les méthodes utilisant les statistiques de second ordre et les statistiques d'ordre supérieur, pour la séparation aveugle de signaux de parole. Nous avons décrit trois méthodes de séparation aveugle de signaux de parole, ICA, HOS et OPCA. Nous avons ensuite comparé leurs performances en utilisant des méthodes subjective (telle que le MOS : *Mean Opinion Score*) et objective (tel que le calcul du rapport SIR : *Signal-Interference Ratio*), et la méthode OPCA qui a été appliquée pour la première fois dans la séparation aveugle de signaux de parole, s'est avérée très performante. Notons que l'approche HOS a aussi donné des résultats satisfaisants. L'article : «A Comparative Study of Blind Speech Separation using subspace methods and Higher Order Statistics», nous a permis de participer à la conférence SIP 2009 (2009 International Symposium on Signal Processing, Image Processing and Pattern Recognition). Cette conférence est la plus grande axée sur les différents aspects des progrès en traitement du signal, traitement d'image et reconnaissance de formes. De plus cet article a été publié par LNCS/CCIS dans Lecture Notes on Computer Science, chez Springer.

OPCA est une méthode statistique du second ordre et représente l'extension de

la méthode de l'analyse en composantes principales qui vise à maximiser le rapport de puissance d'une paire de signaux. Cette approche a l'avantage de ne pas nécessiter de prétraitement (blanchiment). Les résultats de cette méthode dans le cas instantané ont été présentés à la conférence IEEE/ISIEA 2009 (2009 IEEE Symposium on Industrial Electronics and Applications). Cette conférence réunit les chercheurs des milieux universitaires et industriels.

Concernant la méthode HOS, le critère de séparation est basé sur les statistiques d'ordre supérieur avec l'hypothèse que les sources sont statistiquement indépendantes. Nous proposons de simplifier et d'améliorer l'algorithme classique Hérault-Jutten [5] [14] en choisissant une fonction non linéaire d'ordre supérieur à deux adéquate pour l'adaptation de l'algorithme. Les résultats expérimentaux montrent l'efficacité de l'approche proposée en termes de rapport signal-à-interférence, par rapport à l'algorithme Fast-ICA qui est largement utilisé. Ce travail a fait aussi l'objet d'un article intitulé « Blind Speech Separation using High Order Statistics » et a été présenté lors de la conférence IEEE/CCECE 2008 (2008 Canadian Conference on Electrical and Computer Engineering).

Nous avons réalisé des tests avec plusieurs signaux de parole à partir de deux bases de données : TIMIT, NOIZEUS et d'autres signaux enregistrés avec des durées plus longues et les résultats sont très prometteurs.

Chapitre 2

La Séparation de Sources : Problématique et état de l'art

2.1 Problématique

Un message est généralement porteur de sens et la parole n'échappe pas à ce principe. Mais ce sens peut être partiellement ou totalement caché par du bruit ou par un autre signal de parole.

Une opération importante dans les systèmes de transmission de la voix est l'extraction du signal de parole désiré à partir du signal corrompu par un bruit additif ou par d'autres signaux de parole.

Dans un grand nombre de situations, le signal reçu sur un ensemble de capteurs est la somme de plusieurs contributions appelées "signaux sources". En général, il est réaliste d'estimer que les sources émettent des signaux stochastiques indépendants. Pour étudier les effets de chaque source sur tous les signaux, il est nécessaire de séparer les sources par une procédure de filtrage. Pour réaliser cette séparation, une identification partielle des sources (en général pas totale) doit être faite.

D'importants travaux ont été faits sur ce sujet en utilisant les propriétés des statistiques du deuxième ordre des signaux qui sont représentés dans la matrice interspectrale. Plusieurs méthodes utilisant les statistiques du second ordre du signal ont été développées pour identifier et détecter le nombre de sources. En général, cette identification n'est pas réalisable et d'autres hypothèses doivent être ajoutées.

En réalité, seule la sortie du capteur est observée, on ne connait ni les signaux primitifs ni le mélange. On parle, pour ces raisons, de séparation aveugle de sources : il s'agit de restituer simultanément, à partir des observations, tous les signaux primitifs (sources) sans connaissance a priori, ni sur les signaux ni sur les mélanges. La seule hypothèse est l'indépendance statistique des sources.

La solution proposée initialement par Hérault et Jutten [5], dans l'hypothèse d'un mélange linéaire instantané de sources indépendantes a suscité de nombreuses autres méthodes. Ces méthodes sont parfois très différentes au niveau des hypothèses et des principes de séparation des signaux, mais le critère général de séparation de sources est toujours basé sur l'indépendance des signaux de sortie. L'originalité de ces différentes méthodes mises en œuvre, se trouve dans l'utilisation des "statistiques d'ordre supérieur", qui permet l'exploitation d'une information inaccessible par les méthodes classiques basées sur le critère du "deuxième ordre".

L'approche aveugle de la séparation, à laquelle nous nous intéressons, offre en outre l'avantage de ne requérir que des hypothèses faibles sur le mélange : en dehors de sa structure générale, souvent supposée linéaire, aucun paramètre n'est supposé connu.

Depuis plusieurs années, des méthodes efficaces existent pour la séparation des mélanges linéaires instantanés de sources [11], [12]. Le cas des mélanges convolutifs demeure en revanche plus difficile.

Ce deuxième chapitre présente un aperçu général sur la séparation aveugle de sources. Nous commencerons par présenter un bref historique non exhaustif sur les méthodes de séparation aveugle de sources existantes. Ensuite nous modéliserons ce problème dans le cas des canaux instantanés et convolutifs. Puis, nous donnerons quelques concepts fondamentaux de la séparation de sources.

Bien souvent, nous avons affaire à des signaux constitués d'un mélange de plusieurs sources élémentaires que nous souhaiterions voir séparées. Même s'il est parfois possible d'essayer d'en isoler une composante en modifiant la configuration de l'appareillage de mesure - position du capteur ou détection synchrone par exemple - il subsiste un grand nombre de situations sans solution évidente, avec pour conséquence de rendre délicate l'interprétation des résultats. Par exemple, la multiplication des capteurs, adoptée dans le milieu médical avec les électroencéphalogrammes ou les électrocardiogrammes, permet de fournir différentes mesures que le médecin compare entre elles pour en extrapoler des résultats en se basant sur l'expérience acquise au fil des années.

Cette absence de moyens face à des situations où les observations disponibles sont constituées d'un mélange de nombreux signaux sources est restée un obstacle majeur durant de très nombreuses années. Notons cependant, qu'en tant qu'êtres humains, nous possédons tous un dispositif capable de telles prouesses. En effet, nous pouvons tous nous concentrer sur le discours d'un seul orateur, même en présence d'un fort bruit environnant. L'appareil auditif humain couplé au cortex cérébral est capable de séparer plusieurs signaux à partir de deux observations, fournies par nos oreilles, afin de se focaliser sur un seul en particulier.

Historiquement, la recherche de la compréhension du transport et du décodage de l'information dans nos communications nerveuses a posé les premières bases de la

séparation de sources [5], [13], [14]. Ce problème est devenu depuis, un domaine de recherche majeur en traitement du signal avec de nombreux champs d'application cités dans le chapitre précédent.

2.2 État de l'art et recherches dans le domaine de la séparation de sources

La séparation aveugle de sources a été formalisée pour la première fois par des chercheurs français en traitement de signal au milieu des années 80 [5], [14], [15]. Depuis lors, beaucoup de chercheurs dans le monde entier se sont intéressés à ce problème vu la diversité de ses domaines d'application. Les recherches dans ce domaine ont été initiées en 1985 par C. Jutten et J. Hérault [14]. Ces derniers ont développé, à partir d'un problème de décodage du mouvement en neurobiologie, un algorithme de séparation à base d'un réseau neuromimétique. Cet algorithme cherche à résoudre le problème du mélange instantané sans bruit avec une matrice de mélange carrée et des signaux à valeurs réelles.

Depuis, une multitude d'algorithmes a été proposée. Ils peuvent être classés selon différents critères : mélange instantané ou mélange convolutif, linéaire ou non linéaire, statistiques d'ordre deux ou d'ordre supérieur.

En 1988 J. L. Lacoume et P. Ruiz [15][16] ont proposé un algorithme minimisant une fonction objective basée sur les cumulants croisés des sorties d'un filtre inverse, qui sont les outils naturels de mesure d'indépendance.

En 1989, J. F. Cardoso et Souloumiac proposent de considérer les propriétés algébriques des cumulants d'ordre quatre en utilisant l'information d'ordre deux pour rendre unitaire la matrice de mélange puis de l'identifier par des techniques de diagonalisation [17]. En revanche, ils montrent que les cumulants d'ordre quatre sont suffisants pour permettre, à eux seuls, d'estimer la matrice de mélange. Les deux auteurs présentent également une version non itérative de leur méthode reposant sur la diagonalisation conjointe [18] des matrices propres de la quadri-covariance des observations blanchies.

L'approche développée par P. Comon, en 1990, élargit la théorie de l'analyse en composantes principales (en anglais : PCA, *Principale Components Analysis*) à l'analyse en composantes indépendantes (en anglais : ICA, *Independent Components Analysis*) et introduit la notion de fonction de contraste. L'auteur propose comme contraste à maximiser la somme des cumulants d'ordre quatre. Comon montre par ailleurs, qu'il suffit de reconstruire des sources indépendantes deux à deux pour estimer, à une matrice triviale près, des sources mutuellement indépendantes [19]. Il propose alors la

méthode COM2 (COntrast Maximization 2) [19] reposant sur :

i- la propriété précédente ;

ii- la maximisation d'un contraste défini comme la somme des modules au carré des kurtosis (cumulants d'ordre 4 normalisés) des sources estimées.

Indépendamment des travaux de P. Comon, M. Gaeta et J. L. Lacoume [20] proposent de résoudre le problème de séparation par une technique du maximum de vraisemblance approchée.

B. Laheld et J. F. Cardoso [21] ont développé en 1993 une famille d'algorithmes adaptatifs basés sur la notion de gradient matriciel relatif. En absence de bruit, ces algorithmes ont l'avantage de fournir des performances indépendantes de la matrice de mélange et donc de son conditionnement. Toujours en 1993, survient la méthode JADE (Joint Approximate Diagonalization of Eigen-matrices), au travers de laquelle Cardoso et Souloumiac [17] présentent une solution algébrique à la maximization de leur contraste basé sur les cumulants d'ordre 4. Ils étendent par la même occasion l'algorithme de Jacobi dans le but de diagonaliser conjointement un ensemble de matrices [22]. En 1995, Delfosse et Loubaton [23] introduisent le concept de déflation en proposant une manière originale d'optimiser le contraste uni-source d'ordre 4 sur lequel s'appuie leur méthode. Deux années plus tard, Hyvärinen et Oja présentent l'algorithme Fast-ICA. La version initiale [24] de cet algorithme exploite pleinement le principe de déflation et la mesure de non-gaussianité, plus connue sous le nom de néguentropie, approximée par la valeur absolue du kurtosis de la source estimée (voir *Annexe A* pour le rappel sur les kurtosis).

En 2001, Comon trouve une solution analytique, nommée COM1 (COntrast Maximization 1), à la maximisation du contraste défini, au signe près, comme la somme des kurtosis des sources estimées. Ce contraste a été initialement présenté par Moreau et al. dans [25], [26], [27]. En outre, Moreau [26], montre que ce critère est un contraste à la condition que les kurtosis des sources soient de même signe.

En parallèle, nombre de méthodes exploitant les propriétés algébriques des cumulants voient le jour. Parmi elles, on compte les méthodes exploitant les cumulants d'ordre deux. Ces dernières nécessitent que les sources soient colorées, autrement dit temporellement auto-corrélées. Ainsi Fety [28], Tong et al. [29], [30] sont les premiers à proposer de telles approches basées sur l'exploitation des propriétés algébriques de deux matrices de covariance des observations, associées respectivement à un retard nul et non nul. Comon et Lacoume démontrent quelques années plus tard le bien-fondé de ces méthodes. De ces travaux va naître l'algorithme SOBI (Second Order Blind Identification) [31], [32] dans lequel Belouchrani et al. exploitent non plus une, mais plusieurs matrices de covariance des observations associées à des retards non nuls. Plus exactement, ils montrent qu'après blanchiment des observations, une diagonalisation

conjointe des matrices en question permet d'estimer le mélange à une matrice triviale près. Cette approche nécessite cependant que les sources aient des spectres distincts. Un algorithme similaire, baptisé TDSEP (Temporal Decorrelation Source SEParation), est proposé durant cette période par Ziehe et Mûller [33]. Quelques années plus tard, Belouchrani et Cichocki améliorent l'algorithme SOBI, et plus particulièrement l'étape de blanchiment. La méthode obtenue [34], nommée RSOBI (R pour Robust), a l'avantage d'être insensible asymptotiquement à un bruit temporellement décorrélé et de cohérence spatiale inconnue. Des méthodes plus récentes exploitant les propriétés algébriques des cumulants d'ordre deux permettent de traiter et de tirer avantage de la présence de sources non stationnaires. L'idée générale est de diagonaliser conjointement un ensemble de matrices de représentations temps-fréquence spatiales (RTFS) en utilisant la classe de Cohen [35]. La difficulté réside dans le choix de ces matrices de RTFS. En effet, on s'attend à ce que ces dernières offrent les mêmes propriétés que les matrices de covariance des observations. Fevotte et al. proposent alors un critère permettant de sélectionner automatiquement les bonnes matrices de RTFS et présentent ainsi la méthode TFBSS (Time-Frequency Blind Source Separation) [36], [37], [38]. Une autre idée consiste à supposer que les sources non stationnaires sont stationnaires par morceaux. Une diagonalisation conjointe des matrices de covariance estimées à partir de chaque bloc de données permet de ce fait de répondre au problème.

Les méthodes énumérées ci-dessus présentent entre elles des liens très forts. En effet, Comon a montré dans [39], des connexions fondamentales entre les cumulants, la néguentropie et l'information mutuelle. L'équivalence entre les approches INFOMAX et maximum de vraisemblance a été montrée dans [40], [41]. De plus, Cardoso [40] a montré que les contrastes sur lesquels reposent ces deux approches sont étroitement liés à la divergence de Kullback-Leibler (mesure de distance statistique [42]) elle-même étroitement liée à l'information mutuelle [39]. D'autres approches basées cette fois sur la représentation géométrique des mélanges ont été proposées dans la littérature. Citons par exemple les travaux de Puntonet et al. [43], [44] ainsi que ceux de Diamantaras [45]. Nous présenterons plus en détails quelques algorithmes de séparation aveugle de sources telles que SOBI, JADE, COM1, COM2 en *Annexe B.*

Depuis ces travaux pionniers en séparation de sources, ce domaine de recherche a pris une importance grandissante au point où un workshop annuel lui a été dédié (Latent Variable Analysis). Ainsi, une multitude de méthodes et d'algorithmes ont été proposés depuis, incluant celle basée sur les statistiques d'ordre deux proposée par A. Belouchrani et al. en 1997 dans [46] et celle pour la séparation de sources non-stationnaire proposée par D. T. Pham et J. F. Cardoso dans [47], [48] en 2000 et 2001.

Un certain nombre d'auteurs parmi eux K. Kokkinakis en 2006 [49], ont utilisé

des critères en combinant le domaine temporel et domaine fréquentiel avec la mise en œuvre de la rapidité des calculs. Un des moyens les plus efficaces de la séparation aveugle de sources dans le cas des mélanges convolutifs utilise les relations statistiques de la densité de puissance du signal, au lieu de résoudre la séparation de sources dans chaque bande de fréquence pour le cas du mélange instantané. Kim (2006) [50] propose une version d'algorithme multidimensionnel pour l'estimation de la densité (2007) [51] et en exploitant les dépendances en fréquence d'ordre supérieur.

La plupart des algorithmes appartiennent à la catégorie des méthodes exploitant des statistiques d'ordre supérieur implicitement ou explicitement. Sous certaines conditions, les statistiques d'ordre deux peuvent cependant suffire. Dans ce cas, les méthodes tirent alors partie soit de la diversité spectrale des signaux sources [52] soit de leur non stationnarité [53]. Un survol bibliographique est détaillé dans [54].

Sachant que l'étude est axée sur la séparation de sources audio, en particulier les signaux de parole, les propriétés des signaux audio ont été exploitées afin de mettre en œuvre des méthodes de séparation plus efficaces. En particulier, les propriétés de non-stationnarité et de parcimonie dans différents domaines ainsi que du caractère harmonique des signaux, permettent d'améliorer les performances de séparation, de réduire le coût de calcul, d'améliorer la vitesse de convergence et traiter le cas plus délicat de la séparation de sources sous-déterminée (c'est à dire le cas où l'on a moins de capteurs que de sources). Une telle configuration (sous-déterminé) est plus plausible dans le cas d'un enregistrement dans un environnement réel, étant donné que le choix du nombre de sources n'est pas un paramètre contrôlable. On peut très facilement généraliser ces méthodes pour le cas des mélanges sur-déterminés, ce qui offrira un grand choix d'applications.

La première méthode de séparation de sources audio a été effectuée en utilisant des algorithmes de décomposition modale [53] [55]. Les avantages de cette approche résident dans le fait qu'elle permet de traiter le cas des mélanges instantanés et convolutifs, et elle permet en particulier, de traiter le cas sous-déterminé. Cette approche se base sur le fait que les signaux audio peuvent être bien modélisés localement par une somme de signaux périodiques. Ces signaux seront donc décomposés en utilisant les algorithmes de décomposition modale et recombinés par classification suivant leurs directions spatiales, regroupant ainsi les composantes de chacune des sources. Cette méthode peut être utilisée dans le cas du mélange convolutif sur-déterminée [56], en lui conjuguant une décomposition en valeurs et vecteurs propres et un critère de parcimonie approprié.

Deux autres méthodes ont été utilisées avec différentes transformées dans le domaine temps-fréquence.

- La première, utilise les distributions temps-fréquence quadratiques DTFQ (ou

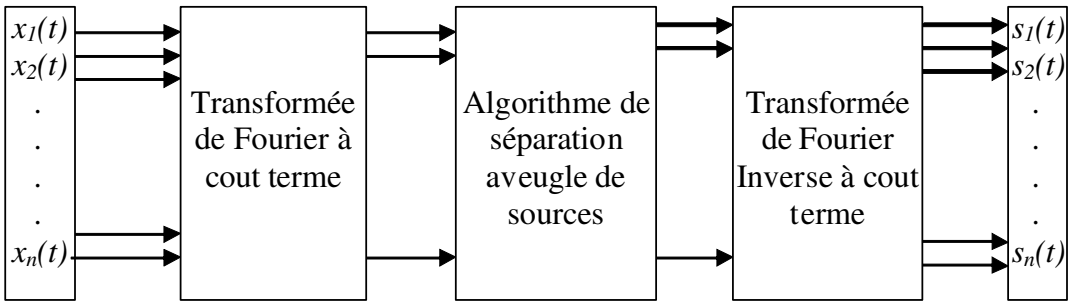

FIG. 2.1 – Schéma synoptique de l'algorithme de séparation aveugle de sources dans le domaine Temps-Fréquence.

STFD pour *Spatial Time Frequency Distribution*),

- la deuxième, utilise la transformée de Fourier à court terme TFCT (ou STFT pour *Short Time Fourier Transform*).

Ces deux méthodes supposent que les sources sont disjointes dans le domaine temps-fréquence ; c'est à dire qu'une seule source est présente dans chaque point temps-fréquence.

Une autre approche basée sur la transformation dans le domaine temps-fréquence a été introduite pour résoudre le problème de séparation de mélange convolutif de sources audio dans le cas sous-déterminé [54]. Cette approche nécessite une pré-estimation du canal de mélange qui sera réalisée en exploitant la propriété de parcimonie temporelle des signaux audio et la structure des canaux acoustiques (Voir figure 2.1).

Une autre méthode itérative de séparation de sources utilise les statistiques d'ordre deux [52]. Les statistiques d'ordre deux des données observées s'avèrent suffisantes pour séparer des sources mutuellement non corrélées à condition que les vecteurs de corrélation temporelle des sources considérées soient linéairement indépendants par paires. En appliquant une technique de gradient relatif, on déduit un algorithme itératif qui possède un certain nombre de propriétés attrayantes, comme sa simplicité et sa généralisation aux cas adaptatifs ou convolutifs.

2.2.1 Application au traitement de la parole

L'application est en général de type convolutif car nous avons le problème d'échos ou de déplacement des personnes qui parlent. Les capteurs sont des microphones et $s_i(t)$ avec $i = [1, 2, 3, ..., N]$ représente le signal de parole (la parole étant de nature non-stationnaire). Le système de propagation peut être à multi-entrées multi-sorties. Les bruits représentent toute perturbation additive reçue au niveau des microphones. La plupart des travaux visant le domaine du traitement de la parole ne s'appliquent pas

dans le cas où l'on dispose d'un seul microphone. Généralement, ces travaux peuvent être appliqués pour résoudre le problème de "*Cocktail Party*". Il s'agit de séparer les voix des différentes personnes qui parlent en même temps dans une salle, tout en utilisant des enregistrements effectués au niveau de plusieurs microphones. Parmi les travaux qui visent le domaine d'application audio, on peut énumérer :

- Dans [57], nous trouvons un algorithme simple s'intéressant aux situations physiques réelles comme le cas d'un environnement acoustique.

- L. Parra et ses collaborateurs ont appliqué concrètement la séparation aveugle de sources pour identifier les signaux acoustiques (non-stationnaires) simultanément enregistrés dans un environnement multi-trajets dans un système de mélange convolutif [58]. De même que dans la majorité des algorithmes convolutifs de séparation des signaux de parole, les auteurs appliquent dans un premier temps la transformée de Fourier pour transformer le modèle convolutif en un modèle instantané. Ils appliquent ensuite la diagonalisation conjointe (DC) sur un ensemble de matrices de même fréquence.

- Le travail proposé dans [59] concerne le problème de la séparation de mélanges convolutifs des sources audio dans une situation réelle. Les auteurs adaptent le principe de l'analyse en composantes indépendantes (ICA) dans le domaine fréquentiel.

- Dans une application réelle concernant des expériences réelles dans des studios d'enregistrement acoustique, A. Westner et V. M. Bove montrent dans [3] que le contraste INFOMAX (INFOrmation MAXimization) est le mieux adapté à la séparation des sons acoustiques.

- En traitement de la parole, un algorithme en ligne appartenant à la classe des approches fréquentielles, proposé dans [60] permet de séparer des sources enregistrées dans un environnement réel. Il s'agit de la situation où des personnes parlent simultanément et en se déplaçant dans une salle.

- Un autre travail qui étudie le problème de séparation et de déréverbération d'un mélange dynamique des sources a été présenté dans [61]. Ce travail prend en considération les caractéristiques acoustiques des sources.

- Toujours dans le cas d'un milieu de réverbération, en 2007, Scott C. Douglas [62] a proposé une approche appelée : Spatio-Temporal Fast-ICA qui représente une extension de l'algorithme très connu Fast-ICA. C'est une méthode temporelle dans laquelle les auteurs combinent un pré-blanchiment spatio-temporel multicanaux via plusieurs étapes de prédiction linéaire des moindres carrés avec des procédures adaptatives.

- La séparation de sources peut être également appliquée au rehaussement de la parole [63].

La solution au problème de la séparation de sources est fortement liée aux modèles du mélange des sources qui représentent les interactions entre les sources et les signaux

reçus à la sortie des capteurs. Les premières solutions de séparation ont été proposées sous l'hypothèse d'un mélange instantané linéaire.

2.3 Principe

Le principe de la séparation aveugle de sources peut être énoncé comme suit : il s'agit de reconstituer les N signaux sources émis à partir de la seule connaissance de M signaux d'observations reçus correspondant à des mélanges de ces sources. Le traitement est alors aveugle, en d'autres termes, on ne possède aucune information sur le système de mélange et les sources sont inobservables. Il est nécessaire d'avoir une hypothèse supplémentaire, sinon la séparation aveugle de sources apparaît comme un problème insoluble. C'est pourquoi la plupart des techniques de la séparation aveugle de sources reposent sur l'hypothèse de l'indépendance des sources. Initialement, elle a été modélisée à base d'un simple produit, autrement dit les capteurs reçoivent à chaque instant une combinaison linéaire des signaux sources. Plus tard, des modélisations proches de la réalité ont été introduites. L'une de ces modélisations considère le canal de transmission comme un système reposant sur une opération de filtrage. En d'autres termes, les signaux captés dépendent linéairement à la fois des sources et de leurs versions retardées. Le système de mélange est dit linéaire convolutif. D'autres modélisations plus complexes et plus réalistes considèrent que les capteurs ont des caractéristiques non-linéaires. Ainsi, la séparation aveugle de sources possède plusieurs degrés de difficulté selon le type du système de mélange envisagé :

- système linéaire ou non-linéaire,
- mélange instantané ou convolutif,
- modèle de mélange sur-déterminé ou sous-déterminé,

et également selon les caractéristiques des sources considérées :

- signaux à valeurs réelles ou complexes,
- aléatoires
- stationnaires, cyclo-stationnaires ou non-stationnaires,
- statistiquement mutuellement indépendants ou corrélés, ou non
- et déterministes, à puissance moyenne finie ou à énergie finie.

Différentes approches ont été développées dans la littérature pour résoudre le problème de la séparation aveugle de sources, dont certaines ont été citées ci-dessus. Une première approche consiste à utiliser la technique d'analyse en composantes indépendantes (ICA) formalisée par P. Comon dans [19]. Les méthodes basées sur la technique d'ICA utilisent les outils d'analyse d'ordre supérieur pour optimiser des fonctions de coût appelées "contrastes", tout en exploitant les propriétés de cumulants des sources supposées indépendantes [17] [82] [64] [65] [66] [67] [68] [69] [70] [71]. La difficulté du

problème de séparation de sources varie selon la nature du mélange, la détermination du mélange et le niveau de connaissance a priori que l'on possède sur les sources et le mélange.

2.3.1 Nature du mélange

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.

2.3.2 Mélange et modèle linéaire instantané

Mélange instantané

C'est le cas le plus simple dans lequel les sources sont uniquement caractérisées par une différence de niveau d'intensité entre les canaux. La séparation de sources est une technique de traitement des signaux (ou images) multi-capteurs dans laquelle on postule un modèle où une séquence d'observations M-dimensionnelles $x(t), t = 1, \ldots, T$ est modélisée par :

$$x(t) = As(t) + n(t) \tag{2.1}$$

où A est une matrice $M \times N$, $s(t)$ est un N-vecteur de sources dont les composantes sont indépendantes pour chaque T et $n(t)$ représente un éventuel bruit additif.

Les hypothèses sont donc très voisines de celles de l'ICA. Cependant, l'existence d'une possible dépendance temporelle entre échantillons successifs permet d'exploiter l'indépendance statistique entre signaux sources autrement que dans un modèle non gaussien.

Cette description est difficile par le fait qu'il n'existe pas d'algorithme uniformément efficace, et ceci pour plusieurs raisons. Tout d'abord, il y a de nombreuses façons d'exprimer l'indépendance. Si l'on choisit de s'appuyer sur une éventuelle structure temporelle (ou spatiale) des observations comme suggéré par l'équation 2.1, il est possible de se contenter de modèles gaussiens. On pourra alors exploiter la diversité

spectrale (respectivement la non stationnarité) des observations à travers la formation des matrices de covariance spectrale (respectivement locale).

Sous certaines hypothèses, on peut alors montrer que l'estimation au maximum de vraisemblance (dans le cas sans bruit) se réduit à une diagonalisation conjointe de ces matrices, opération pour laquelle il existe un algorithme efficace. Lorsqu'il n'est pas possible de s'appuyer sur une éventuelle structure temporelle des observations, l'hypothèse et le principe du maximum de vraisemblance conduisent à la minimisation d'une version empirique de $I(A-1X)$ où I représente l'information mutuelle qui est une mesure naturelle de la dépendance entre les variables aléatoires. La version empirique de $I(A-1X)$ requiert implicitement une estimation de la distribution de chaque composante.

Une autre voie algorithmique consiste à utiliser des approximations de $I(Y)$ en termes de cumulants d'ordre supérieur à deux, obtenues à partir de développements d'Edgeworth ou de Gram-Charlier. La structure des cumulants permet de concevoir des algorithmes algébriques.

Finalement, s'inspirant de l'analogie avec la projection poursuite, il peut aussi être efficace d'extraire les composantes séquentiellement, en utilisant, pour chaque composante, des algorithmes de point fixe tel que Fast-ICA.

Modèle mathématique d'un mélange instantané

On observe un signal $x_i(t)$, superposition des signaux $s_j(t)$. En admettant le mélange linéaire et instantané, un modèle du mélange peut être décrit par la relation suivante :

$$x_i(t) = \sum_{j=1}^{N} a_{ij} s_j(t) + n_i(t) \qquad \text{avec} \quad i \in [1, 2, 3, ..., M] \tag{2.2}$$

L'équation 2.1 donne la forme vectorielle de cette expression, à l'instant t.

Le modèle de mélange instantané traduit donc une relation spatiale entre les sources et les capteurs, à chaque instant, les observations sont des combinaisons linéaires des sources au même instant.

Mélange et modèle de mélanges convolutifs

L'équation 2.1 décrit un mélange instantané. Malheureusement ce modèle est plutôt incomplet dans le cas de sources enregistrées dans un environnement acoustique réel, où les sources sont mélangées convolutivement. En ignorant le bruit $n(t)$, le problème peut-être formulé comme suit :

$$x(t) = A * s(t) \tag{2.3}$$

Le but est de trouver la matrice inverse W telle que :

$$\hat{s}(t) = W * x(t) \tag{2.4}$$

et les composantes de $\hat{s}(t)$ aussi indépendantes que possible. Le mélange étant convolutif, le calcul dans le domaine temporel devient complexe et lent. Mais si l'équation est transformée en domaine des fréquences la convolution se transforme en une multiplication. Cela signifie qu'il est possible d'appliquer les méthodes de mélanges instantanés à chaque fréquence.

$$X(f,t) = A(f)S(f,t), \tag{2.5}$$

$$\hat{S}(f,t) = W(f)X(f,t). \tag{2.6}$$

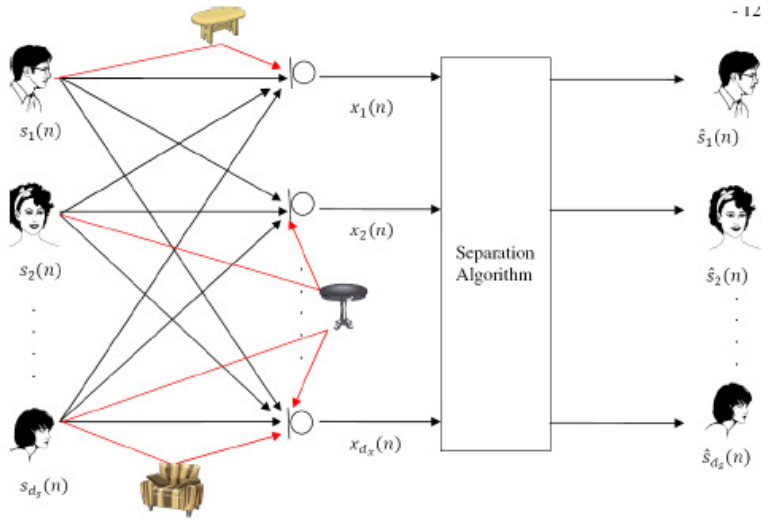

FIG. 2.2 – Problème multiple concernant le mélange convolutif.

Les premiers travaux sur la séparation de sources de mélanges convolutifs ont été réalisés dans le domaine temporel [54]. Inspiré par les méthodes de déconvolution aveugle, K. Torkkola [72] a modélisé le procédé de séparation comme un filtre RIF dont il faut estimer les coefficients w_{ij} :

$$\hat{s}(t) = \sum_{k=0}^{K} W(k)x(t-k), \tag{2.7}$$

Pour résoudre ce problème, l'auteur utilisa une méthode de maximisation de l'information mutuelle [73]. T. W. Lee et al. [74] ont modélisé la procédure de séparation comme un filtre IIR en supposant que l'environnement de l'enregistrement est à phase minimale, ce qui n'est pas toujours valable.

Mukai et al. [61], [75] ont étudié le problème de séparation de mélanges convolutifs en tenant compte des propriétés acoustiques des signaux. Ils traitent le problème en fréquence en appliquant une transformée de Fourier à court terme :

$$S_x(t,f) = A(f)S_s(t,f), \tag{2.8}$$

où $S_x(t,f)$ est le vecteur temps-fréquence (T-F) du mélange, $S_s(t,f)$est le vecteur T-F des signaux sources et $A(f)$ est la transformée de Fourier du filtre de convolution. En appliquant une matrice de séparation $W(f)$ pour chaque fréquence,

$$S_{\hat{s}}(t,f) = W(f)S_x(t,f), \tag{2.9}$$

P. Smaragdis [76], [77] a aussi travaillé dans le domaine temps-fréquence pour transformer le produit de convolution en multiplication. Le problème est défini en tant qu'estimation de la matrice de séparation pour chaque trame de fréquence qui peut avoir une solution au sens du maximum de vraisemblance. Malheureusement, la résolution du problème indépendamment pour chaque trame de fréquence produit des problèmes de permutation. Pour résoudre ce problème P. Smaragdis a proposé un arrangement adaptatif pour appliquer un certain accouplement des fréquences pour les trames voisines.

L. Parra et al. [58] ont également travaillé dans le domaine temps-fréquences en utilisant la non-stationnarité pour effectuer la séparation, leur solution au problème était d'imposer une contrainte à la longueur du filtre de séparation. Enfin, pour résoudre le problème de permutation, les auteurs imposent la contrainte de continuité ("lissage") dans le domaine des fréquences. Une autre solution est d'utiliser la formation de voies (en anglais *Beamforming*) [59] pour aligner les données le long de l'axe de fréquence en utilisant les directions d'arrivée du signal. H. Bousbia et al. [78] ont aussi utilisé la non-stationnarité des signaux comme suit :

soit :

$$\tilde{x}(t) = \left[x(t)^T ... x(t-d)^T\right]^T, \tag{2.10}$$

tel que :

$$\tilde{x}(t) = A\tilde{s}(t), \tag{2.11}$$

où

$$\tilde{s}(t) = \left[s(t)^T ... s(t-d-L)^T\right]^T \tag{2.12}$$

Il a été montré que les matrices d'auto-covariance (sans bruit) du signal $\tilde{x}(t)$ s'écrivent sous la forme :

$$R_{\tilde{x}}(t) = AR_{\tilde{s}}(t)A_d^T, \tag{2.13}$$

où $R_{\tilde{x}}(t)$ est bloc diagonal. Ainsi, trouver une matrice de séparation qui inverserait A revient à bloc-diagonaliser les matrices pour $t = t_1, \ldots, t_K$. Ceci est une généralisation de la méthode de séparation de signaux non-stationnaires par diagonalisation conjointe proposée par D. T. Pham et J. F. Cardoso dans [47].

Mélange anéchoïque

C'est une généralisation du cas instantané, dans lequel on considère que les sources arrivent à chaque capteur à des instants différents, si bien que chaque source est caractérisée par un délai entre les canaux.

Le cas anéchoïque peut être considéré comme un cas de mélange convolutif, dans lequel on considère des chemins multiples entre chaque source et chaque capteur. Les détails d'explication sur les différents types de mélange convolutif seront donnés au cinquième chapitre.

Principes fondamentaux de la séparation de sources en mélanges convolutifs

Nous étudions ici le cas de mélanges convolutifs définis par l'équation 2.3,
à temps discret l'équation 2.3 s'écrit sous la forme :

$$x(n) = [A(Z)] * s(n) \tag{2.14}$$

où la matrice de fonctions de transfert $A(Z)$ de taille $M \times N$ prend la forme dans le cas général :

$$\begin{bmatrix} \sum_{l=-\infty}^{+\infty} a_{11}(l)Z^{-1} & \sum_{l=-\infty}^{+\infty} a_{12}(l)Z^{-1} & . \ . \ . & \sum_{l=-\infty}^{+\infty} a_{1N}(l)Z^{-1} \\ . & . & & . \\ . & & . & . \\ . & & . & . \\ \sum_{l=-\infty}^{+\infty} a_{M1}(l)Z^{-1} & \sum_{l=-\infty}^{+\infty} a_{M2}(l)Z^{-1} & & \sum_{l=-\infty}^{+\infty} a_{MN}(l)Z^{-1} \end{bmatrix} \tag{2.15}$$

où le terme Z^{-1} représente l'opérateur de retard. La notation $[A(Z)]$ de l'équation 2.14 rend compte de l'opérateur temporel associé à la matrice $A(Z)$. Nous obtenons pour chacun des filtres $A_{Mj}(Z)$ la composante :

$$\left[\sum_{l=-\infty}^{+\infty} a_{Mj}\left(l\right) Z^{-l}\right] s_j\left(n\right) = \sum_{l=-\infty}^{+\infty} a_{Mj}\left(l\right) s_j\left(n-l\right) = A_{Mj}\left(n\right) * s_j\left(n\right). \tag{2.16}$$

D'une manière générale, l'obtention d'une relation de convolution entre chaque signal source et chaque signal reçu est une condition nécessaire mais non suffisante pour avoir un mélange qui soit, au sens de la séparation de sources, « convolutif général », c.-à-d. qui inclut véritablement un effet de mémoire. Il est en effet possible d'être dans une configuration où la matrice de mélange peut être ramenée au cas particulier du simple mélange instantané. Prenons l'exemple d'une configuration à deux sources $s_j(n)$ et deux observations $x_i(n)$:

$$x_1(n) = A_{11}(n) * s_1(n) + A_{12} * s_2(n),$$

$$x_2(n) = A_{21}(n) * s_1(n) + A_{22} * s_2(n). \tag{2.17}$$

Dans le cas où la diversité entre les capteurs se résume à un seul facteur d'échelle, éventuellement complexe, $a_j e^{j\phi_i}$ propre à chaque source, nous avons :

$$x_1(n) = A_{11}(n) * s_1(n) + A_{12} * s_2(n),$$

$$x_2(n) = a_1 e^{j\phi_1} A_{11}(n) * s_1(n) + a_2 e^{j\phi_2} A_{12}\left(n\right) * s_2(n). \tag{2.18}$$

Ce qui peut aisément s'écrire :

$$\begin{bmatrix} x_1(n) \\ x_2(n) \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ a_1 e^{j\phi_1} & a_2 e^{j\phi_2} \end{bmatrix} \begin{bmatrix} A_{11}(n) * s_1(n) \\ A_{12} * s_2(n) \end{bmatrix}. \tag{2.19}$$

Ainsi, en l'absence de diversité temporelle entre les différents récepteurs, le mélange devient un mélange instantané au sens de la séparation de sources. L'effet convolutif,

initialement introduit par la propagation, se retrouve factorisé dans une nouvelle définition des sources.

2.3.3 Détermination du mélange

La détermination du mélange représente le rapport entre le nombre de sources et le nombre de capteurs. Si le nombre de sources est égal au nombre de capteurs, on parlera de mélange déterminé. Si le nombre de sources est inférieur au nombre de capteurs, on parlera de mélange sur-déterminé. Enfin si le nombre de sources est supérieur au nombre de capteurs, on parlera de mélange sous-déterminé.

2.3.4 Niveau de connaissance a priori sur les sources et le mélange

Il est plus simple d'estimer le mélange, si l'on connait le nombre de sources, l'emplacement physique des sources et des capteurs, ou bien les caractéristiques et la position relative des capteurs (cas du HRTF : Head Relative Transfer Function). De même un certain nombre de connaissances a priori sur les sources, comme par exemple, des informations de nature spectrale sur les sources, permettent de résoudre le problème de séparation aveugle de sources de façon plus efficace.

2.4 Hypothèses nécessaires

Dans le but de simplifier et permettre de résoudre le problème de la séparation aveugle de sources, les hypothèses suivantes sont habituellement requises :

Hypothèse 1 : Les composantes du vecteur des sources $s(t)$ sont colorées, ce qui signifie que la fonction d'intercorrélation entre une source et sa version retardée ou bien entre deux versions retardées de cette source est non-nulle. Cette hypothèse est importante pour les méthodes de séparation aveugle de sources utilisant les décompositions matricielles.

Hypothèse 2 : Dans le cas d'un système de mélange sous-déterminé, le nombre de capteurs M est inférieur au nombre de sources N ($N \geq M$).

Hypothèse 3 : Les sources sont statistiquement indépendantes. D'un point de vue mathématique, cela signifie que la densité de probabilité conjointe des N sources peut se factoriser comme le produit de leurs densités marginales :

$$p\left(s_1\left(n-l_1\right), s_2\left(n-l_2\right), ...s_N\left(n-l_N\right)\right) = \prod_{i=1}^{N} P\left(s_i\left(n-l_i\right)\right), \tag{2.20}$$

pour tout ensemble d'instants $l_i, i = 1, ...N$.

Cette hypothèse est réduite à la non-corrélation des sources lors de l'utilisation de méthodes de séparation à l'ordre 2. L'hypothèse suivante, renforçant l'hypothèse 3 est également très régulièrement requise et permet de simplifier considérablement le problème.

Hypothèse 4 : Chaque source est un processus indépendant et identiquement distribué (i.i.d.). Concrètement, cela signifie qu'à chaque instant, chaque source présente une distribution statistique indépendante de tous ses instants passés, et futurs et avec l'hypothèse 3 de tous les instants passés, présent et futurs des autres sources. Nous supposons cette hypothèse vérifiée dans cet état de l'art car la plupart des méthodes disponibles la requièrent.

Hypothèse 5 : Tous les signaux considérés sont ergodiques.

Hypothèse 6 : Le vecteur $n(t)$ contenant les bruits est stationnaire au sens strict. Cette hypothèse ne trouve un intérêt que d'un point de vue pratique (comme pour l'hypothèse 5), en d'autres termes elle n'est pas nécessaire en théorie. Le bruit $n(t)$ de la relation 2.1 est considéré nul.

Hypothèse 7 : Les composantes de $s(t)$ sont statistiquement indépendantes des composantes de $n(t)$.

Hypothèse 8 : Dans le cadre des mélanges convolutifs (respectivement instantanés), la matrice A obtenue après reformulation du modèle est de rang plein, c'est à dire $rg\{A\} = N$ ceci garantit l'existence d'une matrice inverse à gauche de la matrice A, et donc l'existence du séparateur.

Hypothèse 9 : Le système de mélange est exponentiellement stable (voir [78] pour plus de détails) :

$$A_{mj}(Z) = \sum_{l=-\infty}^{+\infty} a_{Mj}\left(l\right) Z^{-l} \approx \sum_{l=-L'}^{L} a_{Mj}\left(l\right) Z^{-l}. \tag{2.21}$$

Cette hypothèse, implicitement vérifiée dans le cas de mélanges linéaires instantanés, signifie que chacun des filtres a_{Mj} peut être raisonnablement approché par un filtre RIF pourvu qu'il soit d'un ordre suffisamment élevé.

2.5 Principe de la séparation idéale

Le principe de la séparation peut être résumé par la figure 2.3. Il s'agit de trouver dans le cas idéal un filtre matriciel $W(Z)$ de taille $N \times M$ qui inverse le mélange, c-à-d. qui fournit le vecteur de sortie :

$$\hat{s}(n) = [W(Z)]\, x(n) = [W(Z)]\, [A(Z)]\, s(n) \approx s(n). \tag{2.22}$$

Le filtre $W(Z)$ est donc égal à l'inverse à gauche de la matrice de mélange :

$$W(Z) = A(Z)^{-1}. \tag{2.23}$$

Le principe général utilisé pour déterminer ce filtre est basé sur un critère d'indépendance. En effet, si les sources s_j sont mutuellement indépendantes, alors toute combinaison linéaire d'une version filtrée de deux sources ou plus induit des signaux dépendants entre eux. Ainsi, en considérant les hypothèses 2 à 8 vérifiées, une condition nécessaire mais non suffisante pour que cette séparation idéale soit atteinte est de choisir $W(Z)$ tel que les différentes sorties $y_i(n)$ soient mutuellement indépendantes au sens défini dans l'hypothèse 3. Ce concept est à la base de l'analyse en composantes indépendantes [19] et elle est largement pratiquée en mélanges instantanés [80].

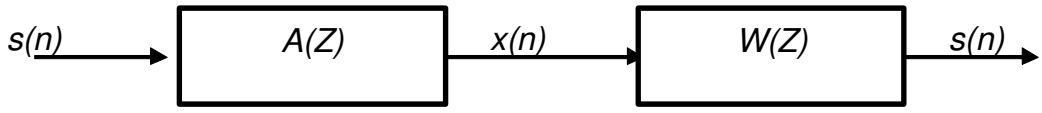

FIG. 2.3 – Principe de la séparation de sources mélangées convolutivement.

2.6 Choix entre domaine fréquentiel et domaine temporel

Dans des domaines d'application tels que l'acoustique, les filtres à estimer comportent plus d'une centaine de coefficients. Les calculs induits par les relations de convolution peuvent alors devenir rapidement coûteux. Cependant, la relation de convolution devient un simple produit dans le domaine fréquentiel, dans lequel il est aisé de se reporter grâce à la rapidité des algorithmes FFT.

Ainsi, après passage dans le domaine fréquentiel, la matrice $A(f)$ de l'équation 2.5 est constante pour f donnée, ce qui permet d'obtenir un mélange linéaire instantané de la contribution des N sources $S_j(f)$ dans chaque bande de fréquences considérée. En séparant les sources dans chacune de ces bandes, il est alors possible de remonter aux signaux d'origine par transformée de Fourier inverse et sommation des sorties

associées à chaque bande. Ce type de problème peut alors être résolu par les méthodes ICA classiques, c.-à-d. applicables à des mélanges linéaires instantanés.

Quelques travaux ont été réalisés en bande étroite [81]. D'autres études portent toutefois sur le problème en large bande. W. Jun [82] propose par exemple une approche bi-spectrale mais ne donne pas de détails concernant la reconstruction des sources une fois la séparation effectuée dans le domaine fréquentiel. V. Capdevielle et al. ([83], [84] et [85]) ont proposé une méthode basée sur la continuité du spectre entre les différentes bandes de fréquences. Cette solution, souvent utilisée dans les quelques approches fréquentielles disponibles, présente certaines restrictions sur les signaux sources. Ceux-ci doivent présenter une certaine continuité dans le domaine fréquentiel. En outre, la transformée de Fourier a tendance à gaussianiser les signaux, ce qui ne facilite pas leur séparation par les méthodes habituelles hormis pour des cas bien particuliers [83], [84] et [85]. Ainsi l'approche fréquentielle, malgré sa simplicité, possède un réel problème d'indéterminations pouvant amener à une reconstruction erronée des sources sans que l'on s'en rende compte.

S. Araki et al. [86] explique qu'en BSS convolutif effectué dans le domaine fréquentiel, il y a des compromis dont il faut tenir compte au niveau de la durée des fenêtres pour la transformation en fréquence et la transformation inverse. Si la fenêtre est trop longue, il y a moins d'échantillons fréquentiels de produits et l'hypothèse d'indépendance n'est plus valide entre les séquences des échantillons spectraux des différentes sources. Si la fenêtre est trop courte, il y a un manque de résolution fréquentielle et les filtres temporels de la matrice de séparation seront trop courts.

H. Attias et C. E. Schreiner [87] ont présenté une approche très intéressante pouvant combiner ou non les informations à la fois temporelles et fréquentielles. Cependant, basée sur des critères Infomax, cette méthode est limitée aux signaux sur-gaussiens, ce qui n'est généralement pas le cas des signaux audio et de parole.

H. C. Wu et J. Principe [88] puis L. Parra et C. Spence [89] présentent également des approches orientées temps-fréquence dans lesquelles les indéterminations sont éliminées par une contrainte sur la longueur des filtres de reconstruction temporels en rapport avec la résolution fréquentielle. Cette condition permet d'éliminer les indéterminations quand les signaux sont bien conditionnés et en particulier non stationnaires, ce qui n'est pas toujours le cas, par exemple dans le cas de signaux de télécommunications pris à la fréquence symbole. Notons toutefois qu'il s'agit d'un cas particulier et que ce type de signaux est plus généralement cyclo-stationnaire.

Toutefois, l'émergence de méthodes basées sur une réelle approche temps-fréquence laisse entrevoir de nouvelles perspectives par le changement des hypothèses nécessaires. Nous développerons une méthode basée sur une approche temps-fréquence pour des mélanges convolutifs nous permettant d'obtenir d'excellents résultats tout en réduisant

fortement les hypothèses habituellement requises par les méthodes classiques, notamment au niveau de la possibilité de séparer entièrement ou partiellement des sources dépendantes, éventuellement gaussiennes.

2.7 Ambiguïté de séparation

Séparer aveuglement les sources contenues dans leurs mélanges est entaché d'ambiguïté. En effet, la séparation peut être obtenue avec une infinité de solutions [90], [91]. Pour chaque cas, instantané ou convolutif, le comportement de l'ambiguïté a une incidence sur la reconstruction des sources.

2.7.1 Ambiguïté de permutation

L'estimation de la matrice inverse des filtres n'est pas unique. Ceci peut être vu sur le modèle de mélange (cas instantané) en permutant par exemple deux colonnes (e.g. d'indice 1 et 2) dans la matrice de mélange et en utilisant la propriété de commutativité de l'opération de multiplication et de l'opération d'addition, nous pouvons écrire :

$$x = \sum_{j=1}^{p} s_j a_j = s_1 a_1 + s_2 a_2 + ... + s_M a_M = s_2 a_2 + s_1 a_1 + ... + s_M a_M, \tag{2.24}$$

Nous constatons, par rapport au mélange initial et par rapport au vecteur initial représentant les sources, qu'il y a eu un changement d'ordre entre les sources d'indice 1 et 2, tout en conservant le même vecteur des mélanges. Ce qui entraîne une permutation de la première ligne avec la deuxième dans la matrice de mélange. Cette permutation est modélisée par une matrice de permutation P qui contient seulement un élément égal à « un » par ligne et par colonne. Nous pouvons conclure que la permutation ne change pas les mélanges, mais implique que l'ordre des sources est indéterminé.

2.7.2 Ambiguïté d'échelle

Nous avons vu dans la sous-section précédente que la permutation de deux colonnes dans la matrice de mélange et de deux sources, ne change en rien les mélanges. Nous allons voir dans cette section, que de même, la multiplication d'une colonne (de la matrice de mélange) et la division d'une source par un scalaire ne changera pas le vecteur mélanges :

$$s_1 a_1 + s_2 a_2 + ... + s_M a_M = \left(\frac{1}{\alpha_1} s_1\right)(\alpha_1 a_1) + \left(\frac{1}{\alpha_2} s_2\right)(\alpha_2 a_2) + ... \left(\frac{1}{\alpha_M} s_M\right)(\alpha_M a_M)\,. \tag{2.25}$$

En factorisant ces scalaires, nous remarquons que la multiplication par un scalaire ne change pas les mélanges, mais montre que l'amplitude des sources est indéterminée.

2.7.3 Ambiguïté de filtrage

Le plus souvent, la séparation de mélanges convolutifs de sources est ramenée à un ensemble de séparations instantanées en passant dans le domaine de la transformée de Fourier. Des ambiguïtés peuvent ainsi se produire à chaque fréquence, ce qui implique un filtrage des signaux dans le temps, ceci bien évidemment sans remettre en cause le procédé de séparation de sources. La méthode de MDP (*Minimal Distortion Principle*) est couramment utilisée pour contrer l'ambiguïté de filtrage pour les approches fréquentielles [92]. En conséquence, la séparation des mélanges convolutifs peut être accomplie à un filtrage près sans remettre en cause la démarche de séparation. Il faut par ailleurs noter que cette ambiguïté est la plus complexe et présente une grande gêne à la reconstruction des sources.

2.7.4 Ambiguïté complète

Dans le cas où les ambiguïtés s'associent, la séparation aveugle de sources, en plus des erreurs d'estimation, ne peut retrouver les signaux sources dans les mélanges linéaires, qu'à une permutation près, à un facteur d'échelle près et/ou à un filtrage près (cas convolutif). Nous pouvons alors conclure que les sources seront estimées à une possible permutation près et/ou un possible facteur d'échelle près. Dans le cas convolutif, toutes les indéterminations peuvent êtres modélisées par une opération complexe de filtrage des signaux. On dit dans ce cas, que les sources seront estimées à un filtrage près.

2.8 Évaluation de la séparation de sources

La définition de critères d'évaluation est une étape importante pour comparer différents algorithmes de séparation de sources.

Il paraît logique d'utiliser des critères du même type que ceux utilisés en débruitage, c'est-à-dire le plus souvent l'évaluation de rapports signal à bruit. Cependant, nous nous trouvons dans une situation légèrement différente du débruitage, car nous avons la possibilité de prendre plusieurs signaux de référence, à savoir les différentes sources de départ, alors que dans le cas du débruitage, seul le signal original importe.

2.8.1 Critères de comparaison

Un message est généralement porteur de sens et la parole n'échappe pas à ce principe. Cependant ce sens peut être partiellement ou totalement caché par du bruit. La qualité d'un message permet de définir un critère de bonne réception du message vis-à-vis des agressions de l'environnement. Cependant la qualité est, en général, une notion assez subjective pour qu'il faille rappeler les méthodes qui permettent de la quantifier. Les critères de qualité d'un message peuvent être divisés en deux groupes que sont les critères objectifs et les critères subjectifs.

2.8.2 Critères subjectifs de qualité

Les critères subjectifs sont des jugements portés par l'auditeur sur le message qu'il reçoit. Ces critères sont dits subjectifs car ils ne font intervenir qu'un seul des deux intervenants de la communication, en l'occurrence le récepteur, qui n'a, a priori, aucune connaissance sur le message qui lui est envoyé. Ces tests de qualité ont été, entre autres, employés pour l'évaluation des méthodes de numérisation de la parole sur des lignes à plus ou moins haut débit et donc avec des compressions plus ou moins fortes. Dans nos expériences, nous avons utilisé une méthode subjective basée sur la note d'opinion moyenne : (*Mean Opinion Score en anglais*, MOS en abréviation) [93].

Lors d'un test subjectif, on demande à des participants de tester un système dans différentes conditions et de noter sur une échelle de qualité, la qualité vocale de ce système. D'une manière générale, la qualité dépend de la personne qui la juge. Sa perception met en jeu l'expérience passée, les attentes et l'humeur de chacun.

La qualité vocale est dépendante de celui qui l'évalue. Ainsi, les notes des participants pour une condition de test donnée sont moyennées pour obtenir la note moyenne d'opinion, qui permet de diminuer l'effet subjectif sur l'évaluation de la qualité vocale. De plus, la perception de la qualité vocale dépend du contexte et de l'environnement dans lesquels est placée la personne qui juge. En effet, si elle est simplement en train d'écouter un message vocal (contexte d'écoute) ou si elle est impliquée dans une conversation avec un interlocuteur (contexte de conversation), les processus d'attention mis en jeu ne sont pas les mêmes et le jugement de la qualité en est affecté. De même, l'environnement (bruit, informations visuelles ou sonores supplémentaires, etc.) influence le jugement de la qualité. Ainsi, les conditions à tester sont définies en fonction de l'objectif visé, le participant étant amené à évoluer dans un ou plusieurs contextes (écoute, locution et conversation).

La notation s'effectue selon l'une des méthodes définies par l'Union Internationale des Télécommunications (UIT) dans la Recommandation P.800 [93], avec les catégories :

5 = Excellente,
4 = Bonne,
3 = Correcte,
2 = Médiocre,
1 = Mauvaise.

2.8.3 Critères objectifs de qualité

Les critères objectifs sont des mesures qui tiennent compte du message tel qu'il est fourni par l'émetteur et tel qu'il est compris par le récepteur. Les critères objectifs permettent donc des calculs exacts bien que certaines erreurs, dues à la méthode de calcul employée ou à de trop fortes approximations, puissent cependant apparaître. Ces critères font appel à des mesures de dissemblance sur le signal et font intervenir le message au niveau de l'émetteur et au niveau du récepteur.

Afin de comparer les différents résultats expérimentaux, comme méthodes objectives, nous avons utilisé quatre métriques différentes : le domaine temporel, le spectrogramme, le rapport signal-à-interférence, ainsi que le PESQ (*Perceptual Evaluation Signal Quality*).

Le Spectrogramme

Le spectrogramme est un diagramme associant à chaque instant t d'un signal, son spectre de fréquence. Il est utilisé pour identifier des sons et il est largement utilisé dans le domaine de la reconnaissance de la parole.

C'est un outil de visualisation utilisant la technique de la transformée de Fourier et donc du calcul de spectres. Il a commencé à être largement utilisé en 1947, à l'apparition du sonagraphe, et est devenu l'outil incontournable des études en phonétique pendant de nombreuses années.

L'apparition de l'informatique puis d'écrans graphiques de bonne qualité a permis d'abandonner tout matériel comme le sonagraphe mais la technique du spectrogramme est encore aujourd'hui largement utilisée du fait de sa simplicité de mise en œuvre et du grand nombre d'études qui ont déjà été réalisées.

Le spectrogramme permet de mettre en évidence les différentes composantes fréquentielles du signal à un instant donné, une transformée de Fourier rapide étant régulièrement calculée à des intervalles de temps rapprochés. Dans cette méthode d'analyse, le signal est considéré comme indéfiniment stable et constitué d'une somme invariable de fonctions sinusoïdales de fréquences différentes. Pour contourner cette contrainte théorique d'invariabilité du signal, il faut convoluer le signal avec une fenêtre temporelle qualifiée de glissante puisque chaque calcul de spectre nécessite de convoluer le

signal avec la fenêtre temporelle à un instant particulier.

Différentes fenêtres temporelles existent mais chacune introduit une erreur résiduelle plus ou moins importante dans le spectre obtenu du fait de la forme choisie qui peut être, dans le pire des cas, triangulaire ou carrée. Le choix de la taille de la fenêtre, en nombre de points de convolution, est également important vis-à-vis de la qualité de l'analyse fréquentielle obtenue. Ainsi, une fenêtre de petite taille (avec un nombre de 128 points, par exemple) permettra d'obtenir une bonne analyse dans le domaine temporel, du fait de son étroitesse, mais ne permettra pas d'obtenir une bonne information fréquentielle, la taille de la fenêtre étant alors trop petite pour ne pas tronquer les phénomènes de basses fréquences. À l'inverse, une fenêtre de grande taille (plus de 512 points) permettra d'obtenir une bonne information fréquentielle mais ne permettra pas d'obtenir une bonne information temporelle car tout événement, même de courte durée, est jugé présent sur l'ensemble du pas de temps analysé puisque la théorie de la transformée de Fourier considère les signaux indéfiniment stables.

Une fois la convolution effectuée, la transformée de Fourier est calculée sur la totalité de la fenêtre, le reste du « signal » étant alors égal à 0. Ce processus permet d'obtenir un spectre qui correspond à une trame, un ensemble de trames calculées à intervalles réguliers permettant d'obtenir le spectrogramme désiré.

Le rapport Signal-à-Interférence

Ce critère est fondé sur l'interférence entre la source originale recherchée et l'autre source originale que l'on voulait atténuée, voire éliminée. Nous appellerons ce critère SIR pour *Source-to-Interference Ratio* (en anglais). Le détail sur le calcul de ce rapport sera donné au chapitre quatre.

Perceptual Evaluation of Speech Quality : PESQ

Les performances des algorithmes en termes de qualité perçue de la donnée reçue sont évaluées au travers d'une mesure objective de la qualité fournie par l'outil PESQ. Cet algorithme compare le signal reçu et le signal original et attribue une note Mean Opinion Score (MOS) entre 0 et 4,5. L'explication détaillée de cette méthode d'évaluation que nous avons utilisée se trouve dans le chapitre quatre.

2.9 Conclusion

Le concept de la séparation aveugle de source a été introduit au milieu des années 1980. Son essor n'a pas cessé de grandir compte tenu de son caractère transversal dans le domaine du traitement de l'information. Dans ce chapitre, nous avons passé

en revue les principaux fondements permettant la mise au point d'approches traitant du problème de la séparation aveugle de sources. Des méthodes spécifiques ont été détaillées et des applications dans le contexte du traitement des signaux de parole exploitant la problématique de séparation aveugle de sources ont été présentées. Nous avons également proposé une formulation du problème de la séparation aveugle de sources et précisant les éléments théoriques qui s'y rattachent.

Nous retiendrons que la solution fournie par les méthodes de séparation aveugle n'est pas unique mais les sources séparées sont estimées à une permutation et à un facteur d'échelle dans les cas de mélanges instantanés ou à un filtrage près dans le cas de mélanges convolutifs.

Chapitre 3

Méthodes de décomposition en sous-espaces appliquées à la séparation de sources

3.1 Introduction

Ce chapitre est essentiellement consacré au développement de méthodes de séparation de sources basées sur des statistiques du second ordre et des statistiques d'ordre supérieur. Ce choix a été effectué après une importante recherche bibliographique menée dans le but de déterminer la méthode la plus adaptée afin de répondre à un de nos objectifs qui consistait à développer un dispositif de séparation aveugle de sources s'appliquant au cas difficile de signaux complexes tels que les mélanges de manière convolutive.

La séparation aveugle de sources possède de nombreuses applications, notamment dans les domaines du traitement du signal audio. Dans ce domaine, la séparation aveugle de sources concerne par exemple la séparation d'extraits musicaux stéréophoniques ou le rehaussement de la parole pour la téléphonie mobile ou pour l'élaboration de prothèses auditives.

La séparation aveugle de sources possède plusieurs degrés de difficulté, selon les caractéristiques des sources et surtout du mélange. Plusieurs paramètres sont à prendre en considération. Le premier paramètre est le nombre M d'observations par rapport au nombre N de sources. Dans cette partie de la thèse, nous détaillerons tous les calculs mathématiques concernant les méthodes du second ordre telles que l'analyse en composantes principales, l'analyse en composantes principales orientées ainsi que tout le détail des calculs concernant les méthodes utilisant les statistiques d'ordre supérieur telle que l'analyse en composantes indépendantes.

3.2 Analyse en composantes principales : PCA

Etant donné un nuage de n points en dimension d, l'analyse en composantes principales (en anglais, PCA : *Principal Components Analysis*) consiste à trouver la direction de l'espace qui porte la plus grande part de la dispersion totale ; puis la direction orthogonale à la précédente portant la deuxième plus grande part, et ainsi de suite. Du point de vue géométrique, c'est un simple changement de base, qui a pour effet une nouvelle représentation dans des directions correspondant à des variables décorrélées (au sens empirique), et obtenant par la même le statut de facteurs, susceptibles d'expliquer l'information contenue dans le nuage sans la redondance qui pouvait caractériser la représentation de départ. La figure 3.1 montre l'analyse en composante principale à deux dimensions. La ligne représente la direction de la première composante principale, qui donne une réduction linéaire optimale de deux à une dimension.

FIG. 3.1 – Analyse en composantes principales : Nuage de données à deux dimensions [94].

L'analyse en composante principale (PCA) est une méthode d'ordination classique. À partir d'un ensemble de m d'objets dans un espace de p descripteurs, son but est de trouver une représentation dans un espace réduit de k dimensions ($k << p$). Cela consiste à rechercher les directions de l'espace représentant les corrélations entre les variables aléatoires. Cette méthode est appelée « transformée de Karhunen-Loève ». Le coefficient de corrélation permet de caractériser les effets conjoints particulièrement lorsque l'on ne considère que deux effets. Son inconvénient, c'est de ne prendre en compte que les effets conjoints linéaires.

Nous pouvons représenter sur un plan à deux dimensions, les points disponibles suivant la loi conjointe de x_1 et x_2. La détermination des deux axes qui expliquent la dispersion des points disponibles, représente le résultat d'une PCA.

Pour le cas de trois effets x_1, x_2, x_3, nous tiendrons compte des trois coefficients de corrélations entre chaque variable c.à.d. entre x_1 et x_2, x_1 et x_3 et entre x_2 et x_3. Dans ce cas-ci, nous sommes en dimension supérieure à deux. La méthode PCA détermine toujours les axes qui montrent la dispersion du nuage des points disponibles et les ordonne par inertie. On peut projeter le nuage de dimension n sur un plan, et le visualiser, si la décision est de ne retenir que les deux premiers axes de la PCA.

Cette approche est utilisée, en particulier pour visualiser des données. Elle permet aussi :

- de les décorréler dans la nouvelle base constituée des nouveaux axes. Dans ce cas-là, les points ont une corrélation nulle ;

- de réduire le bruit qui les affectent, en considérant que les axes que l'on élimine sont des axes bruités ;

- de les classifier en amas corrélés.

3.2.1 Recherche des composantes principales

Soient les composantes : $C_1, C_2, \ldots, C_k, \ldots, C_q$,

C_k = nouvelle variable = combinaison linéaire des variables d'origine $x_1, ..., x_p$:

$C_k = a_{1k}x_1 + a_{2k}x_2 + \ldots ... + a_{pk}x_p$ coefficients a_{jk} à déterminer telle que les C_k soient :

- deux à deux non corrélées,

- de variance maximale,

- d'importance décroissante.

C_1 est la première composante principale qui doit être de variance maximale. Géométriquement, C_1 détermine une nouvelle direction dans le nuage de points qui suit l'axe d'allongement maximal du nuage.

C_2 est la deuxième composante, orthogonale à C_1 et de variance maximale. Géométriquement, C_2 détermine une droite perpendiculaire à C_1, suivant un axe (perpendiculaire au premier) d'allongement maximum.

=>C_1 et C_2 déterminent le plan principal (Voir figure 3.2) : le meilleur plan de projection (de distorsion minimum).

C_1 est telle que la moyenne des $d^2(\lambda_i, \lambda_{i'})$ soit maximum.

C_2 est orthogonal à C_1 et telle que la moyenne des $d^2(\beta_i, \beta_{i'})$ soit maximum.

=> C_1 et C_2 déterminent le plan tel que $d^2(f_i, f_{i'})$ soit maximum.

=> C_3 est la droite orthogonale à C_1 et C_2 telle que la variance des coordonnées soit maximum, et ainsi de suite.

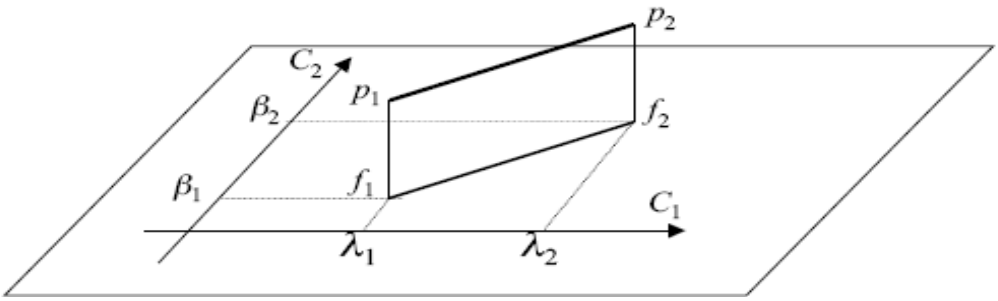

FIG. 3.2 – Plan de projection [95].

3.2.2 Réalisation

Le but est d'extraire le maximum de la variance projetée en trouvant les axes orthogonaux donc indépendants. (Voir Formulaire de l'Analyse en Composantes Principales en *Annexe C*). La procédure s'énonce ainsi :

- déterminer la matrice des données centrées-réduites,
- déterminer la matrice des corrélations entre variables,
- déterminer la matrice des vecteurs propres. Ces derniers sont les coordonnées du point de norme 1 des nouveaux axes sur les anciennes variables,
- déterminer le vecteur des valeurs propres,
- calcul des cordonnées sur les nouveaux axes,
- calcul des corrélations des variables avec les nouveaux axes.

Soit $x_1, \ldots, x_N$ un vecteur d'observations, connu à partir d'un échantillon de K réalisations conjointe, la méthode PCA s'applique comme suit :

$$M = \begin{bmatrix} X_{1,1} & . & . & . & X_{N,1} \\ . & . & & & . \\ . & & . & & . \\ . & & & . & . \\ X_{1,K} & & & & X_{N,K} \end{bmatrix}. \tag{3.1}$$

Les variables aléatoires $X_n = (X_{n,1}, \ldots, X_{n,K})$ ont chacune une moyenne et un écart type σ_{X_n}. Dans le calcul des caractéristiques d'un échantillon chaque élément de la matrice a la même importance $1/n$, si tous les éléments sont équiprobables. Un poids p_i différent à chaque réalisation conjointe des variables peut être appliqué. Ces poids sont représentés par une matrice diagonale D de taille K :

$$D = \begin{bmatrix} P_1 & . & . & . & 0 \\ . & P_2 & & & . \\ . & & . & & . \\ . & & & . & . \\ 0 & & & & P_K \end{bmatrix}. \tag{3.2}$$

Dans le cas où les poids sont égaux, nous aurons :

$$D = \frac{1}{K} I, \tag{3.3}$$

où I est la matrice identité.

La matrice M est généralement centrée :

$$\bar{M} = \begin{bmatrix} X_{1,1} - \bar{X}_1 & . & . & . & X_{N,1} - \bar{X}_N \\ . & . & & & . \\ . & & . & & . \\ . & & & . & . \\ X_{1,K} - \bar{X}_1 & & & & X_{N,K} - \bar{X}_N \end{bmatrix}. \tag{3.4}$$

Elle peut être aussi réduite :

$$\tilde{M} = \begin{bmatrix} \frac{X_{1,1} - \bar{X}_1}{\sigma(X_1)} & . & . & . & \frac{X_{N,1} - \bar{X}_N}{\sigma(X_N)} \\ . & . & & & . \\ . & & . & & . \\ . & & & . & . \\ \frac{X_{1,K} - \bar{X}_1}{\sigma(X_1)} & & & & \frac{X_{N,K} - \bar{X}_N}{\sigma(X_N)} \end{bmatrix}. \tag{3.5}$$

Le nuage de points peut être réduit. Dans ce cas-là une variable qui n'est qu'un bruit se retrouvera avec une variance égale à une variable informative.

Si on transforme la matrice M en $\tilde{M}$ puis on la multiplie par sa transposée on obtient : la matrice variance-covariance des $X_1, \ldots, X_N$. Si on la transforme en $\tilde{M}$, donc M sera centrée et réduite, nous obtenons la matrice de corrélation des $X_1, \ldots, X_N$ après la multiplication par sa transposée.

$$\text{Covariances} = \frac{1}{K \bar{M}' \bar{M}}, \ \text{Corrélations} = \frac{1}{K \tilde{M}' \tilde{M}}.$$

Ces deux matrices sont diagonalisables car elles sont carrées, symétriques et réelles.

Si nous supposons que le nuage est centré et/ou réduit, chaque variable X_n est donc remplacée par :

$$X_n - \bar{X}_n \quad \text{ou} \quad \frac{X_n - \bar{X}_n}{\sigma(X_n)}. \tag{3.6}$$

La matrice M sera notée $\bar{M}$ ou $\tilde{M}$ suivant le cas. Le principe de l'approche PCA est de trouver un axe u. Cet axe est obtenu à partir d'une combinaison linéaire des variables X_n, tel que la variance du nuage autour de cet axe soit maximale (voir décomposition et maximisation de la variance en *annexe D*).

La matrice de covariance sera diagonalisée, et ses éléments représentent les valeurs propres λ_i. La première valeur propre λ_1 représente la variance empirique sur le premier axe u de la PCA. Nous continuons ensuite, la recherche du deuxième axe de projection que nous appellerons w sur le même principe. w est orthogonal à u. la diagonalisation de la matrice de corrélation (ou de covariance), nous permet de dire que le premier vecteur propre est celui qui explique le plus d'inertie du nuage. De même le deuxième vecteur propre représente la plus grande part de l'inertie restante, etc.. . . Finalement, la méthode PCA représente la diagonalisation de la matrice de corrélation.

Si nous nous plaçons dans le cas instantané linéaire et où il y a autant de sources que de capteurs. Les sources sont notées $s_j(t)$ et les observations (mélanges) sont notés $x_i(t)$. On note la matrice de mélange.

$$A = \{a_{ij}\}. \tag{3.7}$$

On peut encore noter l'expression du mélange de l'équation 2.2 sous la forme matricielle suivante mais sans bruit additif $n(t)$:

$$X = AS \tag{3.8}$$

où l'on a noté, pour le mélange :

$$X = \begin{bmatrix} x_1(t_1) & . & . & . & x_1(t_N) \\ . & . & & & . \\ . & & . & & . \\ . & & & . & . \\ x_n(t_1) & & & & x_n(t_N) \end{bmatrix}, \tag{3.9}$$

et de même pour les sources :

$$S = \begin{bmatrix} s_1(t_1) & . & . & . & s_1(t_N) \\ . & . & & & . \\ . & & . & & . \\ . & & & . & . \\ s_n(t_1) & & & & s_n(t_N) \end{bmatrix}. \tag{3.10}$$

On définit alors la matrice de covariance empirique des observations par :

$$C = \frac{1}{N} X X^T. \tag{3.11}$$

La condition de décorrélation des sources implique que la matrice de covariance des sources soit diagonale comme vu précédemment. Il semble alors naturel de diagonaliser la matrice de covariance C et ainsi on peut écrire :

$$CUDU^T, \tag{3.12}$$

où U est une matrice orthogonale, $UUT = I$, et D est une matrice diagonale. Une solution consiste alors à poser :

$$\hat{S} = U^T X, \tag{3.13}$$

et on a :

$$\frac{1}{N} \hat{S} \hat{S}^T = U^T C U = D. \tag{3.14}$$

Les sources estimées $\hat{S}$ sont donc décorrélées, mais comme nous avons posé $A = U$, nous imposons à la matrice de mélange d'être orthogonale, ce qui n'a en fait aucune raison d'être dans la réalité. Nous pouvons d'autre part nous imposer la condition de sphéricité suivante :

$$\frac{1}{N} \hat{S} \hat{S}^T = I, \tag{3.15}$$

qui est une condition a priori plus forte. En fait, les sources ne peuvent être ré-estimées qu'à un facteur d'amplitude près et aux permutations près.

3.3 Analyse en composantes principales orientée (OPCA) et la séparation aveugle de sources

3.3.1 Définition

L'analyse en composantes principales orientée (OPCA : *Oriented Principales Components Analysis*) représente l'extension de l'analyse en composantes principales en maximisant la puissance de deux signaux plutôt que la puissance d'un seul signal. L'OPCA en combinaison avec un filtrage temporel arbitraire de longueur deux peuvent être utilisés pour la séparation aveugle de mélanges linéaires instantanés.

Par rapport à d'autres méthodes utilisant les statistiques du deuxième ordre, cette approche permet d'éviter l'étape de pré-blanchiment (en anglais : *prewhitening* ou

sphering). En outre, il a une convergence rapide avec une approche itérative permettant d'atteindre de meilleurs résultats en combinant plusieurs délais pour l'estimation des paramètres du mélange.

Des méthodes utilisant les statistiques d'ordre deux ont été proposées pour la séparation aveugle de sources par Belouchrani [46], cependant notons que les méthodes d'ordre supérieur ne s'appliquent pas aux signaux gaussiens alors que les méthodes utilisant le second ordre ne présentent pas une telle contrainte.

Diamantaras [96] a montré que pour les données observées et pré-filtrées temporellement, la PCA peut être utilisée comme solution de mélange instantanée dans la séparation aveugle de sources. La méthode a besoin d'une étape de mesure spatiale de pré-blanchiment sur l'observation des données. Diamantaras [97] a montré que, sans pré-blanchiment le problème est un exemple typique de l'OPCA.

Oriented PCA, une extension de la PCA peut être utilisée en opposition au problème de séparation aveugle de sources dans le cas instantané. Les données obtenues par cette méthode doivent-être pré-filtrées temporellement, et n'importe quel filtre sera suffisant.

Les avantages de l'approche OPCA sont résumés comme suit :

i- en utilisant un seul décalage dans le temps, c'est une approche différée sans itérations récursives et lentes.

ii- peut combiner plus de deux décalages en temps pour une meilleure performance dans un algorithme itératif avec une convergence rapide.

iii- ne nécessite pas de pré-blanchiment.

3.3.2 Algorithme OPCA de K. I. Diamantaras et Th. Papadimitriou

L'OPCA est une version étendue de la technique classique d'analyse en composante principale. Les auteurs ont développé un algorithme qui utilise les statistiques d'ordre deux. Comme la PCA, qui correspond à la décomposition en valeurs propres d'une seule matrice de covariance, selon les auteurs, l'OPCA correspond à la décomposition d'une paire de matrices de variance-covariance en valeurs propres. Ainsi, l'objectif de cette technique est de maximiser la fonction de coût donnant le rapport entre une paire de signaux n-dimensionnels u et v :

$$J_{OPCA}(w) = \frac{E(w^T u)^2}{E(w^T v)^2} = \frac{w^T R_u w}{w^T R_v w}, \tag{3.16}$$

où u, v sont deux vecteurs signal, R_u et R_v sont respectivement leurs matrices de covariance et w le « maximiseur » de cette fonction de coût, on l'appelle principale composante orientée (premier vecteur propre).

La PCA d'un vecteur aléatoire $x(k) \in R_m$ est obtenu en retrouvant un système de coordonnées réalisant une opération d'orthogonalisation de sorte que les composantes de x dans ce système deviennent décorrélées, orthogonales et ordonnées en énergie croissante. Afin d'appliquer l'OPCA, l'idée des auteurs est de considérer un nouveau signal (y) issu du filtrage des observations par un filtre $h = [h_0, ..., h_M]$:

$$y(k) = \sum_{m=0}^{M} h_m x(k - l_m). \tag{3.17}$$

Ensuite, sous l'hypothèse de sources colorées avec différentes densités spectrales, ils cherchent la relation entre les observations x et leurs filtrés y à partir de leur matrice de covariance au retard $l_0 = 0$. Ainsi, par des substitutions ils trouvent la relation suivante :

$$R_y(0)A^{-T} = R_x(0)A^{-T}D, \tag{3.18}$$

où

$$D = \sum_{p,q=0}^{M} h_p h_q R_s(l_p - l_q) \qquad (\ l_p,\ l_q \text{ sont deux retard}). \tag{3.19}$$

Désormais, le problème revient à appliquer l'OPCA sur la paire $\{R_y(0), R_x(0)\}$ réalisant leur décomposition généralisée en valeurs propres. Les valeurs propres généralisées de ce problème correspondront aux éléments diagonaux de D et les colonnes de la matrice A^{-T} seront données par les vecteurs propres généralisés. Le détail de la méthode OPCA sera donné dans le chapitre quatre.

3.4 Analyse en Composantes Indépendantes : ICA

L'essor des méthodes à base d'analyse en composantes indépendantes (en anglais ICA : *Independent Components Analysis*), en particulier depuis l'article « An information-maximization approach to blind separation and blind deconvolution » [73] de Bell et Sejnowski en 1995, en font une technique de base incontournable pour la séparation de source. Nous commencerons par présenter, dans le cadre de mélanges linéaires et instantanés, le cas où il y a autant de capteurs que de sources. Nous nous intéresserons ensuite aux cas de mélanges convolutifs.

L'étude de l'analyse en composantes indépendantes représente une introduction dans le cadre du problème posé dans ce mémoire de thèse, et elle est nécessaire pour établir des repères concernant notre travail et d'autre part en raison de son succès pour certains problèmes de séparations de sources, en pratique.

Depuis plus d'une dizaine d'années, une petite communauté du domaine du traitement du signal développe des algorithmes pour l'analyse en composantes indépendantes et la séparation de sources. Il s'agit de méthodes traitant des observations vectorielles afin d'en extraire des composantes linéaires qui soient aussi indépendantes que possible. Cette méthode a des liens évidents avec l'analyse de données (voir ci-dessous une comparaison entre ICA et PCA), avec les méthodes neuromimétiques (codage des flux d'information) et l'analyse harmonique computationnelle (recherche des "bonnes bases" de représentation).

L'analyse en composantes indépendantes est un prolongement de l'analyse en composantes principales, mais dans laquelle la décorrélation des observations est remplacée par une hypothèse plus forte qui est l'indépendance des sources. C'est une méthode d'analyse de données connue dans la séparation aveugle de sources appliquée à divers domaines.

L'exemple emblématique du "*cocktail party*" (Figure 3.3) illustre le point de vue de la séparation de sources : on enregistre N conversations simultanées en plaçant N microphones bien répartis dans la pièce, chacun enregistrant une superposition de toutes les conversations, mais un peu plus nettement celles qui se trouvent à proximité directe. Le problème est de retrouver la voix de chacune des personnes débarrassée des autres voix qui la perturbe donc isoler chacun des discours pour comprendre ce qui s'est dit.

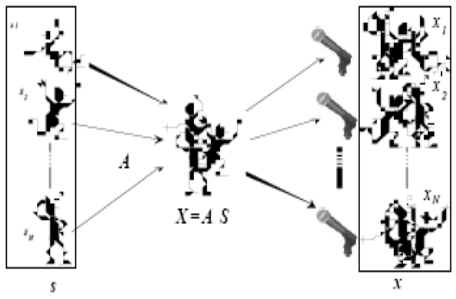

FIG. 3.3 – Problème du "*cocktail party*".

L'ICA permet de résoudre ce problème en supposant que les personnes qui parlent ont des discours indépendants. Ces discours sont considérés comme des signaux aléatoires statistiquement indépendants, ne tenant compte ni de la sémantique de ces discours ni de l'acoustique. L'ICA formulé ainsi, considère uniquement des superpositions linéaires de signaux indépendants, résultant du mélange par A. C'est souvent une restriction acceptable ; par exemple les systèmes de transmission sont des milieux

linéaires où les signaux agissent comme s'ils étaient présents indépendamment les uns des autres, ils n'interagissent pas mais s'additionnent [99].

D'autres formulations prennent en compte des mélanges dits post non linéaires [100] [101] [102] issus de transformations monotones de mélanges linéaires s'exprimant par $X = f(AS)$, f de composantes monotones et A inversible. Il existe aussi des modèles convolutifs, où le mélange par A n'est pas instantané. Enfin, nous pouvons aussi considérer des modèles bruités.

3.4.1 Définition

- On observe grâce à plusieurs capteurs des sources qui sont «mélangées».

- Le mélange a lieu par l'intermédiaire d'une matrice inconnue : la matrice de mélange A tel qu'indiqué dans les équations 3.7 et 3.8.

Le modèle statistique dans l'équation 3.8 est appelée analyse en composantes indépendantes. Ces dernières ne peuvent pas être directement observées. Aussi la matrice de mélange est supposée non connue. Tout ce que nous observons c'est un vecteur aléatoire x et nous devons estimer A et s à partir de ce vecteur, avec la seule connaissance a priori : l'indépendance statistique. Nous devons supposer aussi que les composantes ont des distributions non-gaussiennes.

3.4.2 Indépendance

- Deux variables aléatoires X et Y sont indépendantes si :

$$P(X = x \; et \; Y = y) = P(X = x)P(Y = y), \tag{3.20}$$

cette définition ne peut pas être utilisée directement.

- Des variables indépendantes sont toujours décorrélées (covariance nulle) mais, l'inverse n'est pas toujours vrai :

$$E(XY) - E(X)E(Y) = 0 \tag{3.21}$$

=> Permet une simplification de la recherche des composantes cependant elle n'est pas très intéressante en pratique car il existe trop de possibilités de décorréler des variables.

3.4.3 Principales approches du problème ICA

Des méthodes issues du maximum de vraisemblance et des contrastes basés sur l'information mutuelle ou autre mesures de divergence ont souvent été utilisées. Le

concept de l'ICA est défini comme la maximisation du degré d'indépendance statistique entre sorties.

Non-gaussianité des variables

Pour que l'analyse en composantes indépendantes soit possible, les composantes indépendantes doivent être non-gaussiennes, ce qui représente une restriction fondamentale de l'ICA.

Pour voir cela, nous supposons que la matrice du mélange est orthogonale et les signaux si sont gaussiens, et x_1 et x_2 sont gaussiens, non corrélés, de variance unitaire. Leur densité conjointe est donnée par :

$$p(x_1, x_2) = \frac{1}{2\pi} \exp\left(-\frac{x_1^2 + x_2^2}{2}\right). \tag{3.22}$$

La distribution de toute transformation orthogonale de (x_1, x_2) gaussiens a exactement la même distribution que (x_1, x_2), et X_1et X_2 sont indépendants. Ainsi, dans le cas de variables gaussiennes, le modèle ICA ne peut être estimé qu'avec une transformation orthogonale. En d'autres termes, la matrice A n'est pas identifiable dans le cas des composantes indépendantes gaussiennes. Par contre, le modèle ICA peut être estimé si seulement une des composantes indépendantes est gaussienne. La non-gaussianité représente la clé de l'estimation du modèle ICA et l'analyse est impossible sans son estimation.

Maximisation du caractère non gaussien

Les sommes de variables aléatoires indépendantes ont tendance à devenir gaussiennes. L'idée est de rechercher des sources les moins gaussiennes possibles. En effet un mélange étant une combinaison de plusieurs sources supposées indépendantes, le mélange a tendance à être plus proche d'une gaussienne que les sources. Il faut donc un critère pour mesurer la «gaussianité» d'une variable aléatoire. On cherchera alors à maximiser le caractère non gaussien d'une combinaison linéaire de l'observation x. Il est estimé à partir du kurtosis ou de la néguentropie.

Il s'agit d'une méthode par déflation, c'est-à-dire soit on opère composante par composante, par opposition aux approches simultanées soit on cherche directement W permettant d'extraire toutes les composantes en même temps. Cela entraîne généralement que l'erreur dans l'estimation de la première composante s'accumule et augmente l'erreur des composantes suivantes [103]. C'est également une méthode peu robuste en présence de données aberrantes, en raison d'une croissance cubique en $|y|$ du kurtosis.

La maximisation s'opère par une méthode du gradient simple ou par une méthode de point fixe (Fast-ICA).

La performance d'un algorithme n'est pas mesurée spécifiquement par le degré de suppression de la dépendance, mais par sa capacité à inverser la matrice A.

Mesure de non-gaussianité utilisant le kurtosis

La mesure classique de la non-gaussianité est le kurtosis ou le cumulant d'ordre 4.

- 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 : $s = w^T x$ 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 décroit 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.

* On recommence...

Néguentropie

La deuxième méthode importante de la mesure de la non-gaussianité est donnée par la néguentropie. Cette dernière est basée sur la quantité théorique de l'information de l'entropie et est un concept de base de la théorie de l'information.

L'entropie d'une variable aléatoire peut être interprétée par le degré de l'information que l'observation de la variable donne. Plus la variable est « aléatoire », non structurée et imprévisible et plus l'entropie est grande.

- L'entropie H d'une variable aléatoire est définie par :

$$H(X) = -\int f(X)\log(X)dy. \tag{3.23}$$

- Pour une variable aléatoire discrète :

$$H(x) = -\sum_i p\left(X = a_i\right)\log P\left(X = a_i\right). \tag{3.24}$$

Une variable aléatoire gaussienne a l'entropie la plus grande parmi toutes les variables aléatoires possédant la même variance [104] [105]. Ceci montre que l'entropie

peut être utilisée pour mesurer la non-gaussianité. En fait, ceci montre que la distribution gaussienne, est la « plus aléatoire » de toutes les distributions. L'entropie est petite pour les distributions qui sont concentrées sur certaines valeurs.

Pour obtenir une mesure de non-gaussianité qui est zéro pour une variable gaussienne et toujours positive, nous utilisons souvent une version modifiée de la définition de la dérivée de l'entropie, appelée néguentropie.

On définit donc la néguentropie J pour caractériser le caractère non gaussien d'une variable par :

$$J(x) = H(x_{gauss}) - H(x), \tag{3.25}$$

où x_{gauss} est une variable aléatoire Gaussienne de même matrice de covariance que x. La néguentropie est toujours positive, et elle est nulle si et seulement si x a une distribution gaussienne.

Estimation de la Néguentropie

- Difficile à estimer car le calcul est très compliqué. Il faut une estimation de la fonction de répartition.

- la méthode classique de l'approximation de la néguentropie est l'utilisation des moments d'ordre supérieur comme par exemple :

$$J(x) \approx \frac{1}{12E(x^3)^2} + \frac{1}{48kurt(x)^2}. \tag{3.26}$$

=> Mêmes problèmes que le kurtosis.

- Il existe des estimateurs robustes [105] :

$$J(x) \approx \sum_{i=1}^{p} k_i (E\{G(x)\} - E\{G(\nu)\})^2, \tag{3.27}$$

avec ν suivant une loi normale : $N(0,1)$ et G une fonction bien choisie :

- $G_1(y) = y^4$ on retrouve l'estimateur classique.
- $G_2(y) = 1/a \log(\cosh(ay))$ avec $1 \leq a \leq 2$.
- $G_3(y) = -\exp(y^2/2)$.

Ainsi sont obtenues des approximations de la néguentropie qui donnent un bon compromis entre les propriétés des deux mesures classiques de non-gaussianité donnée par le kurtosis et la néguentropie.

Minimisation de l'information mutuelle

Une autre approche pour l'estimation de l'ICA, inspiré par la théorie de l'information, est la minimisation de l'information mutuelle. Utilisant le concept de la dérivée de l'entropie, on définit l'information mutuelle I entre N variables aléatoires si avec $i = 1 \ldots..N$ comme suit :

$$I\left(s_1, s_2, ..., s_n\right) = \sum_i H(s_i) - H(s). \tag{3.28}$$

Une propriété importante de l'information mutuelle est que nous avons pour une transformation linéaire inversible :

$$I\left(s_1, s_2, ..., s_n\right) = \sum_i H(s_i) - H(x) - \log|\det W\ |, \tag{3.29}$$

avec W la matrice de séparation. Si s_i décorrélés et de variance unitaire :

$$E\left\{ss^T\right\} = WE\left\{xx^T\right\}W^T = I, \tag{3.30}$$

ce qui implique :

$$\det I = 1 = \left(\det WE\left\{xx^T\right\}W^T\right) = (\det W)(\det\left\{xx^T\right\})(\det W^T), \tag{3.31}$$

et donc $detW$ doit être constant. De plus, pour s_i de variance unitaire, l'entropie et la néguentropie diffèrent seulement d'une constante et de signe. Ainsi nous avons :

$$I\left(s_1, s_2, ..., s_n\right) = C - \sum_i J(s_i), \tag{3.32}$$

où C est une constante ne dépendant pas de W. Ceci montre la relation entre la néguentropie et l'information mutuelle.

Maximum de vraisemblance

Estimer un paramètre par le maximum de vraisemblance consiste à proposer la valeur pour laquelle la probabilité de l'observation dans le modèle est la plus forte.

Définition : Soit une distribution continue de densité de probabilité $p(x)$,

* Pour chacune des observations x_i, on mesure la valeur $p(x_i)$ que la densité prend pour cette observation,

* et on fait le produit de ces valeurs, que l'on appelle la vraisemblance de la distribution pour l'échantillon concerné.

Elle est désignée traditionnellement par la lettre "L"en raison de l'appellation anglo-saxonne de la vraisemblance, qui est "*Likelihood*".

Par définition, la vraisemblance d'un échantillon $\{x_1, ..., x_M\}$ est un produit de probabilités ou de densités et s'écrit alors :

$$p\left(x_i\right) L_D\left(x\right) = \prod_i p\left(x_i\right). \tag{3.33}$$

Pour trouver le maximum, il faut déterminer les valeurs pour lesquelles la dérivée s'annule. Or d'après l'équation 3.33, le calcul de la dérivée sera assez complexe. Il est plus simple de dériver une somme, et pour cela on remplace la vraisemblance par son logarithme. Comme la fonction logarithme est croissante, il est équivalent de maximiser :

$$\log(L(x_1, x_2, ..., x_n, \phi)), \tag{3.34}$$

ou

$$L(x_1, x_2, ..., x_n, \phi). \tag{3.35}$$

On détermine une valeur de ϕ *pour* laquelle la dérivée s'annule, et à l'aide de la dérivée seconde il faut s'assurer que ce point est bien un maximum.

Une approche très populaire pour l'estimation d'un modèle ICA est l'estimation du maximum de vraisemblance. Il est possible de formuler directement la probabilité du modèle ICA sans bruit qui a été fait en [80], puis estimer le modèle par une méthode du maximum de vraisemblance.

En notant $W = (w_1, ...w_n)^T$ la matrice A^{-1}, le log de la vraisemblance devient [80] :

$$L = \sum_{t=1}^{T} \sum_{i=1}^{n} \log f_i\left(w_i^T x(t)\right) + T \log\left|\det W\right|. \tag{3.36}$$

où :

- f_i représentent les fonctions de densité de s_i (ici supposés connus) et $x(t), t = 1, \ldots .T$ sont les réalisations de x.

- le terme $log|detW|$ dans la vraisemblance vient des règles classiques pour la transformation linéaire des variables aléatoires et leurs densités [19] : en général, pour chaque vecteur aléatoire x avec une densité p_x et pour chaque matrice W, la densité

de $y = Wx$ est donnée par $p_x(Wxlog|detW|)$.

L'algorithme basé sur un réseau de neurones, est assimilable à une méthode de maximum de vraisemblance.

Infomax (maximum d'information)

Le problème d'inversion du mélange des signaux est considéré ici comme une maximisation de l'entropie, $H(s)$, où s est un signal transformé non-linéaire. L'entropie conjointe est définie par :

$$H(s_1, s_2) = H(s_1) + H(s_2) - I(s_1, s_2). \tag{3.37}$$

La maximisation de l'entropie conjointe est la même que la maximisation des entropies individuellement tout en minimisant l'information mutuelle, $I(s_1, s_2)$, partagée entre les deux signaux. Lorsque l'information mutuelle atteint zéro, les deux variables sont statistiquement indépendantes. Rendre les sorties indépendantes c'est maximiser l'entropie conjointe $H(s)$.

$$H(s) = \sum_{i=1}^{N} H(s_1) - I(s). \tag{3.38}$$

L'entropie des différentes sorties est :

$$H(s_i) = -E\left[\log p(s_i)\right], \tag{3.39}$$

où $p(s_i)$ représente la fonction de densité de probabilité de la sortie. Cependant d'après [107] un moyen plus efficace de maximiser l'entropie conjointe est d'utiliser la pente naturelle, car elle réduit la complexité et accélère la convergence de l'algorithme.

Méthodes tensorielles

Pour supprimer les corrélations jusqu'à l'ordre 4, de $x = (x^1, ..., x^d)$ on considère les tenseurs de cumulants $cum(x^i, x^j, x^k, x^l)$ qui généralisent les matrices de covariance $cov(x^i, x^j)$

Soit F le tenseur d'ordre 4, agissant sur des matrices $d \times d$, défini par :

$$M, d \times d \rightarrow F_{ij}(M) = \sum_{k,l} M_{kl} cum(x^i, x^j, x^k, x^l). \tag{3.40}$$

Une matrice propre de F est une matrice M telle que $F(M) = \lambda M,\ \lambda \in R$. On montre que les matrices orthogonales $W = w_m^t w_m, m = 1, \ldots, n$, sont les matrices propres de F. Les valeurs propres associées sont les kurtosis des composantes indépendantes [106].

3.5 Algorithme ICA

Le premier algorithme d'ICA est basé sur un réseau de neurones récursif [14] dont les poids sont les termes non diagonaux de la matrice de séparation W, les termes diagonaux étant contraints à la nullité. L'estimation des sources est ainsi donnée par :

$$\hat{s} = (I + W)^{-1} x. \tag{3.41}$$

Pour les termes non diagonaux nous avons la règle d'adaptation suivante :

$$\Delta \varpi_{ij} = f(s_i) g(s_j), \tag{3.42}$$

$f(.)$ et $g(.)$ sont des fonctions non linéaires impaires différentes. La mesure d'indépendance est l'annulation des cumulants d'ordre supérieur.

L'analyse en composantes indépendantes se propose d'utiliser des moments d'ordre supérieur à deux dans le but d'obtenir une indépendance statistique des sources et notamment de lever l'indétermination qui existe si l'on se contente de L'analyse en composantes principales. On peut voir l'ICA comme une méthode liée à la théorie de l'information et notamment au concept d'information mutuelle, mais il existe bien d'autres approches pour ce domaine.

Pour illustrer cela, nous nous plaçons dans le cas instantané linéaire où il y a autant de sources que de capteurs N. On suppose cette fois que les sources sont indépendantes, c'est-à-dire que les densités se factorisent, soit :

$$p(s_1(t), ..., s_N(t) = \prod_{i=1}^{N} p_i(s_i(t)). \tag{3.43}$$

Un moyen d'arriver à cette indépendance statistique, consiste à minimiser l'information mutuelle [108], dont l'expression est la suivante :

$$I(s) = \int \log \left[\frac{p(s_1(t), ..., s_N(t)}{\prod_{j=1}^{N} p_j(s_j(t)} \right] p(s_1(t), ..., s_N(t) ds_1 ... ds_N. \tag{3.44}$$

En effet, l'information mutuelle est une quantité toujours positive et lorsque $I(S)$ est nulle, alors les sources sont statistiquement indépendantes : c'est en quelque sorte une mesure de l'indépendance statistique des sources. Il est donc naturel de minimiser cette quantité dans le cadre du problème posé par l'ICA. Nous utiliserons simplement un algorithme du gradient, car le calcul de cette quantité (l'information mutuelle) n'est pas simple, alors qu'au niveau de sa dérivée les choses se simplifient naturellement.

Après calculs [109], on obtient une fonction de contraste à minimiser de la forme :

$$\begin{aligned} c(W) &= \frac{1}{T} f(S) S^T - I, \\ S &= WX, \end{aligned} \tag{3.45}$$

où I est la matrice identité et f est une fonction non linéaire, liée aux densités marginales des sources. On utilise alors la règle de mise à jour de la matrice W suivante (inverse de la matrice de mélange), en utilisant le gradient naturel [107] :

$$W^{k+1} = W^k - \nu_k \left[\frac{1}{T} f(S^k) S^{kT} - I \right] W^k. \tag{3.46}$$

Lorsque le contraste s'annule, c'est-à-dire :

$$\frac{1}{T} f(S^*) S^{*T} = I, \tag{3.47}$$

avec

$$S^* = W^* X, \tag{3.48}$$

un moment d'ordre supérieur (pour des fonctions f_i non linéaires) est diagonal.

Cela généralise d'une part la condition de décorrélation à l'ordre deux et de plus la connaissance exacte des fonctions f_i, qui sont liées aux lois des sources séparées, n'est pas forcément nécessaire [110] sur ces aspects, d'un point de vue théorique.

Notons qu'il existe d'autres méthodes d'ordre 2 permettant d'assurer l'indépendance des sources [32]. Ces méthodes sont fondées sur la diagonalisation jointe de matrices de covariance et sont utilisées notamment dans le cas de sources non stationnaires.

Le détail sur les mélanges instantanés et convolutifs est présenté dans les chapitres suivants.

3.5.1 Prétraitement pour l'ICA

Avant d'appliquer l'algorithme ICA sur les données, il est généralement très utile de faire un prétraitement. Certaines techniques de prétraitement rendent l'estimation du problème de l'ICA plus simple et mieux conditionnée.

Centrage

Le prétraitement le plus élémentaire et nécessaire est de centrer x, c'est-à-dire soustraire sa moyenne $m = E\{x\}$ de manière à rendre la variable x de moyenne nulle. Cela implique que s ait une moyenne nulle. Ce prétraitement sert uniquement à simplifier les algorithmes ICA, cela ne signifie pas que la moyenne ne peut pas être estimée. 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. La valeur moyenne du vecteur s est donnée par $A^{-1}m$, où m est la valeur moyenne qui a été soustraite dans le prétraitement.

Blanchiment

Une autre stratégie de prétraitement utile à l'ICA est d'abord de blanchir les variables observées. Cela signifie qu'avant l'application de l'algorithme ICA (et après centrage), nous transformons le vecteur x observé linéaire 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 :

$$E\left\{\tilde{x}\tilde{x}^T\right\} = I. \tag{3.49}$$

Le blanchiment est toujours possible. Une méthode populaire pour cela est d'utiliser la décomposition des valeurs propres de la matrice de covariance $E\{xx^T\} = EDE^T$, où E représente la matrice orthogonale des vecteurs propres de $E\{xx^T\}$ et D matrice diagonale de ses valeurs propres, $D = diag(\lambda_1, \ldots \lambda_n)$. À noter que $E\{xxT\}$ peut être estimée de façon standard à partir des échantillons $x(1), \ldots, x(T)$. Le blanchiment peut être réalisé par :

$$x = ED^{-1/2}E^T x, \tag{3.50}$$

où la matrice $D^{1/2} = \text{diag}(\lambda_1^{-1/2}, \ldots \lambda_n^{-1/2})$. Le blanchiment transforme la matrice de mélange en une nouvelle matrice . Nous avons alors à partir des équations 3.50 et 3.8 :

$$\tilde{x} = ED^{-1/2}E^T As = \tilde{A}s. \tag{3.51}$$

L'utilité du blanchiment réside dans le fait que la nouvelle matrice de mélange est orthogonale, et ceci peut se voir par :

$$E\left\{\tilde{x}\tilde{x}^T\right\} = \tilde{A}E\left\{ss^T\right\}\tilde{A}^T = I. \tag{3.52}$$

On voit que le blanchiment permet de réduire le nombre de paramètres à estimer. Au lieu d'avoir à estimer n^2 paramètres qui sont les éléments de la matrice originale A, il suffit d'estimer la nouvelle matrice orthogonale mélange contenant $n(n-1)/2$ degrés de liberté. Ainsi, on peut dire que le blanchiment résout la moitié du problème d'ICA, car c'est une procédure standard et très simple, beaucoup plus simple que tout algorithme ICA, ce qui représente une bonne idée de réduire la complexité du problème de cette façon. (Voir l'influence du blanchiment sur les performances de la séparation en *annexe F*)

Il peut également être très utile de réduire la dimension des données en même temps que le blanchiment. Ensuite, prendre les valeurs propres λ_j de $E\{xx^T\}$ et enlever celles qui sont trop petites, comme l'opération réalisée par la technique des statistiques de l'analyse en composantes principales ; cela a souvent pour effet de réduire le bruit. De plus, la réduction de dimension empêche le sur-apprentissage, qui peut parfois être observé dans l'ICA [111]. Pour ces raisons là, certains algorithmes combinent les deux méthodes l'ICA et la PCA [112].

3.6 Principe de base de l'algorithme Fast-ICA

Un algorithme très efficace utilisant la méthode ICA, appelée Fast-ICA (*Fast Independent Component Analysis*), a été développé en 1997 par Oja et Hyvarinen [24]. Il est aussi l'algorithme de l'ICA le plus couramment utilisé pour la séparation de mélanges instantanés de sources. Dans le chapitre suivant, Fast-ICA sera utilisée comme référence pour effectuer des évaluations objectives et qualitatives.

"Fast-ICA" met en évidence des liens forts entre l'analyse en composantes indépendantes et la poursuite de projection. Cette dernière est une méthode d'analyse des données qui relève des statistiques. Elle a été initialement proposée par Jérôme H. Friedman and John Tukey [105] en 1974. Elle consiste à rechercher un sous espace dans lequel un indice de projection est maximisé. L'idée est alors de trouver des directions intéressantes pour la représentation des données ; c.à.d. exprimer les données selon un premier axe dans lequel les données sont bien représentées puis faire de même avec le reste des données sur un nouvel axe et poursuivre itérativement.

Fast-ICA est un algorithme très performant de maximisation de la fonction de contraste pour les sources non gaussiennes [26], [106]. Il est basé sur le principe de l'algorithme d'apprentissage itératif de type point fixe (Fixed point algorithm). L'algorithme d'apprentissage du réseau cherche à trouver une direction qui est en fait un vecteur w tel que la projection $w^T x$ maximise l'aspect non gaussien. Pour cela, l'algorithme utilise le théorème du point fixe. Si on considère que chaque sortie du réseau est donnée sous forme d'une fonction de type g_i $(w^T x)$ tel que g_i est une fonction scalaire

non linéaire, la forme simplifiée de l'algorithme pour l'estimation d'une composante indépendante peut être reformulée de la manière suivante :

- initialiser le vecteur w par de faibles valeurs aléatoires ;
- mettre $w^+ = E\left(xg(w^T x)\right) - E\left(g'(w^T x)\right) w$;
- mettre $w = w^+ / \|w^+\|$;
- si la convergence n'est pas encore atteinte, refaire les étapes 2 et 3.

La convergence signifie que les valeurs de w (anciennes et nouvelles) vont dans la même direction, c'est-à-dire que la différence entre ces deux valeurs est au-dessous du critère de convergence entre deux itérations.

Pour estimer plusieurs composantes, on a besoin d'utiliser l'algorithme suivant avec plusieurs neurones ayant les vecteurs synaptiques $w_1, w_2, ...w_n$. Pour que les vecteurs ne convergent pas vers un même maximum, nous devons décorréler les projections $w_1^T x, w_2^T x, ..., w_n^T x$, après chaque itération. Il existe plusieurs méthodes de décorrélation. La plus populaire est basée sur la méthode de Gram-Shmidt de décomposition de matrices. Avec cette méthode, après l'estimation de N composantes indépendantes $(w_1, w_2, ...w_N\ )$, on utilise l'algorithme du point fixe pour estimer le vecteur w_{N+1} et, après chaque itération, on soustrait du vecteur w_{p+1} les projections $w_{N+1}^T w_j w_j\ (j = 1, ..., N)$, puis on normalise de nouveau le vecteur w_{N+1}.

La performance de l'algorithme Fast-ICA dépend du choix de la fonction g. Les composantes indépendantes peuvent être extraites une à une. Ainsi, on obtient un gain considérable en performance si l'on a seulement besoin de quelques composantes indépendantes à extraire.

Le choix de l'algorithme ICA est essentiellement un choix entre un algorithme adaptatif et un algorithme en mode bloc.

Dans le cas adaptatif, les algorithmes sont obtenus par des méthodes du gradient stochastique. Dans le cas où toutes les composantes indépendantes sont estimées au même instant, l'algorithme le plus populaire dans ce cas là est le gradient naturel de vraisemblance, ou en lien avec les fonctions de contraste, comme InfoMax [107], [73], Dans le cas d'une seule composante, des méthodes du gradient stochastique simple, donnent des algorithmes d'adaptation qui maximisent la néguentropie ou ses approximations [104], [106].

3.6.1 L'algorithme C-FICA

L'algorithme C-FICA est une extension dans le domaine temporel de l'algorithme Fast-ICA. C'est un algorithme temporel rapide effectuant la séparation aveugle dans

le cas de mélanges convolutifs, contrairement à Fast-ICA qui est utilisé dans le cas de mélanges instantanés. Le détail concernant cet algorithme est présenté dans le cinquième chapitre.

3.7 Quelques différences et similarités entre PCA et ICA

• L'analyse en composantes indépendantes ne peut rien apporter à l'analyse de données gaussiennes puisque l'indépendance se réduit alors à la décorrélation. On pourrait dire que l'ICA est l'art d'intégrer l'information non gaussienne dans la recherche de composantes.

• La PCA est une décomposition unique grâce à la combinaison de deux contraintes : $C(A^{-1}X) = 0$ et $A^T A = 0$. Cette dernière contrainte d'orthogonalité n'a rien de physique : si des composantes existent (ont une réalité physique), il n'y a en général pas de raison de supposer leur orthogonalité géométrique.

• On peut voir la PCA comme une décomposition hiérarchique, qui extrait des composantes (décorrélées, orthogonales) par ordre décroissant d'énergie. Par contraste, l'ICA est complètement invariante par le changement d'échelle.

• L'objectif implicite de l'ICA est souvent de trouver des composantes «physiquement significatives». C'est à la fois une force et une faiblesse : les objectifs de la PCA sont plus modestes, mais ils sont toujours atteints : la PCA ne suppose pas, ne cherche pas, ne trouve pas (en général) de composantes ayant une réalité physique.

• L'algorithmique de la PCA est simple à comprendre et à mettre en œuvre. Par contre, il n'y a pas d'algorithme d'ICA « universel » car il n'est pas direct, ne définit pas des versions empiriques du critère $I(S)$ ni les optimise.

• La PCA est une méthode utilisant les statistiques d'ordre deux qui permettent de décorréler les données et réduire la dimension du problème, mais ne permet pas de réaliser l'indépendance totale. Par contre l'ICA a la capacité de réaliser l'indépendance totale des signaux en utilisant les statistiques d'ordre supérieur.

3.8 Comparaison entre PCA et OPCA

L'OPCA est un terme introduit par Kung et Diamantaras [97], comme une généralisation de la PCA. Elle correspond à la décomposition généralisée des valeurs propres

d'une paire de matrices de covariance de la même manière que la PCA correspond à la décomposition des valeurs propres d'une seule matrice de covariance. La fonction coût maximisée par l'OPCA est le rapport signal à signal entre deux signaux u et v de dimension n :

$$J_{OPCA}(w) = \frac{E\left(w^T u\right)^2}{E\left(w^T v\right)^2} = \frac{w^T R_u w}{w^T R_v w}, \tag{3.53}$$

où $R_u = E\{uu^T\}$ et $R_v = E\{vv^T\}, w = e_1$ du J_{OPCA} est appelé composante principale orientée et elle représente la valeur propre généralisée de la matrice $[R_u, R_v]$ correspondant à la valeur propre maximum généralisée λ_1. Comme R_u, et R_v sont symétriques toutes les valeurs propres généralisées sont réelles et, par conséquent, elles peuvent être organisées par ordre décroissant, comme dans le cas du PCA. Les vecteurs propres généralisés $e_1, e_2, ..., e_n$, seront appelés deuxième, troisième, ..., $n^{ème}$ composantes principales orientées. Tous ces vecteurs maximisent la fonction coût J_{OPCA} soumis à la contrainte orthogonalité suivante :

$$e_i^T R_u e_j = e_i^T R_v e_j = 0, \quad i \neq j. \tag{3.54}$$

Le terme "orientée" est justifié par le fait que e_1 est semblable à la composante principale ordinaire pour u à l'exception qu'elle est orientée vers la direction la moins principale de v. En d'autres termes, la distribution de v "conduit" e_1 dans la direction du minimum d'énergie de v. Il est clair que si la distribution de v est isotrope dans toutes les directions, c'est-à-dire que v est un bruit blanc, alors la direction est absente et les composantes principales orientées sont identiques à celles du PCA ordinaire. Ceci est facilement vérifié mathématiquement en posant : $R_v = I$ dans J_{OPCA} en l'identifiant à la fonction coût maximisée par la PCA.

3.9 Conclusion

Dans ce chapitre nous avons présenté le principe des méthodes de décomposition en sous-espaces appliquées à la séparation de sources. Ces méthodes sont : PCA, OPCA et ICA.

Nous avons tout d'abord, présenté le principe de l'analyse en composantes principales linéaires. L'idée de base du PCA est de réduire la dimension de la matrice des données, en retenant le plus possible les variations présentes dans le jeu de données de départ. Cette réduction ne sera possible que si les variables initiales ne sont pas indépendantes et ont des coefficients de corrélation entre elles non nuls. Ces variables initiales sont transformées en de nouvelles variables, appelées composantes principales.

Elles sont obtenues par combinaisons linéaires des précédentes et sont ordonnées et non corrélées entre elles.

L'analyse en composantes principales cherche à identifier les vecteurs propres orthonormaux et les valeurs propres de la matrice de dispersion des variables originelles. Les vecteurs propres orthonormaux sont utilisés pour construire les composantes principales, et les valeurs propres sont les variances des composantes principales correspondantes.

Un modèle PCA est défini globalement par la matrice des premiers vecteurs propres de la matrice de corrélations des données. Cette dernière permet à la fois de définir la projection permettant d'avoir les composantes principales et la projection inverse permettant d'estimer les données.

Cependant, toutes les approches antérieures ont toujours utilisé deux étapes : un prétraitement comme première étape suivie par une deuxième méthode d'analyse telle que la PCA. L'approche OPCA a l'avantage de ne pas nécessiter de prétraitement. Ce dernier est implicitement intégré dans le critère du rapport signal à signal qui est optimisé par l'algorithme OPCA. En outre, cette approche est itérative et améliore les performances en combinant plusieurs décalages dans le temps pour l'estimation des paramètres de mélange.

L'ICA est un concept général avec un très grand nombre d'applications en calcul neuronal, traitement du signal, et des statistiques. L'ICA donne une représentation, ou une transformation des données multidimensionnelles qui semble être bien adaptée pour le traitement de l'information. Les composantes de la représentation doivent être indépendantes entre elles, et en même temps elles doivent être «non-gaussiennes».

L'ICA pourrait être considérée comme un prolongement de l'analyse en composantes principales. Dans ce cas-ci, la décorrelation des observations est remplacée par l'hypothèse d'indépendance des sources. Elle utilise des moments d'ordre supérieur à deux dans le but d'obtenir une indépendance statistique des sources et ainsi, lever l'indétermination qui existe si l'on ne considère que la PCA. C'est une méthode liée à la théorie de l'information et en particulier au concept d'information mutuelle. Chacune des composantes vit dans un certain sous-espace linéaire et l'on requiert l'indépendance linéaire de ces sous-espaces et l'indépendance statistique de ces composantes.

Le principe de la séparation aveugle de source est très intéressant avec l'utilisation de l'algorithme ICA puisqu'aucune forte hypothèse sur les signaux n'est nécessaire à part l'existence de sources indépendantes.

Chapitre 4

Séparation aveugle de sources pour les mélanges instantanés

4.1 Introduction

Au cours de ces dernières décennies, beaucoup d'attention a été accordée à la séparation aveugle des sources. Les processus de mélange sont inconnus et les enregistrements de ces mélanges sont disponibles. Dans plusieurs situations, il est souhaitable de récupérer toutes les sources de ces mélanges enregistrés, ou du moins d'isoler une source particulière. En outre, il est utile d'identifier les processus de mélange lui-même pour révéler des informations sur le système de mélange physique. L'objectif de la séparation aveugle de source est d'extraire les signaux de la source originale de leurs mélanges et éventuellement d'estimer le canal inconnu du mélange en utilisant uniquement les informations du signal observé sans aucune connaissance « a priori » sur les signaux source et sur le mélange. Les techniques de BSS peuvent être généralement divisées en deux, soit en utilisant les méthodes du second ordre [32] ou en utilisant des statistiques d'ordre supérieur [114]. Dans ce chapitre, nous présentons la méthode utilisant les statistiques d'ordre supérieur (en anglais : HOS, *High Order Statistics*), que nous avons développé pour la séparation des signaux de parole. Puis, nous décrivons la mise en œuvre de la méthode OPCA, une méthode utilisant les statistiques d'ordre deux, que nous proposons pour la séparation des signaux de parole pour la première fois.

Les résultats d'une étude comparative sur les trois approches de séparation aveugle de signaux de parole citées ci-dessus, seront présentés. Les trois algorithmes, OPCA, HOS et Fast-ICA sont, comparés de façon objective en termes de rapport Signal-à-Interférence (SIR) et PESQ, et de façon subjective en termes de MOS. Les résultats des expériences ont été effectués en utilisant la base de données « TIMIT ».

4.2 Principe de la séparation aveugle de sources

Considérons un schéma synoptique de la séparation de sources suivant :

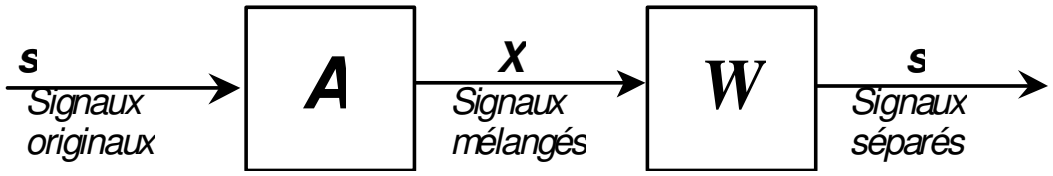

FIG. 4.1 – Système de séparation de sources.

4.2.1 Critère de séparation

Si l'on cherche la matrice W telle que la matrice de covariance R_s des sorties (avec des signaux centrés) soit diagonale [47], on doit avoir :

$$R_s = E\left\{SS^T\right\} = D, \tag{4.1}$$

où D_s est une matrice diagonale.

Compte tenu de la relation 4.1, on obtient alors :

$$D_s = WE\left\{XX^T\right\}W^T = WR_xW^T, \tag{4.2}$$

où R_X est la matrice de covariance des observations.

La matrice de séparation W cherchée est telle que :

$$R_x = W^{-1}D_sW^{-T}. \tag{4.3}$$

Si W^{-1} existe, on peut écrire :

$$R_x^{-1} = W^TD_s^{-1}W. \tag{4.4}$$

décomposant la matrice diagonale D_s, on peut déduire :

$$R_x^{-1} = W^TD_s^{-1/2}D_s^{-1/2}W = \left(W^TD_s^{-1/2}\right)\left(W^TD_s^{-1/2}\right)^T. \tag{4.5}$$

Considérons l'équation 4.3, si l'on pose que :

$$W^* = W^TD_s^{-1/2}. \tag{4.6}$$

La matrice W^* correspond à une normalisation de W. L'équation 4.5 vaut alors :

$$R_x^{-1} = W^* W^{*T}. \tag{4.7}$$

Or cette équation a cependant une infinité de solutions. En effet, la solution de la matrice W^* est une racine carrée de R_x^{-1} : de nombreuses matrices satisfont cette équation. L'utilisation du moment d'ordre deux, qui conduit à la décorrélation des sorties, ne suffit pas pour conduire à une solution unique de séparation de sources [47]. Des contraintes structurelles peuvent être ajoutées pour définir une solution unique, mais elles ne correspondent pas forcément à celle conduisant à la séparation de sources.

4.2.2 Modèle initial

L'algorithme initial a été proposé pour des mélanges instantanés de sources [110]. Le modèle général d'un mélange instantané linéaire de N sources observées par M capteurs a été modélisé par l'équation 2.2 mais sans le bruit additif. En notation matricielle, l'équation s'écrit suivant l'expression 3.8.

4.2.3 Architecture du réseau

L'architecture du modèle de la séparation de sources de Herault-Jutten est basée sur un réseau linéaire. Les entrées du réseau sont les signaux $X_i(n)$ observés à la sortie des capteurs. La sortie $\hat{S}_i(n)$ est connectée à l'entrée de tous les autres avec les poids c_{ij} excepté lui-même : $c_{ii} \approx 0$ (Voir figure 4.2). L'équation de la sortie à l'instant n, s'écrit :

$$\hat{s}_i(n) = X_i(n) - \sum_{\substack{j=1 \\ i \neq j}}^{N} c_{ij}\hat{s}_j(n) \qquad \forall\ i, j\ \in [1, 2, ..., N]\,. \tag{4.8}$$

Les équations matricielles du modèle :

$$\hat{S} = (X - CS) \tag{4.9}$$

et

$$\hat{S} = (I + C)^{-1} X, \tag{4.10}$$

sous réserve d'existence de $(I + C)^{-1}$, où I est une matrice unitaire. La matrice W de séparation des signaux est ici :

$$W = (I + C)^{-1}\,, \quad avec : diag(W^{-1}) = I. \tag{4.11}$$

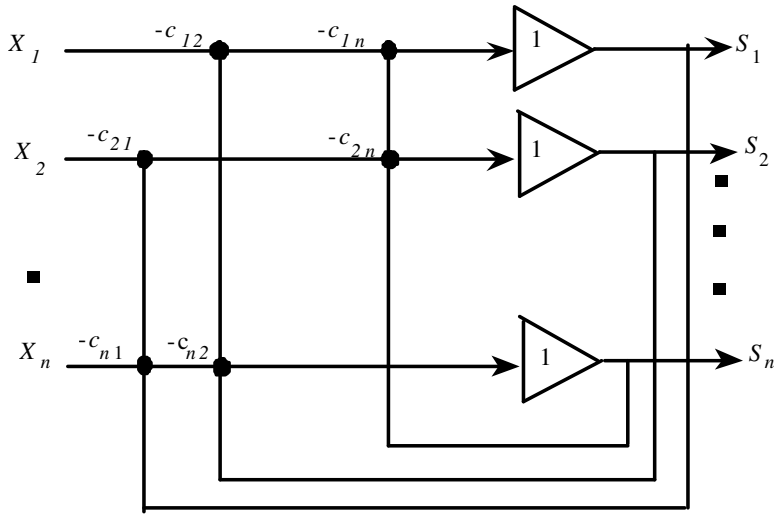

FIG. 4.2 – Architecture du réseau de la séparation de sources de Herault et Jutten.

La solution est obtenue lorsque le vecteur $\hat{S}$ des sorties $\hat{S}_i$ est proportionnel au vecteur S des sources S_i. La solution la plus générale est la suivante :

$$\hat{s}(t) = DPs(t), \tag{4.12}$$

où D est une matrice diagonale et P est une matrice de permutation.

4.3 Séparation aveugle de deux sources utilisant les statistiques d'ordre supérieur

Nous avons déjà proposé cette approche dans le domaine biomédical [114]. L'objectif était de séparer les signaux Électrocardiologiques réels (ECG) d'interférence du secteur (60 Hz) et autres artefacts. Puis, nous l'avons implémenté pour la séparation de deux signaux de parole [116].

Considérons un mélange instantané de deux signaux de parole et un système de capteurs avec la matrice de mélange $A = [a_{11}a_{12}; a_{21}a_{22}]$. L'ICA permet l'estimation de la matrice de séparation $W = [w_{11}w_{12}; w_{21}w_{22}]$, tels que les sorties $\hat{s}_i$ représentent deux signaux séparés qui sont aussi indépendants que possible les uns par rapport aux autres.

Pour arriver à la solution, on s'intéresse au cas de deux sources (voir la figure 4.3). Dans ce cas, les équations du mélange, à l'instant discret n, s'écrivent :

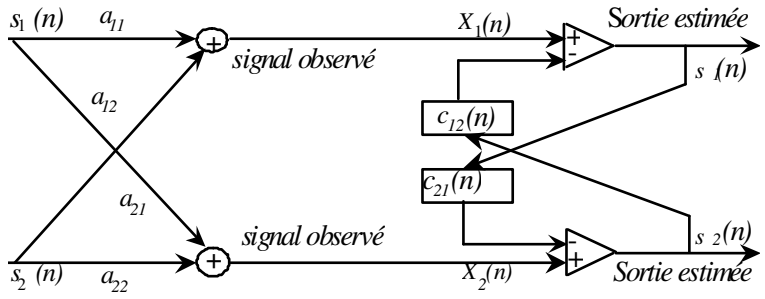

FIG. 4.3 – Schéma synoptique d'une séparation de deux sources

$$\begin{aligned} x_1(n) &= a_{11}s_1(n) + a_{12}s_2(n), \\ x_2(n) &= a_{21}s_1(n) + a_{22}s_2(n), \end{aligned} \tag{4.13}$$

et les équations de la sortie :

$$\begin{aligned} \hat{s}_1(n) &= x_1(n) - c_{12}\hat{s}_2(n), \\ \hat{s}_2(n) &= x_2(n) - c_{21}\hat{s}_1(n). \end{aligned} \tag{4.14}$$

En résolvant le système d'équation 4.14, on obtient alors les sorties :

$$s_1(n) = \frac{(a_{11} - c_{12}a_{21})\, x_1(n) + (a_{12} - c_{12}a_{22})\, x_2(n)}{1 - c_{12}c_{21}}, \tag{4.15}$$

$$s_2(n) = \frac{(a_{22} - c_{21}a_{12})\, x_2(n) + (a_{21} - c_{21}a_{11})\, x_1(n)}{1 - c_{12}c_{21}}. \tag{4.16}$$

Considérons les expressions 4.15 et 4.16, si l'on a :

$$c_{12} = \frac{a_{12}}{a_{22}} \quad et \quad c_{21} = \frac{a_{21}}{a_{11}}. \tag{4.17}$$

Nous arrivons à une solution de séparation de deux sources : $s_1(n)$ est proportionnel à $x_1(n)$ et $s_2(n)$ est proportionnel à $x_2(n)$. La solution duale est a possible, il suffit de prendre :

$$c_{12} = \frac{a_{11}}{a_{21}} \quad et \quad c_{21} = \frac{a_{22}}{a_{12}}. \tag{4.18}$$

En général, il existe des solutions stables si et seulement si la relation suivante est vérifiée [117] et [16] : $c_{12}c_{21} < 1$:

En considérant les équations 4.17 et 4.18 et en appliquant dans chaque cas, la condition $c_{12}c_{21} < 1$:

$$c_{12}c_{21} = \frac{a_{12}a_{21}}{a_{22}a_{11}} < 1, \tag{4.19}$$

$$c_{21}c_{12} = \frac{a_{22}a_{11}}{a_{12}a_{21}} < 1. \tag{4.20}$$

Nous pouvons voir immédiatement que les deux conditions sont incompatibles, un des produits étant l'inverse de l'autre. La solution doit satisfaire la condition de stabilité : par conséquent elle est unique. Cependant, les sources S_i et les coefficients a_{ij} du mélange étant inconnus, à partir des relations 4.15 et 4.16, on ne peut pas trouver la solution 4.17 ou 4.18. Ces coefficients c_{ij} doivent être donc ajustés par un algorithme adaptatif.

4.3.1 Règle d'adaptation

Le critère de séparation des signaux est basé sur l'indépendance des sources qui permet de restituer des sources statistiquement indépendantes. La règle d'adaptation vise donc à rendre les sorties indépendantes. À la convergence des cœfficients c_{ij}, les sorties $S_i(t)$ sont statistiquement indépendantes deux à deux, puisque par hypothèse les signaux primitifs $X_j(t)$ sont supposés indépendants. La condition de non-corrélation est insuffisante. En effet, si deux signaux centrés $u(t)$ et $v(t)$ sont simplement non-corrélés, nous avons :

$$< u(t)v(t) >= 0. \tag{4.21}$$

Si ce sont des signaux indépendants, on doit satisfaire, pour tout i et pour tout j, entiers naturels impairs :

$$< u^i(t)v^j(t) >= 0. \tag{4.22}$$

D'une forme générale, la règle s'écrit :

$$dc_{ij}(t)/dt = af(s_i(t))g(s_j(t)), \tag{4.23}$$

avec les deux conditions suivantes :

i. $f(.)$ et $g(.)$sont des fonctions non-linéaires différentes, ce qui assure la dissymétrie de la règle.

ii. $f(.)$ et $g(.)$ sont des fonctions impaires (positives ou négatives toutes deux pour (.) positif).

La deuxième condition donne au produit $f(S_i)g(S_j - < \hat{S}_j >)$ *où* $< \hat{S}j >$ représente l'estimation de la moyenne de S_j) un signe identique à celui du produit $S_i(S_j - < \hat{S}_j >)$ signe respectant par conséquent la méthode du gradient.

D'après la première condition, si on suppose que f et g sont développables en séries[1] (de Taylor au voisinage de 0, ou en séries entières) on peut écrire :

$$\begin{aligned} f(S) &= \sum_{k=0}^{\infty} f_k S^k, \\ g(T) &= \sum_{k=0}^{\infty} g_k T^k. \end{aligned} \tag{4.24}$$

Compte-tenu de la condition 2 (f et g impaires) seuls les termes de degré k impair ont des coefficients f_k et g_k non nuls et l'on a finalement :

$$f(S_i)g\left(S_j - < \hat{S}_j >\right) = \sum_k f_{2k+1} S_i^{2k+1} \sum_l g^{2l+1} \left(S_j - < \hat{S}_j >\right)^{2l+1}, \tag{4.25}$$

soit :

$$f(S_i)g\left(S_j - < \hat{S}_j >\right) = \sum_{k,l} f_{2k+1} g_{2lk+1} S_i^{2k+1} \left(S_j - < \hat{S}_j >\right)^{2l+1}, \tag{4.26}$$

où $< \hat{S}j >$ représente l'estimation de la moyenne de S_j et l'estimation de la moyenne sur une fenêtre de N échantillons est :

$$< \hat{S}_j >= \frac{1}{N} \sum_{n=0}^{N-1} S_j(n). \tag{4.27}$$

Pendant la phase d'adaptation, les c_{ij} varient, donc les sorties $S_j(t)$ sont non-stationnaires et le calcul d'une moyenne n'a aucun sens, en théorie. En pratique, on ne pourra disposer que d'estimations plus ou moins bonne de $< S_i >$ et $< S_j >$ donc, par voie de conséquence, que d'estimations $\hat{S}_i(t)$ et $\hat{S}_j(t)$ des variables centrées $s_i(t)$ et $s_j(t)$.

Tous les produits $S_i^{2k+1}(t)\left(S_j(t) - < \hat{S}_j(t) >\right)^{2l+1}$ apparaissent dans l'expression. Ainsi la convergence sera atteinte avec cette nouvelle règle,

si la moyenne $< dc_{ij}(t)/dt >=< f(S_i(t))g(S_i(t) - < \hat{S}_j >) >$ est nulle, donc si tous les termes moyens $< S_i^{2k+1}(t)\left(S_j(t) - < \hat{S}_j(t) >\right)^{2l+1} >$ sont nuls. Ces termes sont les

[1] * L'existence d'un développement en séries n'est pas une condition nécessaire pour ces fonctions f et g. L'existence de ces développements permet d'expliquer simplement, de façon formelle, ce qu'apporte la nature non-linéaire des fonctions f et g.

moments impairs croisés entre $S_i(t)$ et $(S_j(t) - < \hat{S}_j(t) >)$, dont le terme du premier ordre $(k = 0, l = 0)$ est la covariance $< S_i(t).(S_j(t) - < \hat{S}_j(t) >) >$. Par ce choix de f et g, on voit que l'on peut considérer cette règle comme un test d'indépendance.

Les coefficients c_{ij} du modèle, à chaque instant n, sont ajustés par la règle de la forme :

$$c_{ij}(n+1) = c_{ij}(n) + \mu f(s_i(n))g(s_j(n)), \tag{4.28}$$

où :

- les deux fonctions f et g sont différentes, non-linéaires et impaires,
- μ est le gain d'adaptation.

Dans le cas le plus simple, on choisit :

$$f = (.)^3 \; et \; g = (.), \tag{4.29}$$

et on obtient alors la règle d'adaptation suivante :

$$c_{ij}(n+1) = c_{ij}(n) + \mu s_i^3(n)s_j(n). \tag{4.30}$$

D'autres fonctions impaires non-linéaires peuvent donner des résultats similaires [118] [47], Sourouchyari [55] a étudié la stabilité de l'algorithme et montre sa dépendance vis à vis du choix des fonctions non-linéaires.

4.3.2 Estimation des pas d'adaptation

En général, les pas d'adaptation des coefficients différents peuvent être différents et variables au cours du temps, pour améliorer la vitesse de convergence pour les signaux non-stationnaires la règle d'adaptation 4.30 du modèle s'écrit :

$$c_{ij}(n+1) = c_{ij}(n) + \mu_{ij} s_i^3(n)s_j(n), \tag{4.31}$$

où $\mu_{ij}(n)$ sont les pas d'adaptation des coefficients c_{ij} à l'instant n. Dans le cas des signaux non-stationnaires, le pas d'adaptation μ est normalisé. Considérons la règle d'adaptaion 4.31, les variations $\Delta c_{ij}\,(n)$ des coefficients $c_{ij}\,(n)$ dépendent des signaux en sortie du modèle selon un terme d'ordre 4 :

$$\Delta c_{ij}(n) = \mu_{ij} s_i^3(n)s_j(n). \tag{4.32}$$

4.4 Séparation aveugle de deux sources utilisant l'analyse en composantes principales orientée (OPCA)

Cet algorithme a été proposé initialement par K.I. Diamantaras et M. Th. Papadimitriou [96], [97], notamment pour séparer quatre signaux aléatoires avec une modulation multi-niveaux en amplitude : PAM (Pulse Amplitude Modulation). Les signaux sont filtrés par un filtre de coloration : ARMA (*Auto-Régressive Moving Average*). Comme la PCA, qui correspond à la décomposition en valeurs propres d'une seule matrice de covariance, selon les auteurs, l'OPCA correspond à la décomposition d'une paire de matrices de variance-covariance en valeurs propres. Oriented PCA décrit une extension du PCA impliquant deux signaux $u(k)$ et $v(k)$. L'objectif est de définir les orientations principales $e_1, ..., .e_N$, qui maximisent le rapport de puissance signal-à-signal $E(e_i^T u)^2/E(e_i^T v)^2$ sous la contrainte d'orthogonalité : $e_i^T R_u e_j = 0,\ i \neq j$. OPCA est une méthode de statistiques d'ordre 2, qui est réduite à la norme de PCA si le second signal est spatialement blanc ($R_v = I$).

L'OPCA en combinaison avec un filtrage temporel arbitraire peut être utilisé pour la séparation aveugle de mélanges linéaires instantanés. Par rapport à d'autres méthodes utilisant les statistiques du deuxième ordre, cette approche permet d'éviter l'étape de pré-blanchiment. En outre, nous avons une convergence rapide avec une approche itérative permettant d'atteindre de meilleurs résultats en combinant plusieurs délais pour l'estimation des paramètres du mélange. La méthode PCA a besoin d'une étape de mesure spatiale de pré-blanchiment sur l'observation des données. Diamantaras [97] a montré que, sans pré-blanchiment, le problème est un exemple typique de l'OPCA. Certaines hypothèses doivent être faites telles que :

i- Les sources sont décorrélées deux à deux, de moyenne nulle et de variance unité :

$$R_x(0) = I, \tag{4.33}$$

ii- Les sources sont colorées temporellement. Il existe $M > 1$ décalages positifs $l_1, \ldots l_M$ tels que :

$$R_s(l_m) := diagonal \neq 0, \tag{4.34}$$

avec $l_0 = 0$, et R_s représentant la matrice de covariance. Les sources ont des fonctions de covariance distinctes, donc leurs densités spectrales ne sont pas identiques :

$$\forall l \neq 0 : r_{ii}(l) \neq r_{jj}(l),\ \text{si}\ \ i \neq j. \tag{4.35}$$

Nous ferons le lien entre la séparation aveugle de sources instantanées avec l'OPCA pour la paire des signaux suivants :

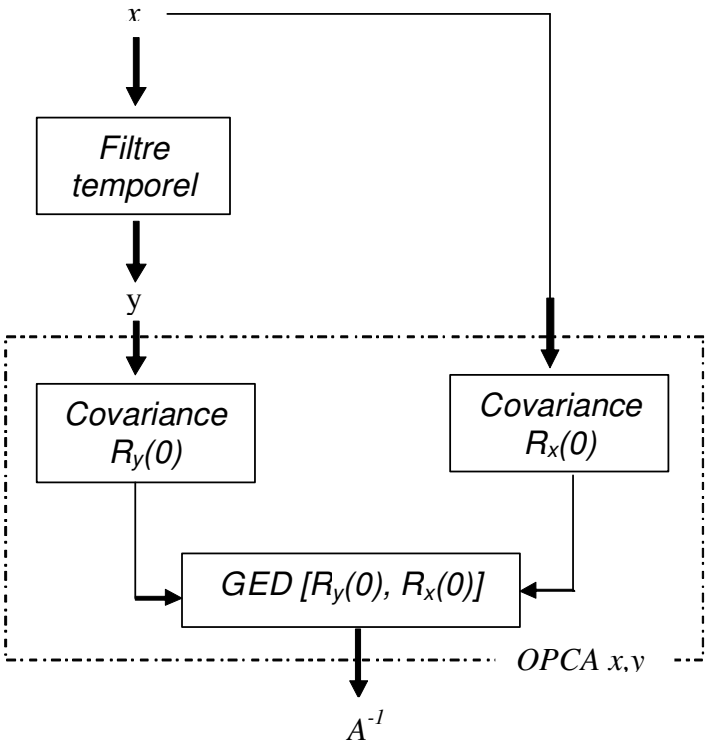

FIG. 4.4 – Schéma synoptique de la séparation aveugle de source utilisant la méthode OPCA appliquée à des signaux aléatoires avec une modulation multi-niveaux en amplitude : PAM [119].

a- le signal observé x

b- toute version filtrée de ce signal y.

Notons que la matrice de covariance de $x(k)$ au temps 0 est de :

$$R_x(0) = AR_s(0)A^T = AA^T. \tag{4.36}$$

Afin d'appliquer l'OPCA, l'idée des auteurs est de considérer un nouveau signal y issu du filtrage des observations par un filtre $h = [h_0, \ldots, h_M]$:

$$y(k) = \sum_{m=0}^{M} h_m x(k - l_m). \tag{4.37}$$

Ensuite, sous l'hypothèse de sources colorées avec différentes densités spectrales, ils cherchent la relation entre les observations x et leurs filtrés y à partir de leur matrice de covariance au retard $l_0 = 0$.

$$R_y(0) = E\left\{y(k)y(k)^T\right\} = \sum_{p,q=0}^{M} h_p h_q R_x(l_p - l_q), \tag{4.38}$$

en supposant que $R_x(l_p - l_q)$ satisfait les conditions 4.34 et 3.34 pour tout p et q.

À partir de l'équation 2.1 (en supposant un siganl sans bruit) nous avons :

$$R_x(l_m) = AR_s(l_m)A^T, \tag{4.39}$$

ainsi nous avons :

$$R_y(0) = ADA^T, \tag{4.40}$$

avec

$$D = \sum_{p,q=0}^{M} h_p h_q R_s(l_p - l_q). \tag{4.41}$$

Comme nous avons supposé que A est carrée et inversible nous avons alors :

$$R_y(0)A^{-T} = AD = AA^TA^{-T}D = R_x(0)A^{-T}D. \tag{4.42}$$

Avec les hypothèses 4.33 et 4.34, D est diagonale. L'équation 4.42 donne la décomposition généralisée des valeurs propres pour la matrice $[R_y(0), R_x(0)]$. Ceci est équivalent au problème de l'OPCA pour les signaux $[y(k), x(k)]$. Les valeurs propres généralisées dans ce cas sont les éléments en diagonal de D. Les colonnes de la matrice A^{-T} sont les vecteurs propres généralisés.

Les vecteurs propres sont uniques jusqu'à une permutation à condition que les valeurs propres soient distinctes (ceci est vrai en général). Dans ce cas, pour toute matrice W nous avons $W = A^{-T}P$ avec P une matrice de permutation, c.-à.-d. chaque ligne et chaque colonne contient un élément non nul. Ainsi les sources peuvent-être estimées comme :

$$\hat{s}(k) = W^{-T}x(k), \tag{4.43}$$

$$\hat{s}(k) = P^TA^{-1}Ax(k) = P^Ts(k). \tag{4.44}$$

Ainsi, les signaux sources estimés sont très proches des signaux originaux avec une permutation et une échelle arbitraire près.

Il se pose après le problème d'initialisation de h et le critère qui détermine optimalement les valeurs de ces retards. Les auteurs proposent un filtre initial de la forme : $h = [h_0, h_1] = [1, \alpha]$, un filtre à deux coefficients où $h_1 = \alpha$, qui est un paramètre

constant . Ils proposent aussi un processus itératif pour calculer α en utilisant un critère $J(\alpha)$ fondé sur la distribution des valeurs propres.

$$D = (1+\alpha^2)I + \alpha R_s(l) + \alpha R_s(-l) = (1+\alpha^2)I + 2\alpha R_s(l). \tag{4.45}$$

Soient d_i représentant les éléments de la matrice diagonale et $r_{ii}(l)$, les éléments de la matrice de corrélation $R_s(l)$, nous obtenons :

$$d_i = 1 + \alpha^2 + 2\alpha r_{ii}(l), \quad i = 1, ..., n. \tag{4.46}$$

Nous pouvons calculer la matrice de corrélation des signaux originaux $R_s(l)$:

$$r_{ii}(l) = \frac{d_i - 1 - \alpha^2}{\alpha}. \tag{4.47}$$

Une fois la corrélation obtenue, nous pouvons l'utiliser pour la conception du filtre temporel optimal h. Le critère d'optimisation sera relié à l'étendue des valeurs propres. Il est préférable d'étendre le plus possible les valeurs propres, pour deux raisons :

i- la convergence de l'algorithme est typiquement rapide quand les valeurs propres sont séparées,

ii- la perturbation des valeurs propres dues au bruit, peut être mieux tolérée.

Ainsi, nous devons définir une métrique tenant compte de la taille des valeurs propres. Les auteurs Diamantaras et Papaditrious [97], [120] proposent le critère de maximisation suivant :

$$J(\alpha) = \min \left[\min_{j \neq i} \frac{(d_i - d_j)^2}{\max_k d_k^2} \right]. \tag{4.48}$$

En tenant compte de l'équation 4.46, nous pouvons exprimer $J(\alpha)$ en fonction de la matrice de corrélation des signaux originaux $R_s(l)$,

$$J(\alpha) = \min \left[\min_{j \neq i} \frac{4\alpha^2 (r_{ii}(l) - r_{jj}(l))^2}{\max_k (1 + \alpha^2 + 2\alpha r_{kk}(l))^2} \right]. \tag{4.49}$$

Posons :

$4\alpha^2 \Delta r^2 = 4\alpha^2 (r_{ii}(l) - r_{jj}(l))^2$ et valeur maximale de $r_{kk}(l)$ dans le dénominateur $(1 + \alpha^2 + 2\alpha r_{kk}(l))^2$, ainsi nous pouvons écrire :

$$J(\alpha) = \frac{4\alpha^2 \Delta r^2}{(1 + \alpha^2 + 2\alpha r_{\max})^2}. \tag{4.50}$$

Le filtre le plus robuste est celui qui maximise $J(\alpha)$. Remarquons que $J(\alpha) \geq 0$ et $lim_{\alpha\to\infty} J(\alpha) = 0$. J sera bornée si $max_k d_k^2 > 0$ et $|d_i| < \infty$ pour tout i. Il résulte que $J(\alpha)$ a au moins un maximum qui est atteint au passage à zéro.

$$\frac{\partial J}{\partial \alpha} = 4\Delta r^2 \left[\frac{2\alpha}{(1+\alpha^2+2\alpha r_{\max})^2} - 2\frac{\alpha^2(2\alpha + 2r_{\max} - 2r'\alpha)}{(1+\alpha^2+2\alpha r_{\max})^3} \right] \quad (4.51)$$

$$= 8\Delta r^2 \frac{\alpha\left(1-\alpha^2-2r'\alpha\right)}{(1+\alpha^2+2\alpha r_{\max})^3} = 0, \quad (4.52)$$

où $r' = \delta r_{\max}/\delta\alpha$. Puisque $J(0) = 0$, la solution $\alpha = 0$ dans l'équation 4.51 ne correspond pas à un maximum. En outre, r_{max} prend un ensemble de valeurs discrètes : $\{r_{11}(l), \ldots, r_{nn}(l)\}$, par conséquent, $J(\alpha)$ n'est pas une fonction continue de α et $r\prime = 0$, excepté pour les points où une discontinuité apparaît. En supposant que $J(\alpha)$ n'est pas maximisé à un des points de discontinuité, sa valeur maximale doit être atteinte pour $(1-\alpha^2) = 0$, c'est-à-dire pour $\alpha = 1$ ou -1.

Contrairement aux filtres de longueur 2 la recherche pour le filtre optimal de longueur 3 est plus complexe. Considérons deux décalages : l_α et l_β . Le filtre de longueur 3 peut être exprimée comme :

$$h = [h_0, h_1, h_2] = [1, \alpha, \beta], \quad (4.53)$$

où α et β sont des paramètres à fixer. Ces paramètres sont optimisés par une nouvelle formulation de la matrice D de l'équation 4.42 comme suit [96] :

$$D = \left(1+\alpha^2+\beta^2\right) I + 2\alpha R_s(l_\alpha) + 2\beta R_s(l_\beta) + 2\alpha\beta R_s(l_\alpha - l_\beta). \quad (4.54)$$

Notons que le critère d'optimalité du pré-filtre est lié à l'étendue des valeurs propres. Le critère de maximisation utilisé pour trouver α et β est donné par :

$$J(\alpha,\beta) = \min_i \left[\min_{j\neq i} \frac{(d_i - d_j)^2}{\max_k d_k^2} \right], \quad (4.55)$$

où $d_{i,j}$ représente les éléments de la diagonale de la matrice D.

Notons que la méthode de l'analyse en composantes principales (PCA) peut être utilisée à la place de la méthode OPCA, mais en utilisant un prétraitement tel que le blanchiment. L'approche OPCA est une méthode qui n'a jamais été appliquée antérieurement pour la séparation des signaux de parole.

4.5 Expériences et résultats

Nous avons mélangé instantanément deux signaux de parole que nous avons enregistrés à partir de la base de données TIMIT. Les coefficients du mélange a_{ij} ont été choisis aléatoirement en utilisant le logiciel Matlab (matrice 4.56). Les signaux du mélange simulé sont calculés par l'équation 4.13. $s_1(n)$ et $s_2(n)$ sont deux signaux de parole respectivement celui d'un homme et d'une femme, échantillonnés à 16 kHz, $x_1(n)$ et $x_2(n)$ les signaux mélangés et a_{11}, a_{12}, a_{21} et a_{22} sont les coefficients de la matrice du mélange A.

Les données d'observation sont de dimension $n = 2$, et la matrice de mélange est choisie telle que :

$$\begin{bmatrix} 1 & 0.6068 \\ 0.5469 & 1 \end{bmatrix}. \tag{4.56}$$

Dans les expériences suivantes la base de données TIMIT est utilisée. Ce corpus contient des enregistrements haut débit sur un total de 6300 phrases, 10 phrases prononcées par chacun de 630 orateurs de dialecte de huit régions majeures des États-Unis, chaque lecture des dix phrases est riche phonétiquement [121]. Certaines phrases de la base de TIMIT ont été choisies aléatoirement pour évaluer nos méthodes de BSS. Nous représentons les résultats dans le domaine temporel ainsi que dans le domaine temps-fréquentiel : le spectrogramme.

4.5.1 Méthode Fast-ICA

La méthode Fast-ICA a été utilisée comme référence et comme une méthode de base pour nos travaux, pour faire une évaluation objective et qualitative. Les figures 4.5 à 4.7 et les figures 4.8 à 4.10, représentent respectivement, les deux signaux de parole originaux, les deux mélanges de ces deux signaux et les signaux estimés par la méthode Fast-ICA dans le domaine temporel et le domaine fréquentiel.

La phrase prononcée par l'homme est :

- "*This brings us to the question of accreditation of art schools in general*".

La phrase prononcée par la femme est :

- "*She had your dark suit in greasy wash water all year*".

Domaine temporel

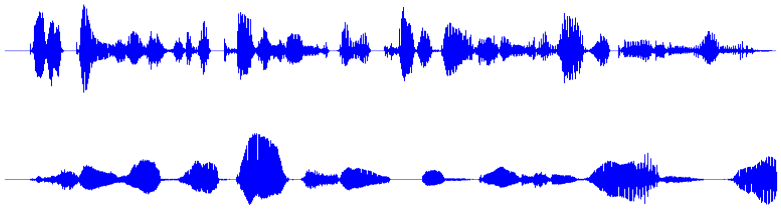

FIG. 4.5 – Signaux de parole originaux, signal de parole de l'homme (en haut) et signal de parole de la femme (en bas).

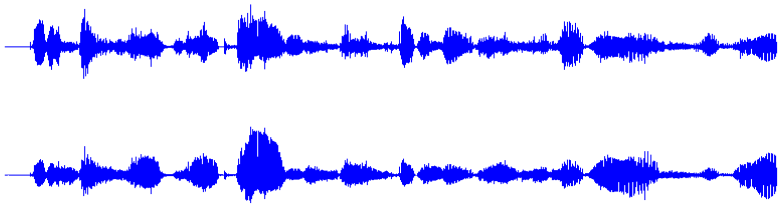

FIG. 4.6 – Signaux mélangés instantanément.

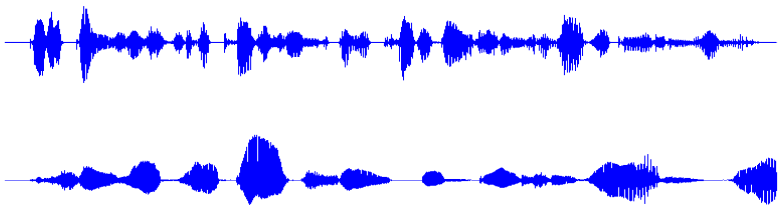

FIG. 4.7 – Signaux séparés avec la méthode Fast-ICA.

Domaine temps-fréquence : le spectrogramme

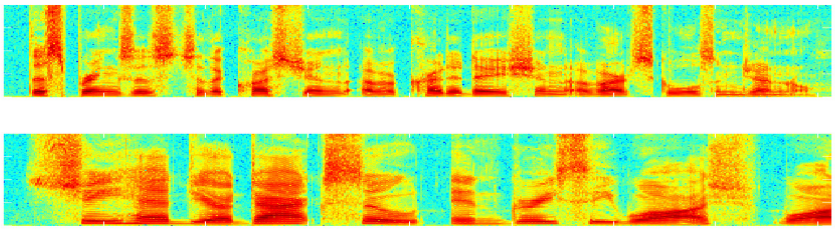

FIG. 4.8 – Spectrogramme des signaux de parole originaux.

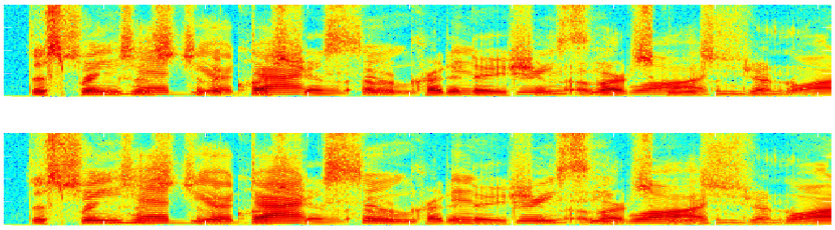

FIG. 4.9 – Spectrogramme des signaux mélangés instantanément.

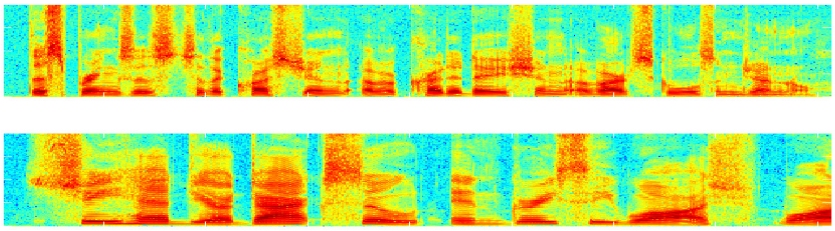

FIG. 4.10 – Spectrogramme des signaux estimés par la méthode Fast-ICA.

4.5.2 Méthode utilisant les statistiques d'ordre supérieur : HOS

Pour la méthode HOS, nous avons choisi : $f = s_i^3(n)$ et $g = s_j(n)$, $\mu_1 = 0,9$, $\mu_2 = 0,8$, $c_{ij}(0) = 0$, la règle d'adaptation de l'équation 4.32 et la matrice de mélange A sont donnés dans le sous-chapitre précédent .

Domaine temporel

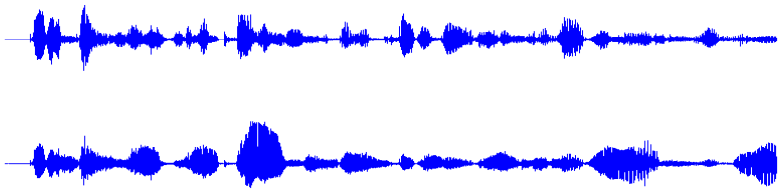

FIG. 4.11 – Signaux de parole estimés par la méthode HOS.

Le spectrogramme

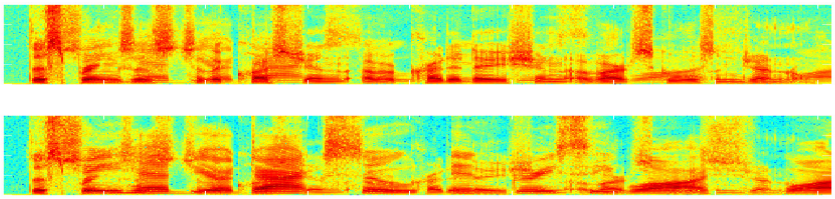

FIG. 4.12 – Spectrogramme des signaux estimés par la méthode HOS.

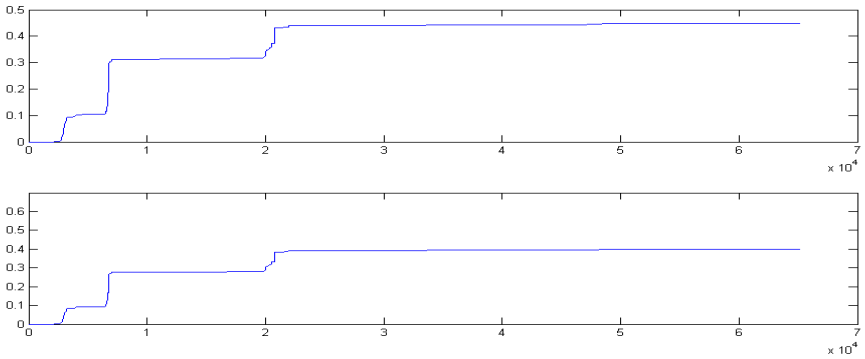

FIG. 4.13 – Évolution des coefficients des mélanges c_{ij}.

Initialement les coefficients $c_{ij}(0)$ sont mis à zéro. À chaque instant n, les coefficients de séparation $c_{ij}(n+1)$ sont mis à jour par la règle 4.32. Les signaux de sortie $s_i(n)$ sont calculés par la relation 4.14. Dans le domaine temporel, nous remarquons que les coefficients avant la convergence 4.13, varient beaucoup puis convergent vers

les valeurs : 0,48 et 0,62 et donc proches de celles qui ont été choisies. Les valeurs de μ_1 et *de* μ_2 ont été choisies empiriquement.

4.5.3 La méthode OPCA

Domaine temporel

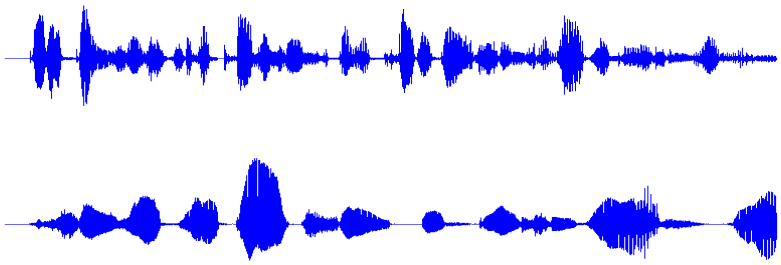

FIG. 4.14 – Signaux de parole estimés par la méthode OPCA.

Le spectrogramme

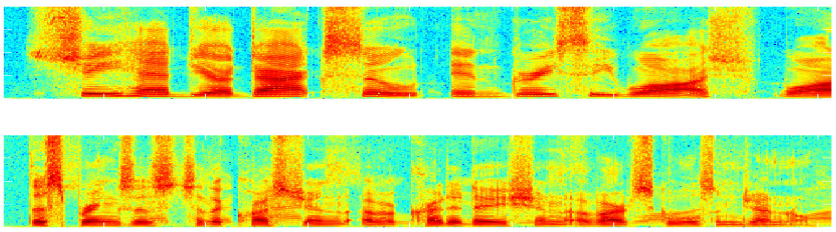

FIG. 4.15 – Spectrogramme des signaux estimés par la méthode OPCA.

4.6 Critères d'évaluation de la qualité du signal de parole

Par cette comparaison, nous cherchons à démontrer l'efficacité de la technique de séparation basée sur la méthode OPCA. Cette approche est efficace, comme on peut le voir dans le domaine temporel, où l'on remarque que les signaux originaux (figures 4.5 et 4.8) et les signaux estimés (figures 4.14 et 4.15) sont très proches. Cette méthode a été vérifiée de manière subjective par l'écoute des signaux originaux, des signaux mélangés et séparées. Nous avons obtenu une très bonne séparation.

4.6.1 Le PESQ

Pour mesurer la qualité de la parole, l'une des méthodes les plus fiables est l'évaluation de la perception de la qualité de la parole. Cette méthode a été standardisée par la recommandation de l'UIT-T P.862 [122]. PESQ fournit une mesure objective et automatisée pour l'évaluation de la qualité de la parole. Comme l'illustre la figure 4.16 [123], la mesure est effectuée en utilisant un algorithme de comparaison d'un échantillon de parole de référence qui est le signal de parole original à un échantillon de parole générée par un système quelconque. Théoriquement, les résultats peuvent être comparés avec pertinence à la moyenne des scores d'opinion en fonction de la dégradation de l'échantillon [124]. L'algorithme PESQ est conçu pour prédire subjectivement les scores opinion d'un échantillon de parole dégradé. Il nous donne un score compris entre 0,5 à 4,5. Les scores les plus élevés indiquent une meilleure qualité.

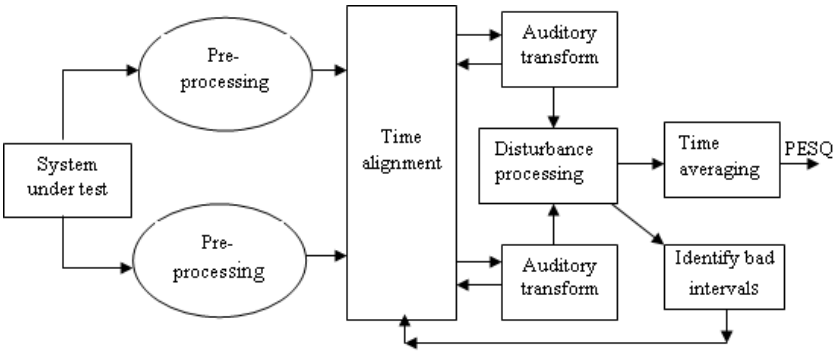

FIG. 4.16 – Schéma synoptique du calcul de la mesure du PESQ [123].

Pour nos expériences nous avons utilisé le code fourni par Loizou [123]. Cette technique est généralement utilisée pour évaluer les systèmes de rehaussement de la parole. Habituellement, le signal de référence est le signal original (qu'on appelle : propre) et le signal dégradé représente le signal de parole prononcé par le même locuteur représentant le signal original, mais qui a été soumis à diverses conditions défavorables. Dans cet algorithme, le signal de référence et les signaux dégradés sont égalisés à un niveau d'écoute standard, grâce à l'étape de prétraitement. Nous n'avons aucune connaissance a priori sur le gain des deux signaux et il peut varier considérablement. Dans l'algorithme original du PESQ, les gains du signal de référence, du signal dégradé et du signal traité, sont calculés à partir de la racine carrée des valeurs moyennes de la bande passante filtrée (350 - 3250 Hz) du signal de parole. La bande de fréquence totale est conservée et normalisée et nous avons supprimé le filtre ayant la réponse en fréquence d'un combiné téléphonique existant dans l'algorithme original PESQ.

PESQ	PESQ (signal de parole : femme)	PESQ (signal de parole : homme)
Fast-ICA	2,5	2,6
HOS	4,2	3,0
OPCA	3,2	3,0

TAB. 4.1 – Comparaison du PESQ des méthodes Fast-ICA, HOS et OPCA.

La méthode PESQ a été utilisée dans toutes nos expériences pour évaluer les trois méthodes. Le PESQ a l'avantage d'être une évaluation indépendante des auditeurs et du nombre d'auditeurs.

4.6.2 Le MOS : Note d'opinion moyenne

Note d'opinion moyenne : MOS est une note caractérisant la qualité de la restitution d'un signal de parole. Cette note varie entre 1 et 5, comparativement à la version originale. Il donne un niveau de qualité subjective. Le principe du calcul est basé sur un sondage d'un échantillon représentatif de la population des utilisateurs. Les personnes qui constituent l'échantillon écoutent un signal, puis son équivalent et après chaque écoute, ils donnent une note sanctionnant la qualité qu'ils ont perçue. La moyenne des notes fournies par cette population constitue le MOS.

Cette méthode présente des inconvénients. La mise en œuvre de cette évaluation coûte chère en coût et nécessite beaucoup d'énergie. Il faut convoquer un nombre relativement grand d'auditeurs, de mettre à leur disposition les équipements audio, de leur expliquer la bonne façon de donner une note qui soit exploitable, et de collecter et consolider les résultats. La nature subjective du résultat représente une autre difficulté. La non-standardisation du MOS représente un inconvénient principal. À cause de l'absence de toute norme méthodologique beaucoup de chiffres sont produits dans de mauvaises conditions expérimentales.

Dans nos expériences, cette méthode subjective a été réalisée par l'écoute des signaux originaux, des signaux mélangés ainsi que les signaux estimés après séparation. Les tests du MOS ont été réalisés avec un groupe de sept auditeurs. Les résultats pour les trois méthodes sont proches et ils sont résumés dans le tableau 4.2.

Méthodes	Fast-ICA	HOS	OPCA
MOS	4,6	4,3	4,1

TAB. 4.2 – Comparaison du MOS pour les méthodes Fast-ICA, HOS et OPCA.

4.6.3 Le rapport signal-interférence : SIR

La mesure de la performance est aussi réalisée grâce au rapport du signal-à-interférence. Le principe de cette mesure est basé sur la décomposition du signal estimé $\hat{s}(n)$ après la séparation, en la somme d'un signal cible et d'un signal interférence tel que représenté par l'équation 4.57 ci-dessous :

$$\hat{s}(n) = s_{t\arg et}(n) + e_{int\,erf}(n), \tag{4.57}$$

où $s_{target}(n)$ est le signal original (cible) et $e_{interf}(n)$ est l'interférence qui représente la différence entre le signal original et le signal estimé. Les signaux originaux et estimés sont d'abord normalisés puis en tenant compte de la décomposition, nous pouvons calculer le critère de performance fourni par le rapport de la puissance du signal original sur la puissance de l'interférence :

$$SIR = 10\log\frac{\|s_{t\arg et}\|^2}{\|e_{int\,erf}(n)\|^2}. \tag{4.58}$$

Pour la mise en œuvre du SIR, nous avons d'abord utilisé l'algorithme fourni dans la "*toolbox C-FICA v1-1*" par J. Thomas et al. [131]. Nous l'avons appelé SIR_{1st} pour représenter le rapport signal-à-interférence après la séparation. Les résultats sont donnés dans le tableau 4.3. Dans ce cas-ci, le rapport signal-interférence (SIR) est calculé en utilisant directement l'équation 4.58. Les sources extraites n'ont pas nécessairement la même amplitude ou le même signe que les sources originelles. À cause de l'ambiguïté sur l'amplitude (et aussi sur le signe) des sources extraites, il n'est pas possible d'utiliser directement cette équation dans le cas de signaux extraits par BSS instantané.

SIR	SIR_{en}	SIR_{en}	SIR_{1st} (locuteur)	SIR_{1st} (locutrice)
Fast-ICA	-4,66 dB	-1,75 dB	3 dB	3 dB
HOS	-4,66 dB	-1,75 dB	11 dB	8 dB
OPCA	-4,66 dB	-1,75 dB	20 dB	17,2 dB

TAB. 4.3 – Comparaison du SIR des méthodes Fast-ICA, HOS et OPCA (calcul du SIR fourni par la "toolbox C-FICA v1-1").

SIR	SIR_{en}	SIR_{en}	SIR_{2st} (locuteur)	SIR_{2st} (locutrice)
Fast-ICA	-4,66 dB	-1,75 dB	45,69 dB	45,45 dB
HOS	-4,66 dB	-1,75 dB	13,26 dB	7 dB
OPCA	-4,66 dB	-1,75 dB	48,83 dB	21,7 dB

TAB. 4.4 – Comparaison du SIR des méthodes Fast-ICA, HOS et OPCA (calcul du SIR fourni par J. Thomas et al. [139]).

Cependant en BSS, une façon pratique de calculer le rapport SIR est d'effectuer une projection ou une prédiction linéaire des sources extraites sur les signaux originaux. Les différences entre les signaux originaux et ces projections deviennent alors les signaux d'erreur. Ces signaux d'erreur sont représentés par e_{interf} dans l'équation 4.57. Les signaux prédits et les signaux d'erreur peuvent ensuite être utilisés pour le calcul du SIR [138]. Dans nos expériences, les mesures du SIR ont été mises en œuvre en utilisant le code MATLAB "*BSS_EVAL*" disponible en ligne sous License GNU Public [139]. Dans ce programme, les mesures de performance de la séparation sont calculées pour chaque source estimée $\hat{s}_j$ en la comparant à une source donnée originale s_j. La source estimée $\hat{s}_j$ est comparée à toutes les sources originales s_j et la "vraie source" donnant les meilleurs résultats sera choisie. Dans ce cas-ci, nous l'avons nommé SIR_{2st} tel qu'indiqué dans le tableau 4.4. Les valeurs du rapport SIR_{1st} et SIR_{2st} différentes dans les deux tableaux 4.3 et 4.4 s'expliquent par la manière de calculer ce rapport tel qu'expliqué dans le paragraphe précédent.

Pour évaluer cette approche dans le cas instantané, nous l'avons comparé avec le célèbre algorithme Fast-ICA et la méthode HOS que nous avons développée. Dans le tableau 4.3 et 4.4, SIR_{en} représente le rapport signal-à-interférence avant la séparation.

4.7 Rehaussement d'un signal de parole en présence d'un bruit avec la méthode OPCA

4.7.1 Bruits enregistrés à partir de la base de données NOISEX : bruit blanc

Nous avons mélangé instantanément un signal de parole et un bruit blanc. Le signal de parole est prononcé par un homme et enregistré à partir de la base de données TIMIT. Nous avons enregistré le bruit blanc à partir de la base de données Noisex. La matrice de mélange choisie aléatoirement est donnée par Matlab. Dans ce cas-ci nous appliquons notre méthode de BSS avec une source de bruit.

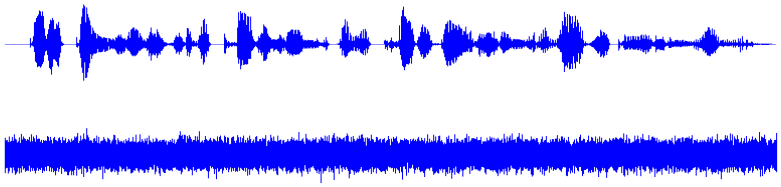

FIG. 4.17 – Signaux originaux : parole et bruit blanc.

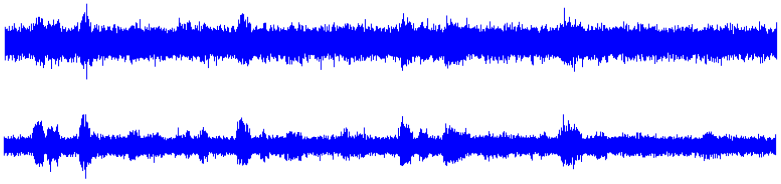

FIG. 4.18 – Signaux de parole et bruit blanc mélangés.

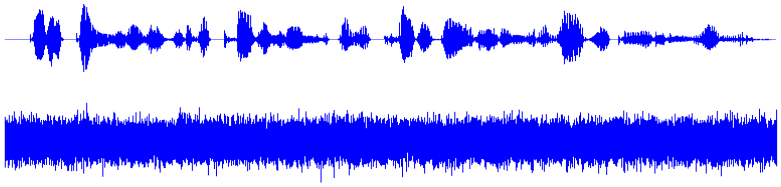

FIG. 4.19 – Signaux estimés par la méthode OPCA : parole et bruit blanc.

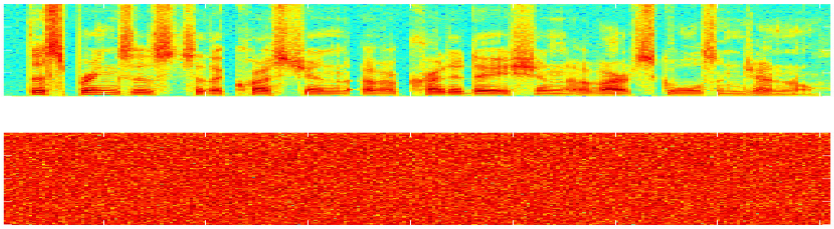

FIG. 4.20 – Spectrogrammes des signaux originaux : parole et bruit blanc.

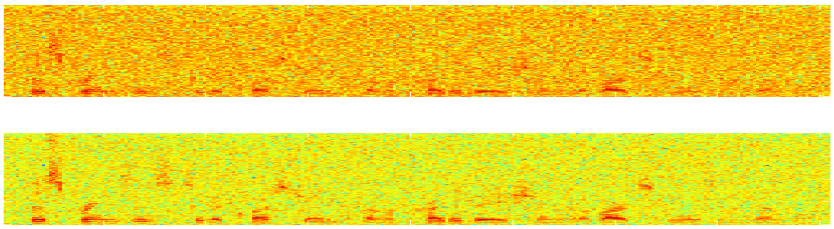

FIG. 4.21 – Spectrogrammes des signaux mélangés.

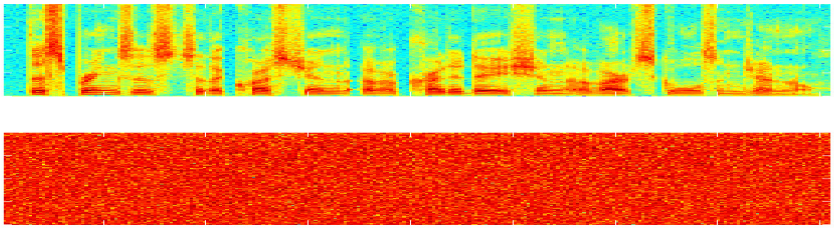

FIG. 4.22 – Spectrogrammes des signaux estimés par méthode OPCA : parole et bruit blanc.

Nous remarquons d'après les figures 4.18 et 4.21 respectivement dans le domaine temps et dans le domaine fréquentiel que le signal de parole est complètement noyé dans le bruit. Après traitement par la méthode OPCA, nous notons une bonne séparation des deux signaux. Nous avons choisi le rapport signal-à-bruit (RSB) avant séparation de 5 dB. Après séparation nous obtenons un RSB de 20,2 dB d'où une amélioration de 15,2 dB. Nous obtenons aussi un PESQ de 4,3 ce qui représente un excellent résultat. Des expériences avec d'autres types de bruit ont été réalisées tels que : brouhaha ("*babble*" en anglais), bruit dans le cockpit d'un avion F16 et aussi dans une voiture. Tous ces bruits ont été enregistrés de la même base de donnés Noisex.

4.7.2 Signaux de parole bruités enregistrés à partir de la base de données : NOIZEUS

NOIZEUS : est un corpus de signaux de parole bruités qui est utilisé pour l'évaluation des algorithmes pour le rehaussement de la parole. La base de données contient 30 phrases bruitées IEEE. Ces phrases sont produites par trois hommes et trois femmes. Elles sont corrompues par huit des bruits réels en différents rapports signal sur bruit. Ces bruits ont été prélevés de la base de données AURORA. Cette base de donnés inclut le bruit de trains de banlieue, le brouhaha ("*Babble*" : en anglais), voiture, salle d'exposition, restaurant, rue, l'aéroport et le bruit de la Gare.

Cas du bruit de train de banlieue

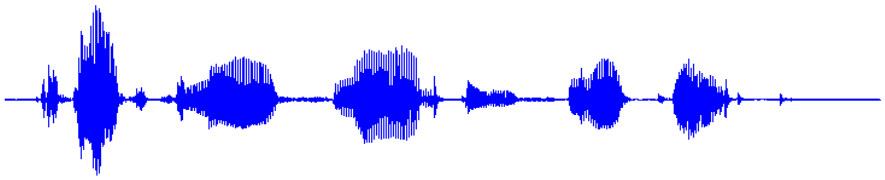

FIG. 4.23 – Signal de parole original.

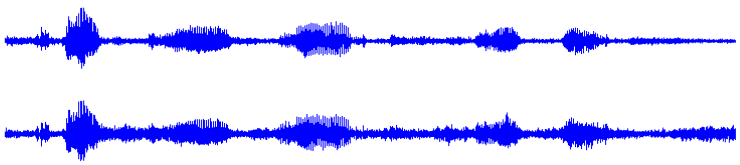

FIG. 4.24 – Signaux de parole et du train mélangés avec des rapports signal-à-bruit de 0 et 5 dB.

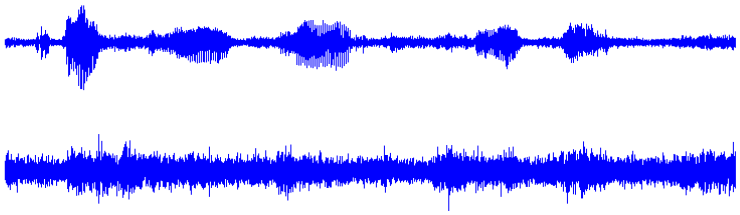

FIG. 4.25 – Signaux de parole et du bruit du train de banlieue estimés par la méthode OPCA.

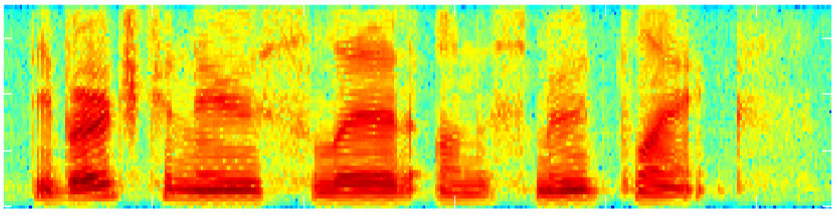

FIG. 4.26 – Spectrogramme du signal de parole original.

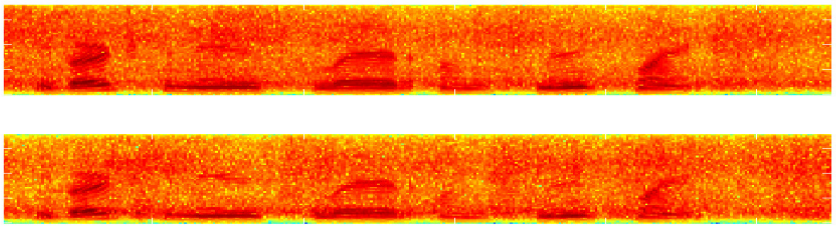

FIG. 4.27 – Spectrogramme des signaux de parole et bruit du train mélangés avec un RSB de 0 et 5 dB.

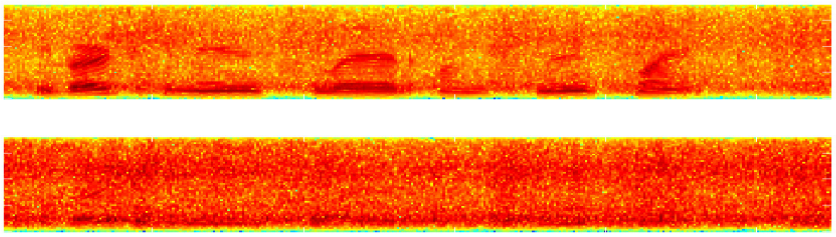

FIG. 4.28 – Spectrogramme des signaux de parole et du train estimés par la méthode OPCA.

RSB (dB) des deux signaux mélangés	0 et 5	0 et 10	10 et 15
RSB (dB) : cas du « train de banlieue »	5,52	8,61	15,40
RSB (dB) : cas du «brouhaha »	5,22	10	15,34

TAB. 4.5 – Résultats du SIR pour différents rapports signal-à-bruit.

Sur l'ensemble des bruits des différents mélanges, le bruit du train mis en évidence dans les figures a été choisi d'une manière aléatoire. Les figures 4.24 et 4.27 montrent un signal de parole mélangé avec le bruit d'un train de banlieue. Les deux signaux mélangés représentés dans ces figures ont un rapport signal-à-bruit (RSB) de 0 dB et 5 dB. Pris de la base de données NOIZEUS, le mélange ainsi que la matrice le déterminant sont inconnus, nous sommes alors dans un vrai cas de séparation aveugle. Nous remarquons d'après les figures 4.25 et 4.28 respectivement dans le domaine temps et dans le domaine temps-fréquence que le signal du bruit est complètement extrait. Pour le signal de parole il y a un résidu de bruit. Dans le tableau 4.5 nous avons représenté trois cas de tranches de rapport RSB : le cas d'une tranche de RSB faible (0 et 5 dB), moyen (0 et 10 dB) et un cas modérément élevé (10 et 15 dB). Les résultats illustrés dans le tableau 4.5 montre une amélioration du RSB de $5,5$ dB pour la première tranche à $15,40$ dB pour la troisième tranche. Le tableau 4.5 récapitulatif nous montre les résultats pour les différents rapports signal-à-bruit pour les cas du bruit du train et du bruit de brouhaha.

4.8 Conclusion

L'algorithme basé sur la méthode HOS a été initialement appliqué à des signaux biomédicaux, pour la suppression de l'interférence 50 Hz (60 Hz en Amérique du nord) des signaux électro-cardiographiques, que nous avons proposé dans un travail antérieur [115].

Nous voyons bien qu'après que les deux signaux étaient complètement mélangés entre eux (figures 4.6 et 4.9), nous avons réussi à réaliser une bonne séparation des deux signaux avec cette approche (figures 4.11 et 4.12). Nous remarquons aussi que les valeurs estimées des coefficients sont plus proches des valeurs théoriques après convergence : $0,48 \approx 0,5$ et $0,62 \approx 0,6$, comme nous pouvons le voir dans la figure 4.13. Nous pouvons aussi constater l'efficacité de la méthode avec un MOS de $4,3$ et un PESQ de $4,2$.

L'approche OPCA s'avère aussi efficace, comme nous pouvons le voir dans le domaine temporel ainsi que dans le domaine temps-fréquence, où l'on remarque que les signaux originaux (figures 4.5 et 4.8) et les signaux estimés (figures 4.14 et 4.15) sont très proches.

Notons que Fast-ICA permet une bonne séparation avec un calcul rapide, mais les signaux sont corrompus par un bruit lors de l'écoute des signaux de paroles dans le contexte de l'évaluation subjective. La méthode OPCA est aussi rapide mais aucun bruit n'est perceptible. Concernant la méthode HOS, le calcul est moins rapide. Pour une évaluation objective du PESQ, HOS a donné des résultats meilleurs par rapport aux deux autres méthodes, comme nous pouvons le voir dans le tableau 4.1. Pour l'estimation du signal de parole du locuteur, nous obtenons le même résultat pour les deux algorithmes HOS et OPCA, mais dans le cas du signal de parole de la locutrice, le PESQ de HOS est le meilleur. Pour la méthode la plus connue et qui nous sert de référence : Fast-ICA, le PESQ est assez mauvais.

Le MOS a donné un résultat très proche pour les trois méthodes comme nous pouvons le voir dans le tableau 4.2. Effectivement, nous réalisons une séparation des deux signaux de parole pour les trois types d'algorithmes, mais les auditeurs n'ont en général pas tenu compte du bruit qui est clairement perceptible dans le cas de l'approche Fast-ICA, ce qui explique un MOS supérieur.

Dans le tableau 4.3, SIR_{en} et SIR_{st} représentent respectivement le rapport signal-interférence avant la séparation et le rapport signal-interférence après la séparation. Si nous regardons le résultat des SIR_1 calculés suivant la méthode proposée par J. Thomas et al. [131], et en compraison avec les deux autres méthodes, OPCA est nettement supérieure comme nous pouvons le constater dans le tableau 4.3. Dans le cas de l'approche Fast-ICA, l'amélioration du rapport SIR pour le signal de parole masculin, est de $1,66\ dB$ et pour le signal de parole féminin est de $7,66\ dB$. Il est de $12,66\ dB$ pour le signal de parole masculin et $15,66\ dB$ pour le signal de parole féminin pour l'algorithme HOS et enfin pour la technique OPCA, de $24,66\ dB$ pour le signal de parole masculin et pour le signal de parole féminin une amélioration de $18.95\ dB$. Considérons maintenant le cas du calcul du SIR_{2st} avec l'algorithme "BSS_EVAL" disponible dans [139], Fast-ICA donne de très bons résultats pour des valeurs de SIR_{2st}

d'environ 45 dB pour les deux locuteurs, tandis que pour l'approche OPCA nous arrivons à une valeur de SIR_{2st} de 48 dB pour le locuteur, une amélioration donc de 3 dB par rapport à celui de Fast-ICA et une valeur de 22 dB pour la locutrice, plus faible donc par rapport à celui de Fast-ICA, mais cela représente tout de même une bonne valeur. Par contre pour la méthode HOS, les résultats sont inférieurs dans ce cas-ci.

Concernant le signal de parole mélangé à un bruit blanc que nous avons enregistré de la base de données NOISEX, nous obtenons une amélioration du rapport signal-à-bruit de $15,2$ dB ainsi qu'un PESQ de $4,3$ après extraction. Pour le cas des signaux enregistrés de la base de données NOIZEUS : un signal de parole mélangé avec le bruit d'un train de banlieue puis avec un bruit représentant le brouhaha, nous avons réussi à extraire le bruit d'une manière totalement aveugle, puisque nous n'avons aucune connaissance a priori ni sur les signaux originaux ni sur le mélange. D'autres expériences ont été réalisés avec les autres types de bruit de cette même base de donnés et les résultats sont similaires.

Le mélange le plus simple est le mélange linéaire instantané : à chaque instant, les observations sont des combinaisons linéaires des sources au même instant. Dans la réalité, les mélanges sont souvent convolutifs : à chaque instant les observations dépendent aussi des valeurs des signaux sources aux instants précédents. Les mélanges sont aussi souvent variant dans le temps. Par exemple lorsque des personnes se déplacent, les fonctions de transfert de la salle entre les émetteurs (les personnes) et les récepteurs (les microphones) changent suivant les positions des émetteurs et récepteurs, produisant donc sur les micros un mélange variant dans le temps.

Dans le chapitre suivant, nous traiterons le cas des signaux mélangés convolutivement.

Chapitre 5

Séparation aveugle de sources pour les mélanges convolutifs

5.1 Introduction

Initialement, la séparation aveugle de sources a été modélisée pour un mélange linéaire instantané, autrement dit les capteurs reçoivent à chaque instant une combinaison linéaire des signaux sources. Plus tard, il s'est avéré que cette modélisation ne convenait pas à toutes les situations rencontrées dans la pratique. C'est pourquoi des modélisations plus réalistes ont été proposées. L'une de ces modélisations considère le canal de transmission comme un système reposant sur une opération de filtrage. En d'autres-termes, les signaux captés dépendent linéairement à la fois des sources et de leurs versions retardées. On suppose que l'environnement est caractérisé par une fonction mathématique dépendante du temps et réalise une opération plus complexe, qui est un produit de convolution, afin d'engendrer les mélanges dits convolutifs. Ceci correspond au cas plus intéressant de la séparation de sources appelé « convolutif ».

Dans ce chapitre, nous présenterons les différents types de modèle de mélange que nous avons utilisés. Nous décrirons brièvement la mise en œuvre de la méthode OPCA que nous proposons pour la séparation des signaux de parole, dans le domaine fréquentiel. À des fins de comparaison, nous présenterons deux autres méthodes que nous avons expérimentées soient : DUET (*Degenerate Unmixing Evaluation Technique*) et la méthode C-FICA (*Convolutive- Fast-ICA*). Les résultats expérimentaux et une discussion suivront. Enfin, nous conclurons et donnerons un point de vue sur notre travail.

5.2 Problématique et éléments classiques de séparation de sources dans le cas convolutif

Nous présentons dans cette section une formulation du problème de BSS convolutif, les techniques permettant de reformuler le modèle convolutif en un modèle de type instantané, les hypothèses classiques, les indéterminations inhérentes au problème et des indices de performance permettant d'évaluer la qualité de séparation.

5.2.1 Formulation du problème de séparation de sources avec mélange convolutif

Rappelons que le problème de BSS peut être modélisé d'une façon simple sous la forme générale donnée par l'équation 2.3 .

Les sources peuvent être déterministes ou stochastiques à des valeurs réelles ou complexes. Elles peuvent également être stationnaires, ou non-stationnaires.

L'application A représente dans ce cas-ci un filtre linéaire et invariant dans le temps, alors pour traiter ce problème nous considérons un système à réponse impulsionnelle finie (RIF) à temps discret décrit par l'équation suivante :

$$x_m(t) = \sum_{n=1}^{N} \sum_{k=0}^{K-1} a_{mnk} s_n(t-k), \tag{5.1}$$

- a_{mnk} représente les coefficients correspondant au filtre de mélange.
- K désigne l'ordre du filtre RIF.

Le signal mélangé est un mélange linéaire des versions filtrées de chacun des signaux source. En pratique, ces coefficients peuvent également changer dans le temps, mais pour simplifier le modèle de mélange nous supposons que nous sommes dans un cas stationnaire. Sous forme de matrice, le modèle convolutif peut être écrit comme suit :

$$x(t) = \sum_{k=0}^{K-1} A_k s(t-k) = \{A * S\}(t), \tag{5.2}$$

dans ce cas-ci, A une matrice de dimension $M \times N$ contenant les coefficients du filtre. La figure 5.1 représente le modèle de mélange convolutif.

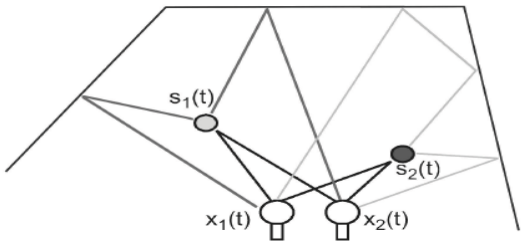

FIG. 5.1 – Modèle de mélange convolutif [125].

En examinant l'équation 5.2, nous remarquons que le mélange des sources est effectué au moyen de trois opérations élémentaires simples qui sont le retard, la multiplication et l'addition. Ainsi, le schéma de réalisation de ce modèle de mélange pour toute observation $x_i(t), i = 1, \ldots, M$, reçue au niveau du $i^{ème}$ capteur, est donné dans la figure 5.2, où z^{-1} : représente le retard, $\times$: la multiplication de deux signaux et $\sum$: l'addition de n signaux [126].

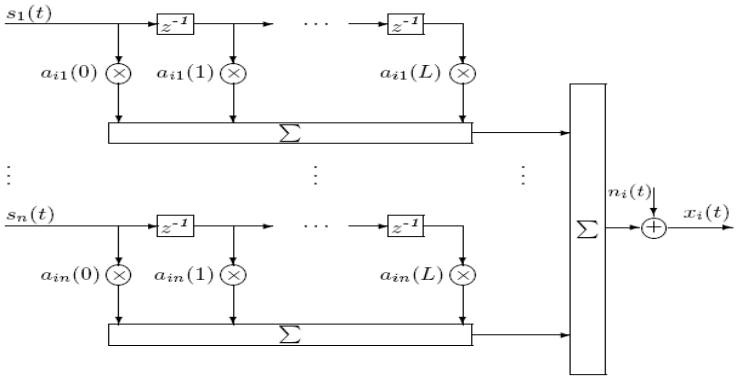

FIG. 5.2 – Schéma de réalisation d'un seul mélange linéaire convolutif de sources [126].

5.2.2 Relation avec le système de mélange instantané

Le système de mélange instantané peut être considéré comme un cas particulier du système convolutif. En fait, il est obtenu lorsque le système $A(t)$ est une matrice de scalaires a_{ij} ne dépendant pas de t,

$$\begin{aligned} A(t) &= 0 \qquad \forall t \neq 0, \\ A(0) &= A, \end{aligned} \tag{5.3}$$

et par conséquent $\mathrm{x}_i(t) = \sum_{j=1}^{n} a_{ij} s_j(t)$.

Une manière de repasser à un modèle de type instantané est d'exploiter la propriété de la transformée de Fourier qui permet de transformer le produit de convolution en un simple produit [3][58][77][88][127][128].

En réalité, la séparation n'est pas complètement aveugle puisqu'on suppose généralement connaître deux données :

i. la nature des mélanges,

ii. le nombre M d'observations.

5.3 Différents modèles de mélanges convolutifs

Le modèle convolutif introduit la relation entre le $M^{ème}$ signal mélangé et les signaux source, en supposant un processus de mélange convolutif y compris le retard suivant l'équation 5.1. Ce processus de mélange convolutif peut être simplifié, en transformant les mélanges dans le domaine fréquentiel. La convolution linéaire dans le domaine temporel peut être écrite dans le domaine fréquentiel comme des multiplications pour chaque fréquence tel qu'indiqué dans l'équation 2.8.

À chaque fréquence f, $A(f)$ est une matrice complexe $N \times M$, $X(f)$ est un vecteur complexe de dimension $M \times 1$, et de même $S(f)$ est un vecteur complexe $N \times 1$. La transformation en fréquence est généralement calculée en utilisant une transformée de Fourier discrète (DFT) dans un délai de taille T à partir de l'instant t :

$$X(f,t) = DFT\left(x(t), ...x(t+T-1)\right), \tag{5.4}$$

et de même pour $S(f,t)$. Souvent, une fenêtre est utilisée pour une transformation de Fourier discrète :

$$X(f,t) = \sum_{\tau=0}^{T-1} w(\tau) x(t+\tau)\, e^{-j2\pi f \tau/T}, \tag{5.5}$$

où la fonction fenêtre $w(\tau)$ est choisie pour minimiser le chevauchement entre les bandes de fréquence. En utilisant la transformée de Fourier rapide (FFT) les convolutions peuvent être mises en œuvre efficacement dans le domaine de la transformée de Fourier discrète.

Nous avons utilisé deux types de mélange convolutif pour valider nos expériences. Le premier mélange utilise les filtres *Head Related Transfer Function* (HRTF), le second représente un mélange anéchoïque.

5.3.1 Modèle de filtres « Head Related Transfer Function » : HRTF

La technique de spatialisation sonore la plus proche de l'écoute naturelle est une technique reposant sur la reproduction au niveau des oreilles d'un auditeur de toutes les informations nécessaires pour la construction d'une image sonore extra-crânienne. Cette synthèse permet une localisation précise des sources sonores en trois dimensions ainsi qu'un rendu fidèle de l'environnement sonore (effet de salle). La diffusion des informations est réalisée grâce à un casque stéréophonique classique contrairement aux autres systèmes qui utilisent des hauts parleurs. L'utilisation d'un casque, et d'un système de suivi de mouvement, permet une immersion totale de l'auditeur dans une scène sonore sans interaction avec le monde extérieur. De plus, cette synthèse est réalisée grâce à des filtres de spatialisation et on peut extraire les indices perceptifs qui gouvernent la localisation sonore. Cette spécificité autorise l'étude psycho-acoustique des indices de localisation.

La technique de reproduction consiste à recréer, directement au niveau de chaque oreille, les signaux correspondants au champ acoustique que l'on souhaite simuler. Du point de vue matériel, cette technique fait donc appel comme nous l'avons cité, à un casque placé sur les oreilles de l'auditeur, permettant de contrôler de façon indépendante les signaux envoyés à chacune des oreilles.

Le principe est d'appliquer à la source sonore deux filtres différents pour construire les deux signaux destinés à chaque oreille ; ces filtres étant eux même dépendants de la position souhaitée de la source. Pour donner une impression réaliste, les signaux reconstruits doivent tenir compte de la présence d'un déphasage entre les deux oreilles, mais également des effets de masquage dus à la tête. Pour cela, il est nécessaire de posséder une banque de données rassemblant les différentes fonctions de transfert vers les deux oreilles, depuis un ensemble de points de l'espace. Cet ensemble de filtres est dénommé HRTF. Le fonctionnement du système est résumé dans la figure 5.3.

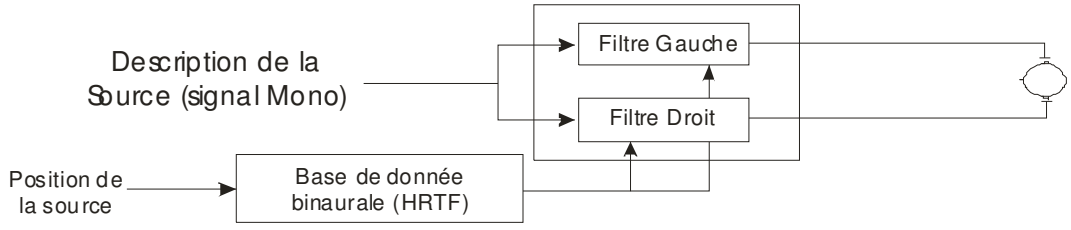

FIG. 5.3 – Schéma de principe de la reproduction des sources au niveau des oreilles du mannequin.

Ces filtres contiennent les informations liées aux phénomènes de diffraction, diffu-

sion et réflexion que subit une onde sonore lors de son trajet entre son point source et l'entrée du canal auditif de l'auditeur. Toutes les informations de spatialisations sont consignées dans les HRTF. Ces derniers sont propres à l'individu et à la position de l'espace simulée. Le principe de la mesure de HRTF est de placer des microphones dans les oreilles et d'enregistrer les signaux qui correspondent à différentes positions de source. Les HRTF sont les fonctions de transfert entre les signaux sources et les signaux au niveau des oreilles.

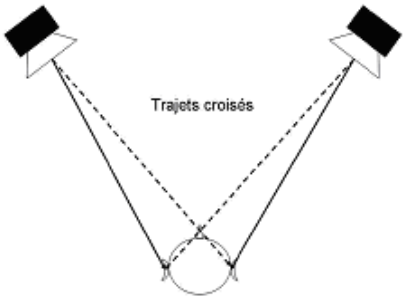

FIG. 5.4 – Représentation des signaux sources et des signaux au niveau des oreilles [129].

Les HRTF peuvent être représentées soit par des fonctions complexes soit par une approche traitement du signal qui les représentent comme des filtres ou des réponses impulsionnelles. Les HRTF sont alors considérées comme des systèmes linéaires et invariants dans le temps. De plus, provenant de mesures physiques, elles sont représentées par des filtres RIF (Réponse Impulsionnelle Finie), ou FIR en anglais (*Finite Impulse Response*), causales et stables.

5.3.2 Modèle de mélange anéchoïque

Un mélange anéchoïque est un mélange à «atténuations et décalages temporels» ; les observations sont des combinaisons linéaires des différentes sources qui sont décalées temporellement.

Afin de modéliser un enregistrement d'une scène auditive effectué à l'aide de microphones placés dans une salle dans laquelle les parois sont supposées anéchoïque, on utilise le modèle de mélange suivant :

$$x_m(t) = \sum_{m=1}^{N} a_m s_n(t - d_{mn}) \qquad 1 \leq m \leq M. \tag{5.6}$$

La seule différence par rapport au modèle instantané, est l'introduction des délais

d_{mn}. Le délai d_{mn} représente l'intervalle de temps entre l'émission du son par la source n et la captation de celui-ci par le microphone m. Ce délai est indéterminable dans l'absolu, mais on peut le caractériser par la différence de temps d'arrivée entre les canaux. Pour fixer cette indétermination, nous supposerons que $d_{1n} = 0$. Ce type de mélange inclut le cas où la propagation des ondes sonores se fait selon un trajet unique des sources aux capteurs et le temps de propagation n'est pas négligeable [125] (voir Figure 5.5).

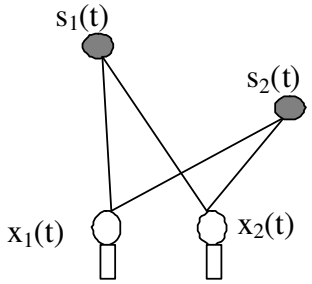

FIG. 5.5 – Modèle de mélange anéchoïque [125].

Nous supposons que nous n'avons aucune réflexion, et donc les ondes provenant directement des sources avec leur affaiblissement et leurs délais respectifs sont les seuls signaux reçus par les capteurs. L'équation 5.6 du mélange anéchoïque peut s'écrire de façon approximative dans le domaine de la transformée de Fourier à court terme par l'expression 2.5.

5.4 Modèles de séparation dans le domaine fréquentiel

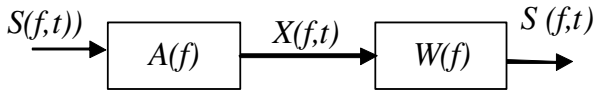

FIG. 5.6 – Système de séparation de sources dans le domaine fréquentiel.

Le modèle est donnée par les équations 2.5 et 2.6. Dans ce cas-ci, le critère de séparation est satisfait si les signaux récupérés sont permutés, multipliés par un facteur d'échelle, et représentent une version filtrée des signaux originaux,

$$W(f)\,A(f) = P(f)\,D(f)\,, \tag{5.7}$$

où $P(f)$ est une matrice de permutation et $D(f)$ une matrice diagonale avec une multiplication par un facteur échelle des coefficients des filtres sur sa diagonale. Si nous

pouvons identifier $A(f)$ et choisir $W(f)$ comme son inverse, alors $D(f)$ est une matrice identité, et ainsi, nous récupérerons des signaux exactement proches des signaux sources originaux. Un survol bibliographique de la séparation aveugle de sources dans le domaine fréquentiel est donné dans [130]. Les avantages et les inconvénients des approches dans le domaine temporel et fréquentiel ont été comparées. L'avantage du domaine fréquentiel, est que la séparation des mélanges convolutifs de sources est ramenée à un ensemble de séparations instantanées. Ceci est réalisé par le passage dans le domaine de Fourier, où nous obtenons un modèle de mélange instantané à chaque fréquence, en plus des gains importants en matière d'efficacité de calcul.

Le problème de mélange convolutif est alors réduit à un mélange «instantané» pour chacune des fréquences et le problème qui se pose dans le domaine fréquentiel, c'est la permutation et l'ambiguïté d'échelle. Si le problème est traité en convolutif pour chaque fréquence comme un problème séparé, les signaux source dans chaque fréquence peut être estimée avec une permutation et une échelle arbitraires. D'après les expressions 2.5, 2.6 et 5.7 nous obtenons :

$$\hat{S}(f,t) = P(f)\, D(f)\, S(f,t)\,. \tag{5.8}$$

5.5 Méthodes de référence de séparation aveugle de sources avec mélange convolutif

5.5.1 La méthode C-FICA

La méthode C-FICA est une méthode de séparation aveugle de sources pour des mélanges convolutifs à moyenne ajustée (MA) de processus MA indépendants. Elles consistent en des extensions dans le domaine temporel de l'algorithme Fast-ICA développé par Hyvärinen et Oja pour les mélanges instantanés. Elles réalisent un blanchiment convolutif dans le but d'utiliser des algorithmes rapides de type point fixe sans paramètre, associés aux critères de kurtosis ou de néguentropie pour estimer les processus d'innovation des sources. Des résultats de tests ont été présentés pour des sources colorées artificielles et pour des signaux de parole.

L'algorithme C-FICA est un algorithme temporel rapide effectuant la séparation aveugle de mélanges convolutifs. Cet algorithme a été proposé par J. Thomas et al. [131] en 2006. Il est basé sur une ortho-normalisation convolutive des observations. Pour l'extraction itérative des processus d'innovation, on utilise des mises à jour de type point fixe. Pour l'estimation des contributions des sources les auteurs utilisent un critère quadratique correspondant à la puissance de la différence entre l'observation et une version filtrée d'un processus d'innovation. Le critère est minimisé par un processus

de filtrage non-causal de Wiener.

La relation globale du système de mélange s'écrit dans le domaine Z :

$$X(Z) = A(Z)S(Z), \tag{5.9}$$

où $X(Z)$ et $S(Z)$ sont respectivement les transformées en Z de $x(n)$ et $s(n)$, et la matrice $A(Z)$ est composée des fonctions de transfert $A_{ij}(Z) = Z[a_{ij}(n)]$ des filtres de mélange. Chaque source $s_j(n)$ est supposée s'exprimer dans le domaine en Z suivant :

$$S_j(Z) = F_j(Z)\, U_j(Z), \tag{5.10}$$

où $F_j(Z)$ correspond à un filtre et $U_j(Z)$ est la transformée en Z d'un processus $u_j(n)$ qui est $U(Z) = [U_1(Z), \ldots, U_N(Z)]^T$, nous pouvons exprimer l'équation de mélange comme suit :

$$X(Z) = G(Z)U(Z), \tag{5.11}$$

où

$$G(Z) = A(Z)F(Z), \tag{5.12}$$

avec

$$F(Z) = diag\left(F_1(Z), ..., F_N(Z)\right). \tag{5.13}$$

Nous faisons les hypothèses suivantes concernant le modèle de mélange présenté ci-dessus :

• Le processus est réel, de moyenne nulle, i.i.d. et spatialement indépendant, c.-à-d. ses composantes $u_j(n)$ sont statistiquement indépendantes entre elles mais n'ont pas nécessairement la même distribution. Nous supposons qu'une composante au plus est gaussienne.

• Les matrices de filtres $F(Z), A(Z)$ et donc $G(Z)$ sont causales, RIF et non singulières. Notons que les systèmes RII peuvent aussi être approximés par des modèles équivalents RIF (d'ordre élevé).

$A_{ij}(Z)S_j(Z)$ représente les contributions de toutes les sources dans chaque observation . Dans les méthodes à déflation telles qu'indiqué dans [131] (voir *Annexe E*), la séparation est réalisé en utilisant la procédure suivante :

i. Extraire le processus d'innovation qu'on appelle $u_j(n)$ d'une source $s_j(n)$ à partir des observations.

ii. Identifier M filtres de coloration et les appliquer à $u_j(n)$ pour estimer les contributions de $s_j(n)$ dans chaque observation, c.-à-d. $A_{ij}(z)S_j(z)$.

iii. Soustraire les contributions de toutes les observations.

iv. Décrémenter N. Si $N \neq 1$, retourner à l'étape (1) pour extraire une autre source.

5.5.2 La méthode DUET

DUET (*Degenerate Unmixing and Estimation Technique*) est une méthode qui peut s'appliquer lorsque les sources sont W-disjoints orthogonales. Cet algorithme a été proposé par A. Jourjine et al. [132] en 2000. Il est basé sur une analyse temps-fréquence des observations. Les auteurs proposent de résoudre le problème de la séparation de sources dans le cas de mélanges anéchoïques (c.-à-d. mettant en jeu des retards purs et atténuations) de N sources et deux observations de la forme :

$$\begin{bmatrix} X_1(Z) \\ X_2(Z) \end{bmatrix} = \begin{bmatrix} 1 & . & . & . & 1 \\ a_{21}z^{-l_1} & . & . & . & a_{2N}z^{-l_N} \end{bmatrix} \begin{bmatrix} S_1(Z) \\ . \\ . \\ . \\ S_N(Z) \end{bmatrix}, \tag{5.14}$$

où z^{-l_i} représente l'opérateur de retard de l_i échantillons. En considérant la transformée de Fourier, le mélange observé peut s'écrire :

$$\begin{bmatrix} X_1(f) \\ X_2(f) \end{bmatrix} = \begin{bmatrix} 1 & . & . & . & 1 \\ a_{21}e^{-j2\pi f l_1} & . & . & . & a_{2N}e^{-j2\pi f l_N} \end{bmatrix} \begin{bmatrix} S_1(f) \\ . \\ . \\ . \\ S_N(f) \end{bmatrix}. \tag{5.15}$$

Ils supposent ensuite que les sources sont W-disjointes-orthogonales dans le plan temps-fréquence associé à la transformée considérée avec une fenêtre de pondération w. Cette forte hypothèse signifie que les sources ne présentent aucun recouvrement dans ce domaine temps-fréquence. Ainsi, pour une fréquence donnée où seule la source $S_i(f)$ est présente, les signaux observés deviennent :

$$\begin{bmatrix} X_1(f) \\ X_2(f) \end{bmatrix} = \begin{bmatrix} 1 \\ a_{2i}e^{-j2\pi f l_i} \end{bmatrix}, \tag{5.16}$$

avec :

$$\frac{X_2(f)}{X_1(f)} = a_{2i}e^{-j2\pi f l_i}. \tag{5.17}$$

L'amplitude et le retard associés à cette source sont alors donnés par :

$$a_{2i}, l_i = \left(\left\| \frac{X_2(f)}{X_1(f)} \right\|, \ \mathrm{Im} \left(\log \left(\frac{X_2(f)}{X_1(f)} \right) \right) / 2\pi f \right). \tag{5.18}$$

Les auteurs proposent de rajouter la dimension temps et réalisent alors une étude temps-fréquence, conduisant à :

$$a_{2i}, l_i = \left(\left\| \frac{X_2(t,f)}{X_1(t,f)} \right\|, \ \mathrm{Im} \left(\log \left(\frac{X_2(t,f)}{X_1(t,f)} \right) \right) / 2\pi f \right). \tag{5.19}$$

Le problème consiste à présent à déterminer les zones temps-fréquence (t, f) où chacune des N sources est présente. Pour cela, les auteurs [132] utilisent l'hypothèse que les sources sont W-disjointes-orthogonales et effectuent un histogramme amplitude/phase de toutes les valeurs de $(a_{2i},\ l_i)$ rencontrées. Sous l'hypothèse annoncée, cet histogramme présente des pics autour des valeurs correspondant aux paramètres de la matrice de mélange vue dans le plan temps-fréquence. Conscient de la limitation de la W-disjointe-orthogonalité, les auteurs ont montré que cette propriété pouvait être légèrement relaxée en considérant une version approximative en particulier pour la parole [133], [134]. Ils proposent alors une méthode de reconstruction des sources d'origine avec deux variantes, suivant le nombre de sources par rapport au nombre d'observations :

- Si le nombre de sources est inférieur ou égal au nombre d'observations, la matrice de mélange de la relation 5.15 peut être directement inversée, ce qui fournit les transformées temps-fréquence des sources, qui sont ensuite reconstruites par transformée temps-fréquence inverse.

- Au contraire, dans le cas sous-déterminé la matrice de mélange ne peut être inversée et les auteurs proposent de reconstruire approximativement les sources à partir des zones temps-fréquence associées à chacune d'entre elles.

Pour chacune de ces deux variantes, leur méthode fonctionne en deux temps :

i- détermination de l'histogramme amplitude/phase ;

ii- à partir des valeurs des coefficients de mélanges obtenues par l'histogramme précédent, refaire un deuxième passage afin de déterminer les zones temps-fréquence associées à chaque source.

Cette dernière opération définit un masque temps-fréquence propre à chaque source. Pour la première approche, il suffit de calculer 5.19 dans une zone de ce masque pour obtenir les paramètres associés à cette source. Pour la deuxième variante, l'estimation de chacune des sources est réalisée en calculant la transformée inverse d'une observation en ne considérant que les zones définies par chaque masque, c.-à-d. en forçant à zéro tout le domaine extérieur au masque considéré.

Afin de ne plus nécessiter ces deux étapes, cette méthode a été légèrement modifiée [134] en introduisant une analyse par maximum de vraisemblance associé à un algorithme de recherche d'extremum.

Pour cette approche, appelée DUET, il est possible de traiter le cas de mélanges sous-déterminés constitués de retards purs, au prix de l'approximation mentionnée sur le caractère incomplet de la carte temps-fréquence associée à chaque source. Ensuite la détermination des zones où chacune des sources est présente et seule s'effectue de manière automatique avec une bonne précision si les sources sont réellement W-disjointes-orthogonales. Enfin, le coût de calcul est extrêmement réduit, jusqu'à pouvoir être implanté en temps réel sans aucune difficulté [134].

Il faut néanmoins préciser que la valeur des retards maximaux qu'il est possible de traiter est étroitement liée à la longueur de la fenêtre temporelle utilisée dans la transformée temps-fréquence. Cette longueur ne peut être choisie librement car elle conditionne la résolution dans le plan temps-fréquence. De plus, l'hypothèse que les sources soient W-disjointes-orthogonales (ou approximativement W-disjointes-orthogonales), est très restrictive et, dans le cas contraire, l'histogramme obtenu ne donne pas de manière très précise la valeur des bons coefficients de mélange, avec la conséquence d'obtenir difficilement de bonnes performances. Il n'est également pas exclu que certaines sources soient cachées par d'autres dans le plan temps-fréquence, ce qui les rend totalement impossible à extraire par cette méthode.

Notons que dans le cas de $M = N = 2$, un algorithme ICA performera normalement mieux que le DUET car il n'y aura pas de "*musical noise*" dans les signaux rehaussés, alors que le masquage utilisé par le DUET introduira du "*musical noise*". Les algorithmes ICA ont une équivalence en terme de "*beamformers*" [135] et peuvent mener à de bonnes performances pour des sources ponctuelles dans le cas où $M \geq N$. Par contre pour le cas sous-déterminé ($M < N$) la performance de l'algorithme ICA se dégrade rapidement, contrairement à la méthode DUET.

5.5.3 Séparation Aveugle de Mélanges Convolutifs par la méthode FOPCA

Nous avons appliqué la méthode OPCA dans le domaine temporel pour le mélange instantané avec succès. Dans ce chapitre, nous appliquons cette méthode dans le domaine fréquentiel pour le cas d'un mélange convolutif de signaux de parole, selon le modèle illustré dans la figure 5.7. Nous l'appellerons alors FOPCA : Frequency OPCA.

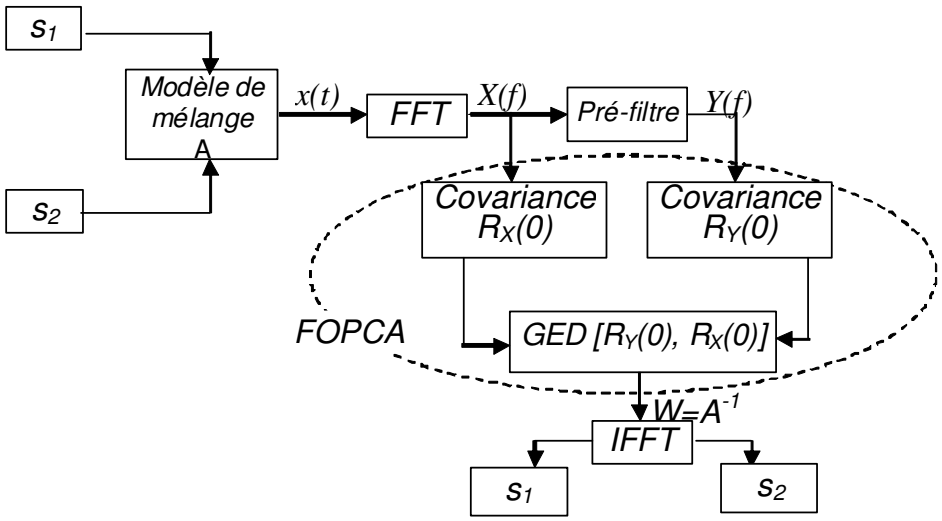

FIG. 5.7 – Schéma synoptique de la BSS utilisant la méthode FOPCA dans le domaine fréquentiel pour les mélanges convolutifs.

Subséquemment, nous considérerons le problème de la BSS avec l'approche OPCA, en analysant les signaux observés x et leur version filtrée. Notons que la matrice de covariance à un retard 0 de $x(k)$ est donnée par l'équation 4.36.

Maintenant, considérons un filtre scalaire et linéaire ayant comme coefficients :

$H = [H_0, \ldots., H_M]$ (désigné sous le nom du pré-filtre dans la Figure 5.7) appliqué à $X(f,t)$:

$$Y\left(t,f\right) = \sum_{m=0}^{M} H_m X(t - l_m, f). \tag{5.20}$$

La matrice de covariance de Y à une valeur de retard 0, s'exprime par :

$$R_y\left(0\right) = E\left\{Y\left(t,f\right)Y\left(t,f\right)^T\right\} = \sum_{p,q=0}^{M} H_p H_q R_x(l_p - l_q). \tag{5.21}$$

De l'équation 2.1 du deuxième chapitre mais sans bruit additif, il s'ensuit que :

$$R_x(l_m) = AR_s(l_m)A^T, \tag{5.22}$$

ainsi,

$$R_y\left(0\right) = ADA^T, \tag{5.23}$$

avec :

$$D = \sum_{p,q=0}^{M} H_p H_q R_s(l_p - l_q). \tag{5.24}$$

Supposant que A est carrée et inversible :

$$R_y(0)\, A^{-T} = AD = AA^T A^{-T} D = R_x(0)\, A^{-T} D. \tag{5.25}$$

L'équation 5.25 exprime une décomposition en vecteurs propres généralisés de la matrice composée $[R_Y(0), R_X(0)]$. Ceci est équivalent au problème de FOPCA pour les paires de signaux $[Y(t,f), X(t,f)]$. Les valeurs propres généralisées pour ce cas là sont les éléments de la diagonale de la matrice D. Les colonnes de la matrice A^{-T} représentent les vecteurs propres généralisés. Les vecteurs propres sont uniques jusqu'à une permutation et à un facteur à condition que les valeurs propres soient distinctes (ceci est vrai en général). Dans ce cas-ci, pour toute matrice propre généralisée W nous avons $W = A^{-T}P$ avec P étant une matrice de permutation avec un facteur d'échelle ; chaque ligne et chaque colonne contiennent exactement un élément non nul. Par conséquent, les sources peuvent être estimées par :

$$\hat{S}(t,f) = W^T X(t,f), \tag{5.26}$$

qui peut s'écrire :

$$\hat{S}(t,f) = P^T A^{-1} A S(t,f) = P^T S(t,f), \tag{5.27}$$

où $\hat{S}(t,f) = [\hat{S}_1(t,f), \hat{S}_2(t,f)]^{\ T}$ représente le vecteur du signal source estimé et $W(f)$ la matrice de séparation à chaque fréquence f. La matrice de séparation $W(f)$ est déterminée de telle sorte que $\hat{S}_1(t,f)$ *et* $\hat{S}_2(t,f)$ deviennent mutuellement non-corrélés, parce que les signaux sources sont supposés être mutuellement décorrélés et de moyenne nulle. Les sources estimées sont égales aux sources originales (non observables) mais avec un ordre et un gain arbitraire. Nous appliquons par la suite, la Transformée de Fourier Inverse (IFFT) de $\hat{S}_1(t,f)$ pour récupérer les signaux dans le domaine temporel.

$$\hat{s}(t) = IFFT\left(\hat{S}(t,f)\right). \tag{5.28}$$

Le pré-filtre indiqué dans la figure 5.7 s'exprime par l'expression 4.53.

Comme nous sommes dans un milieu faiblement réverbéré, l'ambiguïté de permutation de fréquence ne se produit pas. De plus, à cause de l'indépendance des sources, la corrélation croisée entre les sorties dans le domaine fréquentiel correspondant aux différentes sources est nulle, et donc les sorties correspondant à la même source peuvent

être déterminées en maximisant la corrélation croisée entre elles. Ceci ne peut être utilisé que pour les sources qui sont temporellement colorées, ce qui est le cas des signaux de parole, dont il a été pris en compte comme hypothèse.

5.6 Expériences et résultats

Pour évaluer notre approche dans le cas convolutif, nous l'avons comparé avec les techniques très connues telles que C-FICA et DUET, et nous avons utilisé deux types de mélange convolutif, tels que le modèle HRTF et le modèle anéchoïque. FOPCA est efficace, comme nous le voyons dans le domaine temporel et le domaine fréquentiel où l'on note que les signaux d'origine et les signaux estimés par FOPCA sont très proches.

5.6.1 Cas d'un mélange de deux signaux de parole

Dans nos expériences avec la méthode FOPCA, l'ordre du pré-filtre choisi est de trois. La recherche du filtre optimal est transformée en la recherche pour le filtre donnant des valeurs propres les plus étendues que possible. La recherche est exhaustive et est réalisée pour des valeurs de α et β dans un intervalle donné de h ($\forall\ \alpha,\ \beta\ \in [h_{min},\ h_{max}]$). Dans nos expériences, nous avons fixé h_{min}=-5, h_{max}=5, avec un pas de progression de $0,2$. Ces valeurs ont été trouvées empiriquement. Les sources originales et les observations des données sont de dimension $N = 2$. Les signaux de parole ont été enregistrés à partir de la base de données TIMIT, avec des durées typiques de six secondes. Les informations concernant cette base de données se trouvent dans le quatrième chapitre . Les mêmes phrases que pour le mélange instantané de cette base de données TIMIT ont été choisies pour évaluer nos méthodes de BSS. Nous considérons le problème de séparation d'un mélange convolutif à deux-entrées deux-sorties, de deux signaux de parole $s_1(n)$ et $s_2(n)$, respectivement prononcés par une femme et un homme. Nous avons également utilisé une série de six autres signaux, des différents intervenants (trois femmes et trois hommes) lisant des phrases d'une durée d'environ 30 secondes. Ces signaux de parole ont été récoltés par Dimitri Nion, et al. [136]. La fréquence d'échantillonnage des signaux est de 16 kHz. À des fins de comparaison nous avons aussi utilisé les deux signaux de parole en anglais appelés : « *speech*$_1$ » et « *speech*$_2$ » qui ont été proposés par J. Thomas et al. [131] dans l'algorithme C-FICA. Ces signaux sont échantillonnés à une fréquence de 20 kHz et une durée de cinq secondes.

Méthodes de séparation de signaux de parole mélangés convolutivement suivant le modèle HRTF

Pour le mélange utilisant le modèle des HRTF, l'étude expérimentale a été menée sur des mesures effectuées sur un mannequin artificiel avec lequel les HRTF ont été adaptés à n'importe quel auditeur. La tête d'un mannequin avec deux microphones (un sur chaque oreille) a été utilisée au lieu d'un réseau de microphones. Ce type d'enregistrement a été utilisé pour étudier l'efficacité de la BSS au cours d'une configuration plus naturelle des sources [137]. Cette situation tient compte de toutes les modifications dans un champ acoustique en rapport avec la tête, autrement dit les fonctions Head Related Transfert Function (HRTF). Les HRTF influencent à la fois le niveau de pression acoustique et les spectres des signaux source arrivant aux oreilles. Nous avons exécuté les tests dans un cadre global avec un filtre de mélange mesuré aux oreilles du mannequin développé dans [137]. Pour une première expérience, nous avons sélectionné des réponses impulsionnelles associées aux positions des sources définies par les angles 30 et -80 degrés par rapport à la tête du mannequin comme nous pouvons le voir sur la figure 5.8.

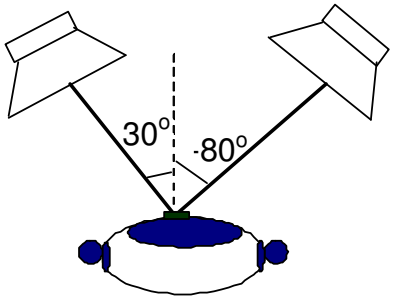

FIG. 5.8 – Le modèle convolutif (HRTF) : positions des sources définies par les angles 30 et - 80 degrés par rapport à la tête d'un mannequin.

Résultats expérimentaux utilisant la méthode FOPCA

Nous avons testé la méthode FOPCA en utilisant un pré-filtre d'ordre 3, comme indiqué précédemment. L'utilisation des matrices de corrélation augmente l'information à l'entrée dans l'estimation du processus et améliore donc la qualité de la séparation. Les données sont pré-filtrées temporellement. FOPCA est une approche utilisant des itérations récursives. Elle a une bonne performance dans un algorithme itératif avec une convergence rapide et ne nécessite pas de pré-blanchiment. Les résultats sont calculés avec un processus en trois itérations choisis après quelques expériences. En outre,

l'approche proposée est itérative et améliore les performances en combinant quelques décalages pour l'estimation des paramètres de mélange. Nous avons aussi expérimenté plusieurs longueurs de la fenêtre d'analyse. Nous avons choisi une longueur de 512 pour les signaux échantillonnés à 16 kHz et une longueur de 256 pour les signaux échantillonnés à 20 kHz. Ces longueurs de fenêtre sont les valeurs qui ont donné les meilleurs résultats où les signaux sont presque stationnaires. Cette structure temporelle est utilisée afin de reconstituer des signaux de source à partir de composantes fréquentielles.

Dans la figure 5.10, nous montrons les signaux mélangés par convolution avec le modèle HRTF dont les positions sont définies par les angles 30 et -80 degrés. Dans les figures 5.9 et 5.12, nous représentons les deux signaux de parole originaux prononcés par un homme et une femme enregistrés à partir de la base de données TIMIT dans respectivement, le domaine temporel et le domaine fréquentiel. Dans le figures 5.11 et 5.14, nous retrouvons les signaux estimés après séparation par la méthode FOPCA.

Les figures 5.15 à 5.20 montrent les résultats des signaux de parole "$speech_1$" et "$speech_2$" utilisés par J. Thomas et al., pour l'algorithme C-FICA.

Les figures 5.21 à 5.26 montrent les résultats des signaux de parole enregistrés par Dimitri Nion et al. [136]. Dans nos expériences, nous avons utilisé six signaux de parole : trois locutrices et trois locuteurs lisant des phrases et nous les avons nommé : s_1, s_3, s_5 pour les signaux vocaux féminins et s_2, s_4, s_6 pour les signaux vocaux masculins. Pour alléger le texte, nous ne représentons que les signaux de parole s_1 et s_2 et leur séparation après mélange avec le modèle HRTF. La même approche a été essayée avec les autres signaux s_3, s_4, s_5 et s_6 et les mêmes conclusions ont été obtenues et présentées dans le tableau 5.8.

Signaux de parole de la base de données TIMIT : Domaine temporel :

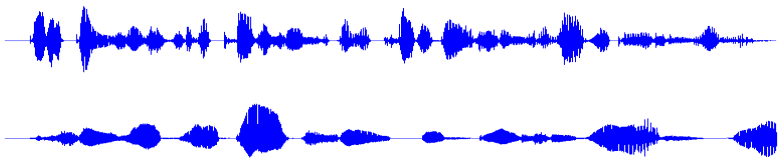

FIG. 5.9 – Signaux de parole originaux de la base de données TIMIT.

FIG. 5.10 – Signaux mélangés par convolution avec le modèle HRTF.

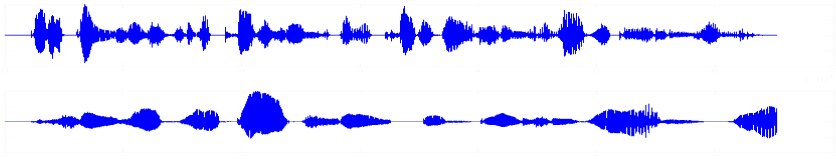

FIG. 5.11 – Signaux estimés par la méthode FOPCA.

Domaine temps-fréquence :

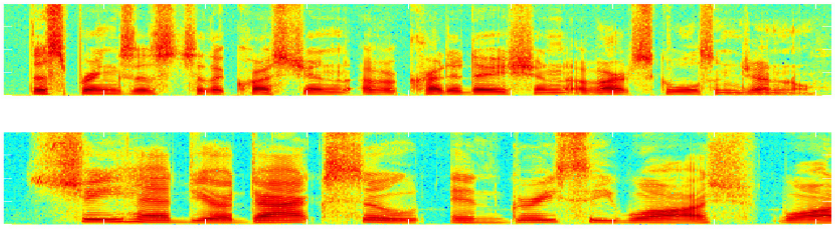

FIG. 5.12 – Spectrogrammes des signaux de parole originaux .

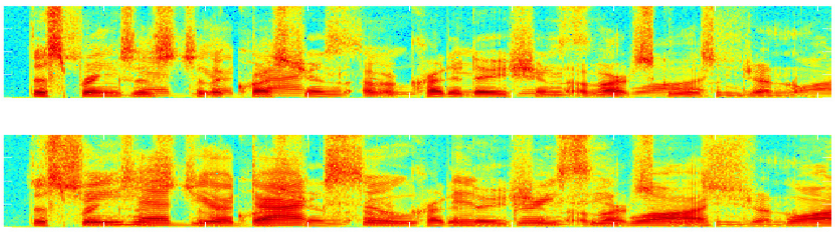

FIG. 5.13 – Spectrogrammes des signaux de parole mélangés convolutivement avec le modèle HRTF.

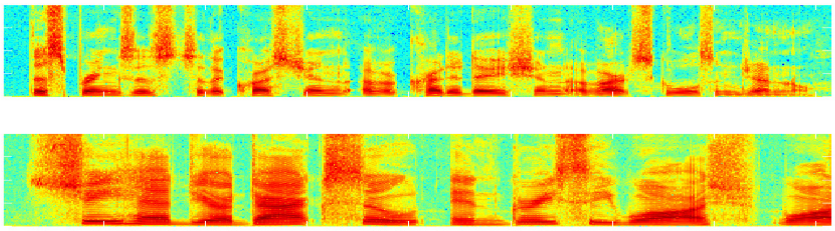

FIG. 5.14 – Spectrogrammes des signaux de parole estimés par la méthode FOPCA.

Séparation des signaux de parole recueillis par J. Thomas et al. [131]
Domaine temporel :

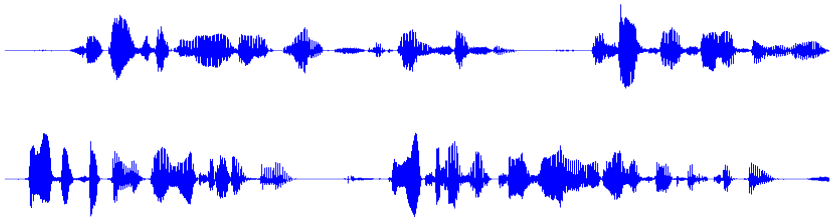

FIG. 5.15 – Signaux originaux : signal "speech$_1$" et signal "speech$_2$ ".

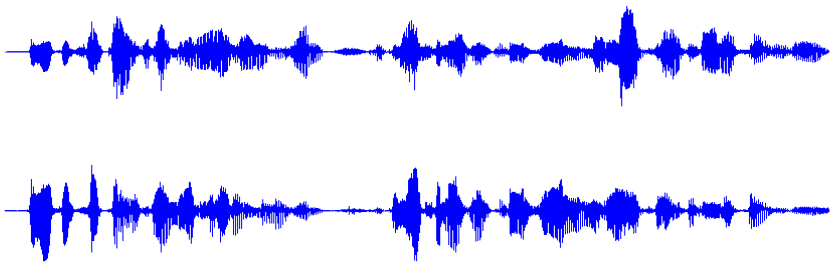

FIG. 5.16 – Signaux mélangés par convolution avec le modèle HRTF.

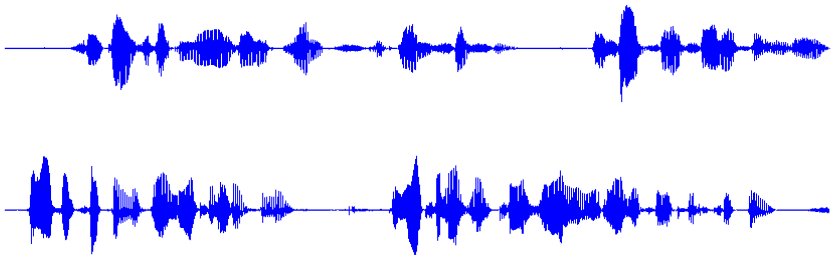

FIG. 5.17 – Signaux estimés par la méthode FOPCA pour les signalux "speech$_1$" et "speech$_2$".

Domaine temps-fréquence

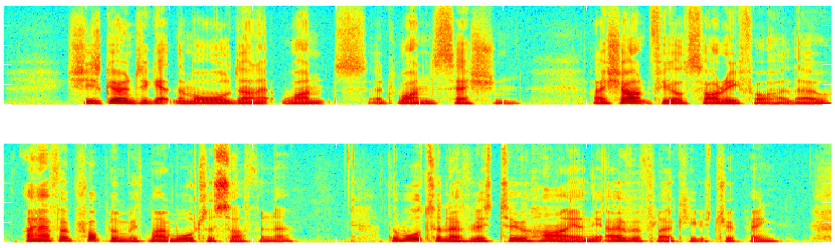

FIG. 5.18 – Spectrogramme des signaux originaux, "speech$_1$" et "speech$_2$".

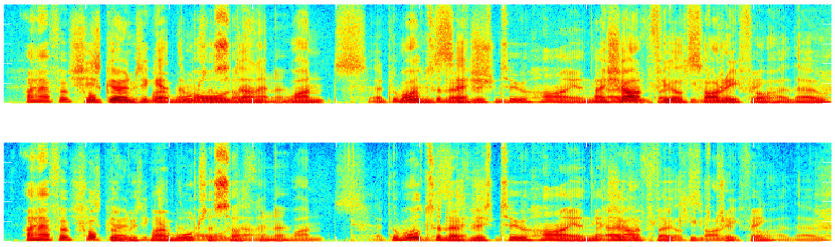

FIG. 5.19 – Spectrogramme des signaux "speech$_1$" et "speech$_2$" mélangés par convolution avec le modèle HRTF.

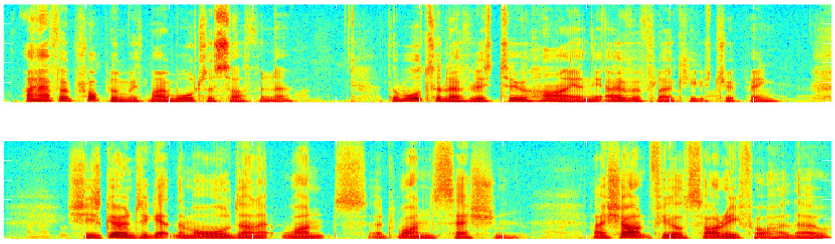

FIG. 5.20 – Spectrogramme des signaux estimés par la méthode FOPCA pour "speech$_1$" et "speech$_2$".

Séparation des signaux de parole recueillis par D. Nion et al. [136]

Domaine temporel :

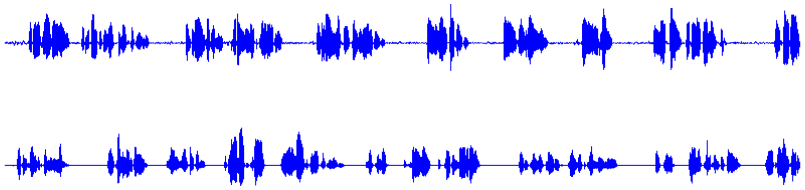

FIG. 5.21 – Signaux de parole originaux : signal original d'un locuteur s_2 et signal original d'une locutrice s_1.

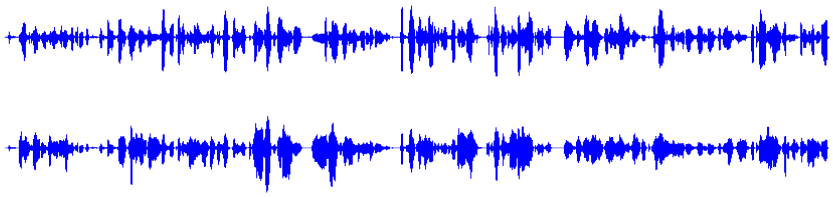

FIG. 5.22 – Signaux s_1 et s_2 mélangés convolutivement utilisant le modèle HRTF.

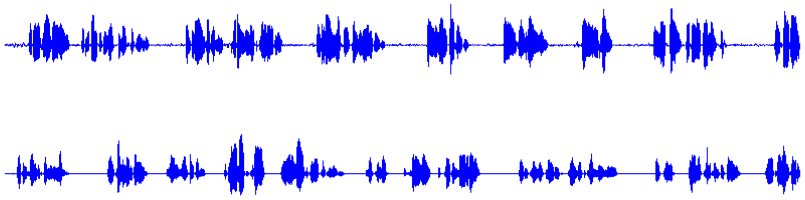

FIG. 5.23 – Signaux estimés $\hat{s}_1$ et $\hat{s}_2$ par la méthode FOPCA.

Domaine temps-fréquence

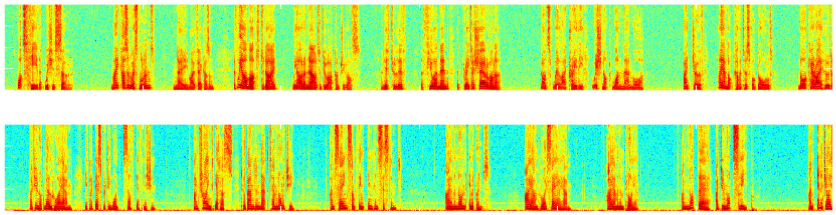

FIG. 5.24 – Spectrogramme des signaux de parole originaux s_1 et s_2.

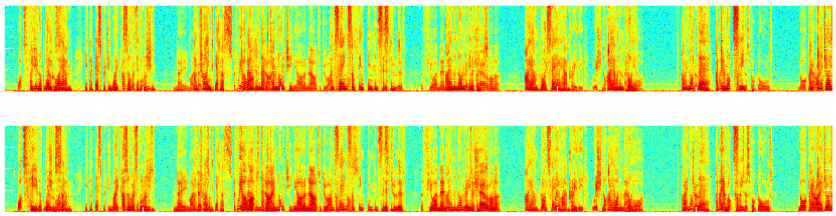

FIG. 5.25 – Spectrogramme des signaux s_1 et s_2 mélangés par convolution utilisant le modèle HRTF .

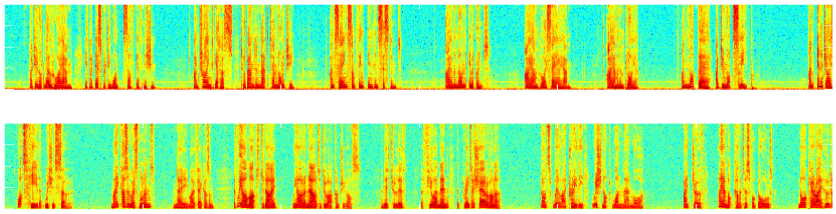

FIG. 5.26 – Spectrogramme des signaux $\hat{s}_1$ et $\hat{s}_2$ estimés par la méthode FOPCA .

Résultats expérimentaux utilisant la méthode C-FICA

Dans le souci d'avoir une bonne comparaison, nous avons effectué les mêmes types de mélanges que précédemment c.à.d. la sélection les filtres associés aux angles 30° et -80°. La séparation a été réalisée en utilisant la néguentropie avec :

$$G(x) = e^{-x^2}. \tag{5.29}$$

Dans l'algorithme une fenêtre d'extraction de processus d'innovation a été choisie là où les deux sources sont stationnaires (les filtres obtenus ont ensuite été appliqués à la totalité des signaux). Pour déterminer cette fenêtre d'extraction, une procédure a été développée qui consiste à effectuer la séparation successivement sur plusieurs fenêtres et à choisir celle donnant des sorties de contributions les plus corrélées.

C-FICA est un algorithme itératif, le nombre d'itérations utilisé est de 10000, et l'ordre des filtres d'extraction a été choisi de 80. L'extraction du processus d'innovation est réalisée sur une fenêtre d'extraction que nous avons choisie d'une longueur de 256, ce qui correspond à 16 ms et où les signaux sont à peu près stationnaires. L'ordre des filtres de recoloration est de 400. Nous avons expérimenté plusieurs données de la longueur de la fenêtre d'extraction de l'innovation de l'ordre du filtre d'extraction et l'ordre des filtres de recoloration et les valeurs choisies ont donné les meilleurs résultats.

Domaine temporel :

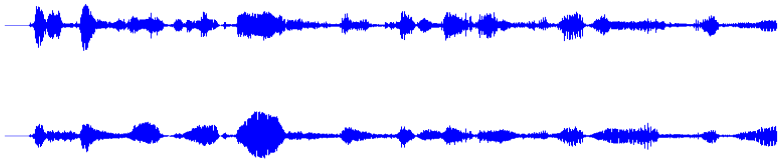

FIG. 5.27 – Signaux de TIMIT estimés par la méthode C-FICA.

Domaine temps- fréquence :

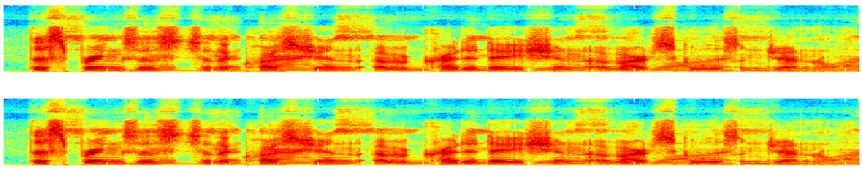

FIG. 5.28 – Spectrogrammes des signaux de parole de TIMIT estimés par la méthode C-FICA.

Résultats expérimentaux utilisant la méthode DUET

L'implémentation pratique de l'approche DUET se résume comme suit :

- Construire une représentation temps-fréquence des deux mélanges de signaux.

- Prendre le rapport des deux mélanges et obtenir des estimations des paramètres locaux des mélanges.

- Combiner l'ensemble des estimations des paramètres locaux des mélanges en N paires correspondant aux vraies paires de paramètres de mélange.

- Générer un masque binaire pour chaque paire de paramètres de mélange correspondant aux points temps-fréquence qui produisent des paramètres de mélange particuliers.

- Séparer les sources en multipliant chaque masque par un des mélanges

- Convertir chaque représentation temps-fréquence dans le domaine temporel.

Un histogramme amplitude/phase est construit. Cet histogramme présente des pics autour des valeurs correspondant aux paramètres de la matrice de mélange vue dans

le plan temps-fréquence. La figure 5.29 représente l'histogramme dans le cas d'un mélange suivant le modèle HRTF. Nous remarquons deux pics prédominants.

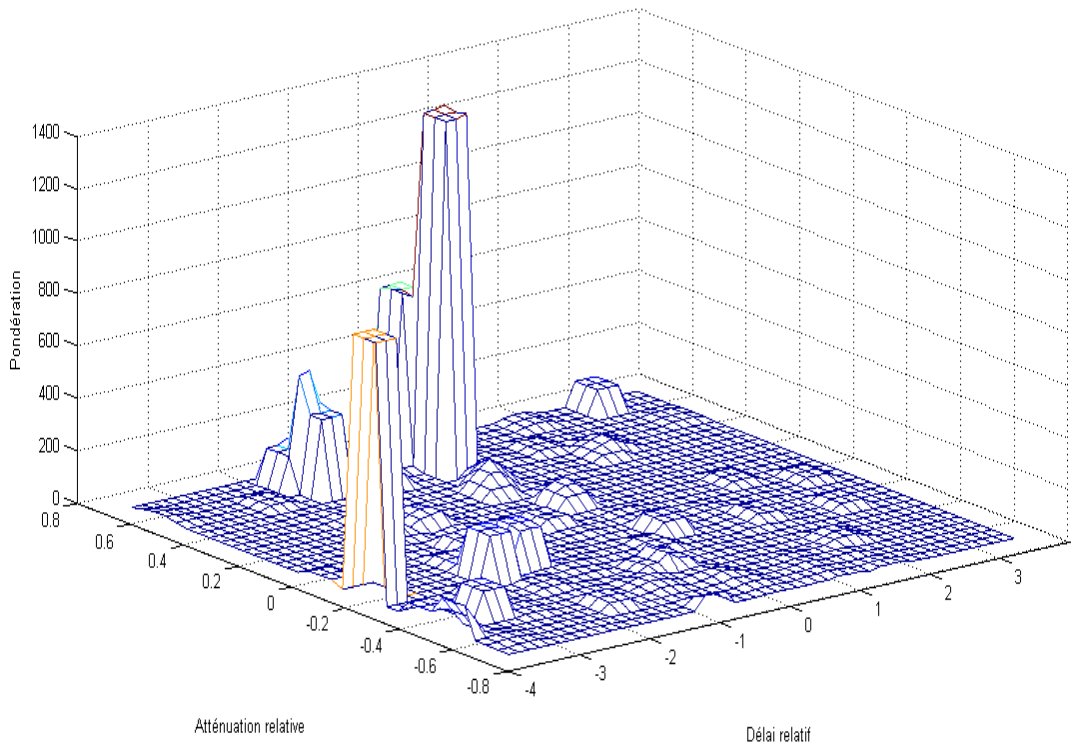

FIG. 5.29 – Histogramme représentant les signaux mélangés suivant le modèle HRTF.

L'extraction des paramètres des mélanges se fait en localisant les pics dans l'histogramme. Une fois les pics identifiés, le but est de déterminer les masques temps-fréquence qui sépareront chaque source à partir des mélanges. Il suffit d'attribuer chaque point temps-fréquence à l'emplacement des pics qui est le plus proche du paramètre local des estimations extraites du point temps-fréquence.

Les paramètres p et q représentent les poids de l'histogramme. On définit l'estimation de l'atténuation symétrique locale par ces deux paramètres. Les valeurs de p et q ont été choisies telles qu'indiquées dans [140], c.-à.-d. que pour les signaux de parole, $p = q = 2$. L'histogramme montre deux pics associés avec les deux signaux sources originaux. Les pics sont exactement les paires des paramètres de mélange pour chacune des deux sources. D'après cet histogramme et les pics prédominants, nous trouvons les atténuations relatives suivantes : -0.1 et 0.44 et les délais relatifs : -3.5 et 0. Les résultats de la séparation sont donnés dans le domaine temporel dans la figure 5.30,

et dans le domaine temps-fréquence dans la figure 5.31.

Domaine temporel :

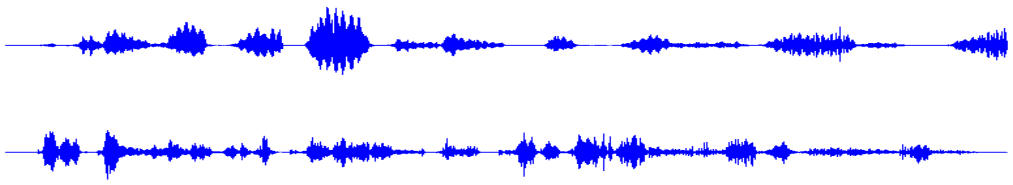

FIG. 5.30 – Signaux estimés par la méthode DUET.

Domaine temps-fréquence :

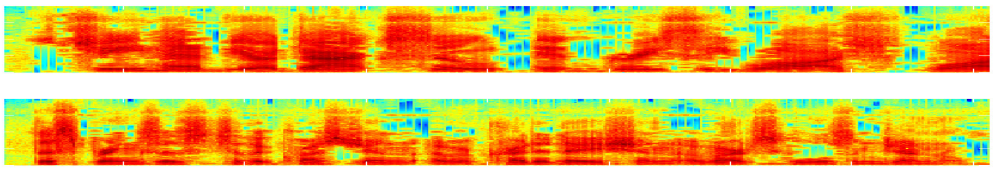

FIG. 5.31 – Spectrogrammes des signaux de parole estimés par la méthode DUET.

Méthodes de séparation de signaux de parole mélangés convolutivement suivant le modèle Anéchoïque

Pour le modèle anéchoïque, les deux mélanges contiennent les deux sources avec des amplitudes relatives de $1,1$ et $0,9$ et avec des échantillons de retards de -2 et 2. Ces mêmes mélanges sont utilisés pour les trois différentes méthodes. Les figures 5.32 et 5.33 montrent respectivement, les signaux mélangés dans le domaine temporel et fréquentiel.

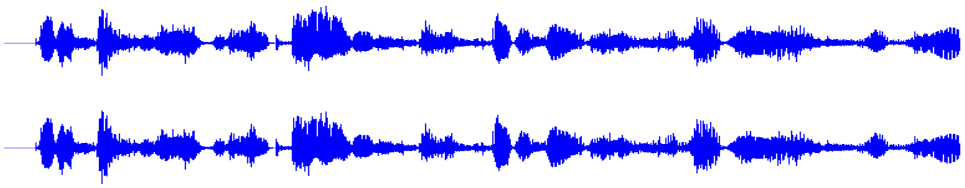

FIG. 5.32 – Signaux mélangés par convolution suivant le modèle anéchoïque.

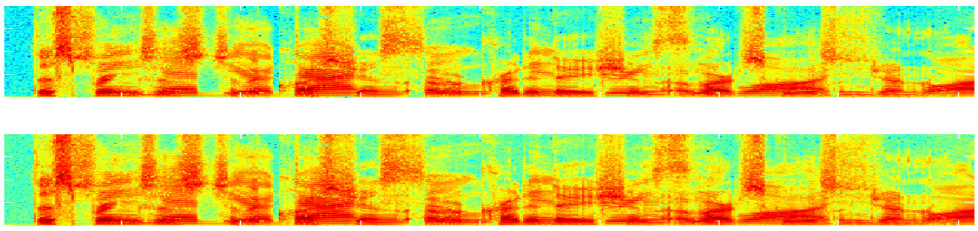

FIG. 5.33 – Spectrogramme des signaux mélangés par convolution suivant le modèle anéchoïque.

Résultats expérimentaux utilisant la méthode FOPCA

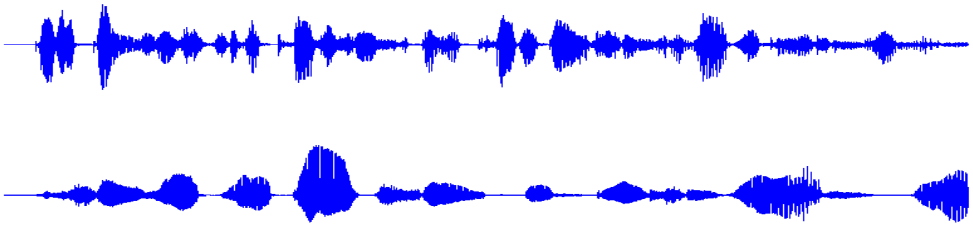

FIG. 5.34 – Signaux de TIMIT estimés par la méthode FOPCA dans le domaine temporel.

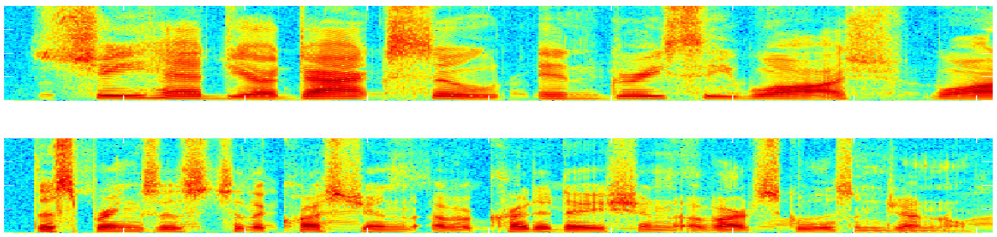

FIG. 5.35 – Spectrogramme des signaux de TIMIT estimés par la méthode FOPCA.

Résultats expérimentaux utilisant la méthode C-FICA

Nous avons testé la méthode C-FICA avec un mélange anéchoïque. Les figures 5.36 et 5.37 montrent respectivement, les signaux estimés dans le domaine temporel et temps-fréquence.

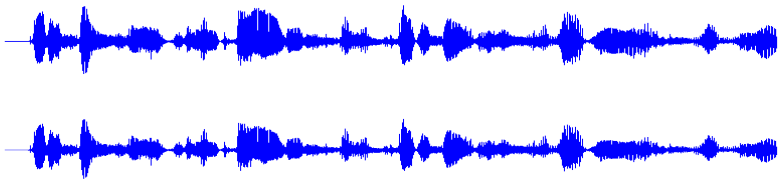

FIG. 5.36 – Signaux de TIMIT estimés par la méthode C-FICA.

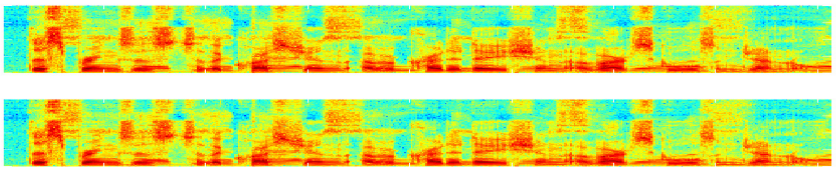

FIG. 5.37 – Spectrogramme des signaux de TIMIT estimés par la méthode C-FICA.

Résultats expérimentaux utilisant la méthode DUET

La méthode DUET, comme expliqué ultérieurement nécessite la construction d'un histogramme. Dans la figure 5.38, l'histogramme montre deux pics associés avec les deux signaux sources originaux. Les pics représentent les paires des paramètres de mélange pour chacune des deux sources, et donnent les atténuations relatives de $0,4$ et $-0,4$ et les délais relatifs de 0. p et q ont été également choisies égaux à 2. Les résultats de la séparation est donnée dans le domaine temporel dans la figure 5.39, et dans le domaine temps-fréquence dans la figure 5.40.

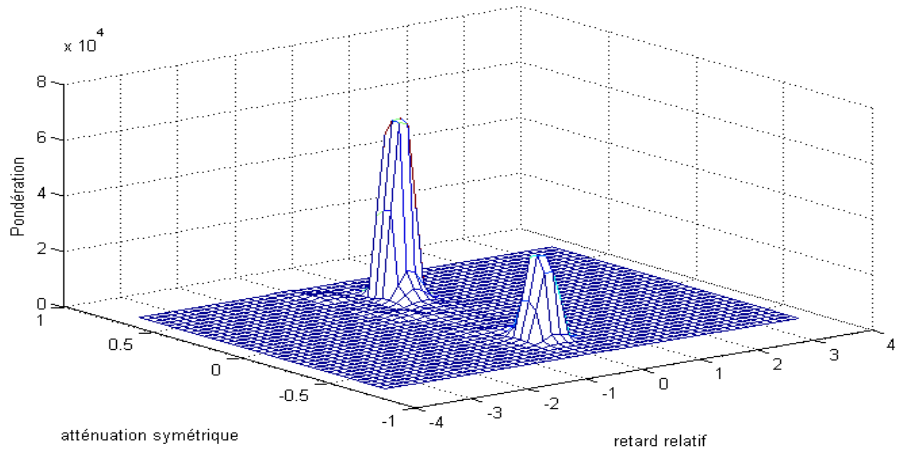

FIG. 5.38 – Histogramme dans le cas d'un mélange convolutif anéchoïque.

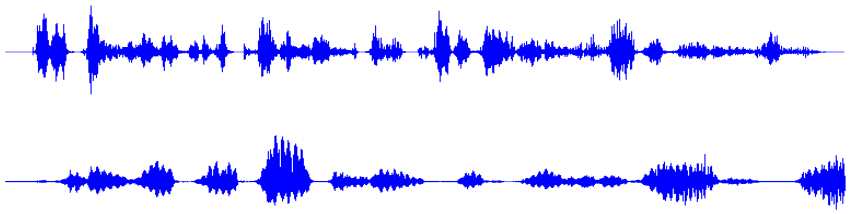

FIG. 5.39 – Signaux de TIMIT estimés par la méthode DUET dans le domaine temporel.

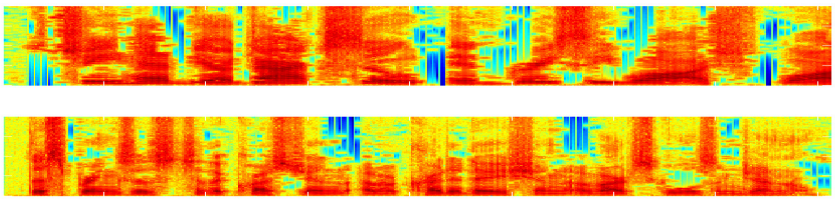

FIG. 5.40 – Spectrogramme des Signaux estimés par la méthode DUET.

Évaluation de la qualité du signal de parole

Le PESQ

Le tableau 5.1 donne les résultats des valeurs du PESQ pour chaque méthode utilisée et pour chaque type de mélange. Rappelons que nous attribuons une note entre 0 et 4,5.

Modèle de mélanges	**HRTF (angles : 30 et -80 degrés)**		**Anéchoïque**	
PESQ / **Méthodes**	Signal de parole de la locutrice	Signal de parole du locuteur	Signal de parole de locutrice	Signal de parole du locuteur
C-FICA	2,6	2,2	2,7	2,3
DUET	3,0	2,9	3,2	2,7
FOPCA	4,3	4,2	3,3	3,7

TAB. 5.1 – Comparaison du PESQ pour les méthodes C-FICA, DUET et FOPCA

Modèle de mélanges	HRTF (angles : 30 et -80 degrés)		Anéchoïque	
SIR_{st} / Méthodes	Signal de parole du locuteur	Signal de parole de la locutrice	Signal de parole de locuteur	Signal de parole de la locutrice
C-FICA	10,56 dB	8,23 dB	10,80 dB	7,10 dB
DUET	9,94 dB	10 dB	10,67 dB	3,10 dB
FOPCA	40,35dB	29,59dB	24,23dB	33,28dB

TAB. 5.2 – Comparaison du SIR des méthodes C-FICA, DUET et FOPCA (calcul du SIR fourni par J. Thomas et al. [139]).

Le Rapport Signal-à-Interférence : SIR

La mesure de la performance est aussi réalisée grâce au rapport signal-à-interférence. De la même façon que pour le cas du mélange instantané nous avons utilisé les deux manières de calcul du SIR : l'algorithme "BSS_EVAL" proposé par C. Fevotte et al. [139] et l'algorithme fourni dans la "*toolbox C-FICA v1-1*" par J. Thomas et al. [131].

Le tableau 5.2 nous donne la comparaison du rapport signal-à-interférence : SIR_{st} (calculé au moyen de l'algorithme "BSS_EVAL") après la réalisation de la séparation des deux sources et cela, pour chaque méthode et chaque type de mélanges. Le tableau 5.3 nous donne les gains entre les rapports SIR après séparation et avant séparation. Les tableaux 5.4 et 5.5 nous donnent respectivement les résultas des SIR calculés au moyen de l'algorithme fourni dans la "*toolbox C-FICA v1-1*", et les gains entre les SIR avant et après séparation.

La méthode DUET devrait donner des résultats probants pour au moins un signal, mais les paramètres inhérents et aléatoires que nous déterminons à partir de l'histogramme provoquent une mauvaise évaluation tel que nous pouvons le constater dans les tableaux 5.2 et 5.4. Nous avons vérifié également l'efficacité de la méthode FOPCA lorsque les réponses impulsionnelles des filtres HRTF sont associées avec des positions des sources définies par différents angles que nous avons sélectionnés aléatoirement. Les résultats sont résumés dans le tableau 5.6. Le tableau 5.7 résume les résultats obtenus avec les signaux "$speech_1$" et $speech_2$".

Mélanges		HRTF	Anéchoïque	
Méthodes	Gain Signal de parole du locuteur	Gain Signal de parole de la locutrice	Gain Signal de parole du locuteur	Gain Signal de parole de la locutrice
C-FICA	15,22 dB	9,98 dB	15,46 dB	8,85 dB
DUET	14,60 dB	11,75 dB	15,33 dB	4,85 dB
FOPCA	45,01 dB	31,34 dB	26,22 dB	26,77 dB

TAB. 5.3 – Comparaison du Gain entre le SIR de sortie et d'entrée pour les méthodes C-FICA, DUET et FOPCA (TIMIT).

Modèle de mélanges	**HRTF (angles : 30 et -80 degrés)**		**Anéchoïque**	
SIR_{st} / **Méthodes**	Signal de parole du locuteur	Signal de parole de la locutrice	Signal de parole de locuteur	Signal de parole de la locutrice
C-FICA	3,7 dB	3,06 dB	1,889 dB	-2,76 dB
DUET	-0,67 dB	-0,09 dB	7,88 dB	0,14 dB
FOPCA	18,34 dB	27,56dB	27,5 dB	1,37 dB

TAB. 5.4 – Comparaison du SIR des méthodes C-FICA, DUET et FOPCA (calcul du SIR fourni par la "toolbox C-FICA v1-1").

Mélanges		HRTF	Anéchoïque	
Méthodes	Gain Signal de parole du locuteur	Gain Signal de parole de la locutrice	Gain Signal de parole du locuteur	Gain Signal de parole de la locutrice
C-FICA	8,36 dB	4,81 dB	6,55 dB	-1 dB
DUET	5,33 dB	1,66 dB	12,54 dB	1,89 dB
FOPCA	23 dB	29,25 dB	32,16 dB	3,12 dB

TAB. 5.5 – Comparaison du gain entre le SIR de sortie et d'entrée pour les méthodes C-FICA, DUET et FOPCA

	SIR_{st} (dB)		
Angles (degrés)	**FOPCA**	**C-FICA**	**DUET**
10	29,60	9,07	-1,63
-10	40,35	8,16	2,07
20	29,60	9,07	-1,63
-120	40,35	8,16	2,07
-20	29,60	9,07	-1,63
50	40,40	8,16	2,07
-30	29,60	9,07	-1,63
80	40,40	8,16	2,07
-10	29,60	9,07	-1,63
60	40,40	8,16	2,07

TAB. 5.6 – Comparaison des SIR des méthodes C-FICA, DUET et FOPCA avec différents angles pour les réponses impulsionnelles des filtres du modèle HRTF.

	SIR_{st} (dB)		Gain (dB)	
Méthodes	***Speech*$_1$**	***Speech*$_2$**	***Speech*$_1$**	***Speech*$_2$**
C-FICA	17,43	20	21,91	21,95
DUET	7	8	11,48	9,95
FOPCA	36,58	34,22	41,06	36,06

TAB. 5.7 – Comparaison du SIR pour les méthodes C-FICA, DUET et FOPCA pour les signaux " speech1 " et " speech2 ".

Signaux de parole	s_1 et s_2	s_3 et s_4	s_5 et s_6	s_2 et s_3
SIR_{st} (dB)	48.70 48.20	40.82 41.16	32.54 31.56	38.88 39.87
$Gain_1$(dB) $Gain_2$(dB)	50 53.41	42.04 47.32	34,56 35.91	43.40 41.78

TAB. 5.8 – Résultats du SIRst pour les signaux de parole recueillis par D. Nion et al. pour la méthode FOPCA.

Concernant les signaux de parole recueillis par D. Nion et al. [136], les résultats sont résumés dans le tableau 5.8. Les signaux sont de l'ordre de 30 s, donc de durée beaucoup plus grande que celle des signaux TIMIT. Nous remarquons de très bonnes valeurs pour le SIR_{st} .

Nous avons de même tracé l'évolution du rapport signal-à-interférence avant et après la séparation : SIR_{en} et SIR_{st}. Les figures 5.41 et 5.42, 5.43 et 5.44 et enfin 5.45 et 5.46 montrent respectivement les SIR après l'application de l'approche FOPCA, DUET et celle de C-FICA. Les courbes en pointillés représentent les SIR_{en} avant la séparation des signaux de parole.

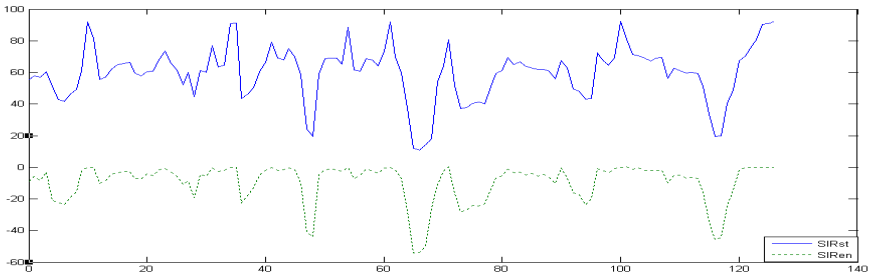

FIG. 5.41 – Comparaison des rapports SIR avant et après séparation par la méthode FOPCA (Signal de parole prononcé par l'homme : TIMIT).

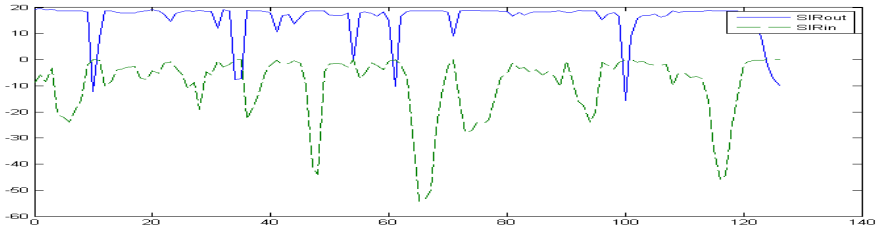

FIG. 5.42 – Comparaison des rapports SIR avant et après séparation par la méthode FOPCA (Signal de parole prononcé par la femme : TIMIT).

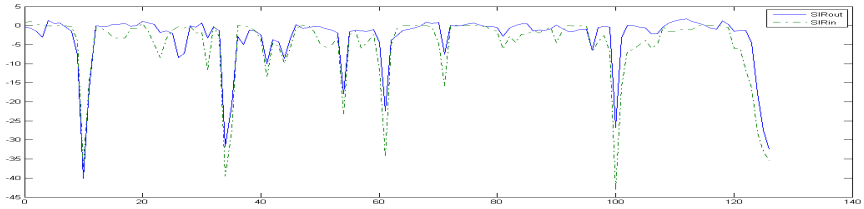

FIG. 5.43 – Comparaison des rapports SIR avant et après séparation par la méthode DUET (Signal de parole prononcé par l'homme : TIMIT).

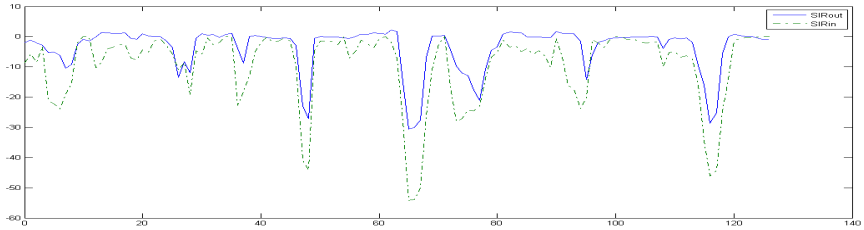

FIG. 5.44 – Comparaison des rapports SIR avant et après séparation par la méthode DUET (Signal de parole prononcé par la femme : TIMIT).

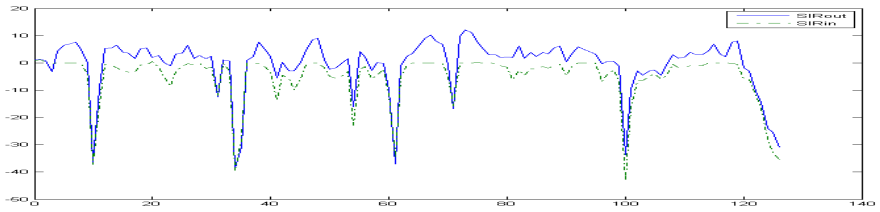

FIG. 5.45 – Comparaison des rapports SIR avant et après séparation par la méthode C-FICA (Signal de parole prononcé par l'homme : TIMIT).

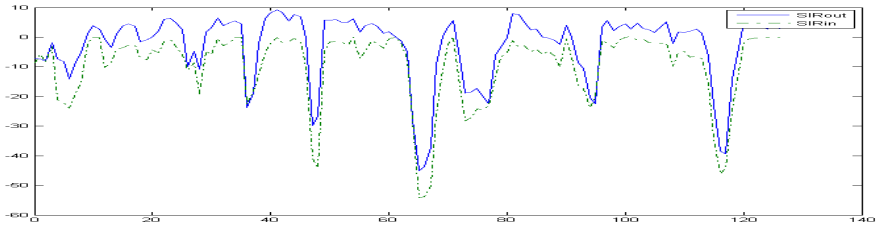

FIG. 5.46 – Comparaison des rapports SIR avant et après séparation par la méthode C-FICA (Signal de parole prononcé par la femme : TIMIT).

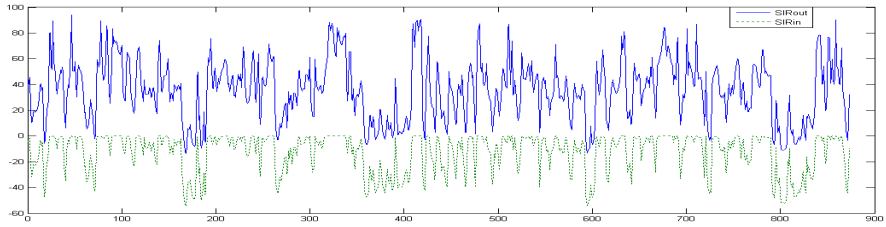

FIG. 5.47 – Comparaison des rapports SIR avant et après séparation par la méthode FOPCA (Signal de parole prononcé par l'homme s_1 et enregistré par D. Nion et al.).

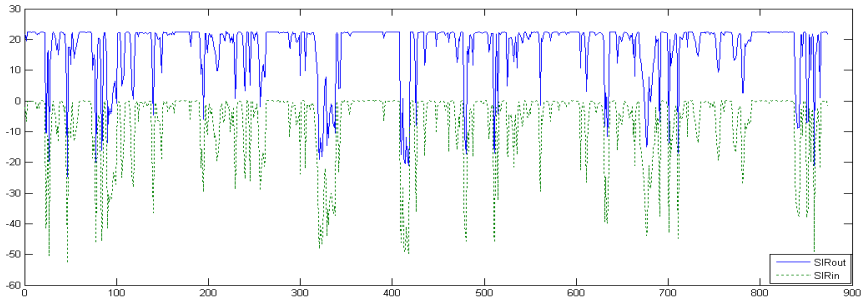

FIG. 5.48 – Comparaison des rapports SIR avant et après séparation par la méthode FOPCA (Signal de parole prononcé par la femme s_2 et enregistré par D. Nion et al.).

Les figures 5.47 et 5.48 représentent les SIR_{en} et SIR_{st} des signaux enregistrés par D. Nion

Nous avons réalisé les expériences avec des signaux de parole à partir des bases de données ayant des puissances comparables. Nous avons effectué d'autres tests à partir d'un ensemble de données et nous avons utilisé différents niveaux des signaux de parole afin de vérifier la robustesse de l'algorithme. Le niveau du signal de parole est calculé en dB par rapport à l'autre signal de parole à l'aide du rapport signal-à-signal avant traitement : SSR. Nous avons calculé le niveau du signal utile avec des valeurs de gain choisies : 0 dB, 5 dB, 10 dB, 15 dB et 20 dB. SIR_{st} représente le rapport signal-à-interférence après la séparation utilisant la méthode FOPCA dans un mélange utilisant le modèle HRTF. Nous arrivons à réaliser la séparation de sources avec un rapport de 29,6 dB et 40,35 dB pour n'importe quelle valeur du SSR.

Le coefficient de corrélation

Le coefficient de corrélation donne l'information sur le lien entre deux variables. La valeur de ce coefficient peut nous informer sur l'existence du lien entre une variable indépendante et une variable dépendante.

Le coefficient de corrélation appelé ρ est la normalisation du coefficient de covariance σ_{xy}. On normalise en divisant la covariance par l'écart type de chacune des variables. Ce coefficient est donc égal au rapport de leur covariance et du produit non nul de leurs écarts types, ce qui nous permet d'avoir une comparaison avec d'autres mesures de corrélation.

$$\rho = \frac{\sigma_{xy}}{\sigma_x \sigma_y} \tag{5.30}$$

	Modèle HRTF					
	FOPCA		C-FICA		DUET	
Corrélation	original Homme	original Femme	original Homme	original Femme	original Homme	original Femme
Estimé Homme	0,99	0,0007	0,95	0,011	0,28	0,02
estimé Femme	0,0009	0,99	0,034	0,92	0,0003	0,19
Modèle Anéchoïque						
Estimé Homme	0,99	0,072	0,95	0,12	0,53	0,024
estimé Femme	0,0011	0,99	0,003	0,89	0,012	0,95

TAB. 5.9 – Coefficient de corrélation entre les signaux originaux et estimés pour les deux types de mélange : HRTF et Anéchoïque.

La valeur du coefficient de corrélation est comprise entre -1 et 1. Les valeurs intermédiaires informent sur le degré de dépendance entre les deux variables. Plus le coefficient est proche des valeurs extrêmes -1 et 1, plus la corrélation entre les variables est forte ; on utilise l'expression « fortement corrélées » pour désigner les deux variables. La valeur de ρ égale à 0 signifie que les variables sont linéairement indépendantes. D'après le tableau 5.9 la méthode FOPCA reste nettement supérieure aux deux autres méthodes, où le coefficient de corrélation a une valeur de 0,9 qui est proche de 1 entre le signal original de l'homme (respectivement de la femme) et estimé de l'homme (respectivement de la femme), et il est très faible de l'ordre de 0,0027 et 0,0035 entre le signal original de l'homme (respectivement de la femme) et estimé de la femme (respectivement de l'homme).

Le MOS

Enfin, nous avons utilisé une méthode subjective basée sur la note d'opinion moyenne. Dans nos expériences, les tests MOS ont été menés avec un groupe de dix auditeurs. Les résultats pour les trois méthodes sont résumés dans le tableau 5.10 . Les tests ont été réalisés pour les deux types de mélange et nous pouvons noter que les MOS de FOPCA sont très supérieurs par rapport à ceux de DUET et C-FICA.

Modèle de mélange	HRTF	Anéchoïque
Méthodes	MOS	
C-FICA	2,6	2,6
DUET	2,3	3,8
FOPCA	4,3	4,4

TAB. 5.10 – Comparaison des MOS pour les méthodes C-FICA, DUET et FOPCA.

5.6.2 Cas de la séparation aveugle de sources sous-déterminée

La séparation aveugle de sources sous-déterminée cherche à estimer N signaux sources inconnus $s_j(n)$ à partir de M signaux observés $x_i(n)$ lorsque le nombre d'observations est inférieur au nombre de sources ($M < N$). Les signaux observés sont dans la plupart des cas des mélanges linéaires instantanés ou convolutifs des signaux sources. La séparation de sources sous-déterminée a été très peu étudiée dans le cas de mélanges convolutifs. On peut citer toutefois [141], [142], [143], [144], [145]. Le modèle linéaire à atténuations et retards a été lui plus étudié, principalement par les approches temps-fréquence telle que la méthode DUET [140].

Mélange sous-déterminé : trois signaux de parole et deux microphones : modèle HRTF

Dans la seconde série d'expériences, nous testons l'algorithme FOPCA dans le cas sous-déterminé $N = 3$ et $M = 2$. Nous étendons l'algorithme FOPCA restreint aux mélanges déterminés ($M = N$), aux mélanges sous-déterminés ($M < N$). Dans cette expérience, nous illustrons les performances de l'algorithme FOPCA en comparaison avec les méthodes DUET et C-FICA.

Dans le cas de l'approche DUET, le calcul des rapports d'observations est couplé à une analyse d'histogramme à deux dimensions (amplitude, retard) pour déterminer le nombre de sources ainsi que les coefficients de mélange. L'algorithme DUET suppose que les sources sont totalement disjointes dans le plan temps-fréquence : dans chaque case du plan temps-fréquence, une seule source doit être présente (ou présente de façon prépondérante). Les méthodes linéaires ne permettent pas une séparation exacte des sources. Dans le cas général, une combinaison linéaire des observations contiendra en effet au minimum $N-M+1$ sources. Cependant la séparation exacte peut être réalisée au moyen d'un masquage binaire du plan temps-fréquence sous réserve que les sources

soient totalement disjointes dans le plan temps-fréquence. L'algorithme DUET utilise cette technique qui consiste pour l'estimation d'une source donnée à masquer les cases temps-fréquences associées aux autres sources [146].

L'algorithme temporel rapide à point fixe C-FICA restreint aux mélanges déterminés a été aussi appliqué aux mélanges sous-déterminés. Cette approche est basée sur le concept de séparation différentielle de sources et sur un processus de blanchiment différentiel convolutif non-causal qui permet l'utilisation de mises à jour de type point fixe pour l'optimisation du kurtosis différentiel. L'estimation des contributions des sources est réalisée au moyen d'un critère quadratique différentiel.

Nous avons encore une fois choisi les signaux sources de parole à partir de la base de données TIMIT disponible et ils ont été mélangés convolutivement en utilisant des réponses impulsionnelles des filtres HRTF qui sont associées avec des positions des sources définies par différents angles que nous avons sélectionnés aléatoirement par rapport à la tête d'un mannequin. Dans ce cas, nous avons trois haut-parleurs et deux microphones. Le tableau 5.11 montre la comparaison des SIR_{st} des trois méthodes.

		SIR_{st} (dB)	
Angles (degrés)	**FOPCA**	**C-FICA**	**DUET**
30	**29**	2.32	-0.38
-80	2,54	14.75	-0.92
10	-2,6	-2.04	-0.85
20	**29**	2.32	-0.38
-50	2,54	14.75	-0.92
40	-2,6	-2.04	-0.85
20	**29**	5.36	0.041
-50	2,54	0.34	-1.04
-10	-2,6	-9.35	-0.21
10	-0.28	5.36	-1.57
-50	**39,26**	0.34	-2.50
-10	-23,14	-9.35	0.83
80	-3	1.20	-0.07
70	3	-6.81	-0.30
-30	**19,6**	10.06	-0.10
60	-2.68	1.20	-0.07
90	3	-6.81	-0.30
-70	**19,5**	10.06	-0.10

TAB. 5.11 – Comparaison des SIRst pour les approches C-FICA, DUET et FOPCA avec différents angles pour les réponses impulsionnelles des filtres du modèle HRTF : cas sous-déterminé.

Les performances de la méthode FOPCA sont mesurées par son rapport signal-à-interférence de sortie, que nous comparons au SIR d'entrée, associé aux observations. Les valeurs des SIR_{en} sont : -1,77 dB, -4,66 dB et -3,45 dB. D'après le tableau 5.11 nous remarquons que quelles que soient les positions des angles, le SIR_{st} de la méthode FOPCA est supérieur par rapport aux deux autres méthodes.

Séparation de trois signaux de parole mélangés suivant le modèle anéchoïque

Dans cette partie, nous avons testé expérimentalement les algorithmes C-FICA DUET et FOPCA en appliquant des mélanges convolutifs sous-déterminés. Après avoir

Méthodes	FOPCA	C-FICA	DUET
SIR_{st} (dB) Des trois signaux de parole	-3,4	3,24	13
	14	3,89	-0,73
	4,27	2,64	1,39

TAB. 5.12 – Comparaison des SIR pour les approches C-FICA, DUET et FOPCA utilisant le modèle de mélange anéchoïque : cas sous-déterminé (trois signaux de parole et deux microphones).

PESQ des trois signaux		
FOPCA	CFICA	DUET
2	1,8	2,8
2	1,9	1,5
2,4	2	1,5

TAB. 5.13 – Comparaison du PESQ pour les méthodes C-FICA, DUET et FOPCA utilisant le modèle anéchoïque : cas sous-déterminé (trois signaux de parole et deux microphones).

illustré le fonctionnement de ces méthodes sur des mélanges de sources de parole, nous avons présenté les résultats dans le tableau 5.12. La méthode FOPCA a encore une fois montré qu'elle était assez efficace par rapport à l'approche C-FICA, mais assez comparable à la méthode DUET. Pour FOPCA, nous réalisons une séparation de l'ordre de 14 dB pour le SIR et de 2,4 pour le PESQ et pour la méthode DUET, 13 dB pour le SIR et 2,8 pour la méthode DUET.

5.7 Séparation aveugle de sources en milieu de réverbération : enregistrement réel

Nous avons réalisé quelques mesures dans la salle d'enregistrement disponible au laboratoire INRS-Télécommunications. C'est une salle dans laquelle nous pouvons réaliser des enregistrements en milieu réverbérant avec un temps $T60 \approx 300\ ms$. La première configuration est représentée dans la figure 5.49 ci-dessous où nous disposons de quatre hauts parleurs et quatre microphones unidirectionnels branchés à deux amplificateurs. Ces derniers sont connectés à un ordinateur par une carte son entrée-multiple sortie-multiple (en anglais : *multiple-input multiple-output*). Nous n'avons réalisé aucun étalonnage sur les différents éléments tels que les microphones et nous nous ne sommes pas intéressés à leur emplacement ni à leur position. Le choix de l'espacement entre les microphones était aléatoire puisque nous n'avons pas besoin de l'information spatiale.

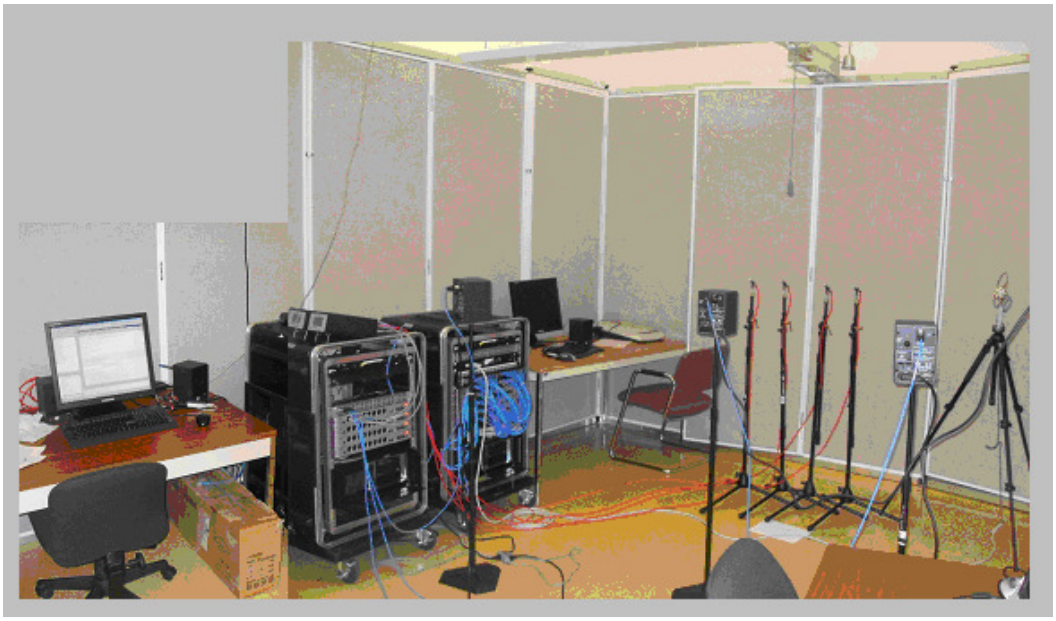

FIG. 5.49 – Salle d'enregistrement en milieu réverbérant : quatre hauts parleurs et quatre microphones.

La deuxième configuration est représentée dans la figure 5.50 ci-dessous où nous disposons, cette fois-ci d'un mannequin avec deux microphones aux deux oreilles. Les positions de mannequin étaient aléatoires.

Malheureusement après une série de mesure utilisant des mélanges de signaux de parole enregistrés en milieu réverbérant, les résultats n'ont pas été très probants. Nous

FIG. 5.50 – Salle d'enregistrement en milieu réverbérant : deux microphones dans les oreilles du Mannequin.

avons obtenu une faible amélioration du rapport signal-à-interférence après traitement : SIR_{st} par rapport à celui avant le traitement : SIR_{en} de l'ordre de 5 dB environ tel que nous pouvons le voir dans le tableau 5.14.

5.8 Discussion

Pour évaluer notre approche dans le cas convolutif, nous l'avons comparé avec les méthodes assez connues et assez récentes telles que les techniques C-FICA et DUET. La comparaison a été faite avec deux types de mélanges convolutifs différents. Nous

	SIR_{en}	SIR_{st}	Amélioration du SIR
Locuteur	-4,8126 dB	-0,0009 dB	4,8117 dB
Locutrice	-1,7615 dB	-0,0011 dB	1,76 dB

TAB. 5.14 – Comparaison des SIR avant et après traitement avec la méthode FOPCA.

avons alors montré l'efficacité de la technique de séparation basée sur la méthode FOPCA. Cette approche est efficace comme nous pouvons le voir dans le domaine temps, dans la figure 5.11 et dans le domaine temps-fréquence dans la figure 5.14 pour le mélange convolutif suivant le modèle HRTF pour les signaux enregistrés à partir de TIMIT ; dans les figures 5.15 à 5.20 pour les signaux "speech$_1$" et "speech$_2$", et enfin dans les figures 5.21 à 5.26 pour les signaux de parole enregistrés par D. Nion et al. Nous pouvons faire les mêmes constatations concernant la robustesse de la méthode pour le mélange anéchoïque. Elle donne des signaux très proches des signaux originaux par rapport aux deux autres techniques et donc une très bonne séparation. C-FICA ne donne pas de très bons résultats en particulier pour les mélanges anéchoïques.

Pour l'évaluation du PESQ, la méthode FOPCA a donné aussi les meilleurs résultats, en comparaison avec C-FICA et DUET et ceci est vrai pour les deux types de mélange comme nous pouvons le voir dans le tableau 5.1. Remarquons que pour le mélange anéchoïque, DUET donne un bon PESQ pour un seul signal. La méthode DUET est connue pour sa capacité de réaliser la séparation d'un seul signal de parole, alors qu'avec la méthode FOPCA, nous arrivons à réaliser la séparation des deux signaux de parole.

Nous savons que l'algorithme PESQ est conçu pour prédire subjectivement les scores d'opinion d'un échantillon de parole dégradé mais de façon objective et qu'il nous donne un score compris entre 0 à 4,5. Le PESQ de la méthode FOPCA est entre 3,3 et 4,3 ce qui est excellent, en particulier pour les signaux de durée assez longue (environ 30 s) où le PESQ dépasse 4, pour des valeurs comprises entre 3,8 à 4,3.

Comme nous avons utilisé les mêmes signaux de parole que pour les mélanges instantanés, les valeurs des SIR avant séparation sont de $-4{,}66\ dB$ pour le signal de parole de l'homme et de $-1{,}75\ dB$ pour le signal de parole de la femme, tel que spécifié dans le tableau 4.3 et 4.4 du chapitre quatre. Dans le cas convolutif nous avons aussi utilisé l'algorithme fourni par la "toolbox C-FICA v1-1" et l'algorithme "BSS_EVAL" pour la mise en œuvre du SIR. Les tableaux 5.2, 5.4, 5.6, 5.7 et 5.8 montrent le SIR issu de la méthode FOPCA supérieur à ceux des deux autres méthodes de référence. Si nous prenons l'exemple de calcul du SIR suivant l'alorithme "BSS_EVAL", le tableau 5.3 montre l'amélioration du rapport SIR$_{st}$ par rapport au SIR$_{en}$ de 15,22 dB et 9,98 dB respectivement pour le signal de parole de l'homme et du signal de parole de la femme pour la méthode C-FICA. Elle est de 14,60 dB et 11,75 dB pour la méthode DUET de 45,01 dB et 31,34 dB pour l'approche FOPCA. Pour les résultats de C-FICA, d'après l'article de J. Thomas et al.[131] qui ont mis au point cette méthode, ont trouvé un SIR de 9,1 dB pour la première source et de 6,9 dB pour la deuxième source et cela dans les mêmes conditions que nous avons utilisées. Pour le mélange anéchoïque, l'amélioration du SIR de C-FICA est de 15,46 dB pour le signal de parole de l'homme et de 8,85 dB

pour le signal de parole de la femme. L'amélioration est de 15,35 dB et 4,85 dB pour la méthode DUET. Nous remarquons dans ce cas-ci que nous arrivons à séparer adéquatement un signal de parole, ce qui confirme le rôle de cette technique qui consiste à extraire un signal source à partir de mélanges anéchoïques de plusieurs sources. Pour la méthode FOPCA, l'amélioration est excellente pour le signal de parole de l'homme, elle est de 26,22 dB et pour le signal de parole de la femme de 26,77 dB. Les figures 5.41 à 5.48 montrent l'évolution des rapports SIR, et à partir desquelles nous pouvons remarquer la nette amélioration du SIR_{st} par rapport au SIR_{en}. Pour confirmer les résultats des expériences, nous avons utilisé plusieurs positions des angles et nous avons trouvé un SIR de 40,40 dB et de 29,60 dB respectivement pour la séparation du signal de parole prononcée par l'homme et la femme. Les performances diminuent un peu lorsque les angles sont très proches et du même côté de la même oreille. Dans ce cas-ci la séparation d'un seul signal source est possible, ce qui semble logique. D'autres tests ont été réalisés pour différents niveaux des signaux de parole avant le traitement.

Le tableau 5.9 montre pour le mélange suivant le modèle HRTF, une forte corrélation entre les signaux originaux et estimés avec un coefficient de corrélation de $0,99$ pour FOPCA et pour C-FICA des coefficients de 0,95 et 0,92. Par contre, pour la méthode DUET, elle ne performe pas très bien pour ce type de mélange, les signaux sont décorrélés. Pour le cas du mélange anéchoïque, nous trouvons aussi 0,99 pour FOPCA, 0,95 et 0,89 pour C-FICA et enfin 0,95 et 0,53 pour DUET. Ce dernier cas confirme que la méthode DUET extrait un seul signal dans le cas d'un mélange anéchoïque. Enfin la valeur du MOS de 4,4 vient confirmer, la robustesse de l'approche FOPCA. Habituellement, les algorithmes classiques dans le domaine fréquentiel sont sujets à des problèmes de permutation. Par contre, l'algorithme FOPCA proposé est similaire à celui décrit dans : M. Kawamoto et Y. Inouye [147], et il a la particularité intéressante que cet inconvénient ne se produit pas, d'où un autre avantage de cette méthode. Nous nous sommes aussi intéressés au cas sous-déterminé. Les résultats sont résumés dans le tableau 5.11, pour différentes positions d'angle pour le modèle HRTF. Nous arrivons à réaliser dans ce cas-ci, la séparation d'un seul signal avec une valeur du rapport SIR_{st} de 39,26 dB, ce qui est à l'état actuel une très bonne valeur. Le tableau 5.12 montre le cas convolutif anéchoïque et dans ce cas-ci FOPCA et DUET sont comparables avec respectivement 14 dB et 13 dB.

Par contre, pour les signaux mélangés dans un milieu réverbérant, les résultats n'ont pas été très probants tel que nous pouvons le constater dans le tableau 5.14, ou l'amélioration du SIR n'est que d'environ 5 dB.

5.9 Conclusion

Dans ce chapitre, nous avons présenté une technique de séparation aveugle de signaux de parole dans le cas de mélanges convolutifs. La méthode est axée sur l'analyse des composantes principales orientées en fréquence (FOPCA). Toutes les approches précédentes ont toujours utilisé deux étapes : une étape de prétraitement suivie par une méthode d'analyse utilisant les moments du second ordre tels que la PCA. L'approche FOPCA a l'avantage qu'aucune étape de prétraitement n'est nécessaire, car le prétraitement est implicitement intégré dans le critère du rapport signal-à-signal, qui est optimisé par l'approche OPCA [97]. Cette technique s'est avéré très efficace comme nous pouvons le voir dans le domaine temporel ainsi que dans le domaine fréquentiel. Une évaluation subjective a été réalisée par l'écoute des signaux originaux avant le mélange, les signaux mélangés ainsi que les signaux estimés. Les résultats sont très satisfaisants ; nous avons obtenu une très bonne séparation, comme peuvent le confirmer les valeurs des trois autres paramètres d'évaluation le PESQ, le SIR et le coefficient de corrélation. Nous avons aussi testé la technique avec plusieurs autres signaux de parole de la base de données TIMIT, NOIZEUS, signaux utilisés par J. Thomas et al. et autres signaux collectés par D. Nion et al. et les résultats obtenus restent aussi performants. Ces derniers confirment les performances de la méthode FOPCA dans les cas des mélanges convolutifs [148] [149] [150], une méthode que nous avions déjà utilisée dans le domaine temporel (OPCA) pour la première fois, dans la séparation des signaux de parole dans le cas de mélange instantané [151] et que nous avons détaillé dans le quatrième chapitre.

Chapitre 6

Conclusion générale

Dans cette thèse, nous nous sommes intéressés au problème de la séparation aveugle de mélanges instantanés et convolutifs de sources ainsi qu'à la réduction de bruit dans un signal de parole. De nombreuses approches ont été développées et étudiées : elles dépendent de la nature des signaux sources considérés.

Nous avons présenté dans ce travail, le principe de l'analyse en composantes principales (PCA) et l'analyse en composantes indépendantes (ICA) qui peut être vue comme un prolongement de l'analyse en composantes principales, mais dans laquelle la décorrélation des observations est remplacée par une hypothèse d'indépendance des sources (hypothèse donc plus forte). L'analyse en composantes indépendantes est une méthode d'analyse de données. Elle est notoirement et historiquement connue en tant que méthode de séparation aveugle de sources.

L'analyse en composantes indépendantes se propose d'utiliser des moments d'ordre supérieur à deux dans le but d'obtenir une indépendance statistique des sources et notamment de lever l'indétermination qui existe si l'on se contente de la PCA. On peut voir l'ICA comme une méthode liée à la théorie de l'information et notamment au concept de l'information mutuelle.

Le principe de séparation aveugle de source est extrêmement intéressant puisqu'aucune forte hypothèse sur les signaux n'est nécessaire à part l'existence de sources indépendantes. L'algorithme de la séparation de sources est capable de restituer simultanément tous les signaux primitifs inconnus dans un mélange inconnu, sans connaissance "a priori" ni sur les signaux ni sur le mélange. En introduisant les fonctions non-linéaires qui génèrent des moments croisés d'ordre supérieur à deux, dans la règle d'adaptation, l'algorithme effectue approximativement un test d'indépendance des sorties deux à deux. Par contre des algorithmes basés sur la décorrélation des signaux et donc utilisant les moments d'ordre deux s'avèrent aussi efficace pour la séparation des signaux de parole.

Dans notre travail, nous avons appliqué l'analyse en composantes indépendantes pour la séparation aveugle de signaux de parole dans le cas d'un mélange instantané, La méthode est appelée HOS. Nous avons montré que l'on pouvait appliquer cette technique de séparation en changeant le critère d'adaptation classique, en choisissant des fonctions non-linéaires adéquates. Nous avons obtenu un bon résultat comme nous le constatons dans le domaine temporel où nous voyons qu'après avoir mélangé instantanément deux signaux de parole l'un d'une femme et l'autre d'un homme, nous sommes arrivés à réaliser la séparation. Le résultat se voit bien aussi dans le domaine temps-fréquence en appliquant le spectrogramme. Nous notons aussi que les valeurs estimées des coefficients de mélange sont plus proches des valeurs théoriques après convergence : $0{,}48 \approx 0,5$et $0{,}62 \approx 0{,}6$. Cependant cet algorithme n'est pas très rapide.

La deuxième méthode que nous avons appliquée c'est la méthode OPCA. Cet algorithme a été proposé initialement pour séparer quatre signaux aléatoires avec une modulation multi-niveaux en amplitude : PAM. Cette approche n'a jamais été appliquée antérieurement pour la séparation des signaux de parole. En comparaison avec les méthodes HOS et Fast-ICA, l'approche OPCA s'avère efficace, comme on peut le voir dans le domaine temporel ainsi que dans le domaine temps-fréquence, où l'on remarque que les signaux et les signaux estimés sont très proches.

L'algorithme très connue, Fast-ICA, a servi de référence pour une évaluation objective. Fast-ICA permet une bonne séparation avec un calcul rapide, mais les signaux sont corrompus par le bruit lors de l'écoute des signaux de parole dans le contexte de l'évaluation subjective. La méthode OPCA est aussi rapide mais aucun bruit n'est perceptible.

Pour une évaluation objective du PESQ, HOS a donné des résultats meilleurs que Fast-ICA. Pour l'estimation du signal de parole de l'homme, nous obtenons le même résultat pour les deux algorithmes HOS et OPCA, mais dans le cas du signal de parole de la femme, le PESQ de HOS est meilleur. Pour Fast-ICA, l'algorithme de référence, le PESQ est assez mauvais. Le MOS, une mesure subjective, a donné un résultat très proche pour les trois méthodes.

Le rapport signal-à-interférence SIR de la méthode OPCA, ainsi que celui de Fast-ICA sont assez proches particulièrement pour le signal de parole de l'homme, par contre celui de la méthode HOS est moyen.

L'avantage de ces méthodes est qu'elles n'introduisent pas de distorsion de phase et conservent au maximum la forme du signal original.

Malgré les nombreux progrès réalisés depuis les premiers travaux des années 80, la séparation de sources en mélanges convolutifs reste un problème complexe et le peu de mises en situations pratiques disponibles dans la littérature montre les difficultés rencontrées lorsque l'on sort du cadre théorique.

La même technique de séparation aveugle de signaux de parole, FOPCA, a été appliquée en fréquence dans le cas des mélanges convolutifs. Toutes les approches précédentes citées dans la littérature ont toujours utilisé deux étapes : une étape de prétraitement suivie par une méthode d'analyse utilisant les moments du second ordre tels que l'analyse en composantes principales. Cette approche a l'avantage qu'aucune étape de prétraitement n'est nécessaire de plus elle a l'avantage que le problème de permutation ne se pose pas. Nous avons utilisé deux types de mélanges : mélange suivant le modèle HRTF et mélange anéchoïque. Une comparaison de cette méthode a été réalisée avec deux algorithmes de référence connus tels que C-FICA et DUET. FOPCA s'est avérée très performante comme nous pouvons le voir dans le domaine temporel ainsi que dans le domaine fréquentiel. Une évaluation subjective a été réalisée par l'écoute des signaux originaux avant le mélange, les signaux mélangés ainsi que les signaux estimés. Les résultats du MOS sont très satisfaisants. Une très bonne séparation a été obtenue, comme peuvent le confirmer les valeurs des trois paramètres d'évaluation le PESQ, le coefficient de corrélation et le SIR. Les graphiques du rapport SIR après et avant traitement montrent une très nette amélioration. Nous avons aussi testé la technique avec plusieurs autres signaux de parole des bases de données TIMIT, NOIZEUS, les signaux recueillis par J. Thomas et al. ainsi que des signaux de parole enregistrés par D. Nion et al. et les résultats étaient similaires. Ces résultats confirment l'efficacité de la méthode FOPCA dans les cas des mélanges convolutifs, qui représentent des mélanges assez complexes. Nous avons aussi testé expérimentalement les algorithmes C-FICA, DUET et FOPCA en appliquant des mélanges convolutifs sous-déterminés (trois sources et deux microphones). Après avoir illustré le fonctionnement de ces méthodes sur des mélanges de sources de parole, la méthode FOPCA a encore une fois montré qu'elle était plus efficace que les deux autres approches en particulier pour le cas du mélange suivant le modèle HRTF. Le SIR_{st} de la méthode FOPCA et celui de la méthode DUET sont comparables pour le cas anéchoïque.

La méthode DUET devrait donner des résultats acceptables pour au moins un signal, mais l'identification des paramètres dépend des identifications des pics de l'histogramme qui se fait de façon manuelle ce qui peut entraîner une mauvaise évaluation. Il y a eu des investigations sur plusieurs méthodes automatiques différentes d'identification des pics, y compris la pondération de l'histogramme, malheureusement aucune technique ne semble être appropriée dans tous les contextes. Pour cela, souvent nous avons recours à l'identification manuelle des pics tel que spécifié par S. Richard dans le chapitre 8 de [140]. Une fois les pics identifiés, l'objectif serait alors de déterminer les masques temps-fréquence qui vont séparer chaque source à partir des différents mélanges.

Nous avons réalisé quelques mesures dans la salle d'enregistrement disponible au

laboratoire INRS-Télécommunications, pour faire l'étude de la séparation de sources dans le cas de mélanges convolutifs générés dans des environnements réels, en particulier, aux cas des salles acoustiques réverbérantes. Actuellement les résultats trouvés sont peu probants, car l'amélioration n'est pas très élevée. D'autres investigations sont nécessaires.

Les principales contributions à la recherche apportées dans ce travail de thèse autant pour les systèmes de mélanges instantanés que pour les systèmes de mélanges convolutifs, sont :

a) *concernant les cas des mélanges instantanés* :

• Application de la méthode utilisant les moments d'ordre supérieur (HOS) avec laquelle nous avons obtenu un bon PESQ et un bon MOS comparativement à la méthode OPCA et la méthode de référence Fast-ICA. Nous avons obtenu une très bonne convergence des coefficients de la matrice de séparation vers les valeurs des coefficients de la matrice de mélange.

• Application de l'approche OPCA pour la première fois dans la séparation aveugle des signaux de parole. Cette méthode s'est avérée très efficace. C'est une approche qui utilise les statistiques d'ordre deux. Le rapport signal-à-interférence SIR de la méthode OPCA, est assez bon par rapport aux deux autres méthodes. Fast-ICA est un algorithme connu pour sa rapidité. Il a donné un bon SIR, mais les signaux sont corrompus par le bruit lors de l'écoute des signaux de parole dans le contexte de l'évaluation subjective. OPCA est aussi rapide que Fast-ICA, mais aucun bruit n'est perceptible.

b) *concernant le cas des mélanges convolutifs :*

• Une des contributions les plus importantes que nous avons réalisées, concerne les travaux sur la séparation aveugle de sources mélangées convolutivement. Ce type de mélange est connu pour être complexe. Nous avons réalisé des tests avec différentes méthodes sur deux types de mélange convolutifs différents.

• Application de la méthode OPCA dans le domaine fréquentiel que nous avons appelée FOPCA, a été très performante comparativement aux méthodes DUET et C-FICA, que nous avons utilisées comme méthodes de référence.

• Dans les algorithmes de domaine de fréquence, le défi est de résoudre l'ambiguïté de permutation. De nombreux algorithmes dans le domaine fréquentiel apportent des

solutions ad hoc, qui ne résolvent l'ambiguïté de permutation que partiellement, ce qui exige une combinaison de différentes méthodes. Cependant, l'algorithme proposé dans le présent document a un grand avantage car l'inconvénient de la permutation ne se pose pas tel que nous l'avons indiqué en discussion dans le cinquième chapitre.

• Les résultats de l'évaluation subjective sont très satisfaisants. Le MOS de FOPCA est très supérieur par rapport aux approches citées ci-haut. Le PESQ avec une moyenne de 4, le coefficient de corrélation d'une valeur de 0,99 et le SIR d'une valeur de 29 dB et 40 dB et même plus pour les signaux de longue durée, confirment la performance de l'algorithme FOPCA par rapport aux deux autres méthodes.

• Pour des fins de vérification de la robustesse de l'algorithme, nous avons utilisé différents niveaux des signaux de parole. Le signal a été calculé en dB à l'aide du rapport Signal-à-Signal (SSR). La séparation des signaux a été réalisée efficacement avec un rapport SIR de 9,26 dB et 40,35 dB et ce, pour n'importe quelle valeur du SSR.

• Traitement du cas sous-déterminé : trois signaux de parole et deux microphones. La séparation d'un seul signal de parole a été effective avec un rapport SIR d'environ 29 dB et $39,26$ dB, ce qui est à l'état actuel une très bonne valeur.

• Un prototype fonctionnel avec interface interactive a également été développé à des fins de démonstration. Il inclut toutes les méthodes développées lors de cette thèse. La réalisation pratique de ce prototype a permis de faire la validation du concept.

Travaux futurs et perspectives

Les études à venir et les extensions possibles sont nombreuses dont voici une liste non-exhaustive.

• il serait intéressant de continuer les investigations dans l'application de la méthode FOPCA, dans le cas de mélanges convolutifs générés dans des environnements réels, en particulier, aux cas des salles acoustiques réverbérantes.

• une étude fine pourrait être menée pour améliorer les résultats de l'algorithme dans le cas bruité.

• le canal MIMO (*Multiple-Input Multiple-Output*) et le traitement de réception dans un système MIMO sont équivalents, respectivement, aux mélange et matrice de

séparation dans la séparation de sources. Le modèle du signal de la séparation aveugle de sources est donc sensiblement le même que celui d'un système de transmission MIMO, son application aux systèmes MIMO serait très prometteuse.

Publications

1- Y. Benabderrahmane, S.A. Selouani, D. O'Shaughnessy, "Frequency Oriented PCA for Blind Speech Separation of Convolutive Mixtures in Multiples Environments", Interspeech 2011, 28-31 août, 2011, Florence, Italy.

2- S. A. Selouani, Y. Benabderrahmane, A. Ben Salem, H. Hamam, D. O'Shaughnessy, "Convolutive Blind Separation Speech Mixtures using Auditory based Subband Model", Electrical Engineering and Applied Computing , Chapitre 14, Éditeur : Springer eds, 29 juin 2011.

3- Y. Benabderrahmane, S.A. Selouani, D. O'Shaughnessy, "Blind Speech Separation using Frequency Oriented PCA method for Convolutive Mixtures", soumis à : EURASIP journal on Audio, Speech, and Music Processing.

4- Y. Benabderrahmane, S.A. Selouani, D. O'Shaughnessy, "Oriented PCA method for Blind Speech Separation of Convolutive Mixtures", Interspeech 2010, 26-30 septembre, 2010, Makuhari, Japan.

5- Y. Benabderrahmane, S.A. Selouani, D. O'Shaughnessy," Blind Speech Separation for Convolutive Mixtures Using an Oriented Principal Components Analysis Method", 2010 European Signal processing Conference (EUSIPCO 2010), 23-27 août, 2010, Aalborg, Nord du Denmark.

6- Y. Benabderrahmane, S.A. Selouani, D. O'Shaughnessy, and H. Hamam, "A Comparative Study of Blind Speech Separation using Subspace Methods and Higher Order Statistics", Lecture Notes in Computer Science, Springer eds., pp. 117-124, 2009.

7- Y. Benabderrahmane, S.A. Selouani, and D. O'Shaughnessy, "Blind speech separation using OPCA method", 2009 IEEE Symposium on Industrial Electronics and Applications (ISIEA 2009), 4-6 Octobre, 2009, Kuala Lumpur, Malaysia, 2009.

8- Y. Benabderrahmane, A. Ben Salem, S. A. Selouani, and D. O'Shaughnessy, "Blind Speech Separation using High Order Statistics", 2009 Canadian Conference on Electrical and Computer Engineering, IEEE/CCECE, St. John's, Newfoundland, Canada, 3-6 Mai, 2009, pp. 670-673.

9- Y. Benabderrahmane, S. A. Selouani, and D. O'Shaughnessy, "Blind Speech Separation", SYTACom, Université Concordia, Montréal, Québec, Canada, 2008.

Bibliographie

[1] F. Berthommier and S. Choi, “Evaluation of CASA and BSS models for cocktail party speech segregation”, ICA Conference, 2001.

[2] N. Mitianoudis and M. Davies, “Audio source separation of convolutive mixtures”, IEEE Trans. on Speech and Audio Processing, 2002.

[3] A. Westner and V. M. Bove, “ Applying blind source separation and deconvolution to real-world acoustic environments”, Proceeding 106th of Audio Engineering Society, 1999.

[4] A. G. Westner, “Object-Based Audio Capture : Separating Acoustically-Mixed Sounds, PhD dissertation, Massachusetts Institute of Technology,USA, 1999.

[5] J. Herault et C. Jutten : “Space or time adaptive signal processing by neural network models”, Ds Actes Intl. Conference on Neural Networks for Computing, p. 206-211, Snowbird, Utah, USA, avril 1986.

[6] Y. Deville, “Towards industrial applications of blind source separation and independent component analysis”, Ds Actes Intl. Conference on Independent Component Analysis and Signal Separation (ICA’99), p. 19-24, Aussois, France, janvier 1999.

[7] Y. Deville, "Panorama des applications biomédicales des méthodes de séparation aveugle de sources", Ds Actes Groupe de Recherche et d’Étude du Traitement du Signal et des Images (GRETSI’03), vol. 1, p. 31-34, Paris, France, septembre 2003.

[8] C. Jutten et M. Babaie-Zadeh, "Source separation : principles, current advances and applications”, Ds Actes Work. on Advanced Control and Diagnosis (ACD’06), Nancy, France, novembre 2006.

[9] C. Jutten, P. Comon et al., "Séparation de Sources", vol. 2, chapitre 13 (Applications des méthodes de séparation aveugle de sources par Y. Deville). Collection Traité IC2, Information-Commande-Communication, Hermès-Lavoisier, 2007,

[10] A. Kachenoura, L. Albera et L. Senhadji, "Séparation aveugle de sources en ingénierie biomédicale", ITBM-RBM, V. 28, 20-34, 2007.

[11] A. Mansour, “Blind separation for instantaneous mixture of speech signals : Algorithms and performances”, IEEE Conference Proceedings TENCON 2000, pages 26-32, 2000.

[12] A. Mansour, “Blind separation of sources : Methods, assumptions and applications”, IEICE TRANS., E83 :1498-1512, 2000.

[13] S. I. Amari, “Adaptive blind signal processing-neural network approaches”, Proceeding of the IEEE, V. 86 :2026-2048, 1998.

[14] J. Herault, C. Jutten, and B. Ans. "Détection de grandeurs primitives dans un message composite pour une architecture de calcul neuromimétique en apprentissage non supervisé", In GRETSI, pages 1017-1022, Nice, France, Mai 1985.

[15] J. L. Lacoume and P. Ruiz, "Sources identification : a solution based on cumulants", in Proc. 4th Annual ASSP Workshop on Spectrum Estimation and Modeling, Minneapolis, MN, USA, août 1988.

[16] J. L. Lacoume and P. Ruiz, "Sources identification : a solution based on cumulants", in Proc. 4th Annual ASSP Workshop on Spectrum Estimation and Modeling, Minneapolis, MN, USA, août 1988.

[17] J. L. Lacoume and P. Ruiz, "Sources identification : a solution based on cumulants", in Proc. 4th Annual ASSP Workshop on Spectrum Estimation and Modeling, Minneapolis, MN, USA, août 1988.

[18] J. F. Cardoso, A. Souloumiac, “Jacobi angles for simultaneous diagonalization”, SIAM Journal on Matrix Analysis and Applications (SIMAX), 161-4, 1996.

[19] P. Comon, “Independent component analysis, a new concept ?”, Signal Processing, Elsevier ; 287–314, 1994.

[20] M. Gaeta and J. L. Lacoume, "Estimateurs du maximum de vraisemblance étendus à la séparation de sources non gaussiennes", Traitement du signal, vol. 7, no. 5, pp. 419-434, 1990.

[21] B. Laheld and J. F. Cardoso, "Séparation adaptative de sources en aveugle, implémentation complexe sans contraintes", in Actes 14ème colloque GRETSI sur le Traitement du Signal et des Images, Juan Les Pins, France, Septembre 1993, pp. 329-332.

[22] J. F. Cardoso, A. Souloumiac, “Jacobi angles for simultaneous diagonalization”, SIAM Journal on Matrix Analysis and Applications (SIMAX), 161-4, 1996.

[23] N. Delfosse, P. Loubaton, “Adaptive blind separation of independent sources : a deflation approach”, Signal Processing, Elsevier ; V. 45, 59-83, 1995.

[24] A. Hyvarinen , E. Oja, "A fast fixed-point algorithm for independent component analysis", Neural Comput 1997, 1483-92.

[25] P. Comon, E. Moreau, “Improved contrast dedicated to blind separation in communications”, ICASSP 97, IEEE International Conference on Acoustics Speech and Signal Processing, Munich, 3453-3456, avril 1997.

[26] E. Moreau, O. Macchi “New self-adaptive algorithms for source separation based on contrast functions”, HOS93, IEEE Signal Processing Workshop on Higher Order Statistics, Lake Tahoe, USA, 215-219, 1993.

[27] N. Thirion, E. Moreau, “New criteria for blind signal separation”, IEEE Workshop on Statistical Signal and Array Processing, Pennsylvania, USA, 344-8, 2000.

[28] L. Fety, "Méthodes de traitement d’antenne adaptées aux radiocommunication", PhD, École Nationale Supérieure des Télécommunications, (ENST), 1988.

[29] L. Tong, R. Liu, V. Soon, Y. Huang, “Indeterminacy and identifiability of blind identification”, IEEE Trans Circuits Systems, 499-509, 1991.

[30] L. Tong, V. Soon, Y. Huang, R. Liu, “Amuse : a new blind identification algorithm”, New Orleans, LA., 1784-1786 1990.

[31] A. Belouchrani "Séparation autodidacte de sources : algorithmes, performances et application à des signaux expérimentaux", PhD, École Nationale Supérieure des Télécommunications (ENST) Paris, 1995.

[32] A. Belouchrani, K. Abed-Meraim, J. F. Cardoso, E. Moulines, “A blind source separation technique using second-order statistics” , IEEE Trans Signal Processing, 434-44, 1997.

[33] A Ziehe, K. R. Müller. “TDSEP : an efficient algorithm for blind separation using time structure”, International Conference on Artificial Neural Networks, 675-680, 1998.

[34] A. Belouchrani, A. Cichocji “Robust whitening procedure in blind source separation context”, Electronics Letters, IEE Journal Electronic Engineering, 2000.

[35] A. Belouchrani, M. G. Amin, “Blind source separation based on time-frequency signal representation” IEEE Trans Signal Processing, 2888-2897, 1998.

[36] C. Fevotte, "Approche temps–fréquence pour la séparation aveugle de sources non-stationnaires", PhD, École centrale de Nantes et de l’université de Nantes, France, octobre 2003.

[37] C. Fevotte, C. Doncarli, “Two contributions to blind sources separation using time-frequency distributions”, IEEE Signal Processing Letter, 386–389, 2004.

[38] C. Fevotte, A. Holobar, C. Doncarli, D. Zazula , “Single autotermes separation based on bilinear time-frequency representations”, EUSIPCO, Toulouse, France, 565–568, 2002.

[39] P. Comon, “Independent component analysis, a new concept ?”, Signal Processing, Elsevier ; 287-314, 1994.

[40] J. F. Cardoso, “Infomax and maximum likelihood for blind source separation”, IEEE Signal Processing Letter, 112-114, 1997.

[41] B. Pearlmutter, L. Parra, “A context-sensitive generalization of ICA”, Internationnal Conference on Neural Information Processing, 151-157, 1996.

[42] M. Basseville, “Distance measures for signal processing and pattern recognition”, Signal Processing, Elsevier, 349–69, 1989.

[43] A. Prieto, B. Prieto, C. G. Puntonet, A. Ganas, P. Martin-Smith, “Geometric separation of linear mixtures of sources : Application to speech signals”, ICA 2004, LNCS 3195, pp. 798–806, 2004, Springer-Verlag Berlin Heidelberg

[44] C. G. Puntonet, A. Prieto, C. Jutten, M. Rodriguez-Alvarez, J. Ortega “Separation of sources : a geometry-based procedure for reconstruction of n-valued signals”, Signal Processing, 267-2814 1995.

[45] KI. Diamantaras “Blind separation of multiple binary sources using a single linear mixture “, ICASSP 2000, IEEE International Conference on Acoustic Speech Signal Processing, Istanbul, Turkey, 2889–2892, June, 2000.

[46] A. Belouchrani, K. Abed-Meraim, J. F. Cardoso, and E. Moulines, “A blind source separation technique using second-order statistics", IEEE Transactions on Signal Processing, vol. 45, no. 2, pp. 434-444, février 1997

[47] D. T. Pham and J. F. Cardoso, “Blind separation of instantaneous mixtures of non stationary sources", in Proc. 2nd Symposium on Independent Component Analysis and Blind Source Separation ICA, Helsinki, Finland, juin 2000, pp. 187-192.

[48] D. T. Pham and J. F. Cardoso, “Blind separation of instantaneous mixtures of non stationary sources", IEEE Transactions on Signal Processing, vol. 49, no. 9, 1837-1848, 2001.

[49] K. Kokkinakis and A. K. Nandi, “Multichannel blind deconvolution for source separation in convolutive mixtures of speech,” IEEE Trans. Audio, Speech, Lang. Proc., vol. 14, no. 1, pp. 200–212, janvier 2006.

[50] I. Lee, T. Kim, and T. W. Lee, “Complex fastica : A robust maximum likelihood approach of mica for convolutive bss,” in ICA’06, pp. 625–632, 2006.

[51] T. Kim, H. Attias, S. Y. Lee, and T. W. Lee, “Blind source separation exploiting higher-order frequency dependencies,” IEEE Trans. Audio, Speech, Lang. Proc., vol. 15, no. 1, janvier 2007.

[52] A. Belouchrani, K. Abed Meraim, J. F. Cardoso, E. Moulines “A blind source separation technique based on second order statistics”, IEEE Transactions on Signal Processing, 434-44, février 1997.

[53] S. J.Qin, R. Dunia, "Determining the number of principal components for best reconstruction", In IFAC DYCOPS’98, Greece, juin1998.

[54] M. S. Pedersen, J. Larsen, U. Kjems, and L. C. Parra, “A survey of convolutive blind source separation methods", Springer Handbook of Speech, Springer Press, septembre 2007.

[55] E. Sourouchyari, "Blind separation of source, Part III : Stability analysis", Signal Processing, Vol. 24, pp. 21-29, 1991.

[56] J. C. Fort : "Stabilté de l’algorithme de séparation de sources de JUTTEN-HERAULT", Traitement du signal, Vol.8, N°1, 35-42, 1991.

[57] A. Weinstein, M. Feder, and A.-V. Oppenheim, “Multi-Channel Signal Separation by Decorrelation,” IEEE Transactions on Speech and Audio Processing, vol. 1, no. 4, pp. 405-413, octobre 1993.

[58] L. Parra and C. Spence, “Convolutive blind separation of non-stationary sources", IEEE Transactions on Speech and Audio Processing, vol. 8, no. 3, pp. 320-327, mai 2000.

[59] N. Mitianoudis and M. E. Davies, “Audio source separation of convolutive mixtures", IEEE Transactions on Speech and Audio Processing, vol. 11, no. 5, pp. 489-497, septembre 2003.

[60] N. Murata, S. Ikeda, and A. Ziehe, “An approach to blind source separation based on temporal structure of speech signals”, Riken Brain Science Institute, Technical Report 98-102, avril 1998.

[61] R. Mukai, S. Araki, and S. Makino, “Separation and dereverberation performance of frequency domain blind source separation”, Proc. Independant Component Analysis (ICA’01), 230-235, décembre 2001.

[62] S. C. Douglas, M. Gupta, H. Sawada, and S. Makino, “Spatio–Temporal FastICA Algorithms for the Blind Separation of Convolutive Mixtures”, IEEE Trans. on Audio, Speech and Language Processing, Vol. 15, no. 5, juillet 2007.

[63] L. Nguyen-Thi, C. Jutten, and J. Caelen, “Speech enhancement : Analysis and comparison of methods in various real situations,” in In J. Vandewalle, R. Boite, M. Moonen, and A. Oosterlinck, editors, Signal Processing VI, Theories and Applications, Elsevier, Ed., 303-306, août 1992.

[64] P. Comon and E. Moreau, “Blind MIMO equalization and joint diagonalization criteria,” Proc. International Conference on Acoustic Speech and Signal Processing (ICASSP’01), vol. 5, mai 2001.

[65] A. Adib, E. Moreau, D. Aboutajdine, “Source separation contrasts using a reference signal,” IEEE Signal Processing Letter, vol. 11, no. 3, pp. 312-315, mars 2004.

[66] M. Castella, J.-C. Pesquet, and A.-P. Petropulu, “A family of frequency and time domain contrasts for blind separation of convolutive mixtures of temporally dependent signals,” IEEE Transactions on Signal Processing, vol. 53, no. 1, 107-120, janvier 2005.

[67] Y. Inouye, “Criteria for Blind Deconvolution of Multichannel Linear Time-Invariant Systems,” IEEE Transactions on Signal Processing, vol. 46, no. 12, 3432-3436, décembre 1998.

[68] R. Liu and Y. Inouye, “Blind Equalization of MIMO-FIR Chanels Driven by White but Higher Order Colored Source Signals,” IEEE Transactions on Information Theory, vol. 48, no. 5, 1206–1214, mai 2002.

[69] E. Moreau and J.-C. Pesquet, “Generalized contrasts for multichannel blind deconvolution of linear systems,” IEEE Signal Processing Letter, vol. 4, no. 6, 182–183, juin 1997.

[70] N. Thirion-Moreau and E. Moreau, “Generalized criteria for blind multivariate signal equalization,” IEEE Signal Processing Letter, vol. 9, no. 2, pp. 72-74, février 2002.

[71] E. Moreau and N. Thirion-Moreau, “Non Symmetrical Contrasts for Source Separation,” in IEEE Transactions on Signal Processing, vol. 47, no. 8, pp. 2241–2252, août 1999.

[72] K. Torkkola, “Blind separation of convolved sources based on information maximization", Proc. IEEE Workshop on Neural Networks and Signal Processing, Kyota, Japan, 423-432, septembre 1996.

[73] A. J. Bell and T. J. Sejnowski, “An information-maximization approach to blind separation and blind deconvolution", Neural Computation, vol. 7, no. 6, pp. 1129-1159, 1995.

[74] T. W. Lee, A. J. Bell, and R. Lambert, “Blind separation of delayed and convolved sources", Proc. Advances in Neural Information Processing Systems 9, 758-764, Cambridge, MA : MIT Press, 1997.

[75] R. Mukai, S. Araki, H. Sawada, and S. Makino, “Evaluation of separation and dereverberation performance in frequency domain blind source separation", Journal of Acoustic Science and Technology, vol. 25, no. 2, pp. 119-126, mars 2004.

[76] P. Smaragdis, “Information theoretic approaches to source separation", Master's thesis, MIT Media Lab, juin 1997.

[77] P. Smaragdis, “Blind separation of convolved mixtures in the frequency domain", Neurocomputing, vol. 22, pp. 21-34, novembre 1998.

[78] H. Bousbia-Salah, A. Belouchrani, and K. Abed-Meraim, “Jacobi-like algorithm for blind signal separation of convolutive mixtures", IEE Electronics Letters, vol. 37, no. 16, pp. 1049-1050, août 2001.

[79] J. K. Tugnait, “Identification and deconvolution of multichannel linear nongaussian processes using higher order statistics and inverse filter criteria”, IEEE Trans. on Signal Processing, 658-672, mars 1997.

[80] J. G. Daugman, "Entropy reduction and decorrelation in visual coding by oriented neural receptive fields", IEEE Trans. on Biomedical Engineering, 36 :107-114, 1989.

[81] L. Lacoume, P. Ruiz “Separation of independent sources from correlated inputs”, IEEE Trans. on Signal Processing, 3074-3078, décembre 1992.

[82] W. Jun, “Wide band blind identification and separation of sources”, Proceedings of Eusipco-96, 2077-2080, Triestle, Italy, septembre, 1996.

[83] V. Capdevielle, Ch. Servière, J. L. Lacoume, “Blind separation of wideband sources in the frequency domain” Proceedings of ICASSP-95, vol. 2080-2083, Detroit (Michigan), USA, mai 1995.

[84] V. Capdevielle, Ch. Servière, J. L. Lacoume, "Blind separation of wideband sources : application to rotating machine signals”, Proceedings of Eusipco-96, 2085–2088, Triestle, Italy, septembre, 10-13, 1996.

[85] Ch. Servière, V. Capdevielle, “Blind adaptive separation of wide-band sources”, Proceedings of ICASSP-96, vol. 5, 2698–2701, Atlanta, Georgia, USA, mai, 7-10, 1996.

[86] S. Araki, R. Mukai, S. Makino, T. Nishikawa, Saruwatari, "The fundamental limitation of frequency domain blind source separation for convolutive mixtures of speech", IEEE Trans. on Speech and Audio Processing, Vol. 11, No. 2, 109 - 116, mai 2003.

[87] H. Attias and C. E. Schreiner, “Blind source separation and deconvolution : The dynamic component algorithm”, Neural computation, V. 10,1373–1424, 1998.

[88] H. C. Wu and J. Principe, “Simultaneous diagonalization in the frequency domain (sdif) for source separation” Proceedings of ICA 99, Aussois, France, janvier, 11-15, 1999.

[89] L. Parra and C. Spence, “Convolutive blind separation of non-stationary sources”, IEEE Transaction on Speech and Audio Processing, 320-327, mai 2000.

[90] L. Tong, R. W. Liu, V. C. Soon, Y. F. Huang, “Indeterminacy and identifiability of blind identification”, IEEE Trans. on Circuit and Systems, 499-509, 1991.

[91] I. Bradaric, A. P. Petropulu, and K. I. Diamantaras, “On resolving the column permutation ambiguity in the estimates of MIMO system response”, Conference on Information Sciences and Systems, 2000.

[92] K. Matsuoka, "Minimal distortion principle for blind source separation", SICE 2002. Proceedings of the 41st SICE Annual Conference, 2138 - 2143 Vol. 4, 5-7 août. 2002, Osaka, Japan.

[93] Recommandation UIT-T P.800. "Méthodes d'évaluation subjective de la qualité de transmission", 1996.

[94] M. Berg and E. Bondesson, “Real-Time Implementation of a Combined PCA-ICA Algorithm for Blind Source Separation”, Rapport de these de Master, Science in Electrical Engineering, Blekinge Institute of Technology avril 2005.

[95] Ch. Decaestecker, M. Saerens, “Analyse en composantes Principales”, ULB (Université Libre de Bruxelles) & UCL (Université catholique de Louvain), Belgique.

[96] K. I. Diamantaras, “PCA neural models and blind signal separation,” in In proc. International Joint Conference on Neural Networks (IJCNN-2001), Washington DC, juillet 2001.

[97] K. I. Diamantaras, Th. Papadimitriou,” Oriented PCA and Blind Signal Separation”, 4th International Symposium on Independent Component Analysis and Blind Signal Separation (ICA 2003), Nara, Japon, avril 2003.

[98] C. Jutten and J. Hérault, “Blind separation of sources, Part I : An adaptative algorithm based on neuromimetic architecture,” Signal Processing, vol. 24, pp. 1-10, 1991

[99] J. Pierce,“An introduction to information theory”, Dover, 1980.

[100] A. Taleb and C. Jutten, "Source Separation in Post-Nonlinear Mixtures", IEEE Transactions on Signal Processing, Vol. 47, No. 10, octobre 1999.

[101] S. Achard, D.T. Pham, C. Jutten, "Mesure de dépendance quadratique pour la séparation aveugle de source dans les mélanges non linéaires", Proc. Gretsi 2003, vol 1, 117-120, septembre 2003.

[102] S. Achard and D.T. Pham and C. Jutten, “Criteria based on mutual information minimization for blind source separation in post nonlinear mixtures”, signal processing, special issue on Information Theoretic Signal Processing, 85 pages 965-974, 2005.

[103] A. Hyvarınen, J. Karhunen. E. Oja, “Independent component analysis”. Inter Wiley Science, 2001.

[104] A. Hyvarınen, ”Survey on Independent Component Analysis”, Neural Computing Surveys 2, 94-128, 1999.

[105] J. H. Friedman and J. W. Tukey, "A Projection Pursuit Algorithm for Exploratory Data Analysis", IEEE Transactions on Computers, C-23, p. 881-890, septembre. 1974.

[106] A. Hyvärinen, “New approximations of dfferential entropy for independent component analysis and projection pursuit,” Advances in Neural Information Processing Systems, V. 10, 273-279. MIT Press, 1998.

[107] S. I. Amari, "Natural gradient works efficiently in learning", Neural Computation, 251-276, 1998.

[108] T. M. Cover and J. A. Thomas, “Elements of Information Theory”, Wiley, 1991.

[109] Te-Won Lee, Mark Girolami, Anthony J. Bell, and Terrence J. Sejnowski, “A unifying information-theoretic framework for independent component analysis”, International Journal of Computers and Mathematics with Applications, 1999.

[110] S. Amari and J. Cardoso. “Blind source separation semiparametric statistical approach”, IEEE Transaction on Signal Processing, 2692-2700, décembre 1997.

[111] A. Hyvärinen, J. Särelä, and R. Vigário. Spikes and bumps : “Artefacts generated by independent component analysis with insufficient sample size”. Proc. Int. Workshop on Independent Component Analysis and Signal Separation (ICA’99), pages 425-429, Aussois, France, 1999.

[112] M. Berg et E. Bondesson, “Real-Time Implementation of a Combined PCA-ICA Algorithm for Blind Source Separation”, thèse de Maîtrise, Bleking Institute of Technology, avril 2005.

[113] S. Amari, A. Cichocki, and H.H. Yang. “A new learning algorithm for blind source separation”, Advances in Neural Information Processing 8, pages 757 763. MIT Press, Cambridge, MA, 1996.

[114] J.-F. Cardoso, “Source separation using higher order moments,” in Proc. IEEE ICASSP, Glasgow, U.K., vol. 4, 2109-2112, 1989.

[115] Y. Benabderrahmane, "Méthodes de Traitement de Signaux Multidimensionnels Appliquées à l’Extraction de Micro-potentiels Electrocardiologiques", PhD, Université Joseph Fourier de Grenoble, France, 1997.

[116] Y. Benabderrahmane, A. Ben Salem, S. A. Selouani, and D. O’Shaughnessy, “ Blind Speech Separation using High Order Statistics”, 2009 Canadian Conference on Electrical and Computer Engineering, IEEE/CCECE, St. John’s, Newfoundland, Canada, 3-6 Mai, 2009, pp. 670-673.

[117] J. C. FORT : "Stabilté de l’algorithme de séparation de sources de JUTTEN HERAULT", Traitement du signal, Vol.8, N°1, pp.35-42, 1991.

[118] J. F. Cardoso, “Blind signal separation : statistical principles”, Proceedings of the IEEE. Special issue on blind identification and estimation : 2009-2026, octobre 1998.

[119] K. I. Diamantaras Th. Papadimitriou, "Temporal Filtering and Oriented PCA Neural Networks for Blind Source Separation", IEEE XI11 Workshop on Neural Networks for Signal Processing, 2003.

[120] K. I. Diamantaras, “Neural Networks and Principal Component Analysis”, in Handbook of Neural Network Signal Processing, Y. H. Hu and J. N. Hwang, Eds., chapitre 8. CRC Press, N.Y., 2002.

[121] W. Fisher, G. Dodington, & K. Goudie-Marshall, "The TIMIT-DARPA speech recognition research database : Specification and status", DARPA Workshop on Speech Recognition, 1986.

[122] ITU, “Perceptual evaluation of speech quality (PESQ), and objective method for end-to-end speech quality assessment of narrowband telephone networks and speech codecs”, ITU-T Recommendation 862, 2000.

[123] P Loizou, “Speech Enhancement : Theory and Practice”, CRC Press LLC, Boca Raton, FL, 2007.

[124] ITU-T Recommendation P.800, “Methods for Subjective Determination of Speech Quality”, International. Telecommunication Union, Geneva, 2003.

[125] Matthieu PUIGT, "Méthodes de séparation aveugle de sources fondées sur des transformées temps-fréquence. Application à des signaux de parole" PhD, Toulouse, Université de Toulouse III-Paul Sabatier, France, 13 décembre 2007.

[126] H. Ghennioui, "Séparation aveugle de mélanges linéaires convolutifs de sources corrélées", PhD, Université Med V-Agdal & Université du Sud-Toulon, 19 juillet 2008, Toulon, France.

[127] T. Nishikawa, H. Saruwatari, K. Shikano, and S. Makino, “Multistage ICA for blind source separation of real acoustic convolutive mixture”, in Proc. Independant Component Analysis (ICA’03), pp. 523-528, avril 2003.

[128] W. Wang, J. A. Chambers, and S. Sanei, “A joint diagonalization method for convolutive blind separation of nonstationary sources in the frequency domain,” Proc. Independent Component Analysis and Blind Sources Separation (ICA’03), 939-944, avril 2003.

[129] S. Busson, “Individualisation d’Indices Acoustiques pour la Synthèse Binaurale», PhD, Université de la Méditérannée Aix-Marseille II, 31 Janvier 2006.

[130] M. S. Pedersen, J. Larsen, U. Kjems, and L.C. Parra, “A Survey of Convolutive Blind Source Separation Methods”, Springer Handbook on Speech Processing and Speech Communication, 2007.

[131] J. Thomas, Y. Deville and S. Hosseini, “Time-Domain Fast Fixed-Point Algorithms for Convolutive ICA”, IEEE Signal Processing Letters, vol. 13, No. 4, pp. 228-231, avril 2006.

[132] A. Jourjine, S. Rickard and Özgür Yilmaz, “Bind Separation of Disjoint Orthogonal Signals Demixing N Sources from 2 Mixtures”, Proceedings ICASSP2000, Volume 5, pp. 2985-2988, juin 2000, Istanbul, Turkey.

[133] R. Balan and J. Rosca, "Statistical properties of STFT ratios for two channel systems and applications to blind source separation", In Proc. of ICA, pages 429-434, Finland, 2000.

[134] S. Rickard and O. Yilmaz. “On the approximate W-disjoint orthogonality of speech”, ICASSP, Orlando, Florida, mai 2002.

[135] Araki et al., "Equivalence between frequency domain blind source separation and frequency-domain adaptive beamforming for convolutive mixtures", EURASIP Journal on Applied Signal Processing, Vol. 2003, janvier 2003.

[136] D. Nion, K.N. Mokios, N.D. Sidiropoulos, and A. Potamianos, “Batch and Adaptive PARAFAC-Based Blind Separation of Convolutive Speech Mixtures”, IEEE Trans. on Audio, Speech and Language Processing, Vol. 18, No. 6, 1193-1207, août 2010. Voir aussi http ://www.telecom.tuc.gr/~nikos/BSS_Nikos.html

[137] B. Gardner and K. Martin,"HRTF Measurements of KEMAR Dummy-Head Microphone", J. Acoust. Soc. Amer., Vol. 97, 3907-3908, 1995. voir aussi http ://sound.media.mit.edu/ica-bench/.

[138] E. Vincent, C. Févotte and R. Gribonval, "Performance measurement in blind audio source separation", IEEE Transactions on Audio, Speech, and Language Processing, Vol. 14, No. 4, 1462 - 1469 , juillet 2006.

[139] C. Févotte, R. Gribonval, and E. Vincent, "BSS_EVAL toolbox user guide," IRISA, Rennes, France, Tech. Rep. 1706, 2005. [On-line], disponible : http ://www.irisa.fr/metiss/bss_eval.

[140] S. Makino, T.W. Lee, H. Sawada, "Blind Speech Separation", Signals and Communication Technology, publié par Springer, 2007.

[141] J. M. Peterson, S. Kadambe, "A probabilistic approach for blind source separation of underdetermined convolutive mixtures", Acoustics, Speech, and Signal Processing (ICASSP'03), vol. 6, 581-584, Hong Kong, Chine, avril 2003.

[142] R. K. Olsson, L. K. Hansen, "Blind separation of more sources than sensors in convolutive mixtures", Acoustics, Speech, and Signal Processing (ICASSP'06), vol. 5, 657-660, Toulouse, France, mai 2006.

[143] A. Blin, S. Araki, S. Makino, "A sparseness-mixing matrix estimation (SMME) solving the underdetermined BSS for convolutive mixtures", Acoustics, Speech, and Signal Processing (ICASSP'04), vol. 4, 85-88, Montréal, Canada, mai 2004.

[144] P. Boll E. Monte, "Underdetermined Convoluted Source Reconstruction Using LP and SOCP, and a Neural Approximator of the Optimizer", Ds Actes Intl. Conf. on Independent Component Analysis and Signal Separation (ICA'06), 569-576, Charleston, Caroline du Sud, mars 2006.

[145] A. Aïssa-El-Bey, K. Abed-Meraim et Y. Grenier, "Blind Separation of Underdetermined Convolutive Mixtures Using Their Time-Frequency Representation", IEEE Trans. on Audio, Speech, and Language Processing, 1540-1550, juillet 2007.

[146] O. Yilmaz et S. Rickard, "Blind Separation of Speech Mixtures via Time-Frequency Masking", IEEE Trans. on Signal Processing,1830-1847, juillet 2004.

[147] M. Kawamoto et Y. Inouye, "Blind separation of multiple convolved colored signals using second-order statistics" In Proc. of ICA'03, Nara, Japan, avril 2003

[148] Y. Benabderrahmane, S.A. Selouani, D. O'Shaughnessy, "Frequency Oriented PCA for Blind Speech Separation of Convolutive Mixtures in Multiples Environments", Interspeech 2011, 28-31 août, 2011, Florence, Italy

[149] Y. Benabderrahmane, S.A. Selouani, D. O'Shaughnessy, "Oriented PCA method for Blind Speech Separation of Convolutive Mixtures", Interspeech 2010, septembre 26-30, 2010, Makuhari, Japan.

[150] Y. Benabderrahmane, S.A. Selouani, D. O'Shaughnessy," Blind Speech Separation for Convolutive Mixtures Using an Oriented Principal Components Analysis Method", 2010 European Signal processing Conference (EUSIPCO 2010), 23-27 août, 2010, Aalborg, Nord du Danemark.

[151] Y. Benabderrahmane, S.A. Selouani, D. O'Shaughnessy, and H. Hamam, "A Comparative Study of Blind Speech Separation using Subspace Methods and Higher Order Statistics", Lecture Notes in Computer Science, Springer eds., pp. 117-124, 2009.

[152] H. BOUMARAF, "Séparation Aveugle de Mélanges Convolutifs de Sources", PhD, Université Joseph Fourier de Grenoble, France.

Annexe A

Rappels sur le kurtosis

Le kurtosis d'une variable aléatoire y est défini par :

$$kurt\,(y) = E\left\{y^4\right\} - 3\left(E\left\{y^2\right\}\right)^2. \tag{A.1}$$

Si on suppose que x a une variance unitaire, nous obtenons :

$$kurt\,(y) = E\left\{y^4\right\} - 3. \tag{A.2}$$

Ceci montre que le kurtosis est simplement une version normalisée du moment d'ordre $4E\left\{y^4\right\}$ Pour une variable y gaussienne, le moment d'ordre 4 est :

$$E\left\{y^4\right\} = 3\left(E\left\{y^2\right\}\right)^2. \tag{A.3}$$

Ainsi, pour une variable aléatoire gaussienne, le kurtosis est égal à zéro, et donc dans plusieurs cas (pas dans tous les cas), les variables aléatoires non gaussiennes, ont un kurtosis différent de zéro.

Le kurtosis peut être positif ou négatif :

i- $Kurt(y) > 0$: super-gaussien,
ii- $Kurt(y) < 0$: sous-gaussien.

Quelques propriétés du kurtosis :

Si X et Y sont des variables aléatoires indépendantes,

iii- $kurt\,(x+y) = kurt\,(x) + kurt\,(y)$,
iv- $kurt\,(\alpha x) = \alpha^4 kurt\,(x)$.

Plusieurs domaines d'applications : reconnaissance de la parole, rehaussement d'images, télécommunications, ...

Annexe B

Algorithmes de séparation aveugle de sources :SOBI, JADE, COM1, COM2

B.1 L'algorithme SOBI

L'algorithme SOBI exploite plusieurs matrices de covariance associées à différents retards non nuls après blanchiment des observations. En outre, les sources doivent être temporellement corrélées, posséder des spectres différents et le bruit doit être temporellement blanc. La matrice de covariance des observations blanchies $\{x[m]\}_{m\in N}$ est alors donnée pour un retard $\tau_j \in IN^*$ par :

$$R_z[\tau_j] = E\left[z[m]\, z[m-\tau_j]^T\right] - E[z[m]]\, E\left[z[m-\tau_j]^T\right] = \tilde{H} R_s[\tau_j]\, \tilde{H}^T \qquad \text{(B.1)}$$

où $\tilde{H}$ représente la matrice de mélange blanchie orthogonale et $R_s(\tau_j)$ désigne la matrice de covariance des sources associée au retard τ_j, qui a la propriété d'être diagonale. Considérons le théorème suivant :

Théorème : soient J retards non nuls τ_j et V une matrice orthonormée telle que :

$\forall\ j, 1 \leq j \leq J,\quad R_z[x_j] = V diag\left(\left[d_1^{(j)},\ ...,\ d_p^{(j)}\right]\right) V^T$ et $\forall\ _{p_1 p_2},\ 1 \leq p_1 \neq p_2 \leq P,\ \exists j,\ 1 \leq j \leq J,\ d_{p_1}^{(j)} \neq d_{p_2}^{(j)}$.

Alors V et $\tilde{H}$ sont égales à une matrice triviale près. Ce théorème est la clé de voûte de l'algorithme SOBI, fournissant une condition d'identifiabilité du mélange $\tilde{H}$. Notons que cette condition ne pourra pas être remplie en présence de sources dont les spectres normalisés sont identiques. En revanche, si ces derniers sont tous différents, il sera toujours possible de trouver un ensemble de J retards τ_j vérifiant la condition

du théorème.

B.2 L'algorithme COM2

La première classe de méthodes est basée sur la technique de l'Analyse en Composantes Indépendantes, utilisant des outils d'ordre supérieur pour construire des fonctions de contraste, lesquelles exploitent les propriétés des cumulants des sources supposées indépendantes. C'est P. Comon qui a étendu la notion de contraste utilisée auparavant dans les systèmes de déconvolution aveugle au cas des vecteurs aléatoires. Il a proposé dans [26] une méthode baptisée COM2 (COntrast Maximization2) consistant à maximiser une fonction de contraste notée $\psi_R^2(y)$ (Il s'agit de l'une des premières fonctions de contraste), et définie comme la somme des modules au carré des cumulants d'ordre $R \geq 3$ des sources estimées $y(t)$:

$$\psi_R^2(y) = \sum_{1=1}^{n} |C_R\{y_i\}|^2, \tag{B.2}$$

où n est le nombre de sources et $CR\{.\}$ représente le cumulant d'ordre R d'un signal stationnaire. La maximisation de la fonction contraste $\psi_R^2(y)$ a conduit à l'algorithme ICA [26].

De même que SOBI, COM2 nécessite un blanchiment des observations. Puis il faut maximiser le contraste. Une optimisation de ce contraste repose sur le théorème suivant :

Théorème : Soit s un vecteur aléatoire de composantes indépendantes, parmi lesquelles au plus une est gaussienne, et dont les densités de probabilité ne sont pas réduites à une masse ponctuelle. Soit T une matrice orthogonale de taille $(P \times P)$ et z le vecteur défini par $z = Ts$. Alors, les trois propriétés suivantes sont équivalentes :

i- les composantes de z sont indépendantes deux à deux,
ii- les composantes de z sont mutuellement indépendantes,
iii- T est une matrice triviale.

Ce théorème nous dit qu'au vu des indéterminations intrinsèques au problème de la Séparation Aveugle de Sources, il suffit juste de chercher à reconstruire des sources indépendantes deux à deux.

Une autre façon d'introduire les cumulants marginaux en tant que critère d'optimisation consiste à remarquer que les variables aléatoires sont indépendantes si tous leurs cumulants croisés sont nuls. On peut donc être tenté de minimiser l'ensemble des cumulants croisés à un ordre donné.

B.3 L'algorithme COM1

Après un blanchiment des observations le contraste suivant est optimisé :

$$\psi_R^2(y) = \lambda \sum_{1=1}^{n} |C_R\{y_i\}|, \text{ où } \lambda = \pm 1. \tag{B.3}$$

La solution itérative proposée pour COM2 peut être utilisée afin d'optimiser le contraste précédent. Cependant, une solution algébrique a été proposée. Notons toutefois que COM1 nécessite que les sources aient des kurtosis de même signe.

L'idée de l'algorithme COM1 (α_1) ou COM2 (α_2) dans $\psi_R^2(y) = \sum_{1=1}^{n} |C_R\{y_i\}|^{\alpha}$ est de remarquer que ce maximum peut être calculé entièrement analytiquement en dimension 2, dans un certain nombre de cas. L'intérêt de l'algorithme COM1 réside dans une propriété [44][45] établissant que la valeur absolue peut être supprimée lorsque les kurtosis des sources sont tous de même signe. C'est ce qui sera toujours implicitement supposé dans toutes les mises en œuvre de COM1 [144].

B.4 L'algorithme JADE

Elle est basée sur l'utilisation de cumulants d'ordre quatre impliquant à la fois des auto-cumulants et des cumulants croisés :

$$J(y) = \sum_{i_1,i_2,i_3=1}^{n} |cum\{y_{i_1}, y_{i_1}, y_{i_2}, y_{i_3}\}|^2, \tag{B.4}$$

où l'opérateur $cum\{.\}$ désigne un cumulant d'ordre quatre. La fonction de contraste précédente est appelée « contraste JADE ». Maximiser cette fonction revient à diagonaliser conjointement un ensemble den^2 matrices.

L'algorithme JADE nécessite aussi un blanchiment des observations, puis la maximisation du contraste.

Une solution algébrique consiste à diagonaliser conjointement l'ensemble des matrices propres M_p construites à partir des vecteurs propres associés aux P plus grandes valeurs propres de la matrice de quadri-covariance.

Annexe C

Formulaire de l'Analyse en Composantes Principales

Matrice des données : m objets (ou individus), p descripteurs (ou variables) :

x_{11}	.	.	x_{1y}
.	x_{ij}	$x_{ij'}$	.
.	$x_{i'j}$	.	.
x_{m1}	.	.	x_{my}
$\bar{x}_1$	$\bar{x}_i$	$\bar{x}_{i'}$	$\bar{x}_y$
s_1	s_i	$s_{i'}$	s_y

- Centrage et réduction des données :

$$X_{ij} = \frac{x_{ij} - \bar{x}}{S_i}. \tag{C.1}$$

- Distances entre objets :

$$d\left(i, i'\right)^2 = \sum_{j=1}^{p} p_j \left(X_{ij} - X_{ij'}\right)^2, \tag{C.2}$$

où p_j est une pondération qui est en général égale à 1.

Inertie totale I du nuage est égale à :

$$I = \frac{1}{m} \sum_{j=1}^{p} \sum_{i=1}^{m} \left(\frac{x_{ij} - \bar{x}_j}{s_j}\right)^2. \tag{C.3}$$

x_{11}	.	.	x_{1p}
.	x_{ij}	$x_{ij'}$	.
.	$x_{i'j}$	.	.
x_{m1}	.	.	x_{mp}
0	0	0	0
1	1	1	1

$\Longrightarrow$

ρ_{11}	.	.	ρ_{1p}
.	ρ_{jj}	$\rho_{jj'}$	.
.	$\rho_{j'j}$	$\rho_{j'j}$	.
ρ_{p1}	.	.	ρ_{pp}

$$\Downarrow$$

U_{11}	.	.	U_{1p}	λ_1
.	.	.	.	.
.	..	..	.	.
U_{Q1}	.	.	U_{Qp}	λ_Q

avec $Q \leq p$.

- Les vecteurs propres sont les coordonnées du point de norme 1 des nouveaux axes sur les anciennes variables. Les valeurs sont proportionnelles à la variance associée à ces axes.
- Calcul des cordonnées sur les nouveaux axes.
- Calcul des corrélations des variables avec les nouveaux axes.

- Relations de transitions :

$$F_q(i) = \frac{1}{\sqrt{\lambda_{ij}}} \sum_{j=1}^{p} \frac{x_{ij} - \bar{x}_j}{s_j} G_q(j), \tag{C.4}$$

où F est la projection de l'individu i sur l'axe factoriel q, G les projections des p variables sur q, et l la valeur propre associée à la composante q et par réciprocité :

$$G_q(j) = \frac{1}{\sqrt{\lambda_q}} \sum_{i=1}^{m} \frac{x_{ij} - \bar{x}_j}{s_j} F_q(j). \tag{C.5}$$

- Nuage des variables :

$$\cos\left(j, j'\right) = \sum_{i=1}^{m} \left(\frac{x_{ij} - \bar{x}_j}{s_j}\right) \left(\frac{x_{ij'} - \bar{x}_{j'}}{s_{j'}}\right) = Cor\left(j, j'\right) \tag{C.6}$$

Annexe D

Décomposition et maximisation de la variance

- Décomposition de la variance

• La variance se décompose de la manière suivante :

$\sigma = \frac{1}{2n(n-1)} \sum_{i=1}^{n} \sum_{j=1}^{n} (x_i - x_j)^T (x_i - x_j)$

$= \frac{1}{2n(n-1)} \sum_{i=1}^{n} \sum_{j=1}^{n} ((x_i - g) + (g - x_j))^T ((x_i - g) + (g - x_j))$

$= \frac{1}{2n(n-1)} \sum_{i=1}^{n} \sum_{j=1}^{n} (x_i - g)^T (x_i - g) + (g - x_j)^T (g - x_j)$
$+2 (x_i - g)^T (g - x_j)$

$= \frac{1}{(n-1)} \sum_{i=1}^{n} (x_i - g)^T (x_i - g)$

- Projection sur une droite

• L'opérateur de projection orthogonale, π, sur une droite de vecteur directeur unitaire v s'écrit :
$\pi = vv^T$,
avec
$v^T v = I$.
• La variance des observations projetées s'écrit alors,
$\sigma_v^2 = \frac{1}{(n-1)} \sum_{i=1}^{n} (\pi (x_i - g))^T (\pi (x_i - g))$.
- Recherche des projections de variance maximale orthogonales au premier axe

Afin de trouver le second axe de variance maximale, nous recherchons :
$\max_v \left(v^T \sum v\right) v^T v = I$ et $v^T v = 0$.

• Avec v_1 étant le premier vecteur propre à valeur propre maximale

• Comme les vecteurs propres de $\sum$ sont naturellement orthonormés, la solution est de choisir le deuxième vecteur propre de $\sum$ (à deuxième valeur propre maximale)

- Matrice variance-covariance

• Notons que si X est la matrice de données qui contient les vecteurs $(x_i - g)^T$ en ligne, la matrice $\sum = (n-1)^{-1} X^T X$.

- Interprétation des valeurs propres

• La somme des valeurs propres correspond à la variance totale

$\mathrm{tr}(\sum) = \sigma^2 = \sum_{i=1}^{n} \lambda_i$.

• Chaque valeur propre mesure la part de variance expliquée par l'axe factoriel correspondant.

Annexe E

Méthode par Déflation

La déflation classique vise à extraire les K sources l'une après l'autre. Cette technique peut être employée avec un critère indépendant de la distribution des sources, tel que les principes de la maximisation du kurtosis. Un seul critère est appliqué pour extraire chaque source des observations. Afin d'éviter d'extraire la même source plusieurs fois, la contribution de la source extraite doit être estimée (par exemple, via des techniques de corrélation) et soustraite de la sortie des capteurs. Cette procédure est répétée jusqu'à ce que les K sources aient été obtenues. La diversité requise pour l'extraction des K sources est limitée par un nombre de capteurs de l'ordre de $P \approx K$. Par ailleurs, les erreurs d'estimation s'accumulent avec le nombre d'extractions, donc la qualité de l'extraction diminue graduellement. La déflation classique est illustrée par la Figure ci-dessous.

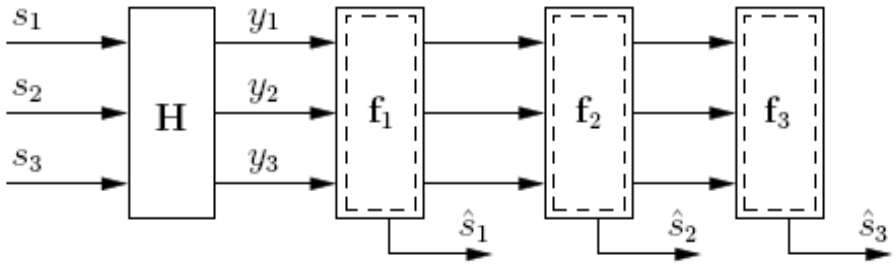

FIG. E.1 – Déflation classique. Extraction de 3 signaux $\{s_p\}_{p=1}^{3}$ayant typiquement (mais pas forcément) la même modulation. La déflation conventionnelle estime les signaux d'entrée un par un.

Annexe F

Influence du blanchiment sur les performances de la séparation

F.1 Blanchiment spatial d'ordre deux

Un vecteur est dit "blanc" si ses composantes sont décorrélées et de puissance égale à un. Ainsi, pour décorréler les composantes du vecteur $X(t)$ tout en imposant en plus que leur puissance soit unitaire, on doit chercher une matrice W dite de "blanchiment" permettant de rendre le vecteur $X_b(t) = WX(t)$ blanc ; en d'autres termes la matrice de covariance du vecteur $X_b(t)$ doit être égale à l'identité. On dit que les composantes de $X_b(t)$ sont blanchies. Cette étape de blanchiment ou de normalisation est nécessaire à de nombreuses méthodes de Séparation Aveugle de Sources. Avant de présenter la procédure de la recherche de la matrice de blanchiment W, on rappelle la définition de la matrice de corrélation d'un vecteur de signaux selon la nature de ces signaux.

F.2 Détermination de la matrice de blanchiment

Nous présentons la procédure de la recherche de la matrice de blanchiment W. Dans le cas où les sources sont aléatoires (resp. déterministes), soit $F = H\bar{R}_s^{1/2}(0)$ (resp. $F = HC_s^{1/2}(0)$). F est une matrice de taille et $\bar{R}_s^{1/2}(0)$

(resp. $C_s^{1/2}(0)$) est la racine de la matrice de corrélation du vecteur S(t) en $\tau = 0$.

La matrice H est régulière, alors la matrice F l'est aussi si $\bar{R}_s^{1/2}(0)$ (resp. $C_s^{1/2}(0)$) est définie positive, ce qui est le cas lorsque les sources possèdent une diversité temporelle (les valeurs propres de la matrice de corrélation sont différentes), elle peut donc se décomposer en valeurs singulières de la manière suivante :

$$F = V\Delta^{1/2}U. \tag{F.1}$$

avec V et U deux matrices unitaires de taille respectivement $M \times M$ et $N \times N$ *et* $\Delta^{1/2}$ est une matrice de dimension $M \times N$, dont le bloc supérieur de dimension $N \times N$ est une matrice diagonale et le bloc inférieur de dimension $(M - N) \times N$ est une matrice nulle. Par conséquent :

$$FF^H = V\Delta' V^H, \tag{F.2}$$

où $\Delta' = \Delta^{1/2}\left(\Delta^{1/2}\right)^H$ est une matrice carrée diagonale de dimension $M \times M$ tel que :

$$\Delta' = diag\left\{\delta_1,\ ..., \delta_N,\ 0, ...0\right\}. \tag{F.3}$$

Cela conduit à avoir :

$$\bar{R}_x(0) = FF'' + \sigma_n^2 I_M = V\Delta'' V'' + \sigma_n^2 I_M = V\Delta'' V'', \tag{F.4}$$

où $\Delta'' = \Delta' + \sigma_n^2 I_M = diag\left\{\delta_1\ + \sigma_N^2,\ ..., \delta_N + \sigma_N^2,\ \sigma_N^2, ...\sigma_N^2\right\}$est une matrice diagonale. La relation donnée par F.4 représente la décomposition en valeurs propres (DVP) de la matrice $\bar{R}_x(0)$. Ainsi, si $M > N$ alors il est possible d'estimer la variance du bruit σ_N^2 par un moyennage des $M - N$ plus petites valeurs propres de $\bar{R}_x(0)$ et la matrice Δ' peut être estimée par les N plus grandes valeurs propres. Lorsque $M = N$ on doit forcément supposer que σ_N^2 est connue, et dans ce cas là, la matrice Δ' est estimée par la DVP de la matrice $\left(\bar{R}_x(0) - \sigma_N^2 I_M\right)$.

La matrice de blanchiment spatiale W de taille $N \times M$ définie par :

$$W = \left(\Delta^{1/2}\right)^{\#} V^H, \tag{F.5}$$

tandis que la matrice unitaire U s'écrit de la façon suivante :

$$U = WH\bar{R}_s(0)^{1/2}, \tag{F.6}$$

où $\bar{R}_s(0)^{1/2}$ est une matrice bloc-diagonale (lorsque les sources sont déterministes, on remplace $\bar{R}_s(0)^{1/2}$ par $C_s^{1/2}(0)$. En fait,

$H = F\bar{R}_s(0)^{-1/2} \implies \left(\Delta^{1/2}\right)^{\#} V^H H = U\bar{R}_s(0)^{1/2} \implies W = \left(\Delta^{1/2}\right)^{\#}$ et $U = WH\bar{R}_s(0)^{1/2}$,
On peut vérifier que la matrice de covariance des observations blanchies $X_b(t) = WX(t)$ est égale à la matrice identité :

$$\begin{aligned}
&\bar{R}_{x_b}(0) = W\bar{R}_x(0)W^H = \left(\Delta^{1/2}\right)^{\#} V^H H\bar{R}_s(0)H^H V\left(\left(\Delta^{1/2}\right)^{\#}\right)^H \\
&= \left(\Delta^{1/2}\right)^{\#} V^H FF'' V\left(\left(\Delta^{1/2}\right)^{\#}\right)^H \\
&= \left(\Delta^{1/2}\right)^{\#} V^H V\Delta' V^H V\left(\left(\Delta^{1/2}\right)^{\#}\right)^H \\
&= \left(\Delta^{1/2}\right)^{\#} \Delta^{1/2}\left(\left(\Delta^{1/2}\right)^H\right)\left(\left(\Delta^{1/2}\right)^{\#}\right)^H = I_N.
\end{aligned}$$

Cette étape de blanchiment permet alors de se ramener au cas d'un mélange unitaire de sources. En modèle instantané ($L = 0, L\prime = 1 et A = H$), la matrice $\bar{R}_s(0)^{1/2}$ n'est rien d'autre que la matrice identité multipliée par la puissance des

sources. Si les sources sont supposées de puissance un (c.à.d. standardisées), alors la matrice unitaire en instantané est la suivante :

$$U = WA. \tag{F.7}$$

Le schéma de la séparation de sources incluant l'étape de blanchiment est donné au niveau de la figure ci-dessous :

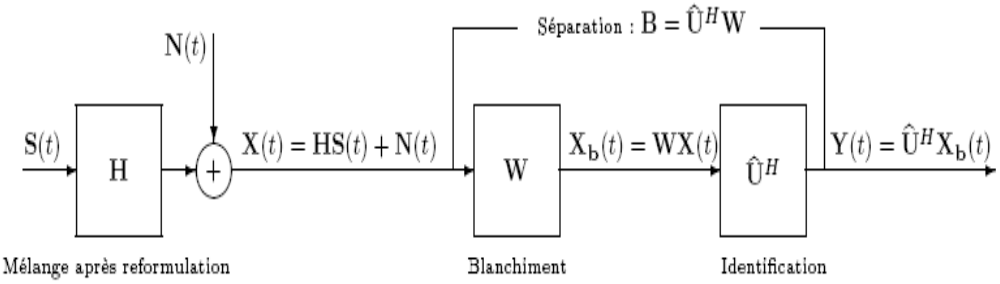

FIG. F.1 – Schéma général de la séparation de sources incluant l'étape de blanchiment.

La Séparation Aveugle de Sources incluant l'étape de blanchiment, est atteinte lorsque la matrice $\hat{U}^H$ estimée de U^H est de la forme PDU^H.

F.3 Influence du blanchiment sur les performances de la séparation

Pour montrer l'influence du blanchiment spatial d'ordre deux, J.-F. Cardoso, établit dans [16] une borne minimale sur les performances de la séparation de mélanges linéaires instantanés de sources. Effectivement, le blanchiment n'a de sens que si la matrice de corrélation des signaux sources à l'origine est à symétrie hermitienne et définie positive d'une part et que les sources sont décorrélées d'autre part. Ainsi, pour obtenir une bonne séparation, la matrice produit $G = BH$ doit être égale au produit d'une matrice de permutation P par une matrice bloc-diagonale D, toutefois la matrice G dépend de la matrice de corrélation des sources à l'origine,

$$\begin{aligned}
&\hat{B}H = \hat{U}^H WH = \hat{U}^H \left(\Delta^{1/2}\right)^{\#} V^H F \bar{R}_s(0)^{-1/2} \\
&= \hat{U}^H \left(\Delta^{1/2}\right)^{\#} V^H V \Delta^{1/2} U \bar{R}_s(0)^{-1/2} \\
&= \hat{U}^H U \bar{R}_s(0)^{-1/2} = PD\bar{R}_s(0)^{-1/2}.
\end{aligned}$$

En utilisant la définition de l'indice de performance décrit au niveau de l'expression ci-dessus et l'inégalité de Schwartz, nous déduisons la relation de la borne minimale de la séparation (en échelle linéaire) :

$$\begin{aligned}
0 \leq I_{conv}(G) &= I_{conv}\left(PD\bar{R}_s(0)^{-1/2}\right) \\
&\leq I_{conv}(PD)\, I_{conv}\left(\bar{R}_s(0)^{-1/2}\right) \leq I_{conv}\left(\bar{R}_s(0)^{-1/2}\right)
\end{aligned}$$

Notons que lorsque $L = 0$, $L\prime = 1$ et $A = H$ ($IConv(G)$ n'est rien d'autre que l'in-

dice de performance introduit dans [86]), nous retrouvons alors la relation de la borne minimale de la séparation de mélanges linéaires instantanés de sources présentée dans [61].

En pratique, il se peut dans le cadre des systèmes de mélanges convolutifs (resp. instantanés) que la matrice de corrélation des sources à l'origine ne soit pas bloc-diagonale (resp. diagonale), et ce même lorsque les sources sont effectivement décorrélées du fait d'une mauvaise estimation de la matrice de corrélation des signaux observés. C'est la raison pour laquelle nous cherchons à supprimer l'étape de pré-blanchiment pour résoudre le problème de la Séparation Aveugle de Sources.