

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE  
MINISTERE DE L 'ENSEIGNEMENT SUPERIEUR  
ET DE LA RECHERCHE SCIENTIFIQUE



N° d'ordre :.....

Série :.....

**Université Mohammed Seddik**  
**BENYAHIA - Jijel**  
**Faculté des Sciences et de la Technologie**  
**Département : Electronique**

**Thèse Présentée**  
Pour l'obtention du Diplôme de Doctorat en Sciences  
Spécialité : Electronique

**Par :**

**KEMIHA Mina**

**Thème :**

**Séparation de mélanges de signaux de parole**

Soutenue le : **11 /10 / 2017** , Devant le Jury composé de :

Mr. BOUDEN Toufik	Professeur	U. M.S.B.Jijel	Président
Mr. KACHA Abdellah	Professeur	U. M.S.B.Jijel	Rapporteur
Mr. BENOUIT Nabil	Professeur	U. Batna 2	Examineur
Mr. BOUGUEZEL Saad	Professeur	U. Sétif 1	Examineur

## Remerciements

Je tiens à exprimer ma gratitude à Monsieur A. KACHA professeur à l'université Mohamed Seddik Ben yahia-Jijel- pour ses qualités de direction, sa constante disponibilité et sa présence bienveillante, sa patience à m'écouter et à expliquer, pour m'avoir poussé à une certaine rigueur scientifique, technique et enfin didactique pour la rédaction de ce document, qui ont fait de ce travail ce qu'il est.

Je remercie Monsieur T. BOUDEN professeur à l'université Mohamed Seddik Ben yahia-Jijel d'avoir apporté son regard personnel et de m'avoir fait l'honneur de présider le jury.

Je tiens à remercier les membres du jury: Monsieur BENOUDJIT Nabil professeur à l'université de Batna 2 et Monsieur BOUGUEZEL Saad professeur à l'université Ferhat Abbas-Sétif 1 qui m'ont fait l'honneur d'évaluer ce travail.

Mes remerciements vont également à Monsieur S. HADDAD maitre conférence classe B à l'université de Mohamed Seddik Ben Yahia-Jijel- pour ces conseils et son aide ainsi que tous ceux qui ont participé de près ou de loin à l'élaboration de ce travail

## Table des matières

Liste des figures.....	i
Liste des tableaux.....	iii
Liste des symboles.....	iv
Liste des acronymes.....	v
Introduction générale.....	1

### Chapitre 1

#### La séparation aveugle de sources :

#### État de l'art

<b>1 Introduction</b> .....	9
<b>2 Formulation du problème de la SAS</b> .....	9
2.1 Mélange instantané.....	10
2.2 Mélange anéchoïque.....	11
2.3 Mélange convolutif.....	11
<b>3 Définitions</b> .....	12
3.1 Blanchiment.....	12
3.2 Circularité.....	12
3.3 Analyse en composantes indépendantes complexes.....	14
<b>4 Hypothèses de base</b> .....	15
4.1 Hypothèse 1.....	15
4.1.1 <i>Cumulant croisé</i> .....	15
4.1.2 <i>Divergence de Kullback-Leibler</i> .....	16
4.1.3 <i>Information mutuelle</i> .....	16
4.2 Hypothèse 2.....	17
4.2.1 <i>L'entropie différentielle marginale et conjointe</i> .....	17
4.2.2 <i>Néguentropie conjointe</i> .....	18
4.2.3 <i>Kurtosis</i> .....	18
<b>5 Ambiguïtés de séparation</b> .....	19
5.1 Ambiguïté de permutation.....	19
5.2 Ambiguïté d'échelle.....	20
<b>6 Domaine temporel vs domaine fréquentiel</b> .....	21
<b>7 Méthodes de séparation de mélanges convolutifs</b> .....	21
<b>8 Méthodes de séparation des signaux à valeurs complexes</b> .....	24
8.1 Méthodes de séparation basées sur la non-circularité.....	25
8.2 Méthodes de séparation basées sur la non-gaussianité.....	25
8.3 Méthodes de séparation basées sur la non-circularité et la non-gaussianité.....	27
<b>9 Conclusion</b> .....	29

### Chapitre 2

#### Décomposition en modes empiriques

<b>1 Introduction</b> .....	30
<b>2 Méthodes conventionnelles de décomposition des signaux</b> .....	30
2.1 Transformée de Fourier.....	30

2.2 Transformée en ondelettes.....	31
2.3 Décomposition par transformée en ondelettes.....	32
.....	
<b>3 Décomposition en modes empiriques.....</b>	<b>34</b>
3.1 Principe.....	34
3.2 Processus de tamisage.....	35
3.3 Définition et caractéristiques de la fonction de mode intrinsèque.....	38
<b>4 Autour de la méthode.....</b>	<b>39</b>
4.1 Critère d'arrêt du processus de tamisage.....	39
4.2 Interpolation.....	40
4.3 Effet des bords.....	40
<b>5 Extension aux signaux à valeurs complexes.....</b>	<b>41</b>
5.1 Décomposition en modes empiriques complexes.....	41
5.2 Décomposition en modes empiriques invariantes en rotation.....	42
5.3 Décomposition en modes empiriques bivariées.....	42
5.3.1 Principe.....	42
5.3.2 Une enveloppe en 3 dimensions.....	43
5.3.3 Algorithme.....	45
<b>6 Conclusion.....</b>	<b>46</b>

### **Chapitre 3**

#### **Déconvolution du signal de parole par décomposition en modes empiriques**

<b>1 Introduction.....</b>	<b>47</b>
<b>2 Déconvolution du signal de parole.....</b>	<b>47</b>
2.1 Principe de la déconvolution par la DME.....	47
2.2 Principe de déconvolution par la transformée en ondelettes.....	52
2.3 Résultats et discussions.....	53
<b>3 Application à l'estimation de la fréquence fondamentale et des formants.....</b>	<b>61</b>
<b>4 Application à l'estimation de la source glottique.....</b>	<b>66</b>
4.1 Estimation de la source glottique par cepstre complexe.....	67
4.2 Estimation de la source glottique par DME.....	67
4.3 Estimation de la source glottique par DMEB.....	68
4.4 Résultats et discussions.....	69
<b>5 Conclusion.....</b>	<b>72</b>

### **Chapitre 4**

#### **Séparation aveugle de mélange de signaux de parole**

<b>1 Introduction.....</b>	<b>73</b>
<b>2 Algorithme conventionnel de l'analyse en composantes indépendantes dans le domaine fréquentiel.....</b>	<b>73</b>
<b>3 Décomposition en composante indépendantes basée sur la minimisation des bornes de l'entropie.....</b>	<b>74</b>
3.1 Principe.....	74
3.2 Principe d'estimation de l'entropie.....	77
3.2.1 Borne 1 de l'entropie.....	78



3.2.2 <i>Borne 2 de l'entropie</i> .....	78
3.3 Estimation d'entropie par la décomposition QR .....	79
3.3.1 <i>Absorption de Q</i> .....	80
3.2.2 <i>Expression explicite de R</i> .....	81
<b>4 Analyse en composantes indépendantes par la minimisation des bornes de l'entropie et décomposition en modes empiriques bivarée</b> .....	82
3.1 Principe .....	82
<b>5 Résolution de l'ambiguïté de permutation</b> .....	85
<b>6 Résultats et discussions</b> .....	85
<b>7 Conclusion</b> .....	101
<b>Conclusion Générale</b> .....	102
<b>Références</b> .....	104

## Liste des figures

1.1	Problème de la SAS .....	10
2.1	Exemple de décomposition en modes empiriques.....	36
2.2	Etapas d'extraction des FMIs par processus de tamisage .....	36
2.3	Décomposition modale empirique du signal de la figure 2.1.....	39
2.4	Illustration de l'interpolation par différentes méthodes.....	41
2.5	Le principe des extractions bivariées.....	44
3.1	Illustration de la déconvolution du signal de parole par la méthode proposée.....	50
3.2	Comparaison entre la méthode de séparation basée sur la TO et celle basée sur la DME appliquées à une voyelle /a/ synthétique pour deux longueurs de la trame.....	55
3.3	Comparaison entre la méthode de séparation basée sur la TO et celle basée sur la DME appliquées à une voyelle /a/ naturelle pour deux longueurs de la trame.....	57
3.4	Deux trames successives extraites du signal de parole continue.....	59
3.5	Comparaison entre les méthodes de séparation basées sur la TO et la DME appliquées à deux trames successives extraites d'un signal de parole correspondant à une phrase prononcée par un locuteur masculin pour une longueur de trame de 1024 échantillons.....	59
3.6	Comparaison entre l'enveloppe spectrale théorique et l'enveloppe spectrale estimée via la méthode de déconvolution basé sur la TO et la DME.....	62
3.7	Erreur quadratique moyenne entre l'enveloppe spectrale théorique et estimée en fonction de la longueur de la trame.....	63
3.8	Schéma bloc de la méthode d'estimation de la source glottique basée sur la DME.....	68
3.9	Schéma bloc de la méthode d'estimation de la source glottique basée sur DMEB.....	68
3.10	Forme d'onde de la source glottique estimée par la DME superposée au modèle de l'excitation glottique pour différentes longueurs L de la trame.	69
3.11	Forme d'onde de la source glottique estimée par la méthode proposée comparée à celle estimée par le cepstre complexe et au modèle théorique pour une trame de longueur L=1024.....	71
3.12	Forme d'onde de la source glottique estimée par DMEB superposée au modèle de l'excitation glottique pour une longueur de trame m1024 échantillons.....	72
4.1	Principe de l'ACIF.....	74
4.2	Séparation aveugle de sources combinant la décomposition modale empirique bivariée (DMEB) et l'analyse en composantes indépendantes complexe par minimisation des bornes de l'entropie (ACI-MBE).....	84
4.3	Exemple d'illustration de la séparation aveugle de mélanges convolutifs de deux signaux de parole par DMEB combinée avec ACI-MBE.....	86
4.4	Parties réelles et imaginaires de la transformée de Fourier des trames de 1024 échantillons extraites du mélange et leurs FMIs correspondantes.....	89
4.5	Estimation des trames des sources 1 et 2 dans le domaine fréquentiel.....	91
4.6	Sources estimées et leurs spectrogrammes.....	92

4.7	Configuration spatiale: localisation des sources et des capteurs dans la pièce (x, y et z en mètre).....	94
4.8	Exemple de séparation aveugle de mélange convolutif dans un milieu réverbérant.....	95
4.9	Comparaison entre la méthode proposée et la méthode d'analyse en composantes indépendantes combinée avec l'algorithme de décomposition en modes empiriques (ACI-DME) et la méthode conventionnelle ACIF en termes de SIR, SAR, SDR et ISNR.....	100
4.10	Comparaison entre les performances de (a) la méthode proposée, (b) la méthode DME-ICA en terme de SIR, SAR, SDR et ISNR .....	100

## Liste des tableaux

1.1	Comparaison entre les méthodes de SAS .....	22
3.1	Fréquences des formants théoriques et estimées pour une voyelle /a/ synthétique en utilisant une longueur de trame de 512 et 1024 échantillons.....	64
3.2	Fréquences fondamentales théorique et estimées pour une voyelle /a/ synthétique utilisant des longueurs de trames de 512 et 1024 échantillons.....	65
3.3	Estimation des fréquences fondamentales pour une voyelle /a/ naturelle en utilisant des longueurs de trames de 512 et 1024 échantillons.....	65
4.1	Valeurs de $p$ du test statistique de Kruskal Wallis pour SIR, SAR, SDR et ISNR de 20 mélanges convolutif.....	101

## Symboles

$\mathbf{s}$	Vecteur de $N$ sources
$\mathbf{A}$	Matrice de mélange
$\mathbf{x}$	Vecteur de $M$ mélanges observés
$\mathbf{z}$	vecteur de sources estimées
$\mathbf{W}$	Matrice de séparation
$\theta$	angle de circularité
$\Sigma_x$	est la matrice de covariance du vecteur $\mathbf{x}$
$J(\mathbf{u})$	La néguentropie est définie par
$\mathbf{P}$	Matrice de permutation
$\mathbf{D}$	Matrice d'échelle
$\psi(t)$	Ondelette mère
$E_\omega(f)$	Spectres de l'excitation glottique.
$V(f)$	Réponse fréquentielle du conduit vocal.
$I(.)$	Information mutuelle
$E[\cdot]$	Espérance mathématique
$H$	Entropie
$neg(.)$	Néguentropie
$H^{[borne, I]}$	borne 1 de l'entropie
$H^{[borne, II]}$	borne 2 de l'entropie
$\det(.)$	Déterminant
$(.)^T$	transpose
$(.)^H$	Conjugué transposé
$/./$	module

## Acronymes

TF	Transformée de Fourier
TFCT	Transformée de Fourier à court terme
TFR	Transformée de Fourier rapide
TFI	Transformée de Fourier inverse
TO	Transformée en ondelettes
TOD	Transforme en ondelettes discrètes
TODI	Transformée en ondelettes discrètes inverses
DME	Décomposition en modes empirique
FMI	Fonction de mode intrinsèque
DEMC	Décomposition en modes empirique complexes
DMEIR	Décomposition en modes empiriques invariantes en rotation
DMEB	Décomposition en modes empirique bivariées
ACI	Analyse en composante indépendantes
ACI-MBE	Analyse en composantes indépendantes par minimisation des bornes de l'entropie
SAR	Signal to artifacts ratio
SIR	Signal to interferences ratio
SDR	Signal to distortion ratio
SNR	Rapport signal bruit (RSB)
ISNR	Improvement signal to noise ratio

# **Introduction générale**

## Introduction générale

La séparation aveugle de sources (SAS) a pour objectif d'estimer un ensemble de signaux appelés sources à partir d'observations qui sont des mélanges de ces signaux et cela sans connaissance a priori. La SAS a beaucoup évolué au cours des deux dernières décennies, et a trouvé un large éventail d'applications dans les principaux domaines de recherche tels que les télécommunications multicanaux [1], l'analyse des signaux biomédicaux [2-4], l'imagerie astronomique multi-spectrale [5], le traitement des données géophysiques [6], la détection et la localisation radar [7] ainsi que de nombreuses autres applications [8-11]. En traitement de la parole, la SAS est utilisée pour le rehaussement et la reconnaissance de la parole et dans le problème du "*cocktail party*" où les signaux de parole d'un ensemble de locuteurs en discours dans une salle sont mélangés et on veut retrouver le signal de chacune des personnes débarrassé des autres signaux qui le perturbent en utilisant des microphones répartis dans divers endroits de la salle.

Le mélange de ces sources s'effectue lors de leurs propagations jusqu'aux capteurs. Dans certaines applications telles que l'analyse du signal biomédical [12], chaque enregistrement du mélange peut être modélisé comme étant la somme pondérée des signaux sources. Ce modèle est connu sous le nom de mélange instantané. Dans de nombreuses applications du monde réel telles que les communications radio [13] et la propagation multi-chemins des signaux acoustiques [14], chaque enregistrement du mélange est modélisé comme étant la somme pondérée des versions retardées des signaux sources. Ce modèle est appelé mélange convolutif.

Le problème de la déconvolution aveugle est étroitement lié à la SAS. Dans la déconvolution aveugle, on dispose d'un seul signal observé modélisé comme étant la convolution du signal source avec la réponse impulsionnelle d'un système linéaire



## Introduction générale

temps invariant. Le signal observé est un mélange du signal source avec lui-même à différents délais. La déconvolution aveugle a pour objectif d'estimer le signal source à partir du signal observé sans connaître le système de mélange, les délais et les coefficients.

Dans la production de la parole, les sons sont considérés comme le résultat de l'action d'un filtre qui modélise le conduit vocal sur une source acoustique qui modélise la source glottique. Ce modèle est appelé modèle source-filtre. La source est une séquence périodique pour les sons voisés et une séquence de bruit blanc pour les sons non voisés. Le spectre d'amplitude d'un signal de parole voisée peut être considéré comme la combinaison de deux composantes à savoir une contribution qui représente le conduit vocal et une autre qui représente la source glottique. La séparation de ces deux composantes est très attrayante pour le traitement de la voix car elle fournit un moyen d'analyser, d'étudier et de comprendre les propriétés de production de la voix, ainsi que pour d'autres applications de la parole telles que la reconnaissance de locuteur où la déconvolution est utilisée pour l'estimation des formants qui sont l'effet des modes propres qui caractérisent le conduit vocal [15]. Dans les applications médicales de la parole, l'effet du conduit vocal doit être éliminé et les paramètres de l'excitation sont utilisés pour calculer les mesures objectives pour évaluer la qualité de la voix des locuteurs dysphoniques [16, 17].

En outre, une voix peut être transformée et synthétisée en utilisant la manipulation indépendante de ses éléments. Malgré le grand nombre de méthodes proposées dans la littérature, l'estimation de l'enveloppe spectrale et de la composante harmonique est le plus souvent considérée comme deux problèmes indépendants. Si la composante harmonique (fréquence fondamentale) est connue, l'enveloppe spectrale peut être estimée de façon fiable et inversement si l'enveloppe spectrale est connue, la

## Introduction générale

composante harmonique peut être estimée avec précision. Les méthodes conventionnelles d'estimation de l'enveloppe spectrale sont basées sur le codage prédictif linéaire (CPL) [15] ou sur le cepstre réel [17]. L'estimation de la composante harmonique ou la fréquence fondamentale (pitch) est basée sur des paramètres qui exploitent l'hypothèse de périodicité locale de la parole voisée soit dans le domaine temporel soit dans le domaine fréquentiel [18-20]. Un procédé qui a été proposé comme alternative pour estimer simultanément l'enveloppe spectrale et la composante harmonique est la déconvolution par la transformée en ondelettes (TO) [21]. L'inconvénient de la méthode de décomposition basée sur la transformée en ondelettes est l'utilisation des fonctions de base définies a priori faisant de cette approche non optimale pour tous les types de signaux de parole.

Le problème de la SAS peut être formulé dans le domaine temporel ou dans le domaine fréquentiel. Les méthodes temporelles de SAS sont complexes en termes de calcul. Les méthodes fréquentielles de SAS sont plus efficaces en termes de complexité de calcul. De plus, le mélange convolutif peut être transformé en un mélange instantané dans le domaine fréquentiel.

L'hypothèse la plus importante dans la SAS est que les sources sont mutuellement indépendantes ou mutuellement non corrélées. Cette hypothèse est exploitée par les méthodes de SAS pour extraire les sources. La plupart des techniques de SAS utilisent une fonction coût comme mesure de l'indépendance statistique des sources. L'analyse en composantes indépendantes (ACI) est la méthode statistique conventionnelle couramment utilisée dans la SAS. La technique ACI transforme le vecteur des données en un vecteur dont les composantes sont statistiquement indépendantes. Les techniques ACI pour la SAS sont basées sur les statistiques d'ordres supérieurs pour optimiser la fonction coût en utilisant les

méthodes du maximum de vraisemblance, l'information mutuelle, l'entropie et la négentropie ainsi que la non-gaussianité [22-28].

Récemment, un nouvel algorithme basé sur l'ACI par minimisation de la borne de l'entropie (ACI-MBE) a été introduit. La technique utilise la projection du gradient conjugué et des estimations précises de l'entropie et exploite efficacement à la fois la non-circularité et la non-gaussianité pour résoudre le problème de la SAS en minimisant l'information mutuelle [29]. Le principal avantage de la méthode ACI-MBE est qu'elle ne nécessite aucune connaissance préalable de la distribution des sources. Dans [30], une extension de l'algorithme ACI-MBE appelée analyse en composantes indépendantes par minimisation du taux de la borne d'entropie (ACI-MTBE) a été proposée. La méthode est basée sur l'ACI-MBE et un modèle de corrélation flexible. La méthode ACI-MTBE a été utilisée dans [31] pour séparer et classer les données électro-encéphalographiques (EEG) de fatigue.

Un inconvénient des techniques de SAS basées sur l'ACI est leur application restreinte à des signaux réels. En outre, elles ne peuvent pas être directement utilisées pour résoudre le problème de SAS de signaux à valeurs complexes. Le problème de la séparation des signaux à valeurs complexes se rencontre dans de nombreuses situations. En effet, la séparation des mélanges convolutifs de signaux à valeurs réelles peut être transformée en une séparation de mélanges instantanés de signaux à valeurs complexes via la transformée de Fourier.

Plusieurs algorithmes ont été proposés pour résoudre le problème de SAS dans le cas des signaux à valeurs complexes [32-34]. En général, les algorithmes ACI complexes exploitent deux caractéristiques du signal qui sont la non-gaussianité et la non-circularité. Dans la littérature, certains algorithmes ignorent la non-gaussianité et n'exploitent que la non-circularité en utilisant la diagonalisation des matrices de

covariance [35, 36] tandis que d'autres algorithmes exploitent uniquement la non-Gaussianité comme dans [37]. Dans [38], les signaux EEG ont été décomposés en utilisant l'algorithme ACI-infomax complexe afin de capter et comprendre leur dynamique.

Plusieurs méthodes qui exploitent à la fois la non-gaussianité et la non-circularité des sources ont été développées. Dans [39], l'amplitude de la corrélation a été utilisée comme une fonction de contrainte non linéaire pour la maximisation de la négentropie afin d'effectuer une correction de permutation. Dans [40], une extension de l'algorithme proposé dans [35] a été présentée. La technique utilise le coefficient d'aplatissement (kurtosis) normalisé adapté aux sources non circulaires en tant que fonction coût. Ces algorithmes ne peuvent pas résoudre le problème de SAS pour les sources gaussiennes non circulaires. En outre, la présence de valeurs aberrantes entraîne une performance de séparation dégradée. Dans [41], l'algorithme ACI-MBE développé dans [29] a été étendu aux signaux à valeurs complexes et une comparaison de performance de huit algorithmes ICA complexes différents en termes de pourcentage d'échecs, de rapport interférence-signal moyen et de temps CPU moyen a été effectuée. Il a été démontré que l'ACI-MBE offre un bon compromis entre la performance de séparation et la complexité de calcul. La technique est basée sur le gradient conjugué et fournit des estimations précises de l'entropie en délimitant ces estimations; De plus, cette technique convient pour une large gamme de fonctions de densité de probabilité.

La décomposition en modes empiriques (DME) [42] est une méthode de décomposition auto-adaptative locale qui permet d'analyser les données provenant de processus non stationnaires et / ou non linéaires. Son principe est basé sur l'extraction séquentielle de différentes contributions sous forme de composantes modulées en

## Introduction générale

amplitude et en fréquence (AM-FM) appelées fonctions de modes intrinsèques (FMI). La DME a donné lieu à de nombreux travaux de recherche [43-47]. Plusieurs versions améliorées de l'algorithme de DME ont été proposées dans la littérature. La décomposition en modes empiriques d'ensemble (DMEE) [48] et la décomposition en modes empiriques d'ensemble complet avec bruit adaptatif (DMEECBA) [49, 50] ont été utilisées pour décomposer des signaux bruités. L'algorithme DME a été combiné avec l'ICA dans plusieurs études pour séparer les signaux sources à partir d'enregistrements de mélanges instantanés. Dans [51] une méthode basée sur la DME-ACI a été utilisée pour séparer les signaux sources à partir d'enregistrements monocanal et à deux canaux. Dans [52] la DMEE combinée avec l'algorithme ACI a été utilisée pour classer les troubles neuromusculaires à partir d'un enregistrement monocanal du signal électromyographique (EMG). Un algorithme adaptatif appelé décomposition moyenne locale qui ressemble dans son principe à la DME a été proposé dans [53] pour démoduler des signaux modulés en amplitude et en fréquence; de même, il a été utilisé dans [54, 55] pour résoudre le problème de SAS.

En se basant essentiellement sur les extrema locaux, la DME conventionnelle se limite à l'analyse des signaux scalaires car la notion d'extrema n'existe pas pour les signaux vectoriels. Ainsi, cette décomposition n'est pas directement applicable aux signaux à valeurs complexes ou bivariés. La décomposition en modes empiriques invariante en rotation (DMEIR) [56] et la décomposition en modes empiriques bivariée (DMEB) [57] sont deux variantes de la DME complexe proposées pour décomposer des signaux à valeurs complexes. Ils ont l'avantage de donner le même nombre de FMIs pour la partie réelle et la partie imaginaire des signaux à valeurs complexes. La décomposition en modes empiriques complexes (DMEC) [58] utilise la relation intrinsèque entre les composantes fréquentielles positives et négatives d'un

## Introduction générale

spectre complexe et la transformée de Hilbert (HT) pour estimer les FMIs. Cependant, cette méthode ne garantit pas le même nombre des FMIs pour la partie réelle et la partie imaginaire. Des extensions de la DME au cas plus général de signaux multivariés ont également été proposées [59, 60].

Dans cette thèse, nous nous intéressons au problème de séparation de mélanges convolutifs. Notre contribution consiste à combiner l'algorithme de la DMEB qui offre le même niveau de décomposition de la partie réelle et la partie imaginaire du signal à décomposer avec l'algorithme ACI-MBE qui exploite à la fois la non-gaussianité et la non-circularité du signal. Le mélange convolutif est d'abord décomposé dans le domaine fréquentiel en un ensemble des FMIs en utilisant l'algorithme DMEB. Ensuite l'algorithme de l'ACI-MBE est appliqué pour l'extraction des sources indépendantes. Pour résoudre le problème de permutation nous avons utilisé l'algorithme récursif proposé par [61] qui consiste à utiliser la solution finale de la matrice de séparation dans la bande de fréquence précédente comme valeur initiale d'itération dans la bande de fréquence courante.

Nous avons consacré une bonne partie du travail à la déconvolution homomorphique. Le spectre d'amplitude du signal de parole voisé peut être considéré comme la combinaison de deux composantes ; une composante qui représente le conduit vocal et une composante qui représente la source glottique. L'originalité de la contribution est l'utilisation de l'algorithme DME comme alternative pour décomposer le signal de parole en ces deux composantes. Les composantes sont ensuite exploitées pour l'estimation conjointe des formants et de la fréquence fondamentale. L'information de phase est exploitée à son tour pour estimer la source glottique.

## **Organisation de la thèse**

La thèse est organisée en quatre chapitres.

Dans le premier chapitre, nous présentons un état de l'art de la séparation aveugle de sources. Nous présentons les différents types de mélanges et les méthodes de séparation conçues pour les signaux à valeurs complexes.

La décomposition en modes empiriques est introduite dans le deuxième chapitre. Dans ce chapitre, nous nous intéressons aux extensions de la DME proposées dans la littérature et conçues pour les signaux à valeurs complexes.

Le troisième chapitre est consacré à notre contribution pour résoudre le problème de déconvolution du signal de parole. La DME est proposée comme alternative pour séparer la composante harmonique et l'enveloppe spectrale du signal de parole. Une procédure adaptative est utilisée dans le domaine log spectral pour sélectionner les fonctions de modes intrinsèques appropriées qui constituent l'amplitude de la source glottique. Cette décomposition est ensuite exploitée pour l'extraction des formants et la fréquence fondamentale.

Le quatrième chapitre est consacré à l'application de l'analyse en composantes indépendantes basée sur la minimisation des bornes de l'entropie combinée avec la DMEB pour la séparation aveugle de mélanges convolutifs des signaux de parole. L'approche proposée est comparée à d'autres algorithmes de références.

# **Chapitre 1**

**La séparation aveugle de sources :**

**État de l'art**



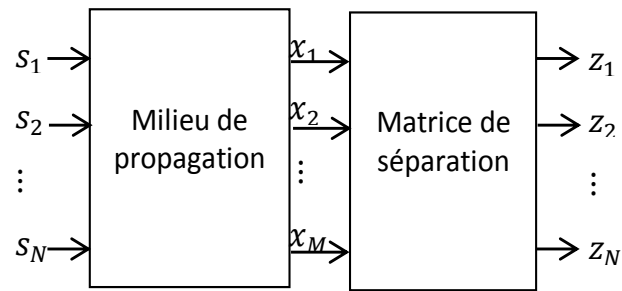
## **1 Introduction**

Depuis que le problème de la SAS a été formalisé, les outils théoriques pour le résoudre n'ont cessé de croître. Ces outils dépendent des hypothèses faites sur la nature des sources et sur le modèle de mélange et de séparation [32-34].

Le problème de la SAS possède plusieurs degrés de difficulté selon la nature du mélange et selon le critère utilisé, et peut être traité dans le domaine temporel ou fréquentiel. Dans ce chapitre nous présentons un état de l'art de la SAS. Nous nous intéressons particulièrement à la séparation des signaux à valeurs complexes. Avant de décrire les méthodes de séparation, nous présentons les fondements de la séparation aveugle de sources. Quelques définitions utiles telles que la gaussianité, l'indépendance, la circularité sont présentées.

## **2 Formulation du problème de la SAS**

Le problème de la SAS représenté par la figure 1.1 peut être décrit comme suit: Soit  $\mathbf{s}(t) = [s_1(t), \dots, s_N(t)]^T$  un vecteur de  $N$  sources indépendantes à l'instant discret  $t$ . Le vecteur  $\mathbf{x}(t) = [x_1(t), \dots, x_M(t)]^T$  de  $M$  mélanges observés est modélisé (dans le cas du mélange instantané) par  $\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t)$ , où  $\mathbf{A}$  est la matrice de mélange de dimension  $(M \times N)$  qui définit le milieu de propagation des sources. L'objectif est de récupérer les  $N$  signaux sources à partir des  $M$  mélanges observés en calculant la matrice de séparation  $\mathbf{W}$  dont la sortie  $\mathbf{z}(t)$  est une estimation du vecteur  $\mathbf{s}(t)$  des signaux sources  $\mathbf{z}(t) = \mathbf{W}\mathbf{x}(t)$ , où  $\mathbf{z}(t) = [z_1(t), \dots, z_N(t)]^T$  et  $\mathbf{W} = [\mathbf{w}_1, \dots, \mathbf{w}_N]^T$ .



**Figure 1.1** Problème de la SAS

La détermination du mélange représente le rapport entre le nombre de sources et le nombre de capteurs. Les mélanges peuvent être classés en trois catégories :

- Les mélanges dits déterminés : si le nombre d'observations  $M$  est égal au nombre de sources  $N$ .
- Les mélanges dits surdéterminés : si le nombre d'observations  $M$  est supérieur au nombre de sources  $N$ .
- Les mélanges dits sous-déterminés : si le nombre d'observations  $M$  est inférieur au nombre de sources  $N$ .

La nature du mélange représente les hypothèses que l'on fait sur l'environnement acoustique dans lequel l'enregistrement est effectué, ou le milieu de propagation. On distingue les types de mélanges suivants :

## 2.1 Mélange instantané

Les sources se mélangent mutuellement selon un gain propre à chaque liaison source-capteur et ce mélange est observé au niveau des différents capteurs. L'observation  $x_i(t)$  du capteur  $i$  à l'instant  $t$  s'écrit sous la forme d'une combinaison linéaire des différentes sources au même instant  $t$  :

$$x_i(t) = \sum_{j=1}^N a_{ij}s_j(t), \quad i = 1, \dots, M \quad (1.1)$$

et peut s'écrire sous forme matricielle comme

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t) \quad (1.2)$$

où  $\mathbf{A}$  est la matrice mélange ayant pour éléments les coefficients de la combinaison linéaire  $a_{ij}$ .

## 2.2 Mélange anéchoïque

Ce type de mélange est une extension du mélange instantané, où le temps d'arrivée des signaux sur les différents capteurs est pris en considération. Le temps d'arrivée ou le délai de retard dépend de la position de chaque source par rapport à chaque capteur. Le mélange de sources est défini par:

$$x_i(t) = \sum_{j=1}^N a_{ij}s_j(t - \delta_{ij}), \quad i = 1, \dots, M \quad (1.3)$$

où  $a_{ij}$  représente les coefficients de la combinaison linéaire et  $\delta_{ij}$  représente le retard.

## 2.3 Mélange convolutif

Le mélange convolutif se produit dans des scénarios acoustiques en raison de la propagation du son dans l'espace et des trajets multiples générés par les réflexions du son par les différents objets, en particulier dans les salles et les environnements fermés. Le modèle de mélange convolutif est donné par

$$x_i(t) = \sum_{j=1}^N \sum_{k=0}^{\infty} a_{ij}(k)s_j(t - k) \quad i = 1, \dots, M \quad (1.4)$$

Toutes les équations reliant les  $M$  observations aux  $N$  sources peuvent s'écrire sous forme matricielle comme :

$$\mathbf{x}(t) = \mathbf{A}(t) * \mathbf{s}(t) \quad (1.5)$$

### 3 Définitions

#### 3.1 Blanchiment

La procédure de blanchiment est tout simplement une décorrélation suivie par un changement d'échelle. Nous soustrayons d'abord la moyenne pour déplacer le centre de localisation vers l'origine, puis nous faisons pivoter les données afin de décorréler conjointement les variables et enfin de redéfinir les variables pour qu'elles présentent des variances unitaires et une matrice de covariance égale à matrice identité.

#### 3.2 Circularité

La notion de circularité et l'importance de cet attribut dans le traitement du signal ont été mises en évidence dans [62]. Ces concepts ont été étendus dans [63] aux vecteurs aléatoires et aux processus aléatoires complexes. La circularité concerne les statistiques d'un élément complexe d'un vecteur  $z$ , c'est-à-dire la relation entre sa partie réelle et sa partie imaginaire. Le degré de non circularité peut être quantifié par la mesure de circularité  $r$  définie comme l'amplitude du quotient de circularité donnée par :

$$\rho(z) = r e^{j\theta} \triangleq \frac{\tau_z^2}{\sigma_z^2} \quad (1.6)$$

où  $\sigma_z^2$  et  $\tau_z^2$  sont respectivement la variance complexe standard formée par les éléments diagonaux des matrices de covariance et la pseudo-variance formée par les éléments diagonaux de la matrice de pseudo-covariance. Le degré de non circularité dans le signal complexe [64] est donné par :

$$r = |\rho(z)| = \frac{|\tau_z^2|}{\sigma_z^2} \quad (1.7)$$

où  $\theta$  est l'angle de circularité donné par :

$$\theta = \arg(\rho(z)) \quad (1.8)$$

L'angle  $\theta$  indique l'orientation de la distribution. Pour un signal purement circulaire,  $r = 0$ , et  $\theta$  ne fournit pas d'informations supplémentaires sur la distribution. D'autres mesures de non-circularité sont également définies et peuvent être utilisées. Les mesures proposées dans [65] sont bornées entre 0 et 1. Une mesure similaire donnée par  $1-r$  a été définie dans [66], le test GLRT (circularity generalized likelihood ratio test) a été aussi proposé. Ce test considère les hypothèses suivantes :

$$\begin{cases} pcov(\mathbf{x}) = 0, \text{ le vecteur est circulaire} \\ pcov(\mathbf{x}) \neq 0, \text{ le vecteur est non circulaire} \end{cases} \quad (1.9)$$

où  $pcov$  est la matrice pseudo-covariance symétrique complexe du vecteur  $\mathbf{x}$ .

Les auteurs dans [67] ont défini la même mesure mais avec une terminologie différente. Enfin, une mesure illimitée dans  $[1, \infty]$  basée sur le rapport des écarts types des composantes réelle et imaginaire de la variable aléatoire complexe a été introduite dans [68]. La mesure de circularité peut également être interprétée graphiquement en utilisant une ellipse (centrée dans le plan complexe) d'excentricité  $\epsilon$  et d'orientation  $\alpha$ , tel que  $r = \epsilon^2$  et  $\theta = 2\alpha$  [8]. Pour  $\epsilon = 0$ , la forme devient un cercle, qui indique également un signal circulaire avec  $r = 0$ , tandis que le cas extrême  $\epsilon = 1$

correspondant à un signal fortement non circulaire avec  $r = 1$ , l'ellipse devient allongée. La pseudo-variance d'une distribution gaussienne complexe générale est alors liée à la forme elliptique par

$$\tau^2 = \epsilon^2 e^{j2\theta} \quad (1.10)$$

### 3.3 Analyse en composantes indépendantes complexes

L'analyse en composantes indépendantes à valeurs complexes est largement utilisée dans un certain nombre d'applications telles que les communications, les radars et le traitement des signaux biomédicaux. En général, on peut obtenir une ACI complexe (ACIC) en exploitant les deux types suivants de propriétés du signal: la non-gaussianité ou la non-circulaire. Mais, tous les algorithmes d'ACIC existants exploitent seulement un des types. Le problème d'ACI dans le domaine complexe suppose que  $N$  sources sont combinées linéairement à travers une matrice de mélange complexe  $\mathbf{A}$  et sont disponibles via le vecteur  $\mathbf{x}$  observé :

$$\mathbf{x} = \mathbf{A}\mathbf{s} \quad (1.11)$$

La matrice de mélange  $\mathbf{A}$  est supposée inversible. L'objectif est de trouver une matrice de séparation  $\mathbf{W}$  de sorte que les sources peuvent être estimées à partir des données observées. Les sources  $\mathbf{s}(t) = [s_1(t), \dots, s_N(t)]^T$  sont supposées non-gaussiennes et mutuellement indépendantes, avec des variances unitaires et des moyennes nulles. En d'autres termes, la matrice de covariance  $E[\mathbf{s}\mathbf{s}^H] = \mathbf{I}$  où  $H$  dénote l'opérateur transposé complexe conjugué. Cependant, aucune hypothèse n'est faite sur la circularité des sources. Dans les méthodes ACI rapide standard [69], les sources sont considérées circulaires, avec une pseudo-covariance  $E[\mathbf{s}\mathbf{s}^T] = \mathbf{0}$ . Il est courant

d'orthogonaliser initialement les données par une transformation  $\mathbf{V}$  de blanchiment telle que

$$\mathbf{x} = \mathbf{V}\check{\mathbf{x}} = \mathbf{V}\mathbf{A}\mathbf{s} \quad (1.12)$$

Le vecteur sources estimées est  $\mathbf{z} = \mathbf{W}^H \mathbf{x}$  avec les composantes  $z_i$  données par :

$$z_i = \mathbf{w}_i^H \mathbf{x}, \quad i = 1, \dots, N \quad (1.13)$$

où  $\mathbf{w}_i$  représente la  $i^{\text{ème}}$  colonne de  $\mathbf{W}$ . La solution optimale une seule composante complexe non nulle avec une amplitude unitaire et une phase inconnue :

$$\mathbf{u}_i = [0, \dots, e^{j\varphi}, 0, \dots, 0]^T \quad (1.14)$$

où  $u_{ij}$  est le  $j^{\text{ème}}$  élément de la colonne du vecteur  $\mathbf{u}$ .

## **4 Hypothèses de base**

### **4.1 Hypothèse 1**

Toutes les procédures de séparation aveugle de sources supposent que, statistiquement, les sources sont mutuellement indépendantes. Une notion plus faible que l'indépendance est la non-corrélation. Il est important de noter que deux variables indépendantes sont forcément non-corrélées. Par contre, deux variables aléatoires non-corrélées ne sont pas forcément indépendantes. Différents critères de dépendance peuvent être trouvés dans la littérature [70, 71] :

#### **4.1.1 Cumulant croisé**

Soient  $\Phi_u(v)$  et  $\psi_u(v)$  la première et la deuxième fonctions caractéristiques d'une variable aléatoire  $u$  définies respectivement par  $\Phi_u(v) = E[e^{iuv}]$  et  $\psi_u(v) = \ln(\Phi_u(v))$ . Les cumulants (d'ordre  $q$ ) de la variable aléatoire  $u$  sont définis par

$$\text{cum}_q(u) = (-j)^q \left. \frac{d^q \psi_u(v)}{dv^q} \right|_{v=0} \quad (1.15)$$

Les cumulants sont bien adaptés pour indiquer la dépendance entre les variables  $u_i$ ,  $i = 1, \dots, p$  du vecteur  $\mathbf{u}$ . En effet, l'indépendance revient à avoir tous les cumulants croisés égaux à zéro.

En réalité, il est impossible d'utiliser les cumulants pour mesurer la dépendance dans le sens absolu, car cela demande d'inspecter une infinité de cumulants.

#### **4.1.2 Divergence de Kullback-Leibler**

Soient  $U$  et  $V$  deux lois de probabilité d'une même variable aléatoire dont les densités de probabilité marginales sont, respectivement,  $p_u(x)$  et  $p_v(x)$ . La divergence de Kullback-Leibler est définie par

$$\delta(p_u, p_v) = \int_{-\infty}^{+\infty} p_u(x) \log \frac{p_u(x)}{p_v(x)} dx \quad (1.16)$$

En utilisant les propriétés de la fonction log, nous démontrons que  $\delta(p_u, p_v) \geq 0$  et qu'elle s'annule si et seulement si  $p_u(x) = p_v(x)$ . Ainsi, la divergence de Kullback-Leibler mesure l'écart entre les deux lois de probabilité.

#### **4.1.3 Information mutuelle**

L'information mutuelle est une mesure de l'information que les membres d'un ensemble de variables aléatoires ont sur les autres variables aléatoires de l'ensemble. Elle peut être exprimée comme une mesure de distance en utilisant la divergence de Kullback-Leibler. L'information mutuelle est toujours positive et elle est nulle si et seulement si les variables aléatoires sont indépendantes.



## 4.2 Hypothèse 2

Parmi les sources, au plus une peut avoir une distribution gaussienne. Intuitivement, la distribution gaussienne est trop simple, car ses cumulants d'ordre supérieur à deux sont nuls. Dans les algorithmes qui utilisent la non-gaussianité, la fonction coût de l'algorithme sera optimisée de telle sorte que les signaux séparés soient aussi non-gaussiens que possible, ce qui les rend aussi indépendants que possible. Par conséquent, cette technique nécessite des méthodes pour mesurer la non-gaussianité des signaux après chaque itération. Les différentes mesures utilisées pour mesurer la non-gaussianité sont :

### 4.2.1 L'entropie différentielle marginale et conjointe

L'entropie différentielle conjointe  $H_{\mathbf{u}}$  est la quantité d'information (scalaire) d'un vecteur aléatoire  $\mathbf{u} = (u_1, \dots, u_k)^T$  de densité de probabilité conjointe  $p_{\mathbf{u}}(\mathbf{u})$  définie par  $H(\mathbf{u}) = - \int_{-\infty}^{+\infty} p_{\mathbf{u}}(\xi) \ln[p_{\mathbf{u}}(\xi)] d\xi$

où  $\xi = (\xi_1, \dots, \xi_k)^T$  est un vecteur de variables d'intégration muettes.

De la même manière, chacune des composantes  $u_i$  du vecteur  $\mathbf{u}$ , de densité de probabilité marginale  $p_{u_i}(u_i)$ , apporte une quantité d'information, quantifiée par l'entropie différentielle marginale  $H(u_i)$  donnée par

$$H(u_i) = - \int_{-\infty}^{+\infty} p_{u_i}(\xi_i) \ln[p_{u_i}(\xi_i)] d\xi_i \quad (1.17)$$

Soient  $\mathbf{u}_g$  et  $\mathbf{u}$  deux vecteurs aléatoires de distributions gaussiennes et quelconques.

L'entropie conjointe  $H(\mathbf{u})$  vérifie l'inégalité :

$$H(u) < H(\mathbf{u}_g) = \frac{1}{2} \ln\{(2\pi e)^n \det(\Sigma_{\mathbf{u}_g})\} \quad (1.18)$$

où  $\Sigma_{\mathbf{u}_g}$  est la matrice de covariance du vecteur  $\mathbf{u}_g$

La définition de l'entropie marginale et conjointe peut être utilisée pour définir une mesure de non-gaussianité d'un vecteur aléatoire de distribution quelconque par rapport à un vecteur de distribution gaussienne.

#### **4.2.2 Néguentropie conjointe**

Pour apporter une information scalaire sur la non-gaussianité, la notion de néguentropie conjointe que nous notons  $neg(\mathbf{u})$  est utilisée. La néguentropie est définie par :  $neg(\mathbf{u}) = H(\mathbf{u}_g) - H(\mathbf{u})$ . De la même manière, on définit la néguentropie marginale entre une variable aléatoire quelconque  $u_i$  et une gaussienne  $u_{gi}$  par  $neg(u_i) = H(u_{gi}) - H(u_i)$ . La néguentropie est une information scalaire, positive et qui s'annule seulement si le vecteur aléatoire suit une distribution gaussienne. Dans le cas contraire, le vecteur aléatoire suit une distribution non gaussienne.

#### **4.2.3 Kurtosis**

Le kurtosis appelé aussi cumulants d'ordre 4 d'une variable aléatoire centrée ( $X = Y - E[Y]$ ) est défini par :

$$kurt(X) = E[X^4] - 3(E[X^2])^2 \quad (1.19)$$

Pour simplifier encore plus la forme du kurtosis, on peut supposer que X a été normalisée pour avoir une variance unitaire. Le kurtosis s'écrit alors

$$kurt(X) = E[X^4] - 3 \quad (1.20)$$

Le kurtosis vaut zéro pour une variable gaussienne et est non nul pour la grande majorité des variables aléatoires. Il est important de noter que le kurtosis peut être positif ou négatif. Les variables aléatoires avec un kurtosis négatif sont appelées sous-gaussiennes et celles avec un kurtosis positif sont appelées sur-gaussiennes. Une

mesure de non-gaussianité est donnée par la valeur absolue du kurtosis. Le carré du kurtosis peut aussi être utilisé. Cette nouvelle mesure est nulle pour une variable gaussienne alors qu'elle est positive pour la plupart des variables non-gaussiennes.

## 5 Ambiguïtés de séparation

Idéalement, les algorithmes de SAS supposent que peu de connaissances est disponible. Ce manque de connaissance conduit à plusieurs ambiguïtés concernant les solutions possibles fournies par un algorithme de résolution du problème de la SAS. Parmi ces ambiguïtés nous citons :

### 5.1 Ambiguïté de permutation

L'ordre des sources estimées peut être différent à des fréquences adjacentes. Ceci est modélisé par la multiplication des sources originales dans le domaine fréquentiel par une matrice de permutation. Si le problème de permutation n'est pas résolu, la reconstruction dans le domaine temporel d'une source donnée contiendra des contributions des autres sources.

Soit le mélange instantané défini par

$$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_M \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1N} \\ a_{21} & a_{22} & \cdots & a_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ a_{M1} & a_{M2} & \cdots & a_{MN} \end{bmatrix} \begin{bmatrix} s_1 \\ s_2 \\ \vdots \\ s_N \end{bmatrix} \quad (1.21)$$

en utilisant la propriété de commutativité des opérations de multiplication et d'addition, nous pouvons écrire :

$$\begin{bmatrix} x_1 \\ \vdots \\ x_M \end{bmatrix} = \begin{bmatrix} a_{12}s_2 \\ \vdots \\ a_{M2}s_2 \end{bmatrix} + \begin{bmatrix} a_{11}s_1 \\ \vdots \\ a_{M1}s_1 \end{bmatrix} + \cdots + \begin{bmatrix} a_{1N}s_N \\ \vdots \\ a_{MN}s_N \end{bmatrix} \quad (1.22)$$

$$= \begin{bmatrix} a_{12} & a_{11} & \cdots & a_{1N} \\ a_{22} & a_{21} & & a_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ a_{M2} & a_{M1} & \cdots & a_{MN} \end{bmatrix} \begin{bmatrix} S_2 \\ S_1 \\ \vdots \\ S_N \end{bmatrix}$$

Par rapport à la matrice de mélange initiale et par rapport au vecteur de sources initial, il y a eu une permutation de la première ligne avec la deuxième et un changement d'ordre entre les sources d'indice 1 et 2, tout en conservant le même vecteur des mélanges. Cette permutation est modélisée par une matrice de permutation  $\mathbf{P}$  qui contient seulement un élément égal à un par ligne et par colonne. En notant la nouvelle matrice de mélange  $\tilde{\mathbf{A}}$  et le nouveau vecteur sources  $\tilde{\mathbf{S}}$ , le vecteur des mélanges s'écrira :

$$\mathbf{X} = \tilde{\mathbf{A}}\tilde{\mathbf{S}} = \mathbf{A}\mathbf{P}\tilde{\mathbf{S}} = \mathbf{A}\mathbf{P}\mathbf{P}^{-1}\mathbf{S} = \mathbf{A}\mathbf{I}\mathbf{S} = \mathbf{A}\mathbf{S} \quad (1.23)$$

où  $\mathbf{I}$  est la matrice identité. La multiplication à droite de la matrice de mélange par une matrice de permutation ne change pas les mélanges mais implique que l'ordre des sources est indéterminé.

## 5.2 Ambiguïté d'échelle

A chaque fréquence  $f$ , la multiplication des sources estimées par des constantes n'affecte le résultat de l'algorithme de séparation que par un filtrage des signaux estimés une fois convertis dans le domaine temporel mais montre que l'amplitude des sources est indéterminée

Prenant par exemple la source 1 ( $i=1$ ), on peut écrire

$$\begin{aligned} x_1 &= s_1 a_1 + s_2 a_2 + \cdots + s_N a_N \\ &= \left(\frac{1}{\alpha_1} s_1\right)(\alpha_1 a_1) + \left(\frac{1}{\alpha_2} s_2\right)(\alpha_2 a_2) + \cdots + \left(\frac{1}{\alpha_N} s_N\right)(\alpha_N a_N) \end{aligned} \quad (1.24)$$

où  $\alpha_i$  représentent des scalaires.

La forme matricielle est donnée comme suite

$$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_M \end{bmatrix} = \begin{bmatrix} \alpha_1 a_{12} & \alpha_2 a_{11} & \dots & \alpha_N a_{1N} \\ \alpha_1 a_{22} & \alpha_2 a_{21} & \dots & \alpha_N a_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_1 a_{M2} & \alpha_2 a_{M1} & \dots & \alpha_N a_{MN} \end{bmatrix} \begin{bmatrix} s_1/\alpha_1 \\ s_2/\alpha_2 \\ \vdots \\ s_N/\alpha_N \end{bmatrix} \quad (1.25)$$

Le vecteur des mélanges s'écrira :

$$X = \tilde{A}\tilde{S} = A\tilde{D}\tilde{S} = ADD^{-1}S = AIS = AS \quad (1.26)$$

où  $D$  représente la matrice d'échelle

## 6 Domaine temporel vs domaine fréquentiel

La SAS de mélanges convolutifs dans le domaine fréquentiel présente un avantage par rapport à la séparation de sources dans le domaine temporel du fait que les algorithmes de séparation sont plus simples et donc plus rapides. La relation de convolution dans le domaine temporel devient un simple produit dans le domaine fréquentiel. Le passage dans le domaine fréquentiel permet d'obtenir un mélange linéaire instantané de la contribution des  $N$  sources dans chaque bande de fréquences considérée, et alors les sources peuvent être séparées dans chacune de ces bandes. Le tableau 1.1 récapitule les avantages et les inconvénients des méthodes SAS temporel et le domaine fréquentiel.

## 7 Méthodes de séparation de mélanges convolutifs

Dans la littérature, les méthodes de résolution du problème de la SAS dans le cas de mélanges convolutifs peuvent être classées selon plusieurs critères tels que le domaine d'application temporel ou fréquentiel ou selon le principe utilisé à savoir les statistiques d'ordre 2 ou les statistiques d'ordres supérieurs ; chaque principe est

défini à son tour sur plusieurs bases, à savoir, moment croisé non-linéaire, non stationnarité, non-gaussianité, hypothèse de minimum de phase, etc.

**Tableau 1.1** Comparaison entre les méthodes de SAS [72]

Domaine temporel		Domaine fréquentiel	
Avantages	Inconvénients	Avantages	Inconvénients
L'hypothèse d'indépendance est meilleure.	Dégradation de la convergence dans un environnement réverbérant	Le mélange convolutif peut être transformé en mélange instantané pour chaque bande de fréquence	Pour chaque bande de fréquence, il existe une permutation et une ambiguïté de mise à l'échelle qui doivent être effectuées.
	De nombreux paramètres doivent être ajustés pour chaque étape d'itération	En raison de l'utilisation de la FFT, les calculs sont moins complexes rapport à une implémentation dans le domaine temporel La convergence est plus rapide	Le problème avec trop peu d'échantillons dans chaque bande de fréquence peut entraîner l'échec de l'hypothèse d'indépendance

Dans cette section, nous présentons un panorama des méthodes existantes dans la littérature classées selon le critère du domaine d'application.

Les premiers travaux sur la séparation de mélanges convolutifs ont été réalisés dans le domaine temporel. Dans [73] la procédure de séparation de source est modélisée comme étant un filtrage à réponse impulsionnelle fini (RIF) des mélanges

et a utilisé une approche de maximisation de l'information. Dans [74] la procédure de SAS a été modélisée comme un filtrage a réponse impulsionnelle infini (RII).

Les méthodes de séparation de sources opérant dans domaine fréquentiel divisent l'ensemble des fréquences en plusieurs bandes. Le problème de SAS de mélanges convolutifs est transformé alors en un problème de séparation de mélange instantané dans chaque bande de fréquences [75]. Dans [76, 77], les auteurs se sont intéressés au problème de séparation des signaux acoustiques simultanément enregistrés dans un environnement réverbérant. La méthode de séparation aveugle de mélanges convolutifs proposée prend en considération la non-stationnarité des signaux. En effet, cette propriété permet de construire un ensemble suffisant d'informations conduisant à la séparation. Les sources sont reconstruites par un filtrage inverse RIF. Comme la majorité des algorithmes de séparation de mélanges convolutifs, cet algorithme transforme le problème dans le domaine fréquentiel en un ensemble de mélanges instantanées, la séparation est effectuée fréquence par fréquence. En se basant sur le même principe, les auteurs dans [78], on résolu le problème de séparation de mélanges convolutifs par la diagonalisation conjointe des matrices de covariance. Dans [79] les auteurs ont suggéré un complément à la méthode proposée dans [75]. Leur contribution concerne le problème de SAS audio mélangées dans un contexte réel. Pour renforcer sa contrainte d'anti-permutation qu'ils qualifient d'insuffisante, les auteurs ont adopté le principe de l'ACI dans le domaine fréquentiel, et ils ont proposé un algorithme basé sur le passage au domaine fréquentiel. Dans [80], l'auteur a utilisé une version modifiée de l'algorithme proposé dans [75] dans laquelle une étape de prétraitement du bruit a été utilisée. Selon son expérience dans l'application des algorithmes de séparation de sources aux

environnements réels, l'auteur conclut que les algorithmes basés sur l'infomax sont les mieux adaptés à la séparation des sons acoustiquement mélangés.

La résolution du problème de SAS dans le domaine fréquentiel exige une solution au problème de permutation. Les méthodes de résolution du problème de permutation pour la SAS de mélanges convolutifs peuvent être classées en 4 catégories [72], à savoir :

- i. Les méthodes basées sur la cohérence des coefficients de filtrage, ce sont les méthodes qui exploitent le lissage du spectre [37], l'emplacement des sources [81], le modèle de la directivité [82], la localisation des vecteurs [83], la direction d'arrivée [84] et la distance de la matrice adjacente [85].
- ii. Les méthodes basées sur la cohérence du spectre du signal récupéré exploitent la fréquence fondamentale [86], la périodicité des signaux [87], la corrélation croisée [88], les cumulants croisés [89], le kurtosis [92] et la classification [91].
- iii. Une classe de méthodes utilise l'information temps-fréquence en se servant d'une fonction coût [92] ou par application de l'ACI au spectrogramme complet [93].
- iv. Les approches combinées telles que [94, 95].

## **8 Méthodes de séparation des signaux à valeurs complexes**

Dans de nombreuses applications pratiques telles que le traitement des signaux audio dans le domaine fréquentiel, les signaux sont complexes et, par conséquent, nous avons besoin d'algorithmes pour résoudre le problème de la SAS à valeurs complexes.



## **8.1 Méthodes basées sur la non-circularité**

L'algorithme de transformation non corrélative généralisée [96] est une généralisation de la transformée fortement corrélative [97, 98]. Il permet d'utiliser des fonctions plus robustes pour générer les matrices de dispersion, mais les sources doivent avoir des spectres distincts comme dans l'algorithme de la transformée fortement corrélative. Tout vecteur aléatoire complexe  $\mathbf{x}$  peut être transformé en utilisant une matrice carrée non singulière  $\mathbf{C}^{-1}$  telle que

$$\mathbf{s} = [s_1, \dots, s_N]^T = \mathbf{C}^{-1}\mathbf{x} \quad (1.27)$$

où la matrice  $\mathbf{C}^{-1}$  représente la transformée de contrainte forte [97, 98]. Qui est une extension de la transformation de blanchiment pour des vecteurs aléatoires complexes. Elle consiste à effectuer une transformation linéaire basée sur la factorisation de Takagis qui diagonalise à la fois la matrice de covariance et la matrice de pseudo-covariance [97,98]. Ces méthodes utilisent uniquement la non-circularité en se servant des statistiques de second ordre.

## **8.2 Méthodes de séparation basées sur la non-gaussianité**

Ce sont les méthodes qui exploitent uniquement la non-gaussianité en supposant que la source est circulaire explicitement ou implicitement, parmi ces méthodes, on distingue :

La diagonalisation conjointe des matrices propres [99].

Elle est basée sur l'utilisation de cumulants d'ordre quatre impliquant à la fois les auto-cumulants et les cumulants croisés. La fonction coût est donnée par :

$$J(y) = \sum_{ijkl \neq iikl} |\text{cum}\{y_i, y_j, y_k, y_l\}|^2 \quad (1.28)$$

## Chapitre 1 : La séparation aveugle de sources ; Etat de l'art

Maximiser cette fonction revient à diagonaliser conjointement un ensemble de matrices. L'algorithme de diagonalisation conjointe nécessite aussi un blanchiment des observations. Une solution algébrique consiste à diagonaliser conjointement l'ensemble des matrices propres construites à partir des vecteurs propres associés aux plus grandes valeurs propres de la matrice de quadri-covariance. Les méthodes qui reposent sur la diagonalisation des matrices cumulatives sont robustes. Cependant, leur performance se dégrade lorsque le nombre de sources augmente, et le temps de calcul de la diagonalisation des matrices cumulants pourrait devenir prohibitif pour séparer un grand nombre de sources.

### Analyse en composantes indépendantes complexes rapide

L'algorithme d'analyse en composantes indépendantes complexes (ACIC) rapide est l'une des méthodes les plus populaires pour l'exécution de l'ACI lorsqu'il s'agit de sources à valeurs complexes [100]. Contrairement aux études effectuées dans [40, 101], l'ACIC utilise des non linéarités générales moins sensibles aux valeurs aberrantes. Dans [102], les auteurs ont présenté un algorithme à point fixe et ont donné les conditions de stabilité locale en supposant des sources circulaires, c'est-à-dire que la fonction coût utilise le module du signal. Cependant limiter l'analyse aux sources circulaires, aboutit à un algorithme à point fixe qui ne fonctionne pas bien avec des sources non circulaires [40]. Il est important de noter que les fonctions coût choisies pour l'algorithme d'analyse en composantes indépendantes complexes rapide sont basées sur la robustesse et la densité car l'utilisation du module du signal limite déjà la fonction coût aux sources circulaires. Autre algorithmes utilisant des fonctions non linéaires ont été présentés dans [37, 103].

### **8.3 Méthodes de séparation basées sur la non-circularité et la non-gaussianité**

Ce sont des méthodes qui prennent en compte à la fois la non-gaussianité et la non-circularité.

#### Analyse en composantes indépendantes rapide non circulaire.

Cette méthode étend l'algorithme d'analyse en composantes indépendantes rapide [104] en exploitant les informations de second ordre contenues dans la matrice de pseudo-covariance dans l'optimisation à point fixe et en utilisant les mêmes fonctions que l'algorithme d'analyse en composantes indépendantes rapide, mais en modifiant la procédure d'optimisation pour tenir compte des sources non circulaires. L'ajout de l'information dans la matrice de pseudo-covariance à l'algorithme à point fixe affecte également le paysage de la surface d'optimisation de manière à garantir la convergence. Ensuite, une analyse rigoureuse de la stabilité locale est effectuée pour quantifier les effets de la non-circularité sur la performance. La fonction coût de l'algorithme d'analyse en composantes indépendantes rapide et l'algorithme d'analyse en composantes indépendantes rapide non-circulaire est donnée par :

$$J(w) = E\{G(|\mathbf{w}^H \mathbf{x}|^2)\} \quad (1.29)$$

où  $G: \mathbb{R} \rightarrow \mathbb{R}$  est une fonction uniforme et lisse et  $w \in \mathbb{C}^N$  avec  $\|w\| = 1$ . La fonction coût donnée n'utilise pas la phase des sources indiquant que toute information de non-circularité est perdue comme c'est le cas pour l'algorithme d'analyse en composantes indépendantes rapide. Les informations de phase sont récupérées dans la mise à jour du point fixe de l'algorithme d'analyse en composantes indépendantes rapide non-circulaire par l'utilisation explicite de la matrice de pseudo-covariance.

Algorithme complexe à points fixes.

Une autre mise à jour du point fixe pour l'ACI complexe, proposée par les auteurs dans [40], est l'algorithme ACI rapide à point fixe basé sur la fonction coût du kurtosis qui utilise la forte transformation non corrélative; Aucune hypothèse de circularité n'est nécessaire car les matrices de covariance et de pseudo-covariance sont diagonalisées en utilisant la forte transformation non corrélative plutôt que le blanchiment conventionnel de la matrice de covariance seule.

Maximisation du kurtosis.

Dans [105], l'auteur a étudié des algorithmes de séparation de sources à valeurs complexes basés sur le kurtosis en utilisant la méthode d'optimisation du gradient et de Newton. L'algorithme a été conçu pour des sources à valeurs complexes et a ainsi surpassé l'algorithme d'analyse en composantes indépendantes complexes rapide avec la non-linéarité basée sur le kurtosis.

Minimisation de l'entropie [106]

L'algorithme ACI-MBE complexe utilise le principe de la minimisation de l'information mutuelle pour effectuer la séparation des sources. L'ACI-MBE complexe estime la limite la plus étroite de l'entropie. L'ACI-MBE complexe permet obtenir une estimation fiable des bornes de l'entropie en résolvant la distribution d'entropie maximale qui maximise l'entropie sous certaines contraintes. Parmi toutes les estimations de l'entropie, seul le minimum est utilisé comme estimation finale de l'entropie. Autre algorithmes utilisant des fonctions non linéaires ont été présentés par [107,108,39]

## **9 Conclusion**

Dans ce chapitre, nous avons présenté une description des modèles de mélange et des hypothèses nécessaires pour résoudre le problème de la séparation de mélanges convolutifs. Dans cette thèse, nous optons pour la séparation de mélange des signaux audio, nous pensons donc judicieux de nous concentrer sur les méthodes de SAS conçu pour les signaux à valeurs complexes en se basant sur les méthodes exploitant la non-circularité et la non-gaussianité des signaux.

## **Chapitre 2**

### **Décomposition en modes empiriques**

## 1 Introduction

Dans ce chapitre, nous présentons le formalisme de la décomposition en modes empiriques (DME). La DME a été proposée dans [42] pour l'analyse des signaux non stationnaires et/ou non linéaires. Elle décompose d'une façon adaptative un signal en une somme de composantes oscillantes. La base de décomposition de la DME est intrinsèque au signal contrairement aux méthodes conventionnelles telles que la transformée de Fourier (TF) et la transformée en ondelettes (TO). Nous nous sommes intéressés aux extensions de la DME proposées dans la littérature pour l'analyse des signaux à valeurs complexes.

## 2 Méthodes conventionnelles de décomposition des signaux

### 2.1 Transformée de Fourier

Le principe fondateur de l'analyse des signaux dans le domaine fréquentiel est l'analyse de Fourier largement répandue. La transformée de Fourier d'un signal  $x(t)$  est définie par :

$$X(f) = TF[x(t)] = \int_{-\infty}^{+\infty} x(t)e^{-j2\pi ft} dt \quad (2.1)$$

La TF présente des limitations dans le cas des signaux non stationnaires où on perd toute information relative au temps. Pour remédier à la limitation de la TF, la transformation de Fourier à courte terme (TFCT) a été introduite. La TFCT d'un signal  $x(t)$  est donnée par :

$$X(\tau, f) = \int_{-\infty}^{+\infty} x(t)h^*(\tau - t)e^{-j2\pi ft} dt \quad (2.2)$$

où  $h$  est une fenêtre glissante. L'inconvénient principal de cette méthode est le compromis entre la résolution fréquentielle et la résolution temporelle.

## 2.2 Transformée en ondelettes

La transformée en ondelette peut être définie comme étant la projection sur la base des fonctions d'ondelette. L'approche par ondelettes est essentiellement une analyse spectrale de Fourier à fenêtre réglable, avec la définition générale suivante : Une fonction intégrable  $\psi$  est dite fonction d'ondelette de base si elle satisfait la condition de moment nul:

$$\int_{-\infty}^{\infty} \psi(t) dt = 0 \quad (2.3)$$

La condition de moment nul est valide lorsque

$$\int_{-\infty}^{\infty} \frac{|\hat{\psi}(\omega)|^2}{|\omega|} d\omega < \infty \quad (2.4)$$

sachant que  $\psi(t)$  représente l'ondelette mère et  $\hat{\psi}(\omega)$  sa transformée de Fourier. Le sous-espace d'échelle  $a$  et du décalage  $b$  est engendré par changement d'échelle et décalage de  $\psi(t)$  :

$$\psi_{a,b}(t) = |a|^{-\frac{1}{2}} \psi\left(\frac{t-b}{a}\right) \quad (2.5)$$

où  $a, b \in \mathbf{R}$  et  $a \neq 0$

La transformée en ondelettes continue est définie par :



$$W_{\psi}x(a, b) = \langle x, \psi_{a,b} \rangle = |a|^{-1/2} \int_{-\infty}^{\infty} x(t) \psi^*\left(\frac{t-b}{a}\right) dt \quad (2.6)$$

L'ondelette mère doit satisfaire un critère d'admissibilité pour obtenir une transformée stablement inversible. Les paramètres  $a$ ,  $b$  couvrent différentes plages de fréquences lorsque  $b$  change. Les grandes valeurs du paramètre de dilatation  $a$  correspondent aux basses fréquences et les petites valeurs de  $a$  correspondent aux hautes fréquences. Le centre de localisation temporelle peut être déplacé en changeant le paramètre de translation  $b$ . Par conséquent, la transformée en ondelettes offre une description temps-fréquence parfaite du signal  $x(t)$ .

En traitement numérique du signal ou d'image, les ondelettes continues doivent être discrétisées. Par conséquent, le paramètre de dilatation et le paramètre de translation ne prennent que des valeurs discrètes. La version discrète de la TO est donnée par :

$$\psi_{i,j}(t) = 2^{-i/2} \psi(2^{-i}t - j) \quad (2.7)$$

et

$$DW_x(i, j) = \langle x(t), \psi_{i,j}(t) \rangle = \sum_i \sum_j 2^{-i/2} x(t) \psi(2^{-i}t - j) \quad (2.8)$$

### **2.3 Décomposition par transformée en ondelettes**

La décomposition en ondelettes est un processus qui analyse un signal avec les filtres de décomposition passe-bas et passe-haut associés à l'ondelette de décomposition et sa fonction de mise à l'échelle. Une description simple de la décomposition en ondelettes est la suivante:

Niveau 0 :  $a_0(n) = x(n)$

Niveau 1 :  $a_1(n) = a_0(n) * h_1(n)$  et  $d_1(n) = a_0(n) * g_1(n)$

Niveau 2 :  $a_2(n) = a_1(n) * h_2(n)$  et  $d_2(n) = a_1(n) * g_2(n)$

Niveau K :  $a_K(n) = a_{K-1}(n) * h_K(n)$  et  $d_K(n) = a_{K-1}(n) * g_K(n)$

où  $a_k(n)$  et  $d_k(n)$  désignent respectivement l'approximations et le détail du signal dans la décomposition en ondelettes de niveau k, respectivement.  $h_k(n)$  est le filtre passe-bas du niveau k,  $g_k(n)$  est le filtre passe-haut de niveau k, avec  $k \in [1, K]$ . Le symbole \* dénote la convolution.

La transformée en ondelettes nécessite quatre filtres, deux filtres passe-bas et deux filtres passe-haut, outre les  $h_k(n)$  et  $g_k(n)$  pour la décomposition, il existe une autre paire de filtres :  $H_k(n)$  (filtre passe-bas) et  $G_k(n)$  (filtres passe-haut) pour la reconstruction.

La décomposition en ondelettes possède de nombreuses applications et propriétés efficaces dans le traitement des signaux et des images. Cependant, pour l'analyse des signaux non linéaires et non stationnaires, l'ondelette présente encore certaines limitations. Par exemple, une fois que l'ondelette de base est choisie, il faudra l'utiliser pour analyser toutes les données, ce qui signifie qu'elle est non adaptative. Pour définir un changement local, il faut rechercher le résultat dans la gamme des hautes fréquences, car plus la fréquence est élevée, plus l'ondelette de base est localisée. Si un événement local se produit uniquement dans la plage des basses fréquences, on sera toujours obligé de rechercher ses effets dans la gamme des hautes fréquences.

### **3 Décomposition modale empirique**

La décomposition en modes empiriques (DME) est entièrement axée sur les données, elle est conçue pour la décomposition à plusieurs échelles et l'analyse temps-fréquence des signaux du monde réel. Le signal d'origine est modélisé comme une combinaison linéaire d'oscillations appelées fonctions de modes intrinsèques (FMIs). Les FMIs sont définies de manière à localiser le signal dans le temps et à représenter un seul mode oscillatoire; En raison d'hypothèses concernant la nature des données, la DME a donné lieu à de nombreux travaux aussi bien en théorie [43] qu'en application telles que le débruitage [44], l'analyse temps-fréquence [45], et la détection radar [46, 47]. L'algorithme de DME est caractérisé par une décomposition dans une base non fixée du signal ; en ce sens, la DME elle est auto-adaptative et fondamentalement algorithmique [109].

#### **3.1 Principe**

La DME considère un signal à l'échelle de ses oscillations locales. L'idée principale de la DME est alors de considérer, localement le signal est la superposition d'oscillations rapides et d'oscillations lentes. La DME est conçue pour définir une composante locale "basse fréquence" comme un résidu (tendance locale)  $e_1(t)$ , en soutenant une composante locale de "haute fréquence" comme oscillation moyenne nulle ou détail local  $d_1(t)$ , afin que nous puissions exprimer  $x(t)$  comme

$$x(t) = d_1(t) + e_1(t) \quad (2.9)$$

où  $d_1(t)$  est un signal oscillant et, s'il est localement de moyenne nulle, il correspond à ce qu'on appelle une fonction de mode intrinsèque.

Pratiquement, cela implique principalement que tous ses maxima sont positifs et que tous ses minima sont négatifs. Sachant que  $e_1(t)$  oscille localement plus lentement que  $d_1(t)$ . En appliquant la même décomposition, ce qui conduit à  $d_1(t) = e_2(t) + d_2(t)$  et, en appliquant récursivement la même décomposition sur  $d_k(t)$ , on obtient une représentation de  $x(t)$  sous forme d'une somme de FMIs et d'un résidu.

### 3.2 Processus de tamisage

Les différentes étapes du processus de tamisage sont exprimées comme suit :

**Etape 1** : initialisation  $j \leftarrow 1$  ( $j^{\text{ième}}$  FMI),  $r_j(t) \leftarrow x(t)$  (résidu).

**Etape 2** : extraction de la  $j^{\text{ième}}$  FMI :

- (a) Identification de tous les extrema de  $r_j(t)$ .
- (b) Interpolation entre les minima (resp. maxima) de  $r_j(t)$  conduisant à une enveloppe  $e_{\min}(t)$  (resp.  $e_{\max}(t)$ ).
- (c) Calcul de l'enveloppe moyenne  $\mu(t) = (e_{\min}(t) + e_{\max}(t))/2$
- (d) Extraction du détail  $d(t) = r_j(t) - \mu(t)$
- (e) Itération sur le détail  $d(t)$  en répétant les étapes de (a) à (d) jusqu'à ce qu'un critère d'arrêt basé sur la déviation standard entre deux détails consécutifs soit inférieur à un seuil prédéfini conduisant à  $FMI_j$

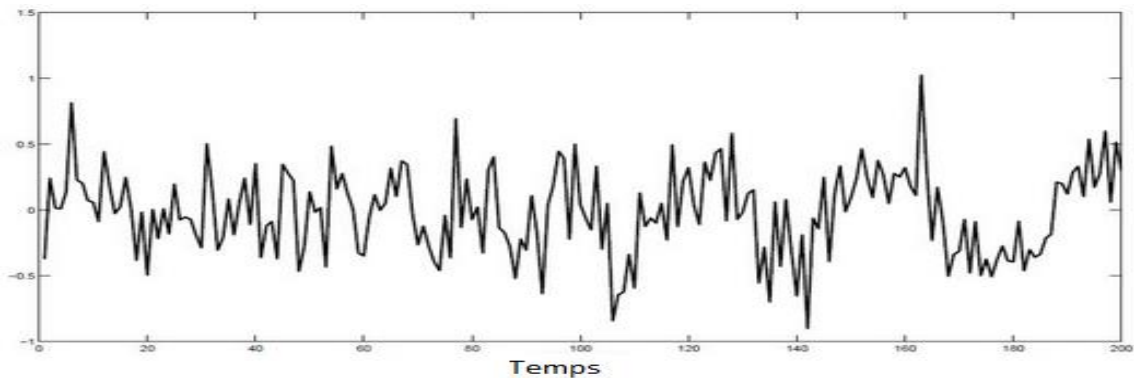
**Etape 3** : mise à jour du résidu  $j \leftarrow j + 1$ ,  $r_j(t) \leftarrow r_{j-1}(t) - FMI_j(t)$ .

**Etape 4** : itération sur le résidu en répétant les étapes 2 et 3 jusqu'à ce que le nombre d'extrema de  $r_j(t)$  soit inférieur à 2.

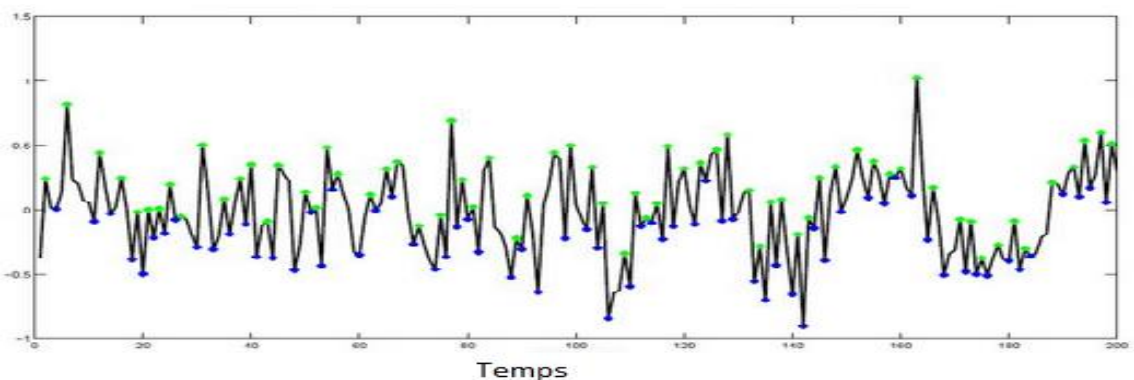
Le processus de reconstruction du signal est donné par :

$$x(t) = \sum_{j=1}^K FMI_j(t) + r_K(t), \quad K \in N^* \quad (2.10)$$

Pour mieux illustré l'algorithme de DME, soit un signal à décomposer défini par la figure 2.1. La figure 2.2 illustre les différentes étapes du processus de tamisage. La décomposition, réalisée sans aucune connaissance préalable des caractéristiques du signal, permet d'extraire correctement les contributions ainsi que la tendance (résidu) du signal.

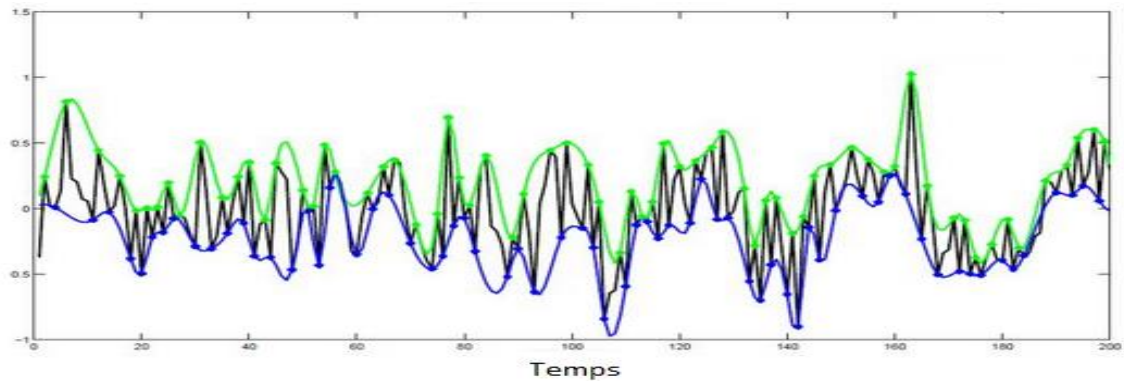


**Figure 2.1** Exemple de décomposition en modes empiriques.

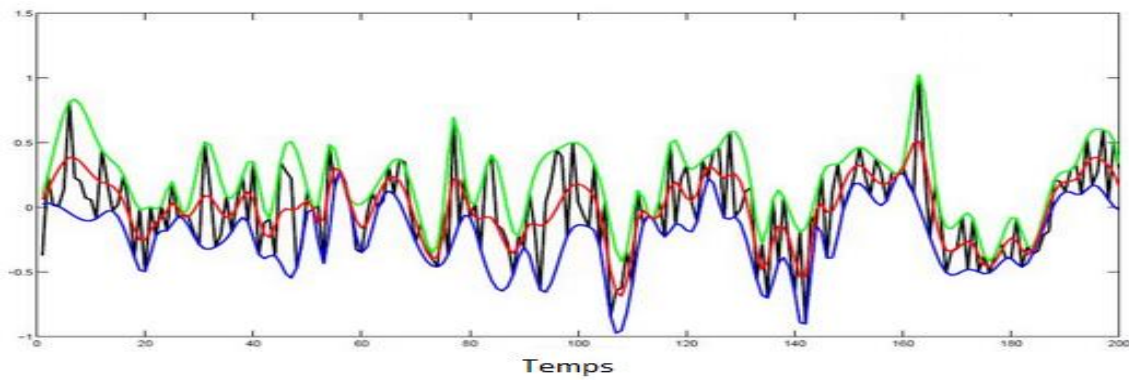


(a) Maximas et minimas

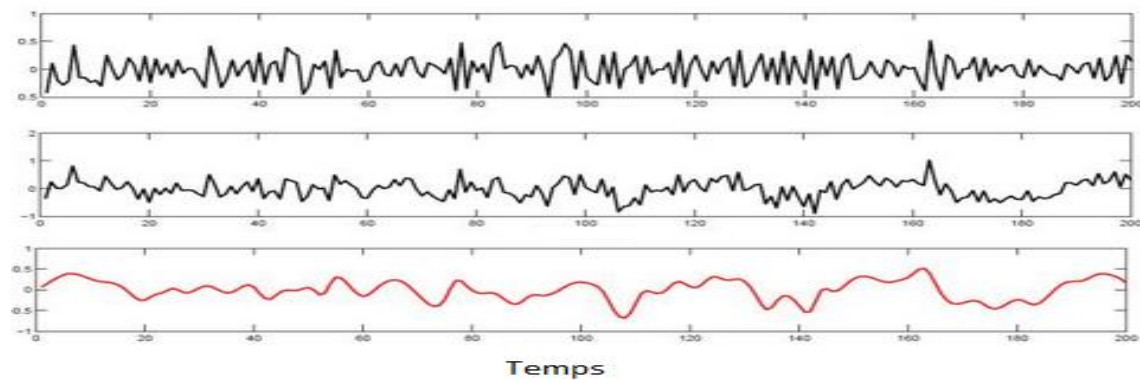
**Figure 2.2** Etapes d'extraction des FMIs par processus de tamisage



(a) Enveloppes supérieure et inférieure.

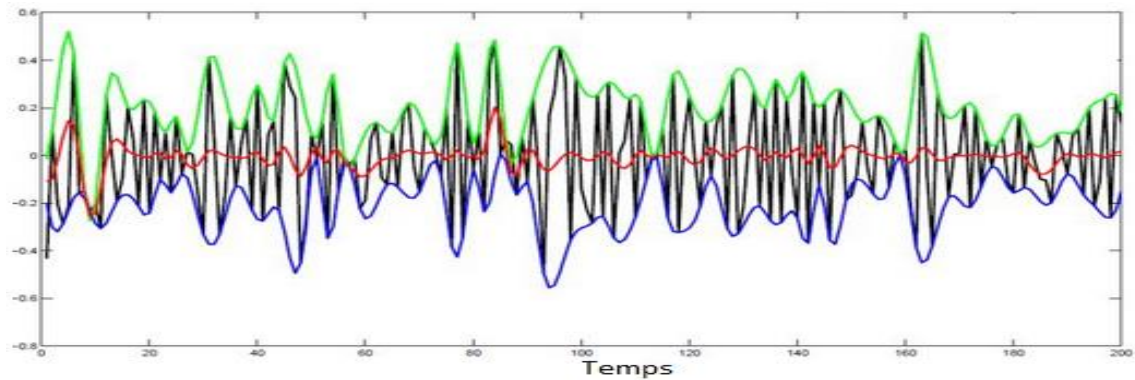


(c) Moyenne des enveloppes.

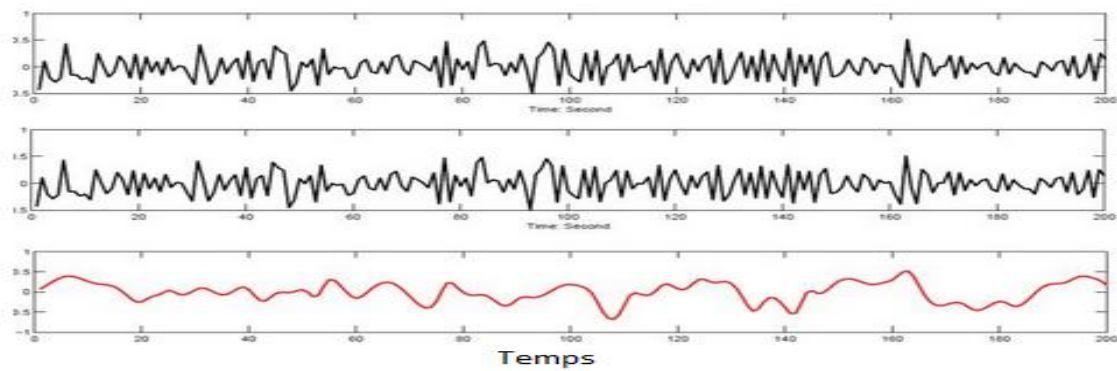


(d)  $FMI_1$  et résidu, itération 0, moyenne non nulle.

**Figure 2.2** (suite)



(e) Itération 1, enveloppe supérieure, enveloppe inférieure et la moyenne des enveloppes



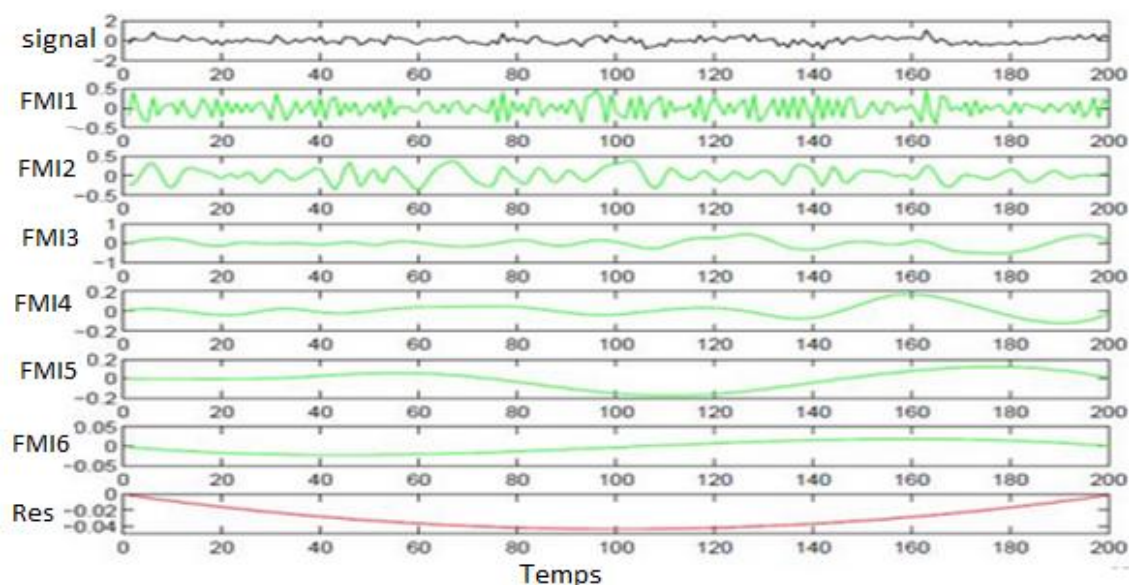
(f)  $FMI_1$ , moyenne non nulle.

**Figure 2.2** (Suite)

L'exécution d'algorithme permet d'avoir 6 FMIs et un résidu comme le montre la figure 2.3

**3.3 Définition et caractéristiques de la fonction de mode intrinsèque**

Une fonction de mode intrinsèque est une fonction oscillante autour de 0, de moyenne locale nulle, modulée en amplitude et en fréquence [109]. Elle possède les caractéristiques suivantes :



**Figure 2.3** Décomposition modale empirique du signal de la figure 2.1

1. De moyenne nulle.
2. Les nombres d'extrema et de passages à zéro diffèrent au plus de un (en d'autres termes, cela signifie qu'entre un minimum et un maximum successifs, une FMI passe par zéro).
3. Elle suit une loi de modulation en amplitude et en fréquence (comportement oscillant) naturellement de type mono-composante.

## 4 Autour de la méthode

### 4.1 Critère d'arrêt du processus de tamisage

Dans les travaux publiés dans la littérature, il apparaît que trop de tamisage nuit au sens physique des modes obtenus, mais que trop peu de tamisage ne permet pas de bien séparer les modes. Ainsi, un critère d'arrêt  $SD(i)$  reposant sur la déviation standard a été proposé il est défini par:



$$SD(i) = \sum_{t=0}^T \frac{|d_{i-1}(t) - d_i(t)|^2}{(d_{i-1}(t))^2} < Q \quad (2.11)$$

La valeur de  $Q$  permettant d'arrêter le tamisage est comprise entre 0.2 et 0.3 [109]. Cette valeur est un compromis et elle est définie de manière empirique. Si  $Q$  est trop grand, alors la méthode ne permet pas de séparer correctement les différents modes présents dans le signal; inversement, si  $Q$  est trop petit, alors le principe risque d'aboutir à des composantes d'amplitudes quasi constantes, modulées par une fréquence unique [109]. D'autres critères d'arrêt ont été proposés, mais ils sont tous basés sur le même concept :  $d_{j,i}(t)$  vérifie bien les propriétés d'une FMI [109].

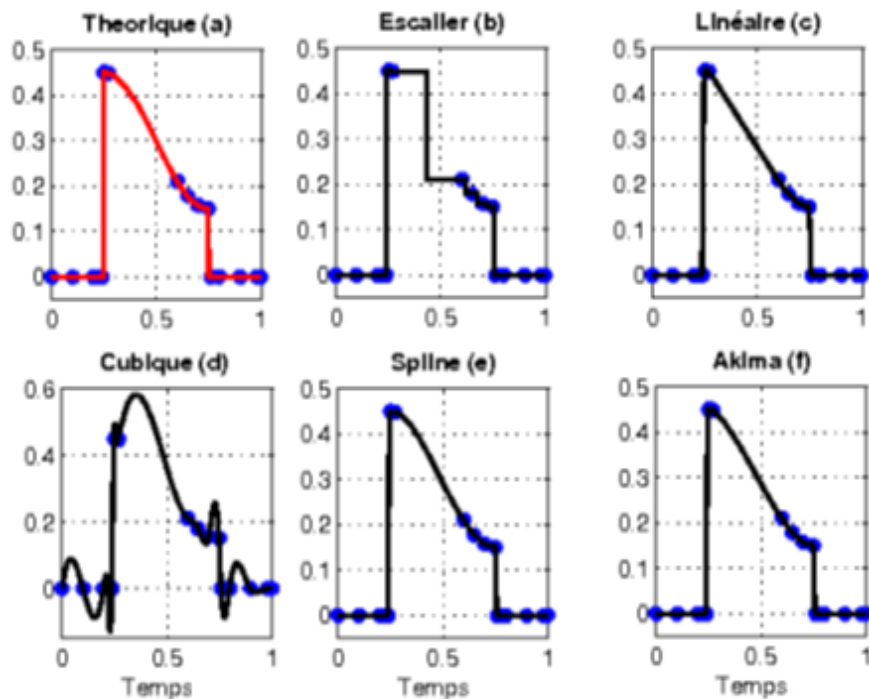
## **4.2 Interpolation**

L'interpolation des enveloppes supérieure et inférieure nécessaire pour estimer les FMIs est un problème non négligeable. De nombreuses approches ont été proposées pour répondre au problème d'interpolation comme le montre la figure 2.4. Les approches 'escalier' et 'linéaire' ne sont pas recommandées pour l'estimation des FMIs. Ces deux méthodes aboutissent à un nombre trop important de modes. L'interpolation par splines cubiques est privilégiée. Elle présente un bon compromis entre la qualité d'interpolation (et non l'approximation) et la simplicité du calcul.

## **4.3 Effet des bords**

De nombreuses approches ont été proposées dans la littérature pour remédier aux problèmes des bords. Parmi ces méthodes, des techniques proposent le prolongement par symétrie miroir des extrema par rapport au dernier échantillon, cependant même avec un bon prolongement on peut créer un autre problème de bord.

Une autre technique proposée, consiste à considérer simultanément les bords comme étant des minimas et des maximas ce qui rend les effets de bords négligeables.



**Figure 2.4** Illustration de l'interpolation par différentes méthodes [109]

## 5 Extension aux signaux à valeurs complexes

De nombreux domaines de traitement du signal utilisent des structures de données à valeurs complexes. La DME dans sa forme classique n'est pas conçue aux signaux à valeurs complexes, et se limite uniquement à l'analyse de données à valeurs réelles. Ainsi, plusieurs extensions de la DME ont été proposées dans la littérature.

### 5.1 Décomposition en modes empiriques complexes

Les auteurs dans [58] ont employé le concept de signal analytique et par la suite, ils ont appliqué la DME standard pour analyser des données complexes ou bivariées; la DME complexe applique la DME univariée séparément aux parties

réelles et imaginaires du signal, ne garantissant pas ainsi le même nombre de FMIs, ce qui est une exigence majeure dans les applications réelles.

## **5.2 Décomposition en modes empiriques invariantes en rotation**

Dans [110], une extension (complexe) entièrement bivariée de la DME est proposée. La méthode opère en prenant des projections du signal d'entrée bivarié le long de deux directions dans le plan complexe pour calculer la moyenne locale. Les enveloppes du signal bivarié sont calculées en interpolant les enveloppes de ces projections univariées (valeurs réelles), et la moyenne locale d'un signal complexe est déterminée en prenant la moyenne des enveloppes. Bien que la décomposition en modes empiriques invariante en rotation (DME-IR) donne le même nombre de FMIs pour les deux composantes du signal, elle n'est pas bien adaptée pour traiter la dynamique des signaux rapides en raison du faible nombre de projections des signaux, ce qui limite son intérêt pratique.

## **5.3 Décomposition en modes empiriques bivariée**

### **5.3.1 Principe**

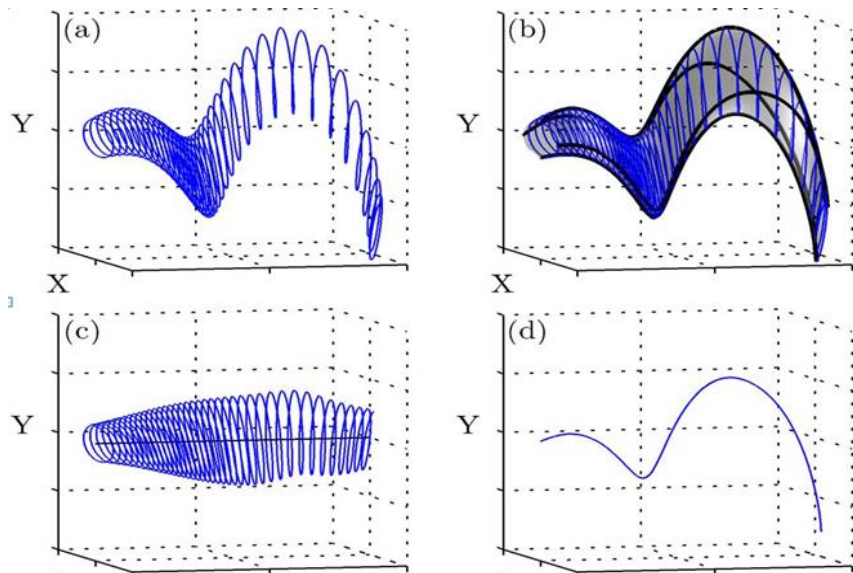
L'algorithme de décomposition en modes empiriques bivariée (DMEB) [57] utilise plusieurs projections (univariées) d'un signal complexe (bivarié) pour déterminer la moyenne locale. Contrairement à la DME-RI, les multiples directions de projection de la DMEB peuvent échantillonner le plan complexe (bivarié) améliorant ainsi la précision et le rendant beaucoup mieux adapté pour les signaux ayant une dynamique de changement rapide que la DME-IR.

La DME s'appuie fortement sur la notion d'oscillation qu'elle définit à partir des extrema locaux du signal. Une oscillation correspond par exemple au signal entre deux maxima locaux. De ce fait, elle est naturellement confinée à l'analyse des signaux scalaires puisque la notion d'extremum local, et donc d'oscillation, n'existe pas pour les signaux vectoriels. En revanche, dans le cas des signaux à deux composantes ou bivariés, on peut considérer la notion de rotation qui est de plus une extension bivariée classique de la notion d'oscillation. Ainsi, le principe sous-jacent à toutes les extensions bivariées proposées [45-47] jusqu'à ce jour est de remplacer dans la DME usuelle la notion d'oscillation par celle de rotation. Le point de vue adopté est alors de considérer que tout signal bivarié peut être décrit comme étant la somme d'une composante (vectorielle) tournant rapidement autour de zéro et d'un autre tournant plus lentement.

### ***5.3.2 Une enveloppe en 3 dimensions***

Techniquement, les extensions bivariées reprennent le principe récursif de la DME mais avec de nouveaux opérateurs de tamisage calqués sur l'original. Ainsi, là où classiquement on soustrait au signal la moyenne des enveloppes supérieures et inférieures, les nouveaux opérateurs soustraient au signal l'« axe central » de son « enveloppe », qui est maintenant un tube tridimensionnel qui enveloppe le signal. En pratique, cet axe est en fait calculé à partir de ce qu'on pourrait appeler des « armatures latérales » du tube de la figure 2.5 (b). Chacune de ces armatures est associée à une direction particulière et forme une ligne de crête de l'enveloppe dans cette direction. Ainsi, l'armature associée à la direction « vers le haut » pourrait être définie à tout instant comme le point de la section du tube dont l'altitude est maximale. Cependant, l'enveloppe n'étant pas connue a priori, les armatures sont

calculées différemment. Dans l'esprit de la Figure 2.5, on peut considérer que le tube enveloppe devrait être en tout point tangent au signal, ce qui implique que chacune des armatures latérales interpole un certain ensemble de points du signal.



**Figure 2.5** Principe des extensions bivariées. (a) Signal rotatif. (b) Signal inclus dans son enveloppe 3D : lignes noires et épaisses représentent les courbes d'enveloppes utilisées pour calculer la moyenne. (c) Composante à rotation rapide. (d) Composante à rotation plus lente correspondant à la moyenne du tube dans (b) [57].

Sachant qu'une armature représente une ligne de crête du tube dans une certaine direction, il apparaît que les points interpolés sont nécessairement des « extrema du signal dans cette direction », c'est-à-dire des extrema du signal projeté sur cette direction. L'armature est alors simplement définie comme la spline cubique interpolant ces points [57].

### 5.3.3 Algorithme

Le processus de la DMEB nécessite de projeter un signal multiple sur un ensemble de directions puis d'appliquer le processus de tamisage de la DME standard aux composantes projetées. Le processus fondamental de la BDME est résumé comme suit:

1. Initialiser le résidu  $k \leftarrow 1$  ( $k^{th}$  FMI);  $r_k(t) \leftarrow x(t)$
2. Extraire la  $k^{ième}$  FMI:

- 2.1. Estimation des courbes d'enveloppe

- 2.1.1. Projeter le signal à valeur bivariée  $r_k(t)$  sur la direction  $\varphi_k$  où  $\varphi_k = 2k\pi/K$  for  $k = 1..K$

$$p_{\varphi_k}(t) = Re \left( e^{-i\varphi_k} r_k(t) \right)$$

- 2.1.2. Extraire l'emplacement  $\{t_j^k\}$  du maxima de  $p_{\varphi_k}(t)$

- 2.1.3. Interpoler l'ensemble  $\{(t_j^k, r_k(t_j^k))\}$  Pour obtenir la courbe d'enveloppe dans la direction  $\varphi_k$

- 2.2. Calculer la moyenne de toutes les courbes d'enveloppe  $\mu(t) = \frac{1}{K} \sum_k e_{\varphi_k}(t)$

- 2.3. Soustraire la moyenne pour obtenir  $d(t) = r_k(t) - \mu(t)$

- 2.4. Itérer sur le détail  $d(t)$  En répétant les étapes 2.1 à 2.3 jusqu'à ce que le critère d'arrêt basé sur l'écart-type entre deux détails consécutifs soit inférieur à un seuil prédéfini conduisant à  $FMI_k$

3. Mettre à jour le résidu  $k \leftarrow k + 1$ ;  $r_k(t) \leftarrow r_{k-1}(t) - FMI_k(t)$

4. Itérer sur le résidu en répétant les étapes 2 et 3 jusqu'à ce que le nombre d'extrema de  $r_k(t)$  soit inférieur à 2.

## **6 Conclusion**

Dans ce chapitre, nous avons étudié l'algorithme de DME. Tous d'abord, nous avons présenté les méthodes conventionnelles conçues pour l'analyse des signaux non-stationnaires telles que la TF et la TO. La DME décompose le signal en un nombre fini de composantes ou oscillations (FMIs) extraites des données. L'extraction des FMIs est non-linéaire, mais leur recombinaison linéaire est exacte. L'algorithme original est conçu pour les signaux réels, nous nous sommes intéressés aux extensions de la DME pour les signaux à valeurs complexes proposées dans la littérature.

## **Chapitre 3**

# **Déconvolution du signal de parole par décomposition en modes empiriques**



## **1 Introduction**

Dans ce chapitre, la décomposition en modes empiriques (DME) est proposée comme alternative pour séparer le signal de parole dans le domaine spectral en sa composante harmonique et son enveloppe spectrale. Les contributions estimées ont été exploitées pour la caractérisation du signal de parole par l'estimation conjointe des formants et de la fréquence fondamentale. L'information de la phase a été exploitée pour l'estimation de la source glottique. Les méthodes proposées sont testées sur des signaux de parole synthétiques et réels et comparées avec les méthodes de références telles que la séparation basée sur la transformée en ondelettes et le cepstre complexe.

## **2 Déconvolution du signal de parole**

### **2.1 Principe de la déconvolution par DME**

Selon le modèle source-filtre de production de la parole, le signal de parole peut être considéré comme étant le résultat de la convolution de l'excitation du conduit vocal (excitation glottique) et de sa réponse impulsionnelle [15] :

$$x(t) = e(t) * v(t) \quad (3.1)$$

où  $x(t)$  est le signal de parole,  $v(t)$  est la réponse impulsionnelle du système modélisant le conduit vocal, et  $e(t)$  est le signal d'excitation ayant pour origine les cordes vocales, et  $*$  désigne le produit de convolution.

En multipliant le signal par une fenêtre de pondération et en prenant sa TF, la convolution se transforme en un produit :

$$X_{\omega}(f) = E(f) \times V(f) \quad (3.2)$$

où  $f$  dénote la fréquence,  $X_\omega(f)$  et  $E(f)$  sont, respectivement, les spectres d'amplitude des trames d'analyse du signal et de l'excitation glottique pondérées par la fenêtre  $w(t)$  et  $V(f)$  est la réponse fréquentielle du conduit vocal.

Le spectre du signal obtenu est donc constitué de raies dont les amplitudes correspondent à la réponse en fréquence du conduit vocal aux différentes fréquences multiples de la fréquence fondamentale (pitch). L'enveloppe de ces raies représente la réponse en fréquence du conduit vocal. Bien que les valeurs des formants varient d'une personne à l'autre, elles sont suffisamment stables pour être caractéristiques du son produit. Le spectre d'amplitude d'une trame pondérée du signal de parole peut être écrit comme

$$|X_\omega(f)| = |E(f) \times V(f)| \quad (3.3)$$

En prenant le logarithme des deux membres de l'équation, il vient :

$$\log|X_\omega(f)| = \log|E(f)| + \log|V(f)| \quad (3.4)$$

Comme le montre l'équation (3.4), le logarithme du spectre d'amplitude d'une trame pondérée du signal de parole est la somme de deux composantes spectrales : le logarithme du spectre d'amplitude de l'excitation pondérée  $\log|E(f)|$ , et le logarithme de l'enveloppe spectrale  $\log|V(f)|$  [111-114].

Le logarithme du spectre d'amplitude du signal de parole voisée peut être considéré comme constitué d'une variation lente, (par rapport à la fréquence) représentant le contour due à la contribution du conduit vocal et d'une série d'harmoniques caractérisées par une structure périodique.

L'algorithme de décomposition modale empirique donne un outil efficace pour séparer les deux composantes du spectre d'amplitude. En effet, l'algorithme de décomposition modale empirique agit comme un banc de filtres [43], de sorte que la décomposition du logarithme du spectre d'amplitude résulte en plusieurs composantes oscillantes (FMIs) qui peuvent être regroupées en deux catégories (classes) où chaque classe de composantes est associée à une partie du spectre d'amplitude [115].

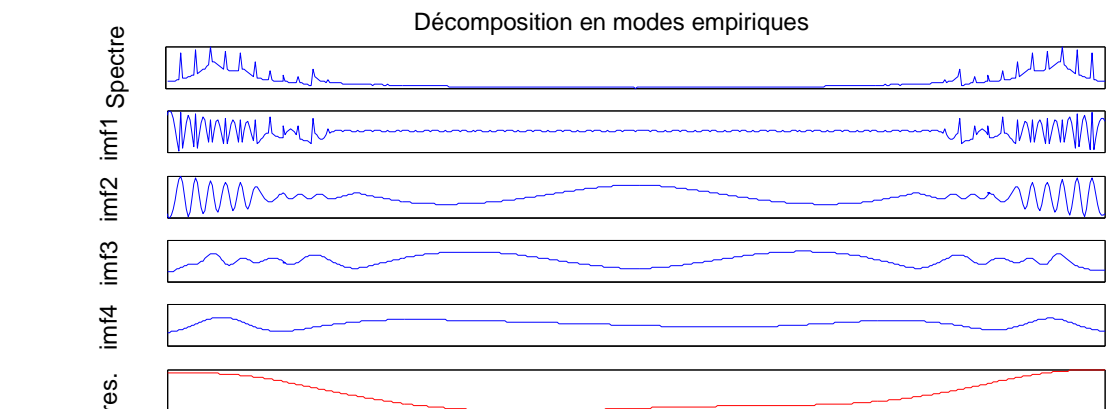
Il a été montré que la variance des FMIs pour les signaux de parole diminue de manière significative après la quatrième FMI au fur et à mesure que l'ordre des FMIs augmente [116]. Il a été constaté expérimentalement que pour le signal de parole, les statistiques des FMIs sont caractérisées par un pic d'énergie à une FMI d'ordre élevé. Cette propriété est utilisée pour sélectionner l'indice optimal qui permet de séparer la composante harmonique et l'enveloppe spectrale. Les différentes étapes de la méthode de séparation de la composante harmonique et de l'enveloppe spectrale illustrées par la figure 3.1 peuvent être résumées comme suit [116] :

1. Décomposer le logarithme du spectre d'amplitude par la DME.
2. Calculer la variance  $V$  de chaque FMI.
3. Identifier l'indice du maximum,  $m_p$  dans  $V$  pour un ordre de supérieur ou égale à 4.
4. Identifier l'indice du minimum  $m_t$ .
5. Calculer  $m_b$  tel que  $m_b = m_p - m_t$ .
6. Déterminer d'indice  $ind = m_p - m_b$ .

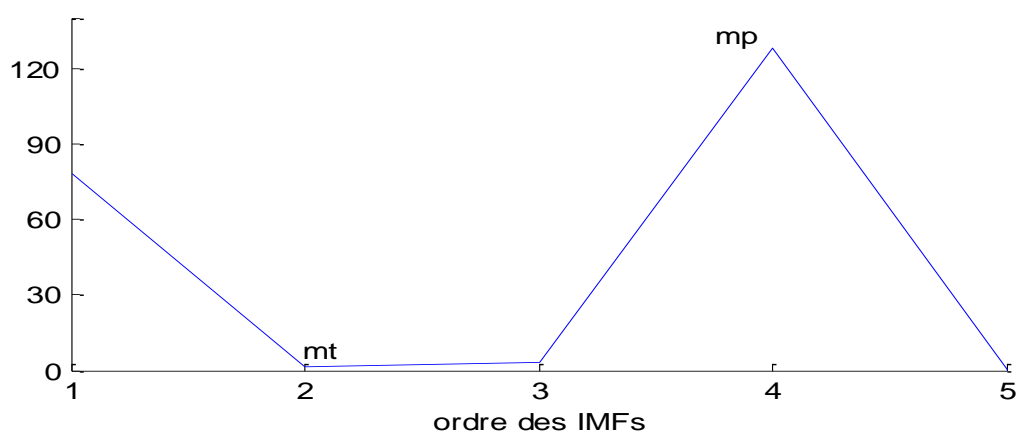
Si  $j < ind$  alors  $FMI_j$  appartient à la composante harmonique

Si  $j \geq ind$  alors  $FMI_j$  appartient à l'enveloppe spectrale

Pour mieux comprendre la méthode proposée, un exemple d'illustration de l'algorithme est présenté. La figure 3.1 montre la décomposition du logarithme du spectre d'amplitude d'une trame extraite d'une voyelle /a/ synthétique en utilisant une fenêtre de Hamming de longueur 1024 échantillons.

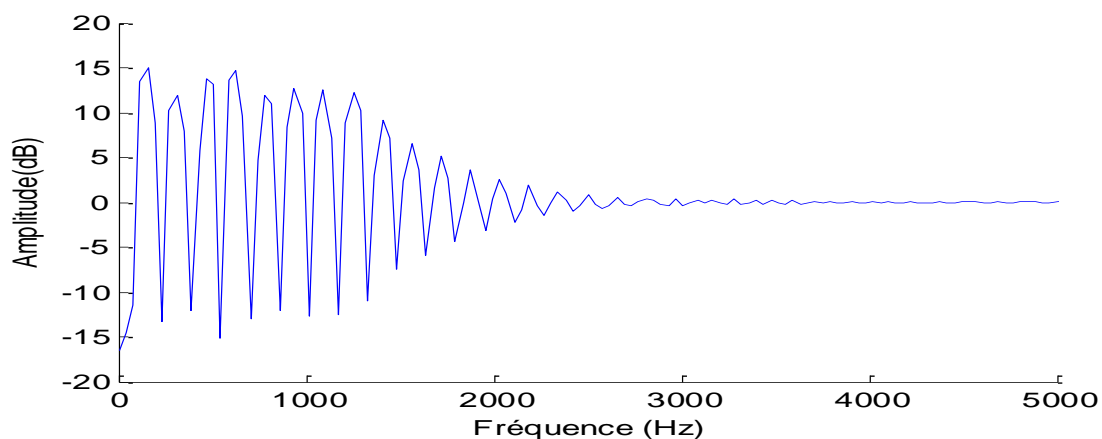


(a)

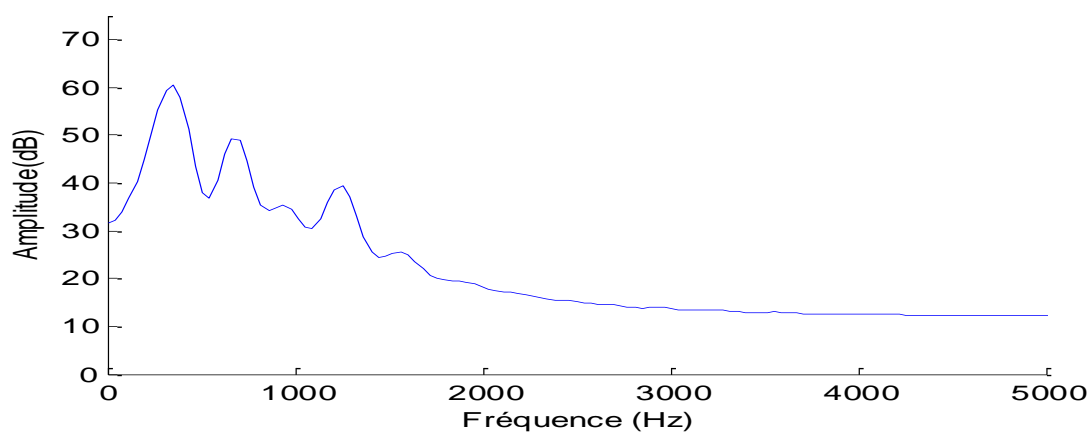


(b)

**Figure 3.1** Illustration de la déconvolution du signal de parole par la méthode proposée: (a) Décomposition modale empirique du logarithme du spectre d'amplitude d'une trame extraite d'une voyelle /a/ synthétique. (b) Variance des FMIs. (c) Logarithme du spectre d'amplitude de la composante harmonique. (d) Logarithme du spectre d'amplitude du conduit vocal.



(c)



(d)

**Figure 3.1** (Suite)

LA DME décompose le logarithme du spectre en  $N = 4$  FMIs, comme le montre la figure 3.1 (a). La variation de la variance des FMIs en fonction de leurs ordres est montrée sur la figure 3.1 (b). Comme le montre la figure, le pic est atteint à  $m_p = 4$  et le creux à  $m_t = 2$  donnant une variance  $m_b = 2$  et un indice  $ind=2$ .

Selon l'algorithme de classification, la composante harmonique est estimée comme étant la première FMI (figure 3.1 (c)), et l'enveloppe spectrale est obtenue comme étant la somme des trois FMIs restantes et le résidu (figure 3.1 (d)).

L'estimation de la composante harmonique illustrée par la figure 3.1 (c) est définie par :

$$\log|E(f)| = \sum_{j=1}^{ind-1} IMF_j(f) \quad (3.5)$$

L'estimation de l'enveloppe spectrale illustrée par la figure 3.1 (d) est donnée par

$$\log|V(f)| = \sum_{j=ind}^K IMF_j(f) + R_K(f) \quad (3.6)$$

## 2.2 Principe de la déconvolution par la transformée en ondelettes

Dans [21], les auteurs ont présentés une méthode de séparation des composantes du signal de parole en utilisant un banc de filtres basé sur la transformée en ondelettes (TO). Les composantes  $|E(f)|$  et  $|V(f)|$  sont séparées en utilisant l'algorithme défini par les étapes suivantes [21] :

1. Décomposition du logarithme du spectre d'amplitude du signal de parole pondéré par la fenêtre de Hamming. Le niveau de décomposition est fixé à  $K=8$

$$\{a_8, d_8, d_7, d_6, d_5, d_4, d_3, d_2, d_1\} = TOD(\log|X_\omega(f)|) \quad (3.7)$$

où TOD désigne la transformée en ondelette discrète,  $a_k$  et  $d_k$ , représentent l'approximation et les détails au niveau de décomposition  $k=1, \dots, K$ .

2. Construction de la matrice D définie par

$$\begin{bmatrix} D_1 \\ D_2 \\ D_3 \\ \vdots \\ D_K \end{bmatrix} = TODI \begin{bmatrix} a_K & 0 & \dots & \dots & \dots & 0 \\ a_K & d_K & 0 & \dots & \dots & 0 \\ a_K & d_K & d_{K-1} & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ a_K & d_K & d_{K-1} & \dots & d_2 & 0 \end{bmatrix} \quad (3.8)$$

où TODI est la TOD inverse. Pour résoudre le problème des dimensions, des zéros sont ajoutés à chaque ligne du vecteur  $D_i, i = 1, \dots, K$ .

3. Calcul de la différence d'énergie  $b_i$  entre chaque deux éléments successifs de la matrice  $D_i$ .

$$b_i = \sum_k \frac{[D_{k+1} - D_k]^2}{N}, i = 1, \dots, K \quad (3.9)$$

4. Identification de l'indice  $j$  qui correspond au premier minimum du vecteur  $b_i$ ,  $b_j = b_{min}$ . L'indice  $j$  permet de séparer les basses fréquences et les hautes fréquences.
5. Utilisation de la composante haute fréquence pour estimer  $|E(f)|$  et la composante basse fréquence pour l'estimation de  $|V(f)|$  :

$$|V(f)| = TODI(a_K, d_K, d_{K-1}, d_{K-2}, \dots, d_j, 0, \dots, 0, 0) \quad (3.10)$$

et

$$|E(f)| = TODI(0, 0, 0, 0, \dots, 0, d_{j+1}, \dots, \dots, d_3, d_2, d_1) \quad (3.11)$$

### 2.3 Résultats et discussions

L'approche proposée a été testée sur des signaux de parole synthétique et naturelle. La fréquence d'échantillonnage est de 20 kHz.

Le synthétiseur est un modèle source-filtre constitué d'une source qui génère un train d'impulsions périodique modélisant le flux d'air glottique et un conduit vocal

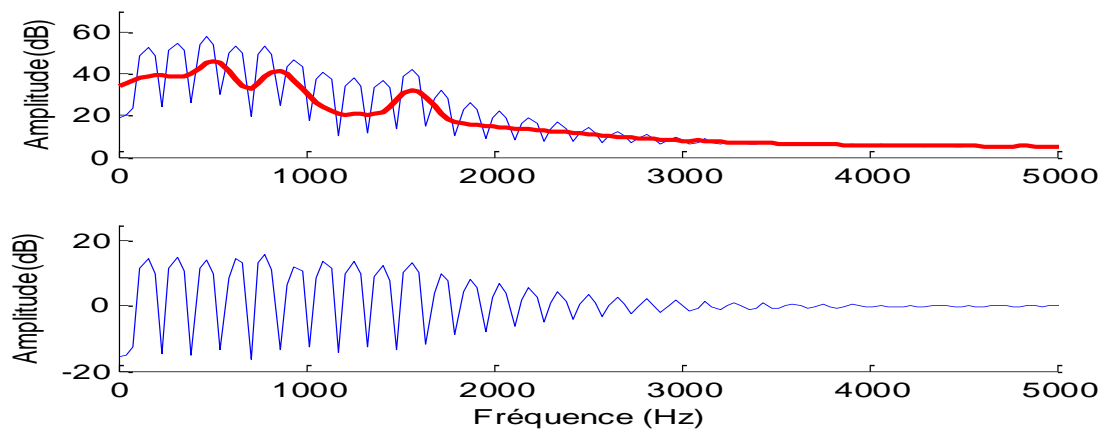
modélisé par un filtre tout-pôle, caractérisé par trois pôles [15, 117] correspondant aux fréquences des formants 800 Hz, 1200 Hz et 2870 Hz présentant des bandes fréquentielles de 90 Hz, 110 Hz et 170 Hz, respectivement. Le rayonnement aux lèvres est modélisé par un dérivateur du premier ordre  $R(z) = 1 - z^{-1}$ .

Le signal de parole est divisé en  $p$  trames sans chevauchement et la composante harmonique et l'enveloppe spectrale sont estimées pour chaque trame en utilisant la méthode de séparation basée sur la transformée en ondelettes et la DME pour différentes tailles de la fenêtre. Cette dernière a pour rôle de réduire la distorsion spectrale causée par la durée finie du signal. Les performances de la méthode proposée en termes de précision ont été comparées avec celles de la séparation basée sur la TO [21].

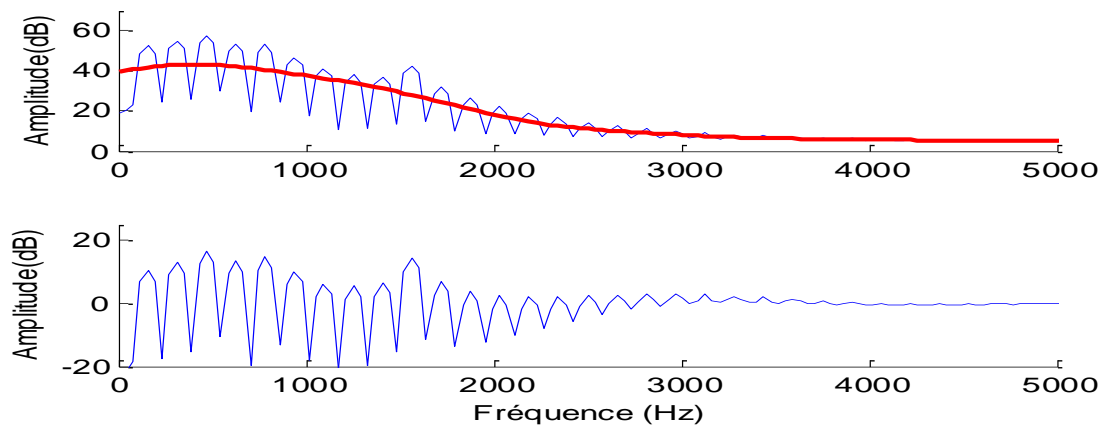
Les figures 3.2 et 3.3 montrent les résultats de séparation par la TO et la DME appliquées à une voyelle /a/ synthétique et une voyelle /a/ naturelle respectivement. Pour les deux voyelles, la longueur de trame est fixée à 256 et 1024 échantillons.

Pour une longueur de trame de 256 échantillons, la méthode de séparation basée sur la TO conduit à une estimation lisse de l'enveloppe spectrale et une composante harmonique qui contient des variations lentes en raison de la contribution du conduit vocal. Pour une longueur de trame de 1024 échantillons, la composante harmonique est estimée avec précision, mais les fréquences des formants ne sont toujours pas distinguées dans l'enveloppe spectrale. La méthode de séparation basée sur la DME donne des estimations précises de la composante harmonique et de l'enveloppe spectrale pour les deux longueurs de trame.





(i) Déconvolution basée sur la décomposition en modes empiriques

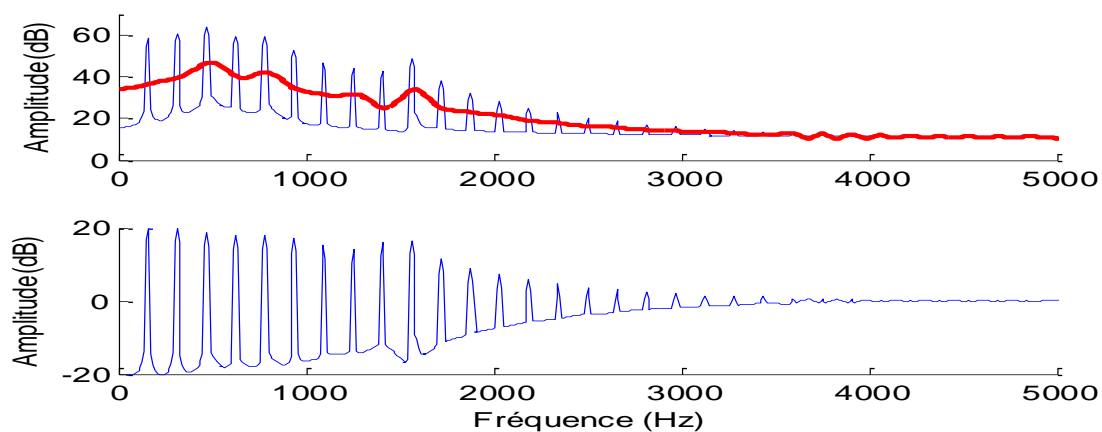


(ii) Déconvolution basée sur la transformée en ondelettes

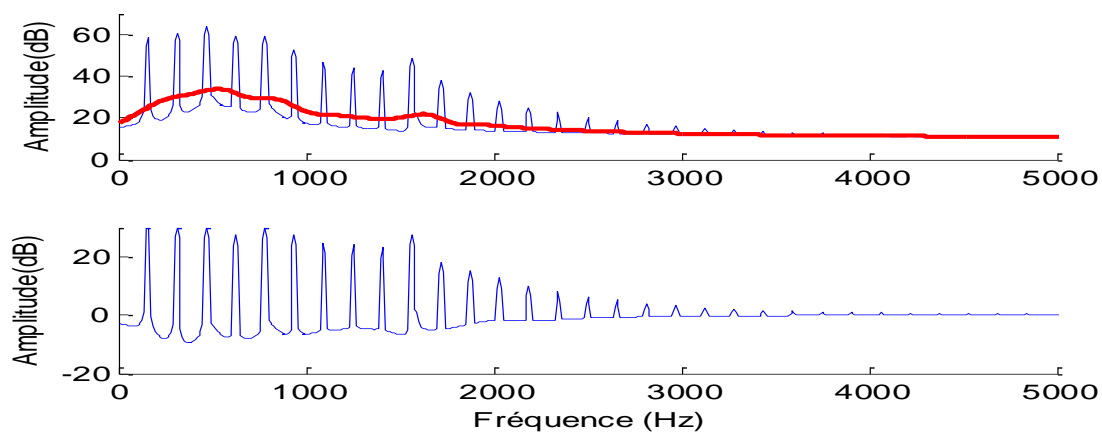
(a)

**Figure 3.2** Comparaison entre la méthode de séparation basée sur la TO et celle basée sur la DME appliquées à une voyelle /a/ synthétique pour deux longueurs de la trame.

(a) Longueur de trame de 256 échantillons. (b) Longueur de trame de 1024 échantillons.



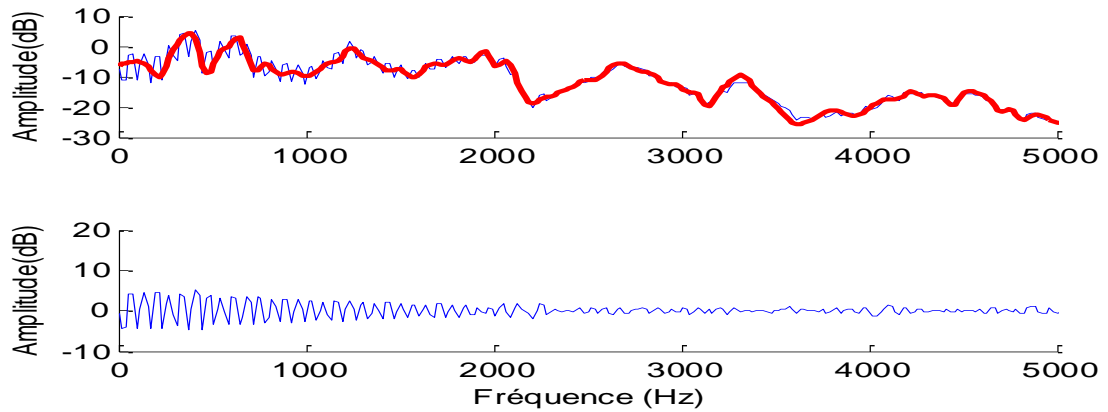
(i) Déconvolution basée sur la décomposition en modes empiriques



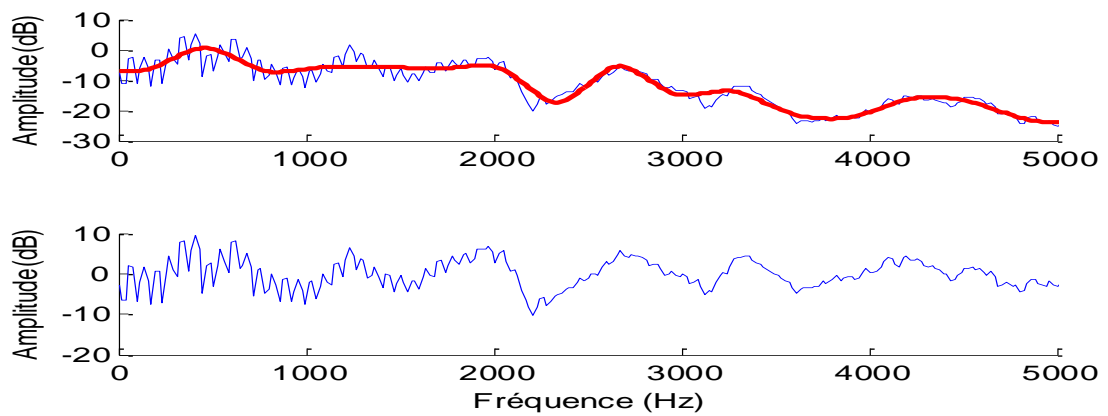
(ii) Déconvolution basée sur la transformée en ondelettes

(b)

**Figure 3.2** (Suite)



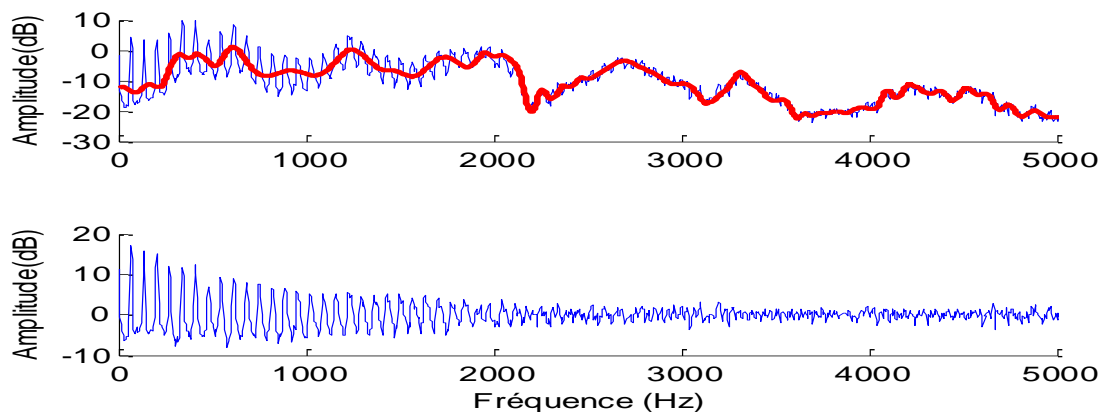
(i) déconvolution basée sur la décomposition en modes empiriques



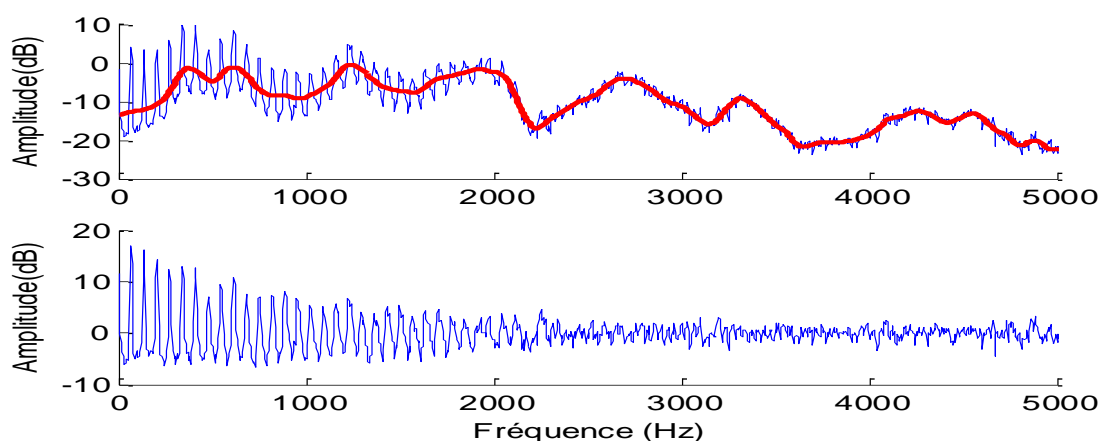
(ii) Déconvolution basée sur la transformée en ondelettes

(a)

**Figure 3.3** Comparaison entre la méthode de séparation basée sur la TO et celle basée sur la DME appliquées à une voyelle /a/ naturelle pour deux longueurs de la trame. (a) Longueur de trame de 256 échantillons. (b) Longueur de trame de 1024 échantillons.



(i) Déconvolution basée sur la décomposition en modes empiriques

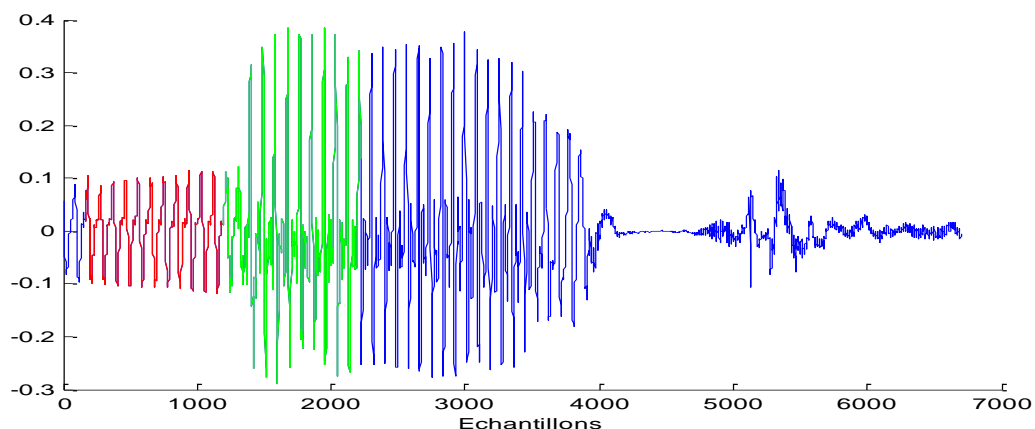


(ii) Déconvolution basée sur la transformée en ondelettes

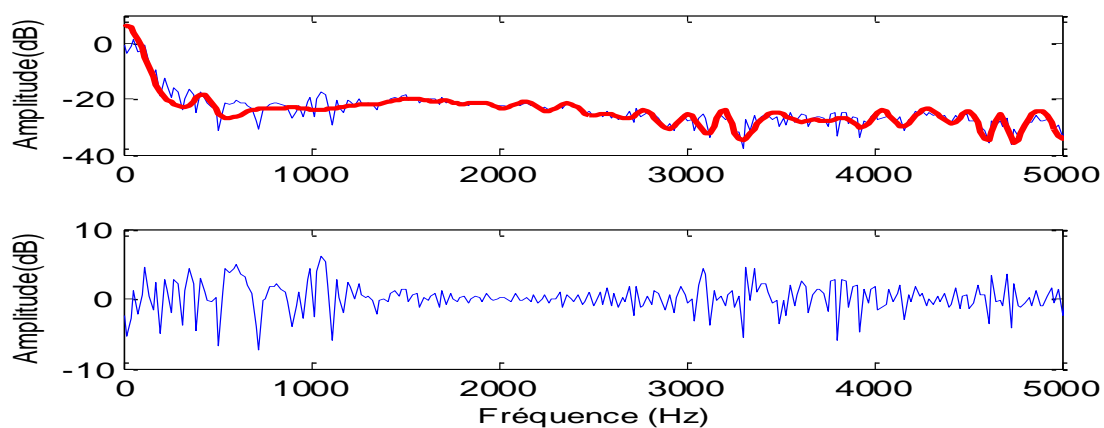
(b)

**Figure 3.3** (Suite)

Une comparaison entre les méthodes de séparation basées sur la TO et la DME appliquées à deux trames successives extraites d'un signal de parole correspondant à une phrase prononcée par un locuteur masculin pour une longueur de trame de 1024 échantillons. Les trames décomposées sont représentées en rouge et en vert sur la figure 3.4. La figure 3.5 montre les résultats de séparations obtenus

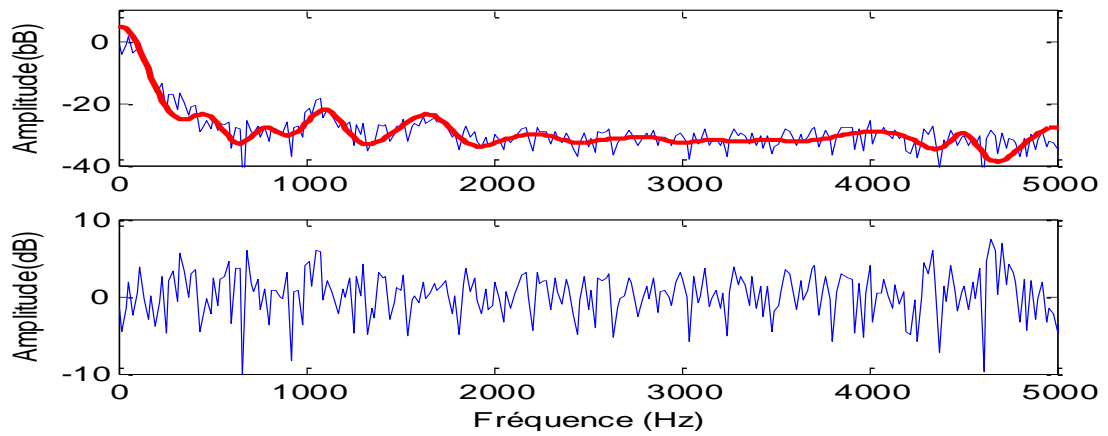


**Figure 3.4** Deux trames successives extraites du signal de parole continue.



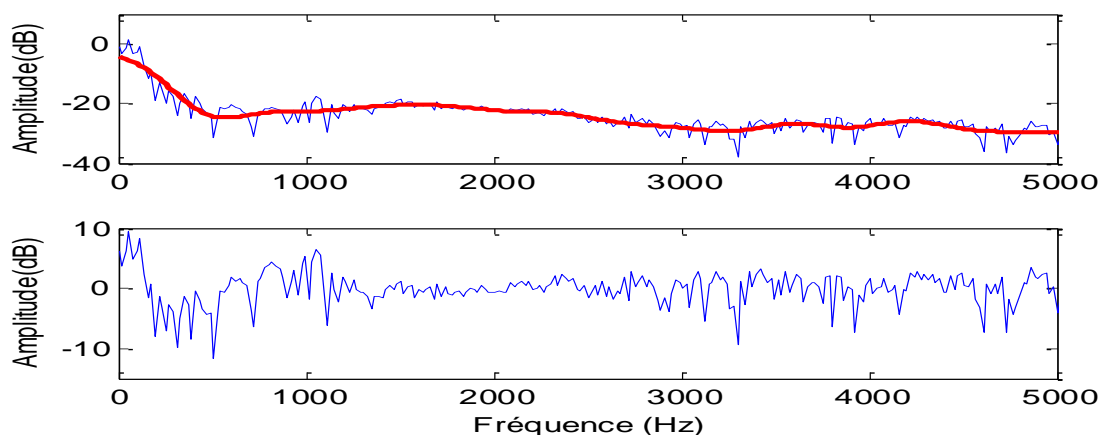
(i) Déconvolution de la première trame basée sur la décomposition en modes empiriques

**Figure 3.5** Comparaison entre les méthodes de séparation basées sur la TO et la DME appliquées à deux trames successives extraites d'un signal de parole correspondant à une phrase prononcée par un locuteur masculin pour une longueur de trame de 1024 échantillons. (a) Séparation basée sur la DME. (b) Séparation basée sur la TO.



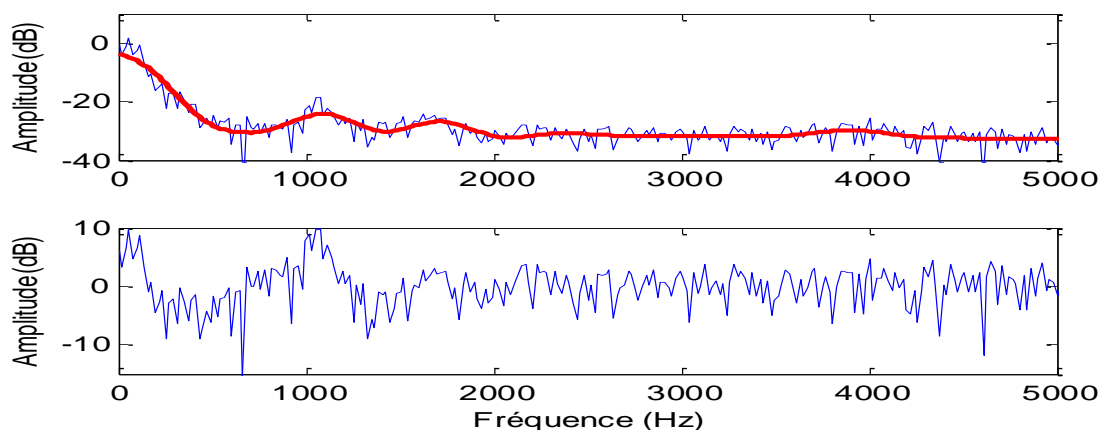
(ii) Déconvolution de la seconde trame basée sur la décomposition en modes empiriques

(a)



(i) Déconvolution de la première trame basée sur la transformée en ondelettes

**Figure 3.5** (Suite)



(ii) Déconvolution de la seconde trame basée sur la transformée en ondelettes

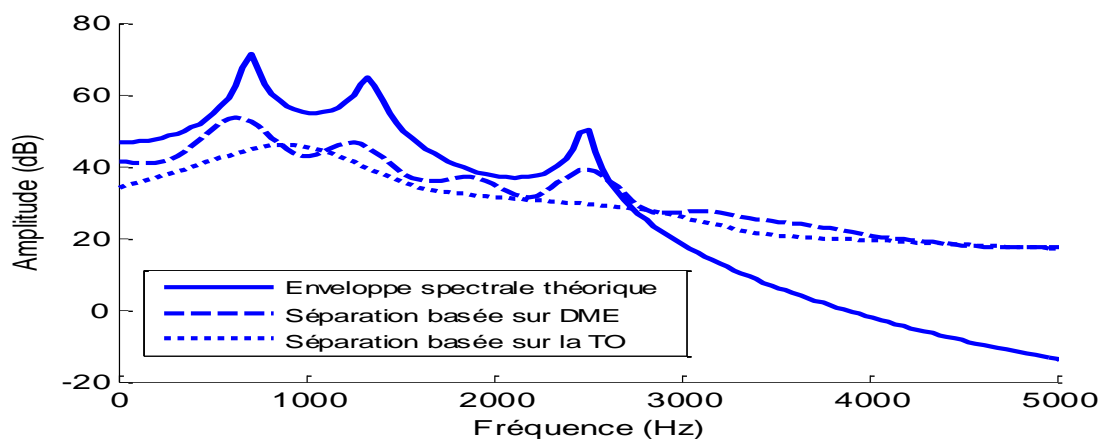
(b)

**Figure 3.5** (Suite)

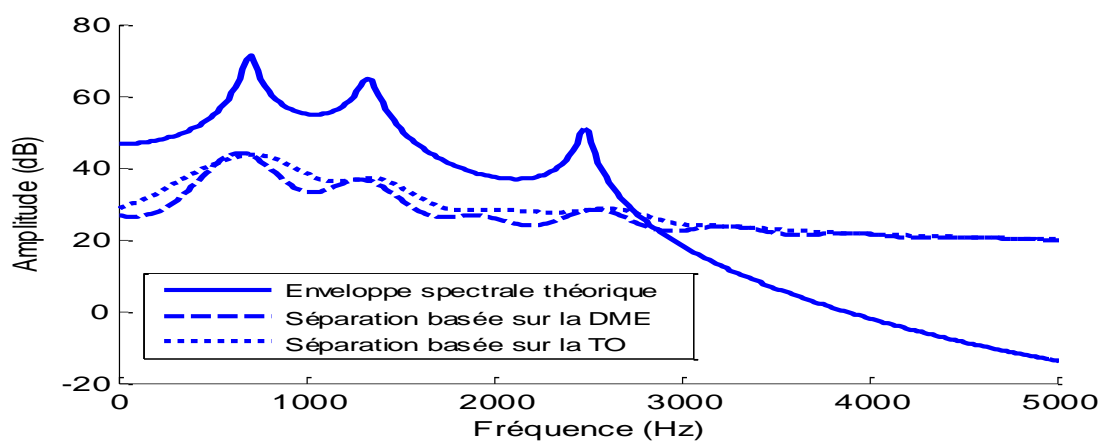
### **3. Application à l'estimation de la fréquence fondamentale et des formants**

L'estimation de la fréquence fondamentale et des formants est très souvent rencontrée dans les applications d'analyse et de synthèse de la parole. Les résultats de déconvolution sont exploités pour l'estimation conjointe des formants qui représentent les pics de la réponse fréquentielle du conduit vocal et de la fréquence fondamentale du signal de parole [118]. La fréquence fondamentale est estimée comme l'inverse de l'intervalle de temps séparant deux pics successifs de la séquence d'autocorrélation de la composante harmonique [119].

La figure 3.6 fournit une comparaison entre l'enveloppe spectrale théorique et celle estimée par déconvolution basée sur la DME et la TO appliquées à une voyelle /a/ synthétique en utilisant des longueurs de trames de 256 et 1024 échantillons.



(a)



(b)

**Figure 3.6** Comparaison entre l'enveloppe spectrale théorique et l'enveloppe spectrale estimée via la méthode de déconvolution basé sur la TO et la DME. (a) Longueur de trame de 256 échantillons. (b) Longueur de trame de 1024 échantillons.

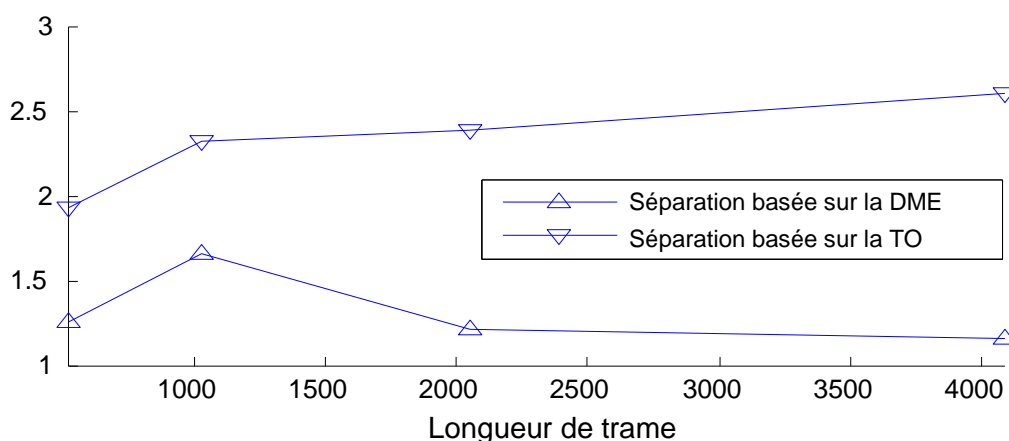
Pour comparer les performances des deux méthodes en termes de précision, l'erreur quadratique moyenne entre l'enveloppe spectrale théorique et celle estimée par DME et TO a été calculée pour différentes longueurs de la trame :



$$e = \frac{1}{L} \sqrt{\sum_i (\log(|H(f_i)|) - \log|V(f_i)|)^2} \quad (3.12)$$

où  $i$  fait référence à l'indice de fréquence et  $L$  le nombre de fréquences,  $\log(|H(f)|)$  désigne l'enveloppe spectrale théorique prise comme référence, et  $\log|V(f)|$  dénote l'enveloppe spectrale estimée.

Comme le montre la figure 3.7, l'erreur quadratique moyenne entre l'enveloppe spectrale théorique et celle estimée par la TO est supérieure à l'erreur quadratique moyenne entre l'enveloppe spectrale théorique et celle estimée par la DME.



**Figure 3.7** Erreur quadratique moyenne entre l'enveloppe spectrale théorique et estimée en fonction de la longueur de la trame.

Le tableau 3.1 montre les valeurs des trois premiers formants ainsi que ceux estimés comme étant la position fréquentielle des pics de l'enveloppe spectrale obtenus par DME et TO pour une voyelle /a/ synthétique en utilisant des longueurs de trame de 512 et 1024 échantillons. Les fréquences des formants correspondant à l'intervalle d'analyse totale sont obtenues en faisant la moyenne des fréquences des formants locales sur toutes les trames.

Pour une longueur de trame de 1024 échantillons, la DME conduit à des fréquences des formants plus précises que la méthode de décomposition à base d'ondelettes. Le troisième formant ne peut pas être calculé à partir de l'estimation de l'enveloppe spectrale à base d'ondelettes. Pour une longueur de trame de 512 échantillons, les 3 formants ne sont pas discernables dans l'enveloppe spectrale estimée par TO.

La fréquence fondamentale du signal de parole a été estimée par une méthode d'extraction de pic appliquée à la séquence d'autocorrélation de la composante harmonique [15] estimée en utilisant la DME ainsi que la technique de décomposition à base d'ondelettes pour des longueurs de trames de 512 et 1024 échantillons. Les résultats sont présentés dans le tableau 3.2. La DME présente une performance supérieure à celle de la décomposition par ondelettes en termes de précision de l'estimation pour les deux longueurs de trame.

**Tableau 3.1.** Fréquences des formants théoriques et estimées pour une voyelle /a/ synthétique en utilisant une longueur de trame de 512 et 1024 échantillons.

			F1 (Hz)	F2 (Hz)	F3 (Hz)
Valeurs théoriques			700	1333	2486
Valeurs estimées	512 échantillons	Séparation par DME	665	1328	2500
		Séparation par TO	-	-	-
	1024 échantillons	Séparation par DME	668	1348	2539
		Séparation par TO	703	1309	-

**Tableau 3.2.** Fréquences fondamentales théorique et estimées pour une voyelle /a/ synthétique utilisant des longueurs de trames de 512 et 1024 échantillons.

		Fréquence fondamentale théorique =125 Hz	
		Séparation par DME	Séparation par TO
Longueur de Trame	512 échantillons	128 Hz	123 Hz
	1024 échantillons	125 Hz	106 Hz

Les fréquences fondamentales estimées pour une voyelle /a/ naturelle sont reportées dans le tableau 3.3. La fréquence fondamentale estimée à l'aide du logiciel Praat [120, 121] a été utilisée comme référence pour comparaison. L'estimation de la fréquence fondamentale à partir de la composante harmonique obtenue par DME est plus proche de la fréquence fondamentale calculée avec Praat comparée à l'estimation de la fréquence fondamentale à partir de la composante harmonique obtenue par TO pour les deux longueurs de la trame.

**Tableau 3.3.** Estimation des fréquences fondamentales pour une voyelle /a/ naturelle en utilisant des longueurs de trames de 512 et 1024 échantillons.

		Fréquence fondamentale obtenue par Praat =136 Hz	
		Fréquence fondamentale estimée par DME	Fréquence fondamentale estimée par TO
Longueur de trame	512 échantillons	128 Hz	333 Hz
	1024 échantillons	136 Hz	190 Hz

## **4 Application à l'estimation de la source glottique**

La séparation des composantes du signal de parole est utilisée dans plusieurs applications, telle que la caractérisation de l'excitation glottique en reconnaissance de locuteur [122], la caractérisation et l'analyse des troubles de la voix [123], la reconnaissance de la parole [124] et la synthèse de la parole [125]. Ces raisons justifient la nécessité de développer des algorithmes capables d'estimer et de paramétrer le signal glottique de manière robuste et fiable. Bien que les techniques de modélisation du conduit vocal soient assez bien établies, ce n'est pas le cas de pour la source glottique. Certains travaux ont abordé le problème d'estimation de la contribution de la glotte directement à partir de la forme d'onde du signal de parole. La plupart des approches s'appuie sur une première modélisation paramétrique du conduit vocal, puis utilise le filtrage inverse afin d'éliminer l'effet du conduit vocal et obtenir une estimation du signal glottique. Dans [126], le modèle discret tout pôle a été utilisé pour modéliser le conduit vocal. La technique itérative adaptative par filtrage inverse décrite dans [127] isole le signal source en estimant de manière itérative à la fois les composantes dues au conduit vocal et à la source. Dans [128], l'estimation du signal glottique est affinée sur plusieurs cycles glottiques. Une technique non paramétrique basée sur les zéros de la transformée en Z (ZZT) et le cepstre complexe a été proposée dans [129, 130]. Cette approche repose sur l'observation que la parole est un signal à phase mixte comprenant une composante causale et une composante anti-causale où la composante anti-causale correspond à la phase d'ouverture de la glotte et la composante causale comprend à la fois la fermeture de la glotte et les contributions du conduit vocal.

#### **4.1 Estimation de la source glottique par cepstre complexe**

Le spectre d'une trame de parole pondérée peut être exprimé comme

$$X_{\omega}(f) = |X_{\omega}(f)|e^{j\angle X_{\omega}} \quad (3.13)$$

où  $\angle X_{\omega}$  dénote la phase de  $X_{\omega}(f)$ . Le cepstre complexe  $\hat{x}(n)$  correspondant est défini par :

$$\hat{x}(n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \log[X_{\omega}(f)]e^{j\omega n} d\omega \quad (3.14)$$

Dans [131, 132] l'auteur propose une méthode d'estimation de l'excitation glottique basée sur la décomposition en phase mixte du signal de parole en utilisant le cepstre complexe. La méthode nécessite une fenêtre de pondération qui vérifie le compris de la décomposition en phase mixte ce qui n'est pas vérifié pour tous les signaux.

#### **4.2 Estimation de la source glottique par DME**

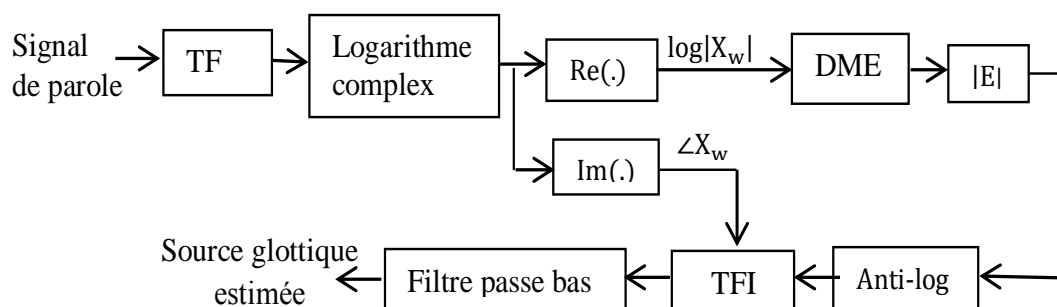
Dans cette partie du travail, l'information de phase est exploitée pour estimer la source glottique par DME. En prenant le logarithme complexe des deux membres de (3.12), il vient

$$\log(X_{\omega}) = \log|X_{\omega}(f)| + j\angle X_{\omega} \quad (3.15)$$

La partie réelle du logarithme complexe représente le logarithme du spectre d'amplitude du signal et la partie imaginaire représente la phase du signal. La phase doit être corrigée en ajoutant des multiples de  $\pm 2\pi$  [133].

L'approche d'estimation de la source glottique illustrée par la figure 3.8 combine le logarithme du spectre d'amplitude de la source glottique et la phase

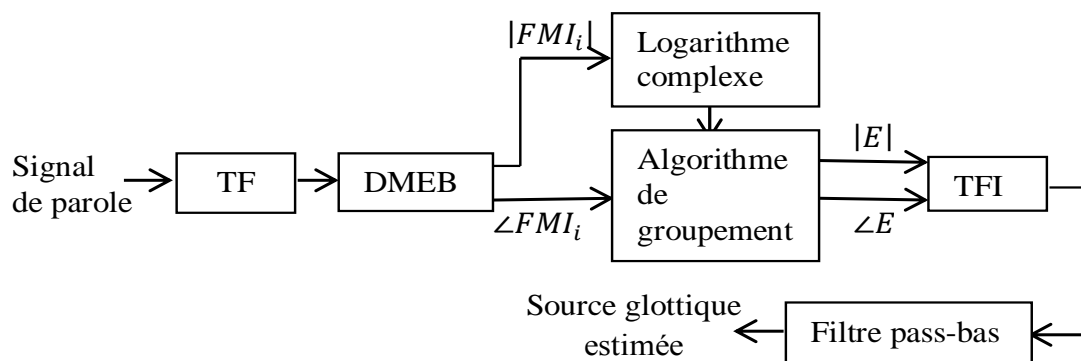
estimée. Le filtre passe bas a pour rôle d'éliminer les oscillations causées par la fenêtre de pondération  $w(t)$  [134].



**Figure 3.8** Schéma bloc de la méthode d'estimation de la source glottique basée sur la DME.

### 4.3 Estimation de la source glottique par DMEB

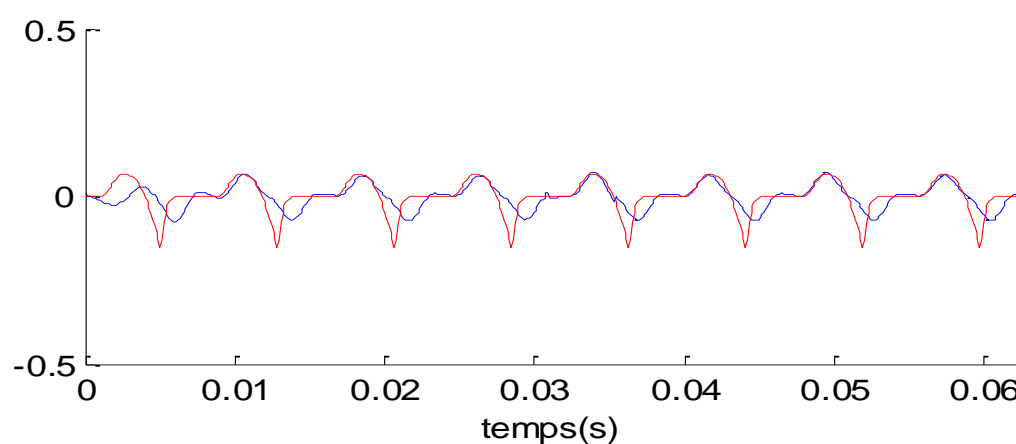
Une autre méthode proposée pour l'estimation de la source glottique exploite les caractéristiques de la DMEB [135]. L'obtention des FMIs présentant la phase du logarithme du spectre permet de sélectionner les IMFs constituant la phase contrairement aux algorithmes existants dans la littérature qui consiste l'ajout de la phase du signal complet. La figure 3.9 illustre l'organigramme de la méthode proposée.



**Figure 3.9** Schéma bloc de la méthode d'estimation de la source glottique basée sur DMEB.

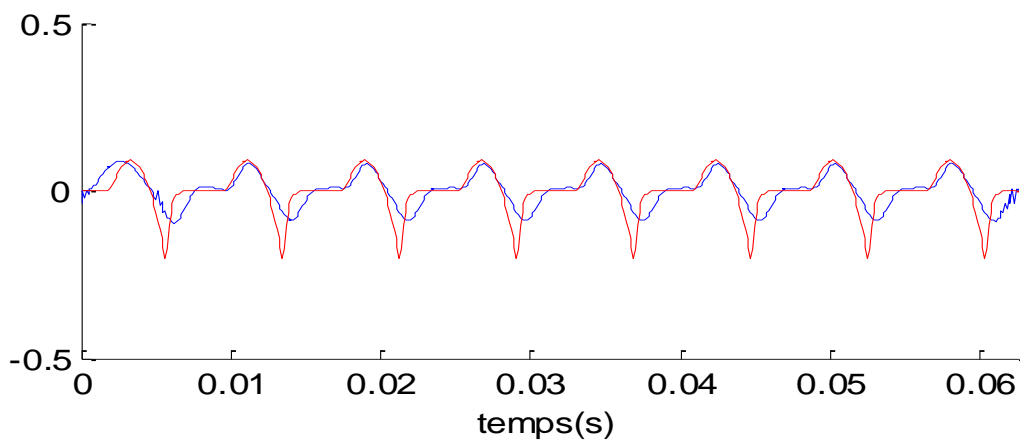
#### 4.4 Résultats et discussions

Le signal utilisé pour tester la méthode d'estimation de la source glottique est une voyelle /a/ synthétique de durée 1 seconde et de fréquence fondamentale  $f_0=128$  Hz générée selon le modèle source-filtre de production de la parole. Afin d'analyser les performances de la méthode proposée, l'estimation de la source glottique a été effectuée pour différentes tailles de la fenêtre de pondération. Les résultats obtenus sont montrés sur la figure 3.10. La décomposition en modes empiriques permet d'avoir des estimations précises de la source glottique quelle que soit la longueur de la fenêtre utilisée.

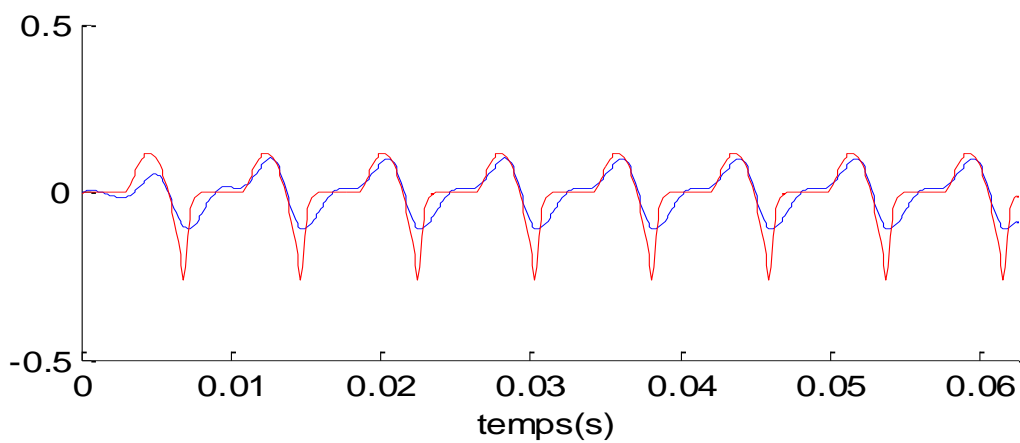


L=256

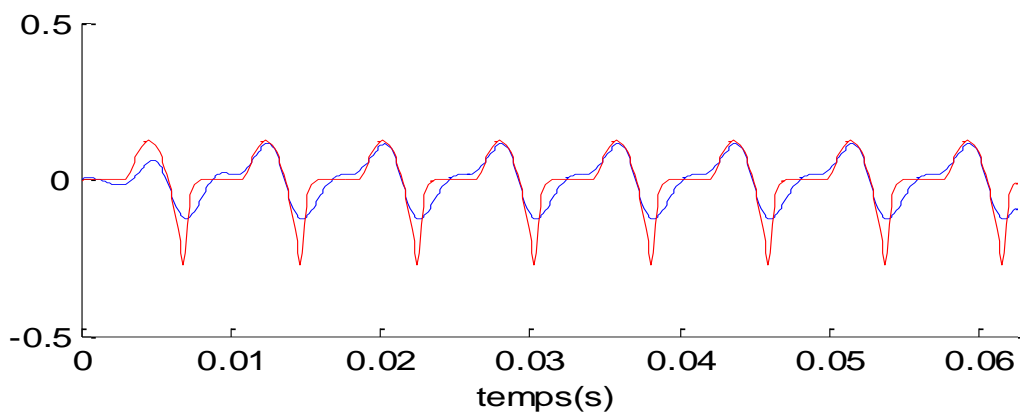
**Figure 3.10** Forme d'onde de la source glottique estimée par la DME superposée au modèle de l'excitation glottique pour différentes longueurs L de la trame.



L=512



L=1024

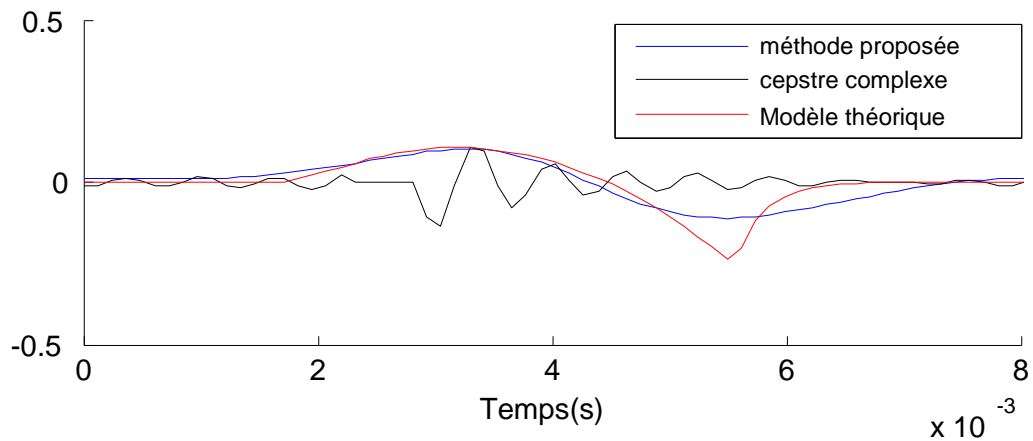


L=2048

**Figure 3.10** (Suite)

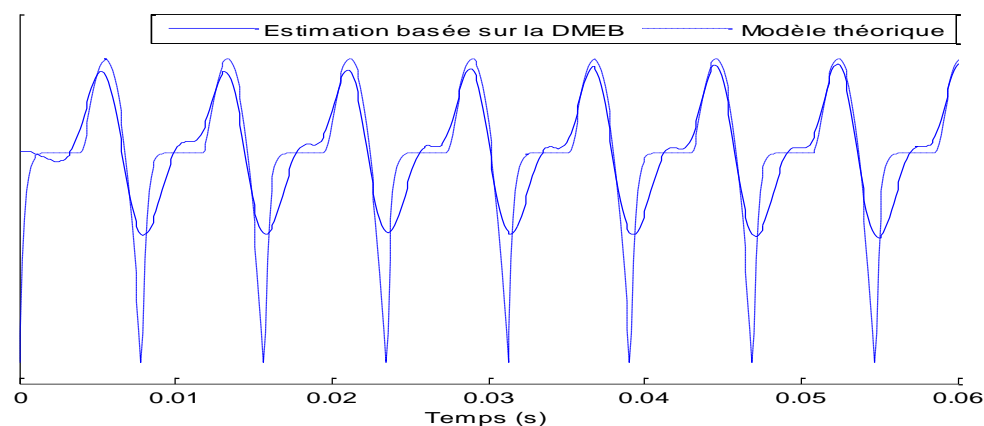


La figure 3.11 fournit une comparaison entre le modèle de la source glottique avec les estimations basées sur la décomposition modale empirique et le cepstre complexe pour une trame de longueur 1024 échantillons. Il est observé que l'estimation par décomposition en modes empiriques est plus précise que celle basée sur le cepstre complexe.



**Figure 3.11** Forme d'onde de la source glottique estimée par la méthode proposée comparée à celle estimée par le cepstre complexe et au modèle théorique pour une trame de longueur  $L=1024$ .

L'application de l'algorithme de DMEB au logarithme complexe du spectre fournit une meilleure estimation de la source glottique comme illustré par la figure 3.12.



**Figure 3.12** Forme d'onde de la source glottique estimée par DMEB superposée au modèle de l'excitation glottique pour une longueur de trame m1024 échantillons.

## 5 Conclusion

Dans ce chapitre, l'algorithme de la décomposition modale empirique a été proposé comme alternative pour décomposer le logarithme du spectre d'amplitude du signal de parole en ses deux composantes qui sont la composante harmonique et l'enveloppe spectrale. Les performances de la méthode proposée ont été comparées à celles de la décomposition par TO. Les résultats de la déconvolution ont été exploités pour effectuer une estimation conjointe des formants et de la fréquence fondamentale. Les résultats montrent que la méthode proposée fournit des estimations précises de la composante harmonique et de l'enveloppe spectrale pour les trames courtes ainsi que pour les trames longues. L'exploitation de l'information de phase permet, à son tour, d'effectuer une estimation de l'excitation glottique. Les résultats obtenus montrent que la méthode proposée fournit une estimation précise de l'excitation glottique.

## **Chapitre 4**

# **Séparation aveugle de mélanges de signaux de parole**

## **1 Introduction**

Les méthodes de résolution du problème de la SAS sont basées sur plusieurs concepts tels que la circularité et la gaussianité. Dans ce chapitre l'algorithme de la décomposition en modes empiriques bivariée est combiné avec l'analyse en composantes indépendantes par minimisation des bornes de l'entropie (ACI-MBE) qui exploite conjointement la non-gaussianité et la non-circularité des signaux pour séparer les mélanges convolutifs des signaux de parole.

Les performances de l'approche proposée sont testées sur des signaux de parole réelle choisis parmi les bases de données disponibles et comparées aux performances de la séparation basée sur l'algorithme conventionnel de l'analyse en composantes indépendantes fréquentielle (ACIF) ainsi qu'aux performances de l'algorithme de l'analyse en composantes indépendantes combiné avec l'algorithme de décomposition en modes empiriques publiés dans la littérature.

## **2 Algorithme conventionnelle de l'analyse en composantes indépendantes dans le domaine fréquentielle**

L'approche conventionnelle à l'égard de la résolution du problème de SAS du mélange convolutif dans le domaine fréquentielle, consiste à effectuer séparément l'ACI pour chaque fréquence, comme le montre la figure 4.1, en utilisant un algorithme d'ACI pour estimer la matrice de séparation  $\mathbf{W}(f)$  des vecteurs  $\mathbf{X}(f, t)$ .

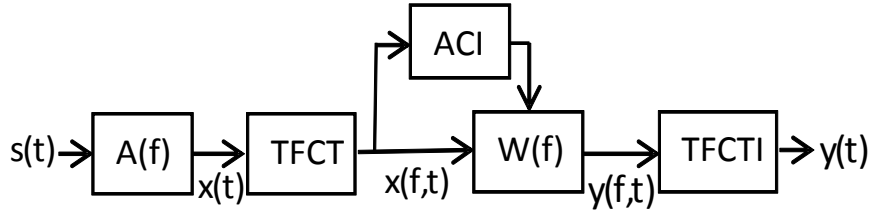


Figure 4.1 Principe de l'ACIF

### 3 Décomposition en composantes indépendantes basée sur la minimisation des bornes de l'entropie

#### 3.1 Principe

Pour effectuer la séparation des  $N$  sources indépendantes, la méthode ACI-MBE minimise l'information mutuelle  $I(z_1; \dots; z_N)$  définie comme

$$I(z_1; \dots; z_N) = \sum_{n=1}^N H(z_n) - \log | \det(\mathbf{W}) | - H(\mathbf{x}) \quad (4.1)$$

où  $H(z_n)$  est l'entropie de la  $n^{\text{ème}}$  source séparée, et  $H(\mathbf{x})$  est l'entropie des signaux observés.

Dans les approches de l'ACI unitaires, le blanchiment des mélanges permet d'avoir  $|\det(\mathbf{W})| = 1$  sous la contrainte d'une matrice de séparation orthogonale. La contrainte d'orthogonalité garantit les propriétés de stabilité et de convergence des algorithmes d'ACI. La fonction coût à minimiser devient :

$$I(z_1, \dots, z_N) = \sum_{n=1}^N H(z_n) - H(\mathbf{x}) \quad (4.2)$$

sous la contrainte unitaire, on a  $\mathbf{W}\mathbf{W}^H = I$ , où  $I$  est la matrice identité. Même si elle est couramment utilisée, la contrainte unitaire limite l'espace de recherche pour la matrice de séparation optimale et peut donc conduire à une performance non optimale [29].

Lors du développement des mises à jour de l'algorithme ACI-MBE complexe, le problème de minimisation  $I(z_1, \dots, z_N)$  par rapport à  $\mathbf{W} = [\mathbf{w}_1, \dots, \mathbf{w}_N]^H$  est divisé en une série de sous-problèmes de minimisation de  $I(z_1, \dots, z_N)$  par rapport à chacun des vecteurs ligne  $\mathbf{w}_n$ ,  $n = 1, \dots, N$ . Ce dernier est un problème plus facile à résoudre et fournit une meilleure performance lorsqu'on estime séparément l'entropie de chaque source. Par conséquent, il faut mettre à jour  $\mathbf{w}_n$ , tandis que  $\mathbf{w}_m$ ,  $m = 1, \dots, n-1, n+1, \dots, N$  sont maintenus constants.

Pour cette tâche, la fonction coût est écrite en fonction de  $\mathbf{w}_n$  seulement, alors  $|\det(\mathbf{W})|$  est exprimé comme :

$$|\det(\mathbf{W})| = |S \mathbf{h}_n^H \mathbf{w}_n| \quad (4.3)$$

où  $S = \sqrt{|\det(\mathbf{W}_n \mathbf{W}_n^H)|}$ , avec  $\mathbf{W}_n$  dénotant la matrice obtenue en supprimant le vecteur du  $n^{\text{ième}}$  rang de  $\mathbf{W}$ . Pour supprimer l'ambiguïté d'échelle inhérente à la solution de l'ACI, dans [29] les auteurs supposent que tous les vecteurs lignes de  $\mathbf{W}$  sont de longueur unitaire ( $\|\mathbf{w}_n\| = 1$ ), ce qui implique que les sources sont de variance unitaire puisque les mélanges sont pré-blanchis.

La fonction coût conduit à un problème d'optimisation restreint, qui n'est qu'une fonction de  $\mathbf{w}_n$

$$\min \hat{H}(z_n) - 2 \log |\mathbf{h}_n^H \mathbf{w}_n| + C \quad (4.4)$$

où  $z_n = \mathbf{w}_n \mathbf{x}$  et  $2 \log |\mathbf{h}_n^H \mathbf{w}_n|$  peut être considéré comme une fonction de pénalité qui essaie de maintenir  $\mathbf{w}_n$  orthogonal à tous les autres vecteurs lignes de la matrice de séparation  $\mathbf{W}$ .

Soit  $J_n(\mathbf{w}_n) = \hat{H}(z_n) - 2 \log |\mathbf{h}_n^H \mathbf{w}_n| + C$ . Le problème d'optimisation peut être écrit comme la fonction lagrangienne suivante :

$$L_n(\mathbf{w}_n, \lambda) = J_n(\mathbf{w}_n) + \lambda(\mathbf{w}_n^H \mathbf{w}_n - 1) \quad (4.5)$$

Les dérivés de Wirtinger sont utilisées pour calculer les gradients conjugués [108] donnés par

$$\frac{\partial J_n(\mathbf{w}_n)}{\partial \mathbf{w}_n^*} = \frac{\partial \hat{H}(z_n)}{\partial \mathbf{w}_n^*} - \frac{\mathbf{h}_n}{\mathbf{w}_n^H \mathbf{h}_n} \quad (4.6)$$

et

$$\frac{\partial L_n(\mathbf{w}_n, \lambda)}{\partial \mathbf{w}_n^*} = \frac{\partial J_n(\mathbf{w}_n)}{\partial \mathbf{w}_n^*} + \lambda \mathbf{w}_n \quad (4.7)$$

où le multiplicateur lagrangien  $\lambda$  est un nombre réel, le terme  $\frac{\partial \hat{H}(z_n)}{\partial \mathbf{w}_n^*}$  dans le gradient conjugué est strictement colinéaire à  $\mathbf{w}_n$  alors, il est supprimé puisqu'il ne contribue pas. Le gradient conjugué projeté sur la surface de contrainte  $\|\mathbf{w}_n\| = 1$  est calculé comme suit

$$\mathbf{u}_n = \frac{\mathbf{u}_n^+}{\|\mathbf{u}_n^+\|} \quad (4.8)$$

avec

$$\mathbf{u}_n^+ = \frac{\partial J_n(\mathbf{w}_n)}{\partial \mathbf{w}_n^*} - \text{Re} \left\{ \mathbf{w}_n^H \frac{\partial J_n(\mathbf{w}_n)}{\partial \mathbf{w}_n^*} \right\} \mathbf{w}_n \quad (4.9)$$

Il est clair que le gradient conjugué projeté normalisé  $\mathbf{u}_n$  est un vecteur de longueur unitaire et il est orthogonal au précédent  $\mathbf{w}_n$ . Finalement, la mise à jour de  $\mathbf{w}_n$  est donnée par

$$\mathbf{w}_n^{[\text{nouveau}]} = \frac{\mathbf{w}_n - \mu \mathbf{u}_n}{\|\mathbf{w}_n - \mu \mathbf{u}_n\|} \quad (4.10)$$

où  $\mu > 0$  représente le pas.

L'algorithme ACI-MBE est résumé comme suit

Pour  $n = 1, \dots, N$

1. Calculer le vecteur  $\mathbf{h}_n$  qui est un vecteur unité obtenu par la procédure d'orthogonalisation de Gram-Schmidt et qui satisfait  $\mathbf{W}_n \mathbf{h}_n = 0$  où  $\mathbf{W}_n = [\mathbf{w}_1, \dots, \mathbf{w}_{n-1}, \mathbf{w}_{n+1}, \dots, \mathbf{w}_N]^H$ .
2. Estimer l'entropie de  $z_n = \mathbf{w}_n^H \mathbf{x}$ . L'estimation de l'entropie des variables aléatoires complexes est basée sur le principe de l'entropie maximale introduit dans [136]. Plusieurs maxima d'entropie sont calculés, puis délimités par des calculs numériques. L'estimation de l'entropie finale est la plus étroite, ce qui est obtenu lorsque la néguentropie estimée est toujours non négative. Pour la procédure d'estimation, un estimateur de l'entropie flexible basé sur la décomposition QR est utilisé.
3. Calculer le gradient conjugué défini par l'équation (4.6)
4. Calculer le gradient conjugué projeté sur la contrainte de surface  $\|\mathbf{w}_n\| = 1$  par l'équation (4.8)
5. Mettre à jour  $\mathbf{w}_n$  par équation (4.10)

L'algorithme ACI-MBE complexe répète la procédure sur différents vecteurs ligne de  $\mathbf{W}$  jusqu'à la convergence donnée par

$$1 - \max(\text{abs}[\text{diag}(\mathbf{W}_n^{\text{nouveau}} \mathbf{W}^H)]) \leq \varepsilon \quad (4.11)$$

### 3.2 Principe d'estimation de l'entropie

Dans le développement de l'algorithme de l'ACI-MBE, les auteurs dans [29] ont proposé une nouvelle procédure pour l'estimation de l'entropie.

Soit la variable aléatoire complexe  $z = z_R + jz_I$ , où  $E[z_R] = E[z_I] = 0$ . Par définition, l'entropie de  $z$  est donné par  $H(z) \triangleq H(z_R, z_I) = -E[\log p(z_R, z_I)]$ . où  $p(z) \triangleq p(z_R, z_I)$  est la densité de probabilité de  $z$  et  $E$  désigne l'espérance.



### 3.2.1 Borne 1 de l'entropie

Les auteurs dans [29] considèrent la décomposition linéaire suivante :

$$\begin{bmatrix} u \\ v \end{bmatrix} = \mathbf{B}^{-1} \begin{bmatrix} z_R \\ z_I \end{bmatrix} \quad (4.12)$$

où  $u$  et  $v$  sont deux variables aléatoires de moyenne nulle. Il est facile d'observer que la décomposition conduit à la limite supérieure suivante :

$$\begin{aligned} H(z) &= \log |\det(\mathbf{B})| + H(u, v) \leq \log |\det(\mathbf{B})| + H(u) + H(v) \\ &= H^{[borne,I]}(z, \mathbf{B}) \end{aligned} \quad (4.13)$$

où l'égalité est valable si et seulement si  $u$  et  $v$  sont statistiquement indépendantes, les bornes de l'entropie sont écrites en fonction de  $\mathbf{B}$ .

### 3.2.2 Borne 2 de l'entropie

Dans ce cas, on considère la décomposition suivante :

$$\begin{bmatrix} z_R \\ z_I \end{bmatrix} = \mathbf{B} \begin{bmatrix} u \\ v \end{bmatrix} = \mathbf{B}r \begin{bmatrix} \cos \theta \\ \sin \theta \end{bmatrix} \quad (4.14)$$

où  $\mathbf{B}$  est une matrice non-singulière,  $r$  et  $\theta$  sont l'amplitude et l'argument de  $u + jv$ , la décomposition conduit à la borne d'entropie suivante :

$$\begin{aligned} H(z) &= \log |\det(\mathbf{B})| + H(u, v) \leq \log |\det(\mathbf{B})| + E[\log r] + H(r, \theta) \\ &= H^{[borne,II]}(z, \mathbf{B}) \end{aligned} \quad (4.15)$$

Dans les deux décompositions données par les équations (4.12) et (4.14), toute mise à l'échelle et changement des signes de  $u$  et  $v$  peut être absorbée par  $\mathbf{B}$ . Ainsi, sans perte de généralité, on suppose toujours que  $\mathbf{B}$  est normalisé tel que

$$\det(\mathbf{B}) > 0, \quad E[u^2] = E[v^2] = 1 \quad \text{et} \quad E[r^2] = 2 \quad (4.16)$$

### 3.3 Estimation de l'entropie par la décomposition QR

Pour une matrice non-singulière  $\mathbf{B}$ , il faut déterminer les borne d'entropie de  $H(u)$  et  $H(v)$ . Comme  $u$  et  $v$  sont normalisées et de moyennes nulles et, on peut utiliser directement la limite d'entropie maximale développée dans [29] qui consiste à calculer plusieurs entropies maximales puis borner l'entropie des estimations en utilisant le calcul numérique.

Le nouvel estimateur de l'entropie suppose qu'il existe  $K^{[I]}$  fonctions de mesure  $G_k^{[I]}(x)$ ,  $k = 1, \dots, K^{[I]}$ , pour la détermination de la première borne d'entropie, et  $K^{[II]}$  fonctions de mesure  $G_k^{[II]}(x)$ ,  $k = 1, \dots, K^{[II]}$ , pour déterminer la seconde.

Pour toute matrice non-singulière  $\mathbf{B}$  satisfaisant la condition d'échelle donnée en (4.16), on peut obtenir les  $(K^{[I]})^2$  bornes d'entropie du premier type et  $K^{[II]}$  du second type par :

$$\begin{aligned} H_{k_1, k_2}^{[borne, I]}(z, \mathbf{B}) \\ = \log \det(\mathbf{B}) + \log(2\pi e) - V_{k_1}^{[I]} \left\{ E \left[ G_{k_1}^{[I]}(u) \right] \right\} \\ - V_{k_2}^{[I]} \left\{ E \left[ G_{k_2}^{[I]}(v) \right] \right\} \end{aligned} \quad (4.17)$$

et

$$H_k^{[borne, II]}(z, \mathbf{B}) = \log \det(\mathbf{B}) + \log(2\pi e) - V_k^{[II]} \left\{ E \left[ G_k^{[II]}(r) \right] \right\} \quad (4.18)$$

où  $k_1, k_2 = 1, \dots, K^{[I]}$ ,  $k = 1, \dots, K^{[II]}$ , et  $V_{k_1}^{[I]} \left\{ E \left[ G_{k_1}^{[I]}(u) \right] \right\}$  et  $V_{k_2}^{[I]} \left\{ E \left[ G_{k_2}^{[I]}(v) \right] \right\}$

représentent la négentropie de  $G_k^{[I]}$  et  $G_k^{[II]}$  respectivement. L'estimation de l'entropie finale est la plus étroite lorsque la négentropie estimée est toujours non négative.

Parmi toutes ces bornes d'entropie possibles, la plus étroite est utilisée comme estimation finale de  $H(z)$  définie par

$$\hat{H}(z) = \min \left( \min_{1 \leq k_1, k_2 \leq K^{[I]}} H_{k_1, k_2}^{[borne, I]}(z), \min_{1 \leq k \leq K^{[II]}} H_k^{[borne, II]}(z) \right) \quad (4.19)$$

où

$$H_{k_1, k_2}^{[borne, I]}(z) = \min_{\mathbf{B}} H_{k_1, k_2}^{[borne, I]}(z, \mathbf{B}) \quad (4.20)$$

et

$$H_k^{[borne, II]}(z) = \min_B H_k^{[borne, II]}(z, B) \quad (4.21)$$

L'estimateur d'entropie donné dans l'équation (4.19) exige d'abord la solution optimale  $\mathbf{B}$ , ce qui n'est pas pratique pour une mise en œuvre de l'algorithme de l'ACI. En considérant la décomposition QR tel que  $\mathbf{B} = \mathbf{Q}\mathbf{R}$ , où  $\mathbf{Q}$  est une matrice orthogonale de dimension 2x2, et  $\mathbf{R}$  une matrice triangulaire inférieure de même dimension, on constate que  $\mathbf{Q}$  peut être absorbée par la matrice de separation et qu'on peut obtenir une expression explicite pour  $\mathbf{R}$ . Cette observation simplifie grandement l'utilisation de l'estimateur d'entropie donné par l'équation (4.19). Les éléments diagonaux de  $\mathbf{R}$  doivent être positifs de sorte que la décomposition QR soit unique puisque  $\mathbf{B}$  est non-singulière.

En outre, on a  $\det(\mathbf{Q}) = 1$ , puisque  $\det(\mathbf{R}) = 1$  et  $\det(\mathbf{B}) = 1$  sont positifs, conduisant à un  $\det(\mathbf{Q})$  positif. Par conséquent  $\mathbf{Q}$  représente aussi une matrice de rotation.

### 3.3.1 Absorption de $\mathbf{Q}$

Sachant que  $\mathbf{z} = \mathbf{W}^H \mathbf{x}$  dénote la source estimée. A partir des équations précédentes, on peut écrire

$$\begin{bmatrix} u \\ v \end{bmatrix} = \mathbf{R}^{-1} \mathbf{Q}^T \begin{bmatrix} z_R \\ z_I \end{bmatrix} \quad (4.22)$$

où  $\mathbf{Q}$  est définie par

$$\mathbf{Q} = \begin{bmatrix} \cos \vartheta & -\sin \vartheta \\ \sin \vartheta & \cos \vartheta \end{bmatrix} \quad (4.23)$$

où  $\vartheta$  est l'angle de rotation. Alors on peut écrire

$$\mathbf{Q}^T \begin{bmatrix} z_R \\ z_I \end{bmatrix} = \begin{bmatrix} \text{Re}[\exp(j\vartheta) z] \\ \text{Im}[\exp(j\vartheta) z] \end{bmatrix} = \begin{bmatrix} \text{Re}\{[\exp(j\vartheta) \mathbf{w}]^H \mathbf{x}\} \\ \text{Im}\{[\exp(j\vartheta) \mathbf{w}]^H \mathbf{x}\} \end{bmatrix} \quad (4.24)$$

On constate que la matrice de rotation orthogonale  $\mathbf{Q}$  peut être absorbée par le vecteur de séparation  $\mathbf{w}$ . Comme  $\mathbf{w}$  est inconnu et à estimer à partir des mélanges, les auteurs dans [29] supposent simplement que  $\mathbf{Q} = \mathbf{I}$  et  $\mathbf{B} = \mathbf{R}$ .

### 3.3.2 Expression explicite de $\mathbf{R}$

La matrice triangulaire inférieure  $\mathbf{R}$  peut être déterminée de façon unique par les statistiques de second ordre de  $\mathbf{z}$  et les conditions de mise à l'échelle données dans (4.16).

Dans le premier type de distribution de l'entropie maximale,  $u$  et  $v$  sont statistiquement indépendants, ce qui implique que  $u$  et  $v$  sont décorrelés.

Dans le second type,  $u + jv$  est strictement circulaire, ce qui implique que  $u + jv$  est circulaire de second ordre. Ainsi, dans les deux distributions de l'entropie maximale,  $u$  et  $v$  sont décorrelés, c'est-à-dire  $E[uv] = 0$  sachant que  $u$  et  $v$  sont également de moyennes nulles. Puis, en considérant la décomposition

$$\begin{bmatrix} z_R \\ z_I \end{bmatrix} = \mathbf{R} \begin{bmatrix} u \\ v \end{bmatrix} \quad (4.25)$$

et la condition de mise à l'échelle donnée par l'équation (4.16), on a :

$$E \left\{ \begin{bmatrix} z_R \\ z_I \end{bmatrix} \begin{bmatrix} z_R & z_I \end{bmatrix} \right\} = \begin{bmatrix} \sigma_R^2 & \rho \\ \rho & \sigma_R^2 \end{bmatrix} = \mathbf{R}\mathbf{R}^T \quad (4.26)$$

Le produit  $\mathbf{R}\mathbf{R}^T$  définit une factorisation de Cholesky puisque  $\mathbf{R}$  est une matrice triangulaire inférieure et la matrice de covariance est une matrice symétrique et définie positive. On peut résoudre (4.26) pour obtenir

$$R^{-1} = \begin{bmatrix} \frac{1}{\sigma_R} & 0 \\ -\frac{\rho}{\sigma_R\sqrt{\Delta_1}} & \frac{\sigma_R}{\sqrt{\Delta_1}} \end{bmatrix} \quad (4.27)$$

où  $\Delta_1 = \sigma_R^2 \sigma_I^2 - \rho^2$ ,

Par conséquent, toutes les expressions explicites de  $v$  et  $r$  peut être facilement obtenues

par  $u = \frac{z_R}{\sigma_R}$ ,  $v = -\frac{\rho z_R}{\sigma_R\sqrt{\Delta_1}} + \frac{\sigma_{RZ_I}}{\sqrt{\Delta_1}}$  et  $r = \sqrt{u^2 + v^2} = \sqrt{\frac{\Delta_1}{\Delta_2}}$  et  $\Delta_2 = z_R^2 \sigma_I^2 + z_I^2 \sigma_R^2 + 2z_R z_I \rho$ .

Les bornes de l'entropie deviennent

$$H_{k_1, k_2}^{[borne, I]}(z) = 0.5 \log(\Delta_1) + \log(2\pi e) - V_{k_1}^{[I]} \left\{ E \left[ G_{k_1}^{[I]}(u) \right] \right\} - V_{k_2}^{[I]} \left\{ E \left[ G_{k_2}^{[I]}(v) \right] \right\} \quad (4.28)$$

et

$$H_k^{[borne, II]}(z) = 0.5 \log(\Delta_1) + \log(2\pi e) - V_k^{[II]} \left\{ E \left[ G_k^{[II]}(r) \right] \right\} \quad (4.29)$$

## 4 Analyse en composantes indépendantes par minimisation des bornes de l'entropie et décomposition en modes empiriques bivariée

### 4.1 Principe

Le mélange convolutif dans le domaine temporel est transformé en un mélange instantané dans le domaine fréquentiel :

$$\mathbf{X}(f) = \mathbf{A}(f)\mathbf{S}(f) \quad (4.30)$$

où  $f$  désigne la fréquence,  $\mathbf{A}(f)$  est une matrice complexe  $M \times N$  dont les éléments  $a_{ij}$  sont les réponses fréquentielles de la source  $i$  au capteur  $j$ ,  $\mathbf{X}(f)$  est un vecteur complexe  $M \times 1$  dont les éléments  $\mathbf{X}_j(f)$  sont les transformées de Fourier des signaux de mélange  $x_j(t)$ . Et  $\mathbf{S}(f)$  est un vecteur complexe de dimension  $N \times 1$  dont les éléments  $\mathbf{S}_i(f)$  sont les transformées de Fourier des signaux de source  $s_i(t)$ .

Etant donné que la séparation du signal est effectuée fréquence par fréquence, on utilise une transformée de Fourier de courte terme (TFCT). Le modèle de mélange de chaque trame peut être écrit comme suit

$$\mathbf{X}(f, m) = \mathbf{A}(f)\mathbf{S}(f, m) \quad (4.31)$$

où  $m$  désigne l'indice de trame,  $\mathbf{X}(f, m)$  et  $\mathbf{S}(f, m)$  sont des vecteurs complexes correspondant à la trame  $m$ :  $\mathbf{S}(f, m) = [S_1(f, m) \ \cdots \ S_N(f, m)]^T$   
 $\mathbf{X}(f, m) = [X_1(f, m) \ \cdots \ X_M(f, m)]^T$ . La solution du problème de SAS dans le cas des mélanges convolutifs consiste à rechercher pour chaque intervalle de fréquences la matrice de séparation  $\mathbf{W}(f)$  qui fournit une estimation des signaux sources:

$$\mathbf{Z}(f, m) = \mathbf{W}(f)\mathbf{X}(f, m), \quad (4.32)$$

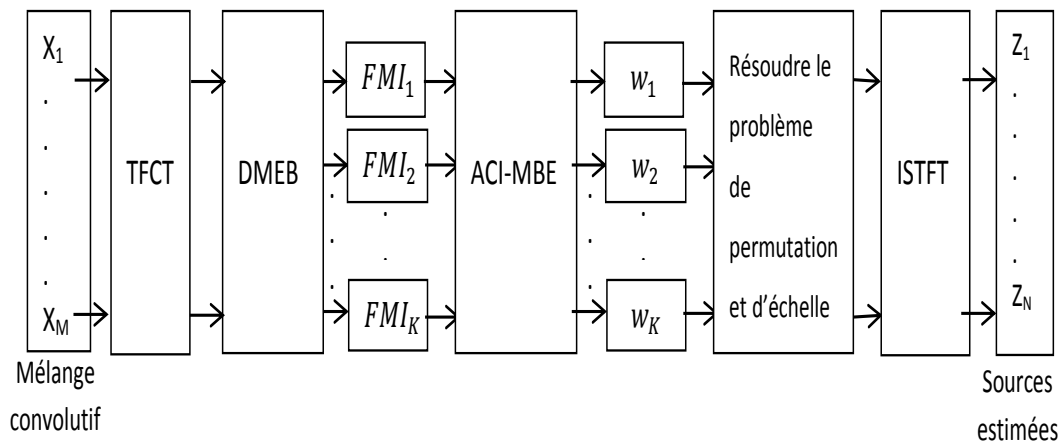
où  $\mathbf{Z}(f, m)$  désigne l'estimation du vecteur  $\mathbf{S}(f, m)$  des signaux sources  $\mathbf{Z}(f, m) = [Z_1(f, m), \dots, Z_N(f, m)]^T$

Le problème de la SAS pour les mélanges convolutifs peut être résolu dans le domaine fréquentiel en utilisant la version complexe de la méthode de l'ACI [100, 102] combinée avec l'algorithme de la DMEB. Tout d'abord, les mélanges convolutifs dans le domaine fréquentiel sont décomposés en un ensemble de FMIs à l'aide de l'algorithme de DMEB, la principale caractéristique de la DMEB est qu'elle aboutit à un nombre égal des FMIs des parties réelle et imaginaire du signal. La DMEB est appliquée dans le domaine fréquentiel pour décomposer les parties réelles et imaginaires de la transformée de Fourier de chaque trame des mélanges observés en un ensemble de FMIs. Puis l'algorithme de l'ACI complexe est appliqué pour extraire les sources indépendantes.

La méthode proposée est résumée par l'organigramme représenté par la figure 4.2. Les différentes étapes de la méthode combinant la DMEB et l'ACI-MBE sont les suivantes [137]:

1. Calculer la transformée de Fourier de chaque trame des signaux de mélange observés. La longueur de trame est fixée à 1024 échantillons.
2. Appliquer la DMEB à la transformée de Fourier de chaque trame des mélanges observés. Ainsi, pour chaque trame, nous obtenons deux ensembles de FMIs correspondant aux parties réelles et imaginaires de la transformée de Fourier.
3. Extraire les composantes indépendantes dans le domaine fréquentiel en utilisant l'algorithme ACI-MBE.
4. Résoudre l'ambiguïté de permutation.
5. Calculer la transformée de Fourier inverse pour obtenir les composantes indépendantes dans le domaine temporel.

Plusieurs algorithmes ont été proposés pour résoudre l'ambiguïté de permutation [68-69,138]. Dans ce travail, nous utilisons la méthode proposée dans [61].



**Figure 4.2** Séparation aveugle de sources combinant la décomposition modale empirique bivariée (DMEB) et l'analyse en composantes indépendantes complexe par minimisation des bornes de l'entropie (ACI-MBE).

## **5 Résolution de l'ambiguïté de permutation**

Afin de surmonter le problème de permutation, la méthode proposée dans [61] est utilisée. Cette méthode permet la résolution de l'indétermination de la permutation dans les bandes de fréquence et permet une convergence rapide et un réajustement facile.

Lorsque les signaux observés sont convertis dans le domaine fréquentiel, leurs spectres changent progressivement le long de l'axe des fréquences, ce qui signifie que si les bandes de fréquences sont suffisamment étroites, les spectres entre les bandes voisines sont fortement corrélés. Par conséquent, on peut s'attendre à ce que les matrices de séparation obtenues dans des compartiments de fréquence adjacents n'aient pas de grands changements dans leurs coefficients, ainsi que leurs ordres de permutation. De ce fait, on utilise la solution finale de la matrice de séparation dans le compartiment de fréquence précédent comme valeur initiale de l'itération dans le compartiment de fréquence courant. Si la résolution dans le domaine fréquentiel est suffisamment élevée, les matrices de séparation dans des bandes de fréquences consécutives tendront à converger dans le même ordre de permutation. Ce qui signifie que l'étape de résolution de la permutation peut être évitée. L'itération peut être effectuée de deux façons toutes aussi efficaces:

1. Commencer à partir de la fréquence la plus basse.
2. Commencez par la fréquence la plus haute.

La méthode 2 est légèrement plus rapide que la méthode 1, car elle converge plus rapidement dans les compartiments des fréquences les plus basses.

## **6 Résultats et discussions**

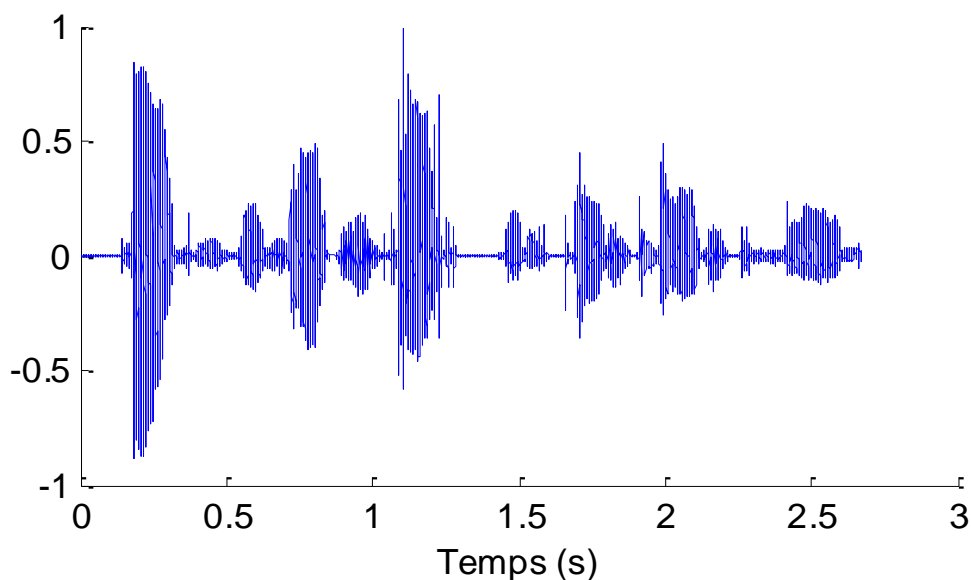
A titre d'exemple d'illustration, la méthode proposée est utilisée pour séparer un mélange convolutif de deux signaux de parole. Les figures 4.3-a et 4.3-b montrent les



signaux sources et les figures 4.3-c et 4.3-d montrent leurs spectrogrammes respectifs. Les sources sont convolutivement mélangées en utilisant la matrice de mélange décrite par sa transformation en  $z$  :

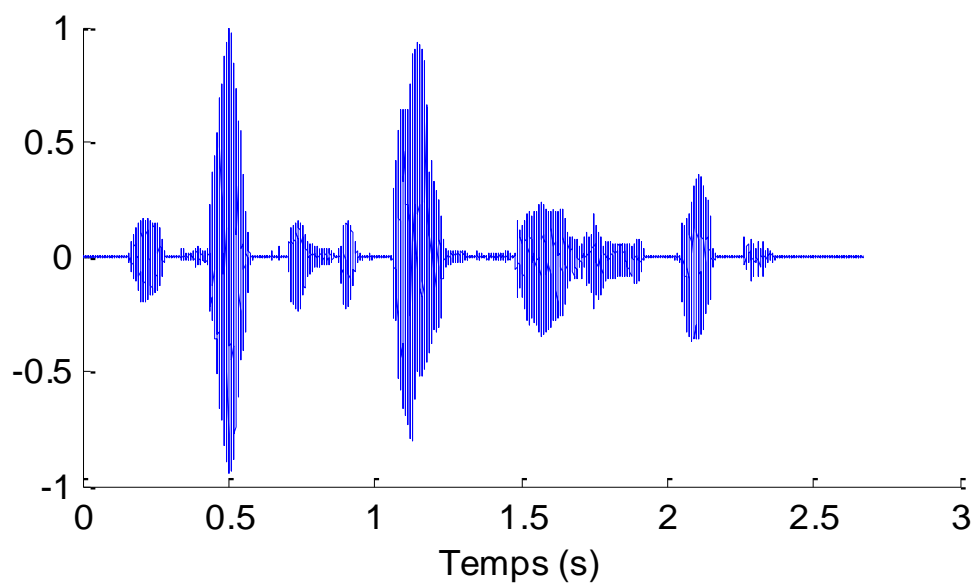
$$\mathbf{A}(z) = \begin{bmatrix} 1 + .08z^{-1} + 0.5z^{-2} & 0.8 + 0.7z^{-1} + 0.4z^{-2} \\ 0.7 + z^{-1} + 0.6z^{-2} & 1 + 0.9z^{-1} + 0.3z^{-2} \end{bmatrix} \quad (4.33)$$

Les signaux de mélange observés sont représentés par les figure 4.3-e et 4.3-f. Les parties réelles et imaginaires de la transformée de Fourier des trames extraites des deux signaux de mélange et leurs FMIs correspondants sont représentées par la figure 4.4. Les parties réelles et imaginaires des sources estimées correspondant aux deux trames de signaux de mélange dans le domaine fréquentiel sont montrées sur la figure 4.5 et les sources estimées sur la figure 4.6. Comme illustré, les signaux estimés sont très semblables aux signaux sources. La DMEB combinée avec l'ACI-MBE complexe fournit une estimation précise des signaux sources.

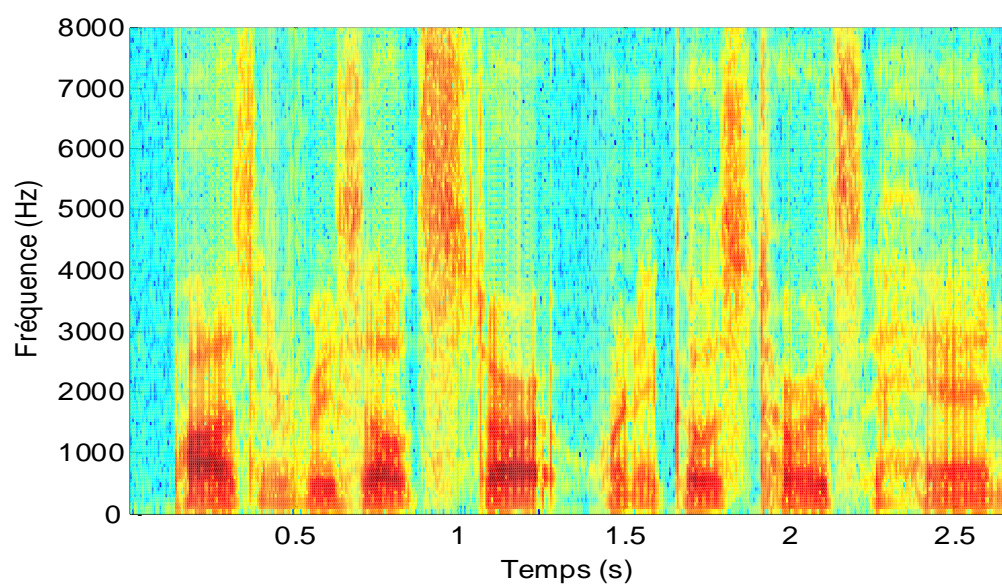


(a) Source 1

**Figure 4.3** Exemple d'illustration de la séparation aveugle de mélanges convolutifs de deux signaux de parole par DMEB combinée avec ACI-MBE

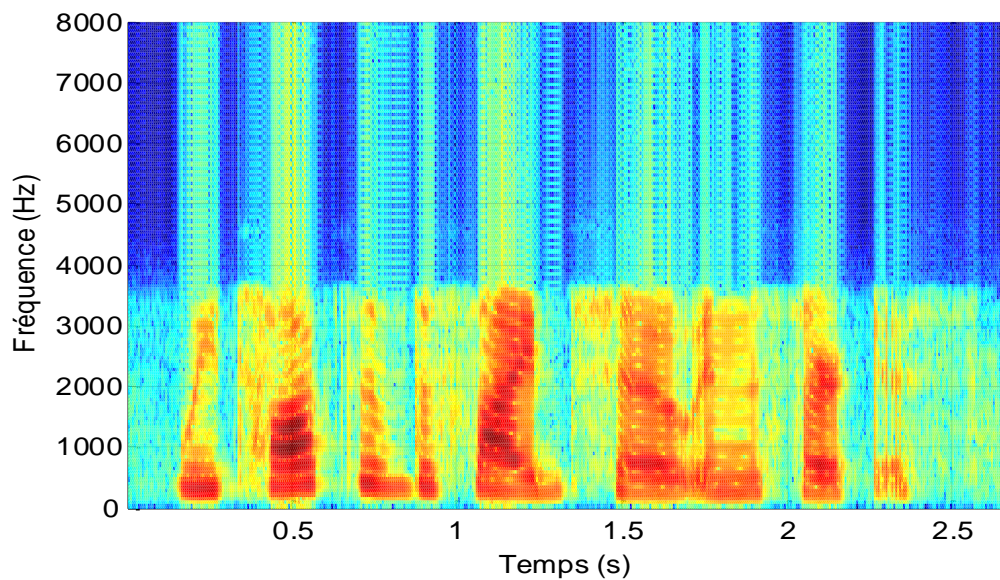


(b) Source 2

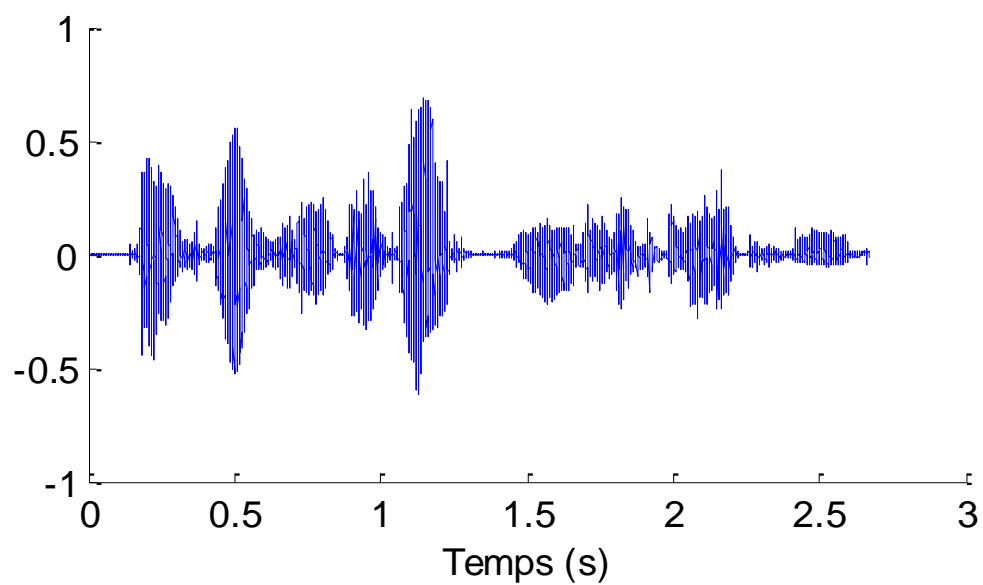


(c) Spectrogramme de la source 1

**Figure 4.3** (Suite)

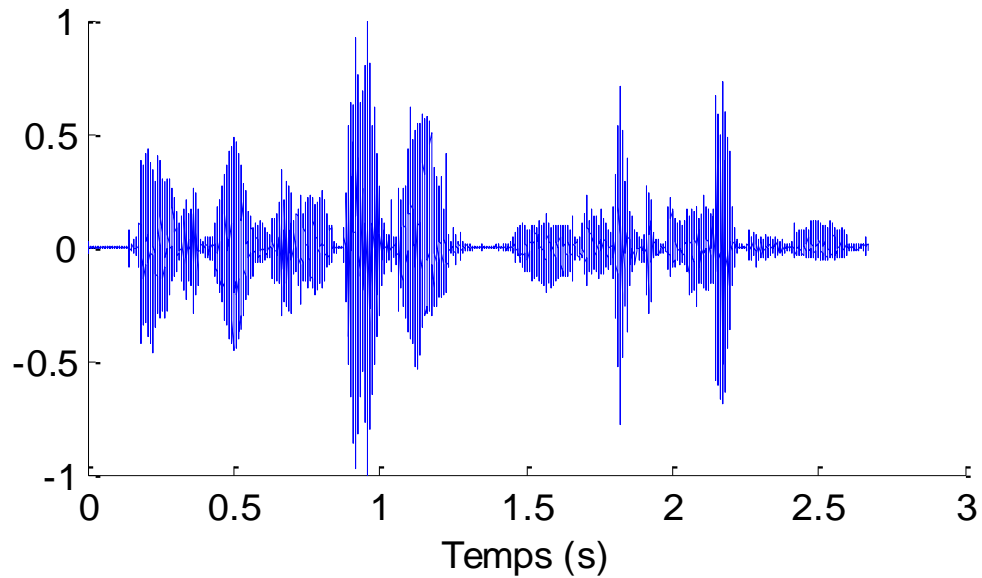


(d) Spectrogramme de la source 2



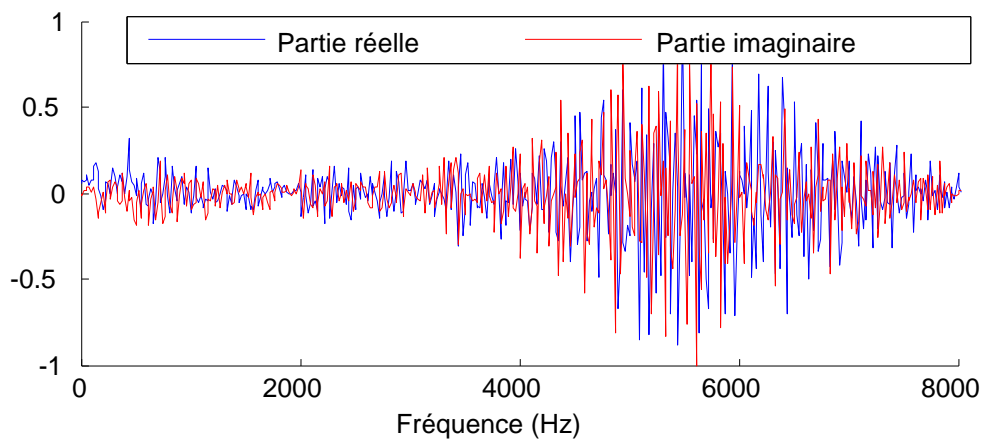
(e) Signal observé 1

**Figure 4.3** (suite)



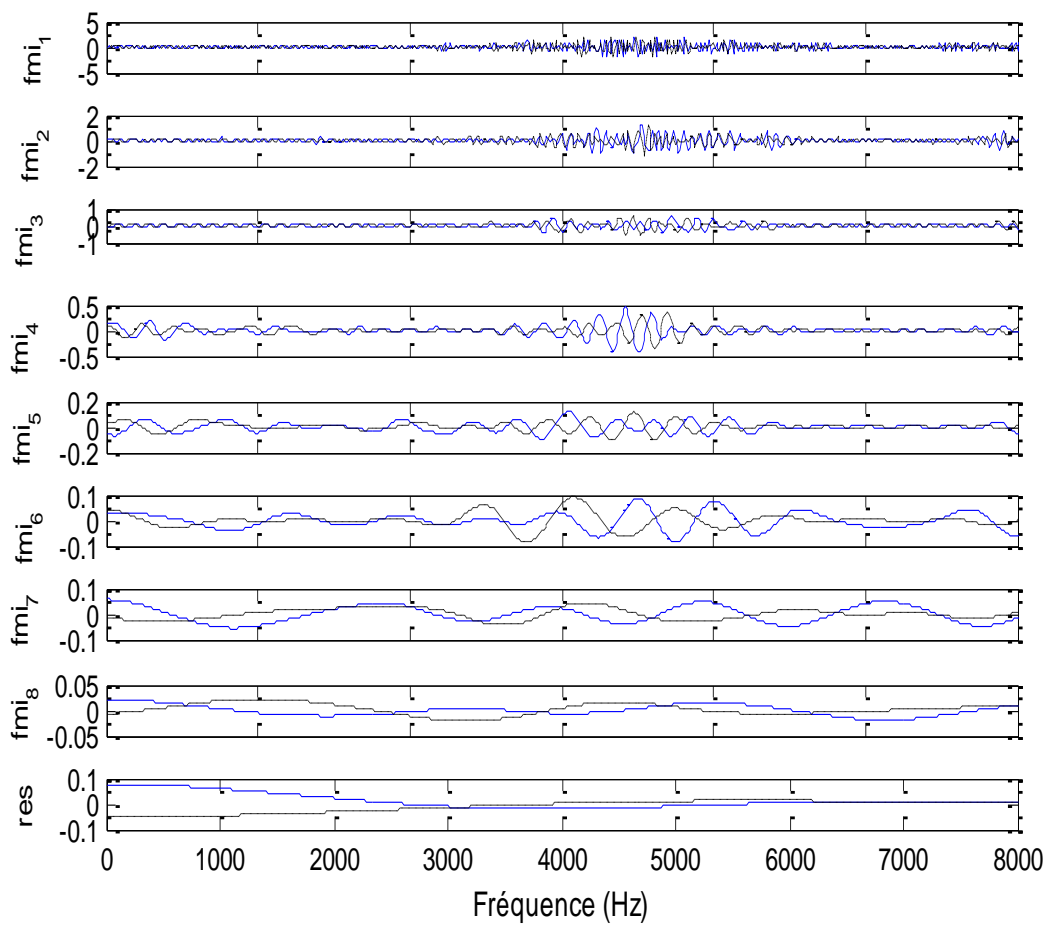
(f) Signal observé 2

**Figure 4.3** (suite)

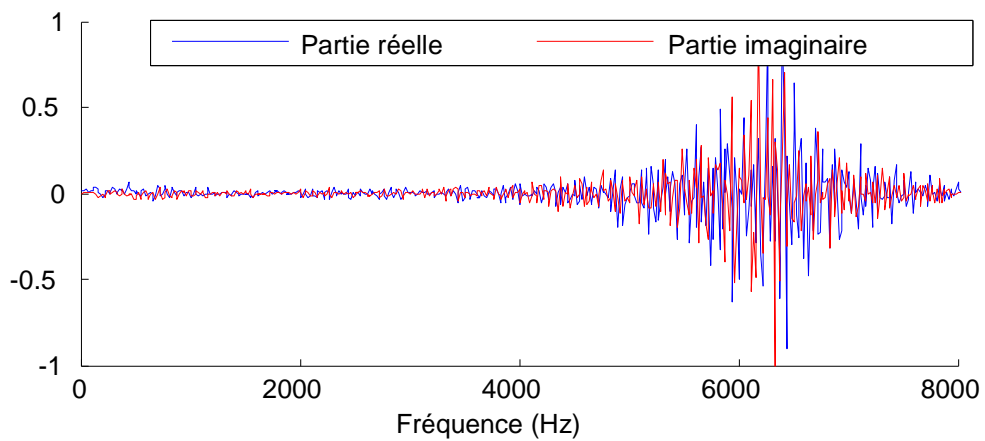


(a) Transformée de Fourier d'une trame du signal observé 1

**Figure 4.4** Parties réelles et imaginaires de la transformée de Fourier des trames de 1024 échantillons extraites du mélange et leurs FMIs correspondantes.

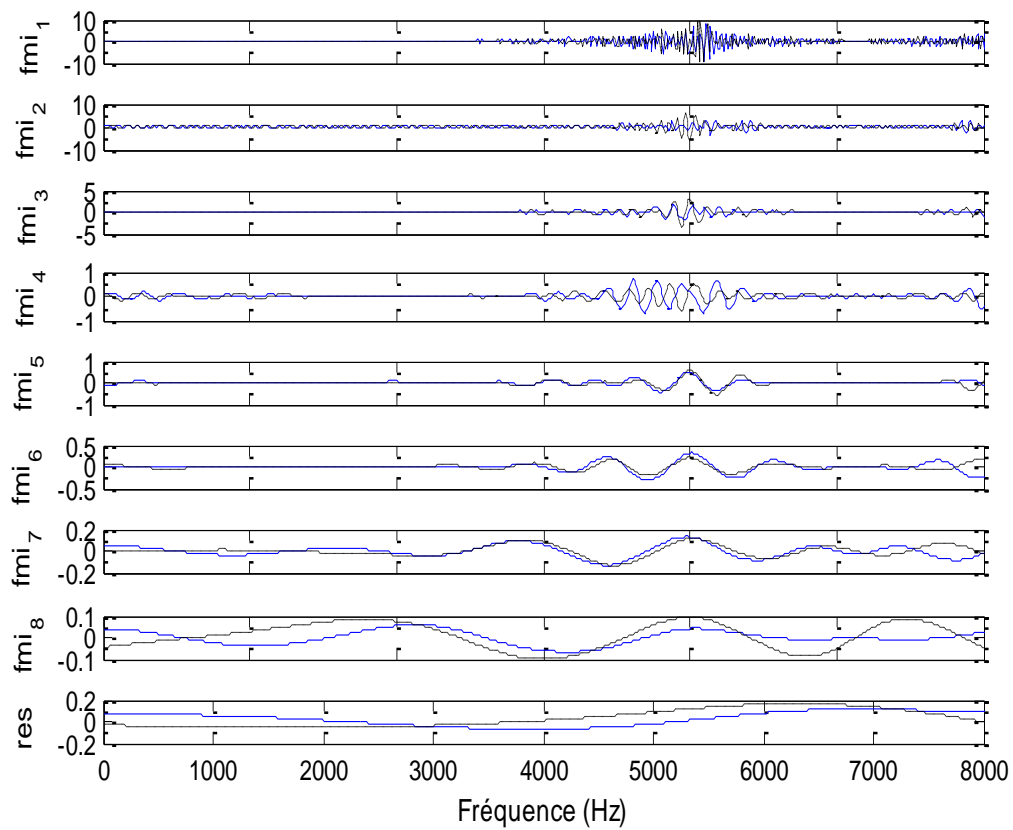


(b) FMIs et résidu de la TF d'une trame du signal observé 1



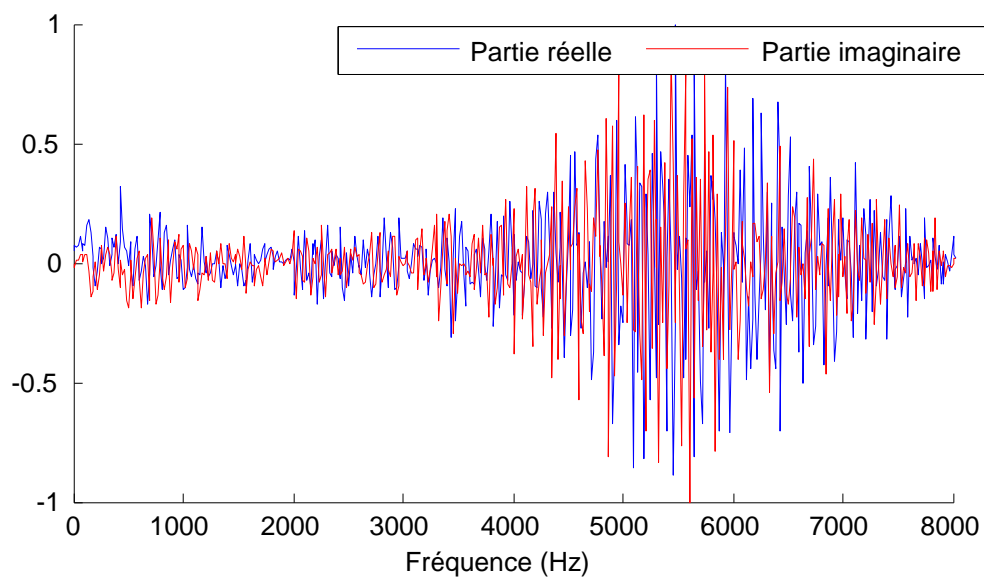
(c) Transformée de Fourier d'une trame du signal observé 2

**Figure 4.4** (suite)



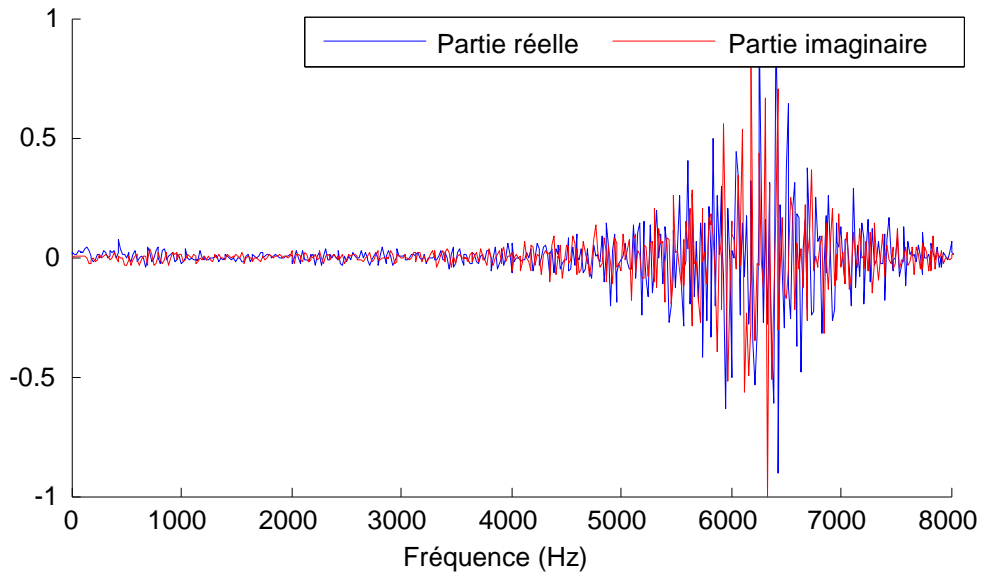
(d) FMIs et résidu de la transformée de Fourier d'une trame du signal observé 2

Figure 4.4 (suite)



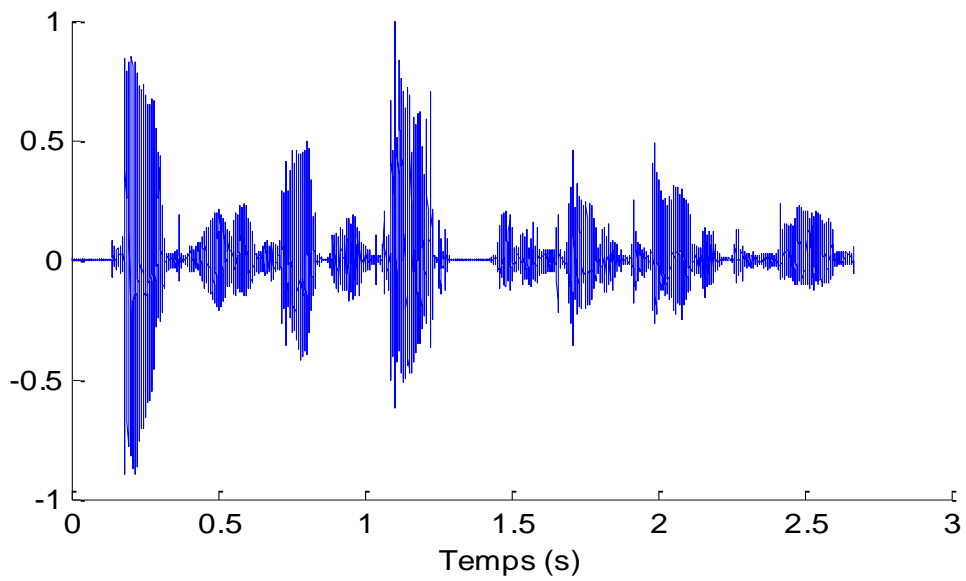
(a) Trame estimée de la source 1

Figure 4.5 Estimation des trames des sources 1 et 2 dans le domaine fréquentiel.



(b) Trame estimée de la source 2

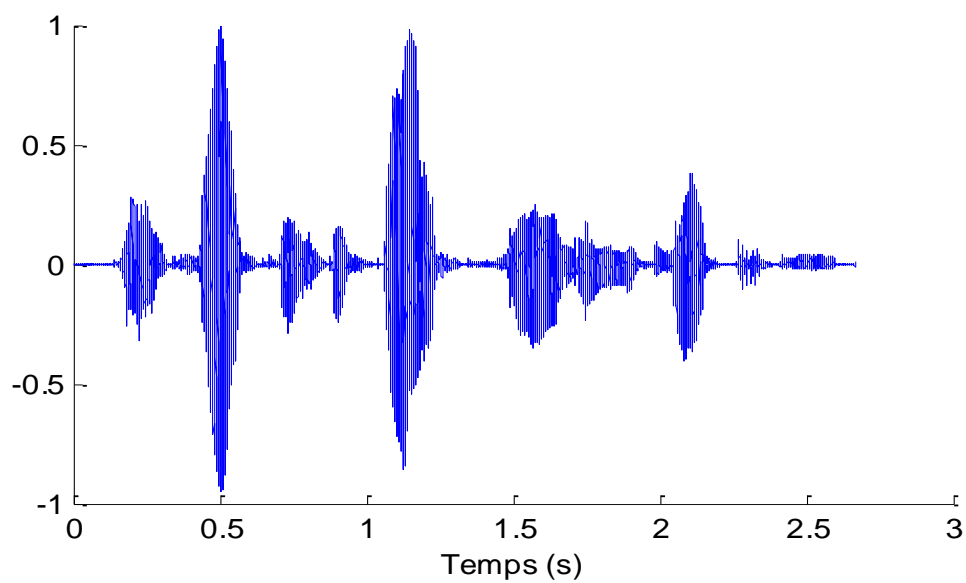
**Figure 4.5** (suite)



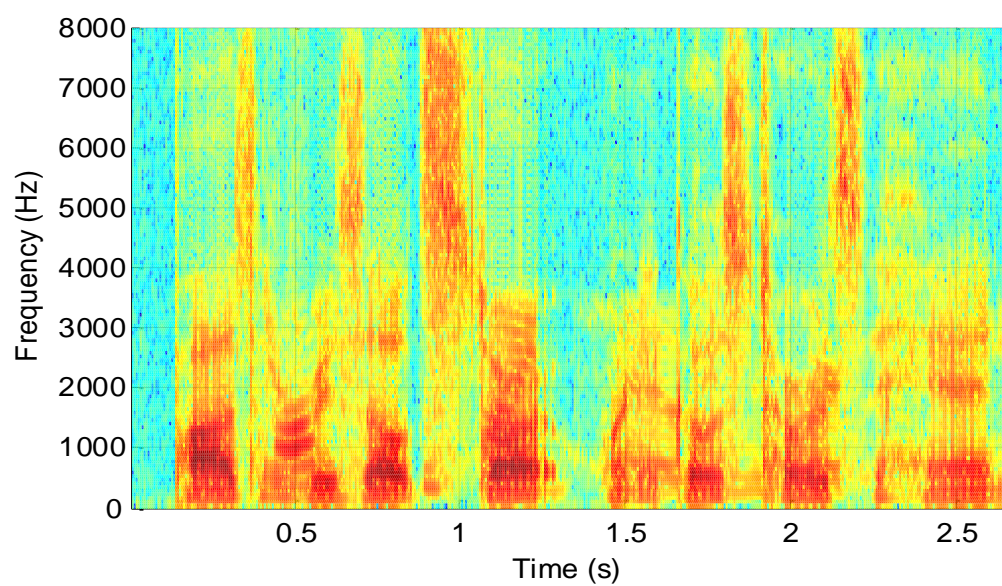
(a) Source estimée 1

**Figure 4.6** Sources estimées et leurs spectrogrammes.





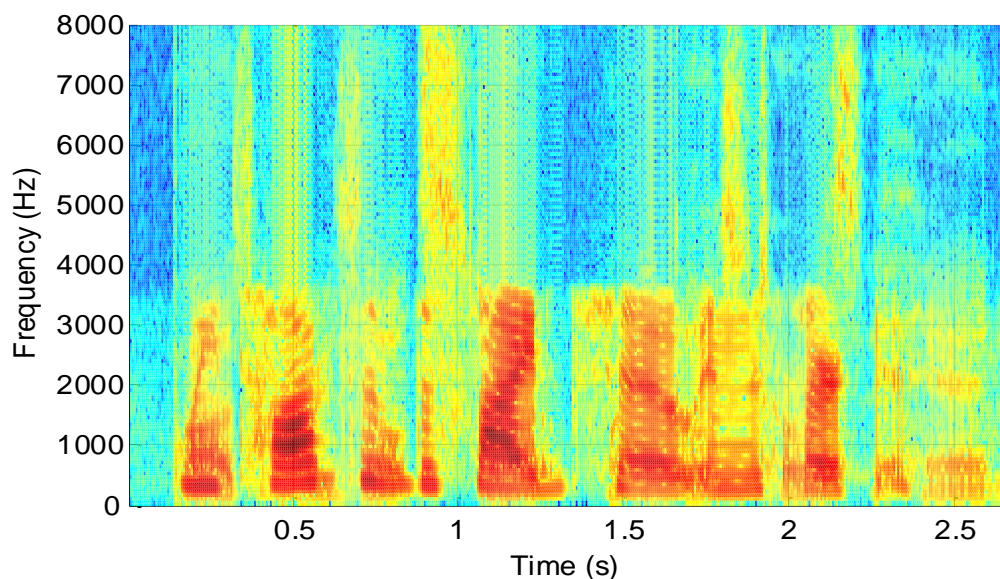
(b) Sources estimée 2



(c) Spectrogramme de la source estimée 1

**Figure 4.6** (suite)

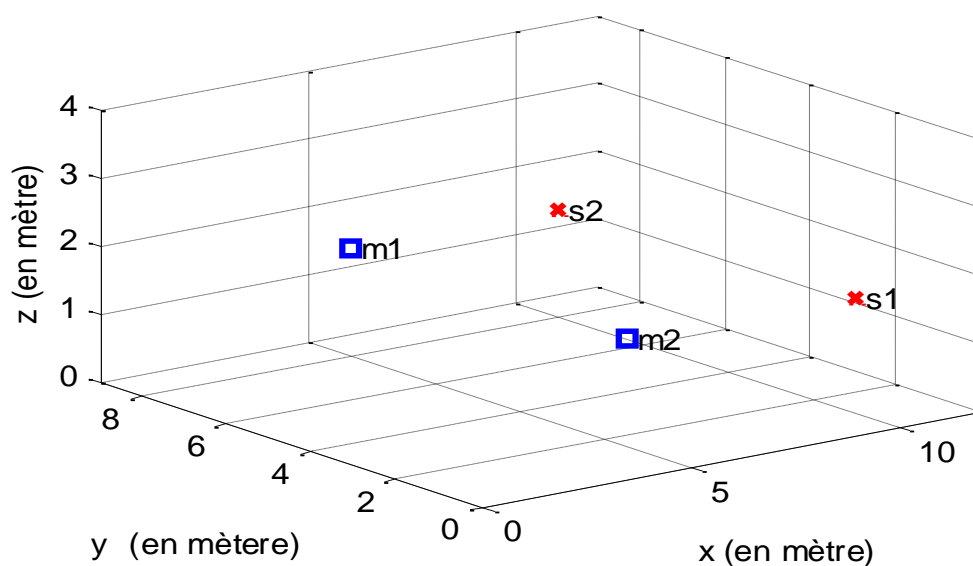




(d) Spectrogram de la source estimée 2

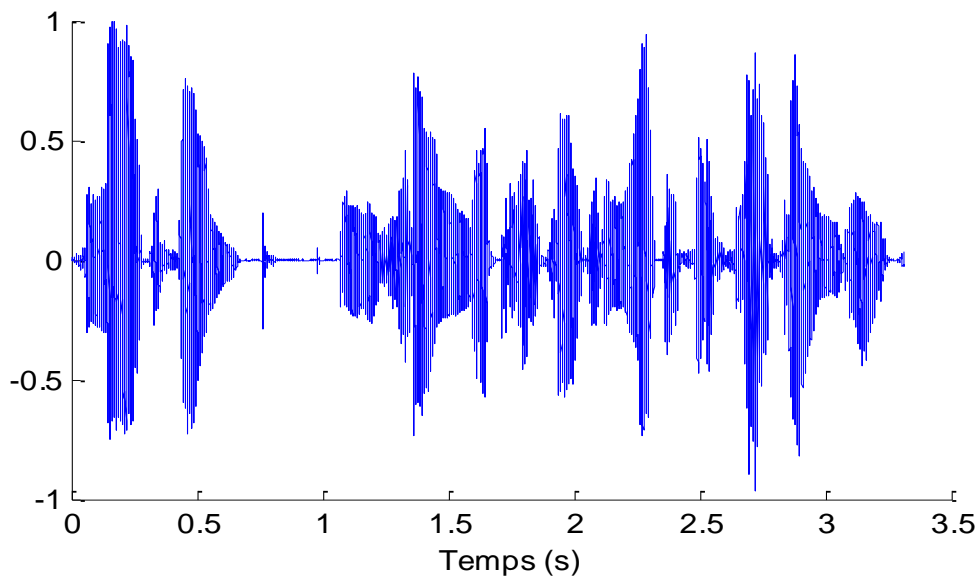
**Figure 4.6** (suite)

Dans cet exemple, les signaux audio sont mélangés selon le modèle décrit dans [139]. Les signaux observés sont acquis par deux microphones repartis dans une pièce. Le milieu de propagation est réverbérant. La localisation des sources  $s_1$  et  $s_2$  et des microphones  $m_1$  et  $m_2$  dans la pièce est illustrée par la figure 4.7

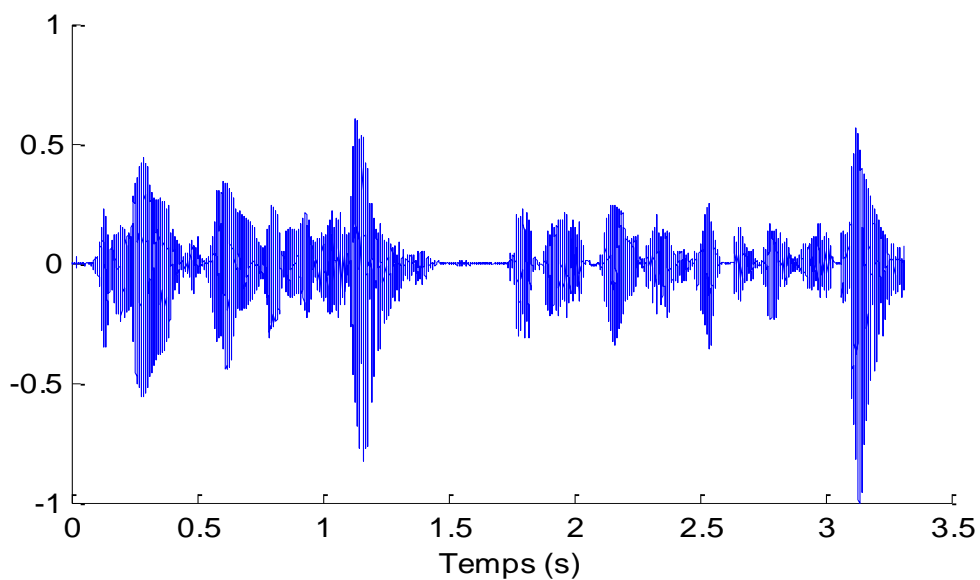


**Figure 4.7** Configuration spatiale: localisation des sources et des capteurs dans la pièce (x, y et z en mètre)

Les figures 4.8-a et 4.8-b montrent les signaux sources, les figures 4.8-c et 4.8-d montrent les signaux observés. Les signaux estimés sont donnés par les figures 4.8-e et 4.8-f.

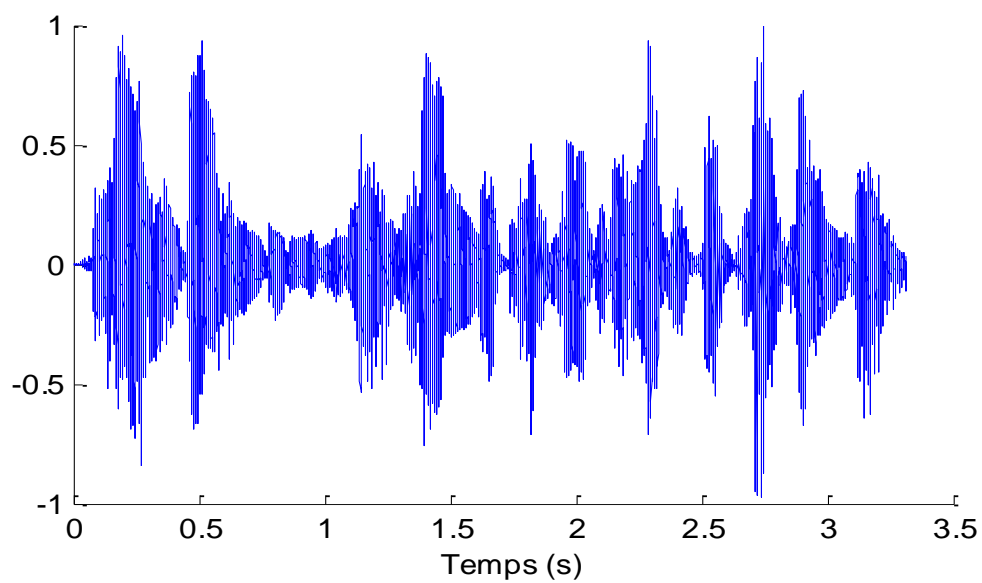


(a) Source 1

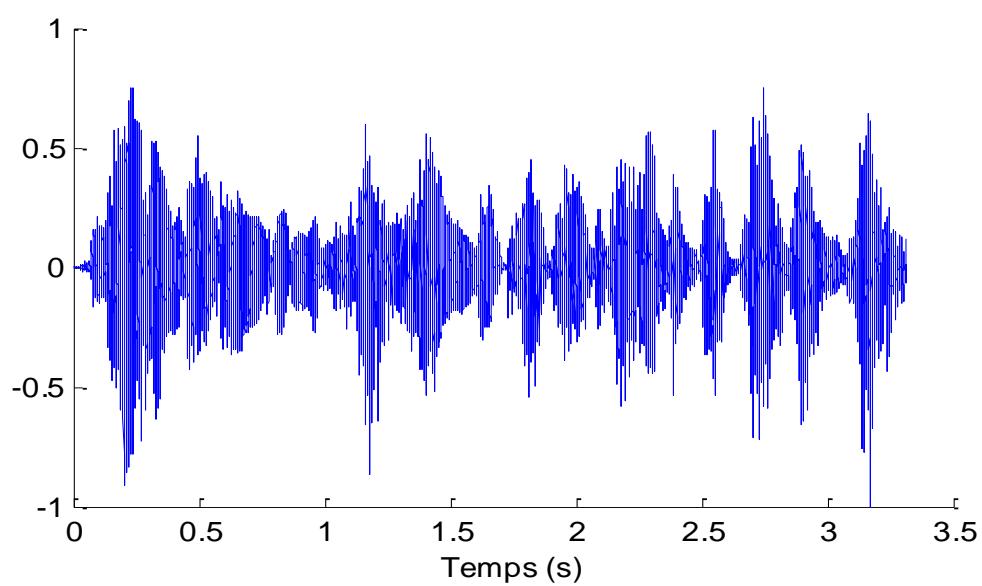


(b) Source 2

**Figure 4.8** Exemple de séparation aveugle de mélange convolutif dans un milieu réverbérant.

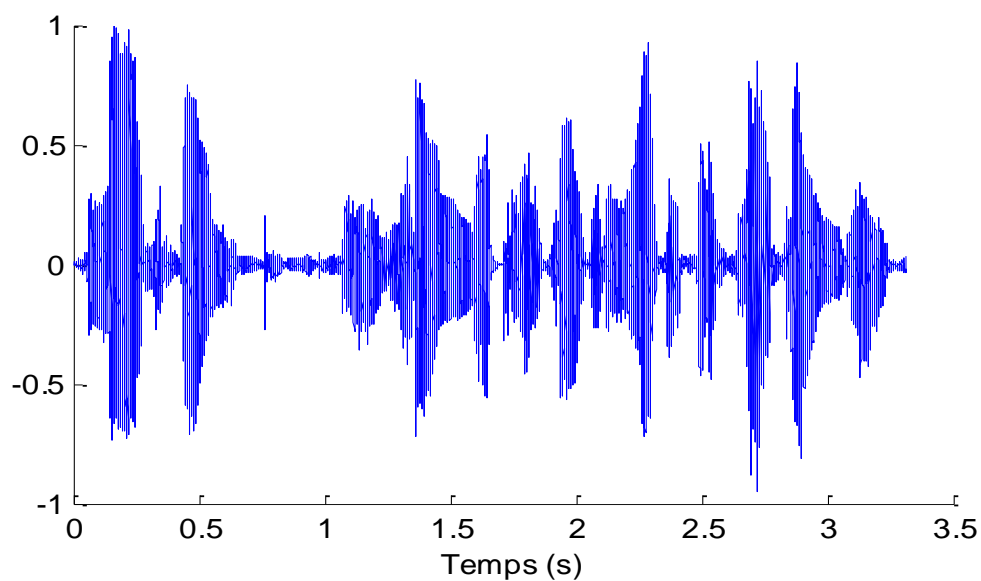


(c) Signal observée 1

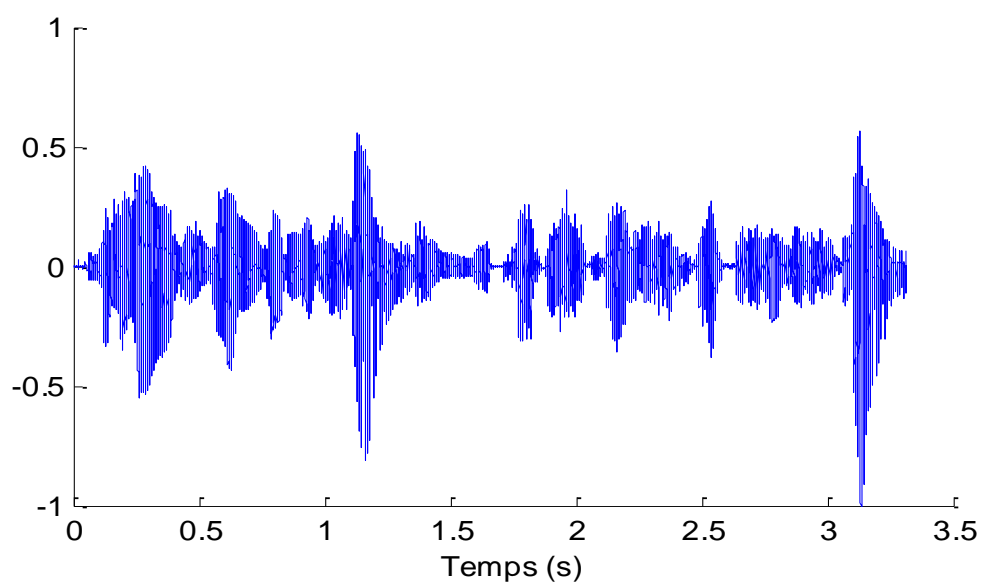


(d) Signal observée 2

**Figure 4.8** (suite)



(e) Source estimée 1



(f) Source estimée 2

**Figure 4.8** (suite)

Les performances de l'approche proposée sont évaluées et comparées aux performances de la méthode conventionnelle de séparation dans le domaine fréquentiel (ACIF) et aux performances de l'algorithme combiné DMEB-ACI proposé dans [140] que

nous avons adaptés aux mélanges convolutifs. Deux ensembles de signaux acoustiques comprenant des mélanges convolutifs sont construits. Le premier ensemble de signaux audio est construit à partir de la base de données TIMIT disponible en ligne [141] en simulant des mélanges convolutifs en utilisant les enregistrements de deux phrases de 3 secondes échantillonnées à 16 kHz prononcées par des locuteurs masculins et féminins.

Un ensemble de 4 mélanges bruyants est simulé en corrompant des mélanges propres à un rapport signal sur bruit (RSB) en anglais signal-to-noise ratio (SNR) allant de 5 dB à 20 dB avec un pas de 5 dB. Le deuxième ensemble de données de parole est construit en simulant des mélanges convolutifs utilisant 20 paires de phrases choisies au hasard dans la base de données NOIZEUS [142]. Les mélanges de la parole sont corrompus par le bruit blanc gaussien additif avec un RSB fixé à 10 dB.

Les performances de la technique proposée sont analysées à l'aide de BSSeval toolbox [143]. Pour la mesure objective des critères de performance, les sources estimées sont exprimées en  $\hat{s} = s_{target} + e_{interf} + e_{noise} + e_{artif}$ , où  $s_{target}$  représente les signaux source, et  $e_{interf}$  désigne les interférences provenant d'autres sources,  $e_{noise}$  est la distorsion causée par le bruit, et  $e_{artif}$  comprend tous les autres artefacts introduits par l'algorithme de séparation. Les critères de performance sont les suivants:

$$SDR = 10 \log_{10} \frac{\|s_{target}\|^2}{\|e_{interf} + e_{noise} + e_{artif}\|^2} \quad (4.34)$$

$$SAR = 10 \log_{10} \frac{\|s_{target} + e_{interf} + e_{noise}\|^2}{\|e_{artif}\|^2} \quad (4.35)$$

et

$$SIR = 10 \log_{10} \frac{\|s_{target}\|^2}{\|e_{interf}\|^2} \quad (4.36)$$

Afin de mesurer la distorsion entre les sources originales et les sources estimées, le rapport signal sur bruit d'amélioration (improvement signal to noise ration, ISNR) est calculé. L'ISNR est défini comme

$$ISNR = SNR_{sortie} - SNR_{entrée} \quad (4.37)$$

Le SNR d'entrée et de sortie est défini comme

$$SNR_{entrée} = 10 \log \frac{\sum_t |s(t)|^2}{\sum_t |x(t) - \hat{s}(t)|^2} \quad (4.48)$$

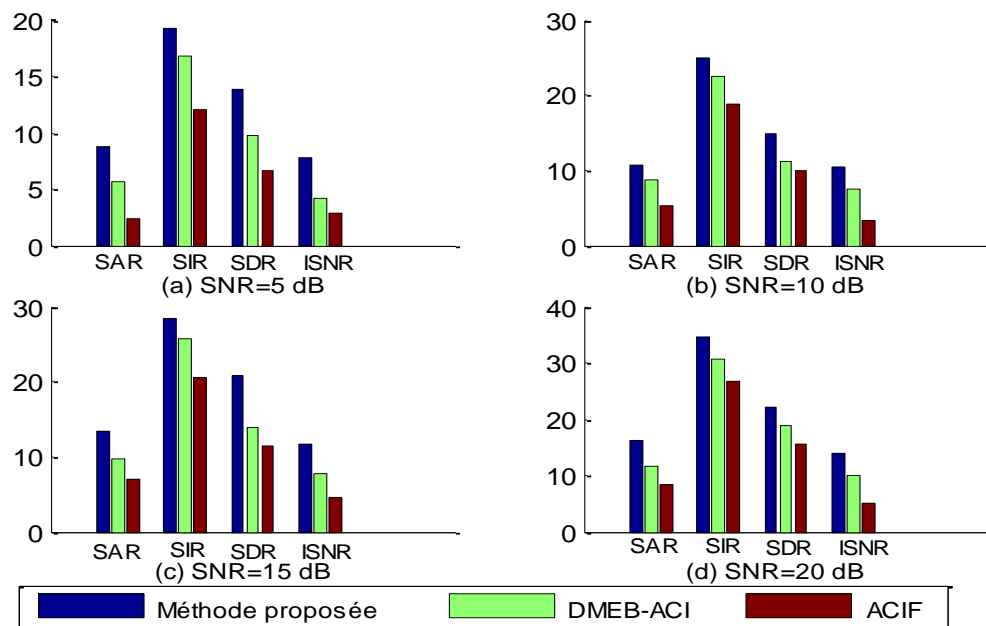
et

$$SNR_{sortie} = 10 \log \frac{\sum_t |s(t)|^2}{\sum_t |s(t) - \hat{s}(t)|^2} \quad (4.39)$$

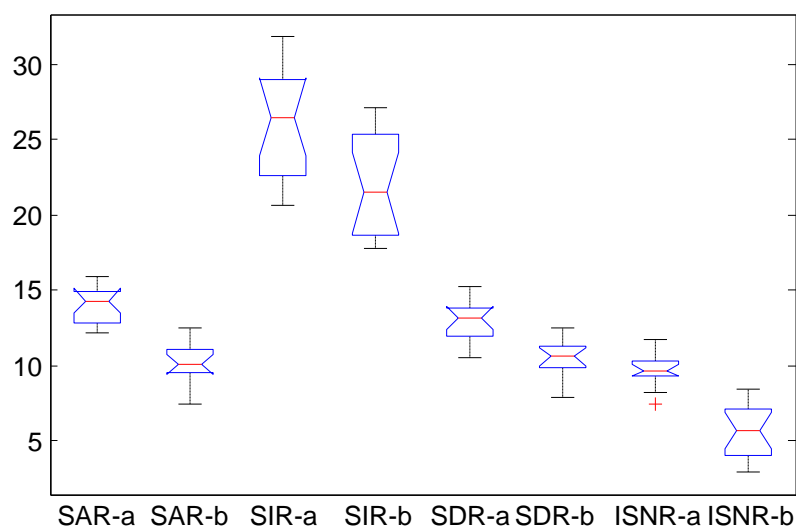
où  $x(t)$  est le mélange observé,  $s(t)$  est la source originale et  $\hat{s}(t)$  est la source estimée.

Les critères de performance des trois méthodes BSS correspondant au premier ensemble de données sont présentés par la figure 4.9 pour différentes valeurs du SNR. La figure 4.10 présente une étude comparative entre la méthode proposée et la méthode ICA-DME pour différentes valeurs du SNR. La méthode proposée se traduit par une meilleure performance en termes des quatre critères par rapport aux méthodes FICA et DMEB-ACI [140] pour les mélanges convolutifs. Pour toutes les valeurs du SNR, la BDME combinée avec la méthode ACI-MBE présente des meilleures performances.

Pour tester si la différence entre les critères de performance de la méthode proposée et ceux des méthodes ACIF et DMEB-ACI est statistiquement significative, le test statistique de Kruskal Wallis [144] a été réalisé sur le deuxième ensemble de signaux acoustiques. Le test statistique de Kruskal Wallis est un test statistique non paramétrique qui ne nécessite pas d'hypothèses sur la distribution des données.



**Figure 4.9** Comparaison entre la méthode proposée et la méthode d'analyse en composantes indépendantes combinée avec l'algorithme de décomposition en modes empiriques (ACI-DME) et la méthode conventionnelle ACIF en termes de SIR, SAR, SDR et ISNR.



**Figure 4.10** Comparaison entre les performances de (a) la méthode proposée, (b) la méthode DME-ICA en terme de SIR, SAR, SDR et ISNR

L'avantage de ce test est qu'il reste aussi puissant que le test statistique ANOVA. Le test statistique de Kruskal Wallis montre que les valeurs moyennes des critères de performance de la méthode proposée diffèrent statistiquement significativement de celles des méthodes ACIF et DMEB-ACI pour les mélanges convolutifs. Les valeurs  $p$  du test sont données dans le tableau 4.1.

**Tableau 4.1** valeurs de  $p$  du test statistique de Kruskal Wallis pour SIR, SAR, SDR et ISNR de 20 mélanges convolutifs.

	SIR	SAR	SDR	ISNR
Valeur de $p$	$1.025^{-10}$	$11.129^{-11}$	$7.143^{-11}$	$0.8692^{-10}$

## 7 Conclusion

Une nouvelle méthode combinant les algorithmes DMEB et l'ACI-MBE pour la séparation aveugle de mélanges convolutifs a été présentée. La méthode opère dans le domaine fréquentiel. Les mélanges convolutifs observés dans le domaine temporel sont transformés en mélanges instantanés dans le domaine fréquentiel. La décomposition en modes empiriques bivariée est utilisée pour décomposer les mélanges complexes en un ensemble de FMIs à partir duquel les composantes indépendantes sont extraites à l'aide de l'algorithme ACI-MBE. La méthode proposée a été testée sur des ensembles de signaux audio construits à partir des bases de données TIMIT et NOIZEUS. Les résultats sont satisfaisants par rapport à la séparation du DME-ACI pour les mélanges convolutifs et l'ACIF conventionnelle en terme de performances évaluées par l'estimation de SIR, SAR, SDR et ISNR et confirmées par le test statistique de Kruskal Wallis.



## **Conclusion générale**

## Conclusion générale

Bien que beaucoup de travaux de recherche sur la séparation aveugle de sources à valeurs réelles ont été publiés dans la littérature, le nombre de recherches ayant traité le cas complexe reste insuffisant. L'analyse en composantes indépendantes s'est révélée utile pour la séparation des signaux à valeurs complexes dans de nombreuses applications telles que les communications, l'analyse des données d'électroencéphalographie, la reconnaissance de visage et la détection radar.

Au cours de cette thèse, nous nous sommes intéressés au développement de nouvelles méthodes de résolution du problème de la SAS pour les mélanges convolutifs. Nous avons considéré le problème de la séparation aveugle de sources et la déconvolution aveugle.

En considérant le problème de déconvolution du signal de parole, la décomposition en modes empiriques a été proposée pour décomposer le spectre d'amplitude du signal de parole en une composante harmonique due à l'excitation glottique et une enveloppe spectrale représentant l'effet du conduit vocal. Les résultats de la déconvolution ont été exploités pour la caractérisation du signal de parole. La composante harmonique et l'enveloppe spectrale obtenues ont été utilisées pour l'estimation de la fréquence fondamentale et les formants, respectivement. Les résultats obtenus sur des signaux de parole synthétique et naturelle montrent que la méthode proposée fournit une estimation conjointe précise de la fréquence fondamentale et des formants. L'information de phase a été exploitée pour l'estimation de la source glottique.

Dans la littérature, les méthodes conçues pour résoudre le problème de la séparation aveugle des signaux à valeurs complexes exploitent soit la non-circularité soit la non-gaussianité des signaux. Nous avons ainsi proposé une solution concrète permettant non seulement la séparation de mélanges convolutifs mais aussi

### Conclusion générale

d'améliorer les performances de la séparation par rapport aux méthodes conventionnelles de séparation aveugle de sources. La méthode proposée consiste à combiner l'algorithme de l'analyse en composantes indépendantes avec l'algorithme de la décomposition en modes empiriques bivariée. La méthode proposée a été testée sur des bases de données de parole construites à partir des bases de données de TIMIT et NOIZEUS; Les résultats obtenus sont très satisfaisants et montrent que la méthode proposée fournit des performances améliorées en termes de SIR, SAR, SDR et ISNR par rapport aux méthodes proposées dans la littérature telles que l'algorithme conventionnel de l'analyse en composantes indépendantes fréquentielles conçue pour les mélanges convolutifs et la combinaison de l'algorithme de l'analyse en composantes indépendantes combinée avec l'algorithme de décomposition en modes empiriques. Les résultats obtenus ont été confirmés par le test statistique de Kruskal-Wallis.

Les travaux présentés dans cette thèse ouvrent la voix à plusieurs développements, nous envisageons comme travaux futurs le développement d'une méthode adaptative pour le choix optimal de l'indice qui permet de séparer la composante harmonique et l'enveloppe spectrale, l'adaptation de l'algorithme de SAS proposé pour la résolution du problème de la séparation aveugle des signaux à valeurs complexes dans les cas de mélanges sous-déterminés et surdéterminés ainsi que la validation en ligne de la séparation de mélanges convolutifs de l'algorithme de DMEB combiné avec l'ACI-MBE.

## Références

1. J. Traa, P. Smaragdis, Multichannel source separation and tracking with RANSAC and directional statistics. *IEEE/ACM Trans. Audio, Speech, Language Process.* **22**(12), 2233-2243 (2014).
2. Y. Deville, Panorama des applications biomédicales des méthodes de séparation aveugle de sources. *In proceeding of GRETSI* (2003), pp. 31-34
3. G. Pendharkar, G.-R. Naik, HT. Nguyen, Using blind source separation on accelerometry data to analyze and distinguish the toe walking gait from normal gait in ITW children. *Biomed Signal Process Control.* **13**, 41-49 (2014).
4. G.-R. Naik, A.-H. Al-Timemy, H.-T. Nguyen, Transradial amputee gesture classification using an optimal number of sEMG sensors: an approach using ICA clustering. *IEEE Trans. Neural Syst. Rehabil. Eng.* **24**(8), 837-846 (2016).
5. D. Nuzillard, A. Bijaoui, Blind source separation and analysis of multispectral astronomical images. *Astron Astrophys Suppl Ser.* **147**(1), 129–138 (2000).
6. A.-K. Takahata, E.-Z. Nadalin, R. Ferrari, L.-T. Duarte, R. Suyama, R.-R. Lopes, J.M.-T. Romano, M. Tygel, Unsupervised processing of geophysical signals: a review of some key aspects of blind deconvolution and blind source separation. *IEEE Signal Process. Mag.* **29**(4), 27-35 (2012).
7. I. Bekkerman, J. Tabrikian, Target detection and localization using mimo radars and sonars. *IEEE Trans. Signal Process.* **54**(10), 3873-3883 (2006).
8. C. Jutten, M. Babaie-Zadeh, Source separation: Principles, current advances and applications. *In proceeding of IAR Annual Meet* (2006).
9. C. Jutten, P. Comon, Séparation de sources - Tome 2 : au-delà de l'aveugle et applications, Chap. 13, ed. Y. Deville. Hermès – Lavoisier (2007).
10. A. Kachenoura, L. Albera, L. Senhadji, Blind source separation in biomedical engineering. *IRBM.* **28**(1), 20-34 (2007) .
11. Y. Guo, S. Huang, Y. Li, G.-R. Naik, Edge effect elimination in single-mixture blind source separation. *Circ Syst Signal Process.* **32**(5), 2317-2334 (2013).

12. D. Farina, C. Févotte, C. Doncarli, R. Merletti, Blind separation of linear instantaneous mixtures of nonstationary surface myoelectric signals. *IEEE Trans. Biomed. Eng.* **51**(9), 1555-1567 (2004).
13. S. Cruces-Alvarez, A. Cichocki, L. Castedo-Ribas, An iterative inversion approach to blind source separation. *IEEE Trans. Neural Netw. Learn. Syst.* **11**(6), 1423–1437 (2000).
14. K. Kokkinakis, V. Zarzoso, A.-K. Nandi, Blind separation of acoustic mixtures based on linear prediction analysis. *In proceeding of the 4<sup>th</sup> ICA2003* (2003), pp 343-348.
15. J. H. Deller, J. G. Proakis, J. H. L. Hansen, Discrete-time processing of speech signals, Prentice-Hall, 1993.
16. M. O. Rosa, R. M., J. C. Pareira, M. Grellet, Adaptive estimation of residue signal for voice pathology diagnosis. *IEEE Trans. Biomed. Eng.*, **47**(1), 96-104 (2000).
17. A. V. Oppenheim and R. W. Schaffer, Digital signal processing, NJ: Prentice Hall, 1975.
18. de Krom, G, A Cepstrum-based technique for determining a harmonics-to-noise ratio in speech signals. *J. Speech and Hearing Res.* **36**, 254-266 (1993).
19. M. Noll, Short-time Spectrum and Cepstrum Techniques for Vocal-Pitch Detection. *J. Acoust. Soc. Amer.* **36**(2), 296-302 (1964).
20. P. Veprek and M. S. Scordilis, Analysis, enhancement and evaluation of five pitch determination algorithms. *Speech Comm.* **37**, 249:270 (2002).
21. H. Weiping and R. Linggard, “Speech Signal Deconvolution Using wavelets Filter Banks”, WAA '01, Lecture Notes in Computer Science (LNCS), Y. Y. Tang et al. (Eds.), LNCS 2552, pp. 248-256, 2001.
22. P. Comon, E. Moreau, Blind mimo equalization and joint diagonalization criteria. *In proceeding of ICASSP* (2001).
23. A. Adib, E. Moreau, D. Aboutajdine, Source separation contrasts using a reference signal. *IEEE Signal Process. Lett.* **11**(3), 312-315 (2004).
24. M. Castella, J.-C. Pesquet, A.-P. Petropulu, A family of frequency and time domain contrasts for blind separation of convolutive mixtures of temporally dependent signals. *IEEE Trans. Signal Process.* **53**(1), 107-120 (2005).

25. D.-T. Pham, P. Garat, Blind separation of mixture of independent sources through aquasi-maximum likelihood approach. *IEEE Trans. Signal Process.* **45**(7), 1712–1725 (1997).
26. J.-F. Cardoso, Higher-order contrasts for independent component analysis. *Neural Comput.* **11**(1), 157–192 (1999).
27. G.-R. Naik, K.-K. Dinesh, Estimation of independent and dependent components of non-invasive EMG using fast ICA: validation in recognising complex gestures. *Comput Methods Biomech Biomed Eng.* **14**(12), 1105-1111 (2011).
28. G.-R. Naik, K.-K. Dinesh, P. Marimuthu, Signal processing evaluation of myoelectric sensor placement in low-level gestures: sensitivity analysis using independent component analysis. *Expert Syst.* **31**(1), 91-99 (2014).
29. X. Li, T. Adali, Independent component analysis by entropy bound minimization. *IEEE Trans. Signal Process.* **58**(10), 5151–5164 (2010).
30. G.-S. Fu, R. Phlypo, M. Anderson, X.-L. Li, T. Adali, Blind source separation by entropy rate minimization. *IEEE Trans. Signal Process.* **62**(16), 4245-4255 (2014).
31. R. Chai, G. Naik, T.-N. Nguyen, S. Ling, Y. Tran, A. Craig, H. Nguyen, Driver fatigue classification with independent component by entropy rate bound minimization analysis in an EEG-based system. *IEEE J. Biomed. Health Inform.* PP(99), DOI: 10.1109/JBHI.2016.2532354 (2016).
32. T. Adali, D.-V. Calhoun, Complex ICA of brain imaging data. *IEEE Signal Process. Mag.* **24**(5), 136-139 (2007).
33. L.-E. Di Persia, D.-H. Milone, Using multiple frequency bins for stabilization of FD-ICA algorithms. *Signal Process.* **119**(c), 162-168 (2016).
34. J. Guo, Y. Deng, A time-frequency algorithm for noisy ICA. In: F. Bian, Y. Xie (eds) *Geo-Informatics in Resource Management and Sustainable Ecosystem. Communications in Computer and Information Science* (2016), pp 357-365
35. L. De Lathauwer, B. De Moor, On the blind separation of non-circular sources. In *proceeding of EUSIPCO* (2002).
36. E. Ollila, V. Koivunen, Complex ICA using generalized uncorrelating transform. *Signal Process.* **89**(4), 365–377 (2009).

37. P. Smaragdis, Blind separation of convolved mixtures in the frequency domain. *Neurocomputing*. **22**(1-3), 21–34 (1998).
38. J. Anemüller, T.J. Sejnowski, S. Makeig, Complex independent component analysis of frequency-domain electroencephalographic data. *Neural Netw.* **16**(9), 1311–1323 (2003).
39. M. Novey, T. Adali, Complex ICA by negentropy maximization. *IEEE Trans. Neural Netw. Learn. Syst.* **19**(4), 596-609 (2008).
40. S.-C. Douglas, Fixed-point algorithms for the blind separation of arbitrary complex-valued non-Gaussian signal mixtures. *EURASIP J Adv Signal Process.* **2007**(1), 83–83 (2007).
41. X.-L. Li, T. Adali, Complex independent component analysis by entropy bound minimization. *IEEE Trans. Circuits Syst. I, Reg. Papers.* **57**(7), 1417 – 1430 (2010).
42. N.-E. Huang, Z. Shen, SR. Long, MC. Wu, HH. Shih, Q. Zheng, NC. Yen, CC. Tung, HH. Liu, The empirical mode decomposition and the Hilbert spectrum for nonlinear and nonstationary time series analysis. *Proc. Roy. Soc. London A*, **454**(1971), 903-995 (1998).
43. P. Flandrin, G. Rilling, P. Gonçalves, Empirical Mode Decomposition as a Filter Bank. *IEEE Signal Process. Lett.* **11**(2) 112-114 (2004).
44. A.-O. Boudraa, J.-C. Cexus, EMD-Based Signal Filtering. *IEEE Trans. Instrum. Meas.* **56**(6), 2196-2202 (2007).
45. JC. Cexus, AO. Boudraa, Non-stationary signals analysis by Teager-Huang Transform (THT). *In proceeding of EUSIPCO* (2006).
46. H. Chunming, G. Huadong, W. Changlin, F. Dian, A novel method to reduce speckle in SAR images. *Int J Remote Sens.* **23**(23), 5095-5101 (2002).
47. L. Du, B. Wang, Y. Li, H. Liu, Robust classification scheme for airplane targets with low resolution radar based on EMD-CLEAN feature extraction method. *IEEE Sens J.* **13**(12), 4648-4662 (2013).
48. ME. Torres, MA. Colominas, G. Schlotthauer, P. Flandrin. A Complete Ensemble Empirical Mode Decomposition with Adaptive Noise. *In proceeding of 36<sup>th</sup> ICASSP* (2011).
49. M.-A. Colominas, G. Schlotthauer, M.-E. Torres, Improve complete ensemble EMD: A suitable tool for biomedical signal processing. *Biomed Signal Process Control.* **14**, 19-29 (2014)

50. B. Mijović, M. De Vos, I. Gligorijević, J. Taelman, S. Van Huffel, Source separation from single-channel recordings by combining empirical-mode decomposition and independent component analysis. *IEEE Trans. Biomed. Eng.* **57**(9), 2188-2196 (2010).
51. Z. Wu, NE. Huang, Ensemble empirical mode decomposition: A noise-assisted data analysis method. *Adv Adapt Data Anal.* **1**(1) 1-41 (2009).
52. G.-R. Naik, S.-E. Selvan, H.-T. Nguyen, Single-channel EMG classification with ensemble empirical mode decomposition based ICA for diagnosing neuromuscular disorders. *IEEE Trans. Neural Syst. Rehabil. Eng.* **24**(7), 734-743 (2016).
53. J.-S. Smith, The Local Mean Decomposition and Its Application to EEG Perception Data, *J R Soc Interface.* **2**(5), 443-454 (2005).
54. Y. Guo, G.-R. Naik, H. Nguyen, Single channel blind source separation based local mean decomposition for Biomedical applications. *In proceeding of 35th Annual International Conference of the IEEE Eng Med Biol Soc* (2013) pp 6812-6815.
55. W. Li, H. Z. Yang, Blind source separation in underdetermined model based on local mean decomposition and AMUSE algorithm. *In Proceeding of CCC* (2014), pp 7206-7211.
56. M.-UB Altaf, T. Gautama, T. Tanaka, D.-P.Mandic, Rotation Invariant Complex Empirical Mode Decomposition. *In proceeding of ICASSP07* (2007), pp 1009-1012
57. G. Rilling, P. Flandrin, P. Gonçalves, J. M. Lilly, Bivariate Empirical Mode Decomposition. *IEEE Signal Process. Lett.* **14**(12) 936-939 (2007).
58. T. Tanaka, DP. Mandic, Complex Empirical Mode Decomposition. *IEEE Signal Process. Lett.* **14**(2), 101-104 (2007).
59. N. Rehman, D.-P. Mandic, Multivariate Empirical Mode Decomposition. *Proc. Roy. Soc. London A.* **466**, 1291-1302 (2010).
60. M.-H. Yeh, The complex Bidimensional Empirical Mode Decomposition, *Signal Process.* **92**(2), 523-541 (2012).
61. P. Xie, SL. Grant, A Fast and Efficient Frequency-Domain Method for Convolutional Blind Source Separation. *In proceeding of Region 5 Conference, IEEE* (2008).



62. F.D. Neeser and J.L. Massey, Proper complex random processes with applications to information theory. *IEEE Trans. Inf. Theory*, **39**,1293–1302 (1993).
63. B. Picinbono, On circularity. *IEEE Trans. Signal Process*, **42**, 3473–3482 (1994).
64. M. Novey and T. Adalı. On extending the complex FastICA algorithm to noncircular sources. *IEEE Trans.Signal Process*. **56**(5) 2148–2154 (2008).
65. P. J. Schreier. Bounds on the degree of impropriety of complex random vectors. *Signal Process. Lett*, **15**,190–193 (2008).
66. E. Ollila. On the circularity of a complex random variable. *Signal Process. Lett*, **15**, 841–844 (2008).
67. P. J. Schreier, L. L. Scharf, and A. Hanssen. A generalized likelihood ratio test for impropriety of complex signals. *Signal Process. Lett*, **13**(7):433–436 (2006).
68. J. P. Delmas and H. Abeida. Asymptotic distribution of circularity coefficients estimate of complex random variables. *Signal Process*. **89**(12), 2670–2675 (2009).
69. E. Bingham and A. Hyvärinen. A fast fixed point algorithm for independent component analysis of complex valued signals. *Journal of Neural Systems*, **10**, 1– 8 (2000).
70. Y. Deville. Actes de l'école de printemps : De la séparation de sources à l'analyse en composantes indépendantes. Méthodes, algorithmes et applications. Ch. Jutten, A. Guérin-Dugué, éditeurs, Villard-de-Lans (Isère), 2 (2001)
71. J.-F. Cardoso. Blind signal separation: statistical principle. *Proceeding of the IEEE* (1998), pp. 2009-2025
72. M.-S. Pedersen, J. Larsen, U. Kjems, L.-C. Parra A Survey of Convolutional Blind Source Separation Methods Springer Handbook of Speech Processing. Editors Jacob Benesty, Yiteng Huang, Mohan Sondhi. Springer Press. 2007
73. K. Torkkola, Blind separation of convolved sources based on information maximization. In *Proceeding. IEEE Workshop on Neural Networks and Signal Processing* (1996).

74. T. W. Lee, A. J. Bell, and R. Lambert, "Blind separation of delayed and convolved sources," in *Advances in Neural Information Processing Systems 9*. Cambridge, MA: MIT Press (1997) pp. 758–764.
75. P. Smaragdis. Blind source separation of convolved mixtures in the frequency domain. In International Workshop on Independence & Artificial Neural Networks (1998).
76. L. Parra and C. Spence. On line blind source separation of non stationary signals. *J. VLSI Signal Proceeding Systems for Signal, Images and Video Tech.* **26**(8) 15-24 (2000).
77. L. Parra and C. Spence. Convolutive blind source separation of non-stationary sources. In *Proceeding of IEEE Trans. on Speech and Audio Process.* **8**(3) 320-327 (2000).
78. H.-C. Wu and J. C. Principe. Simultaneous diagonalization in the frequency domain (SDIF) for source separation. In *Proceeding of ICA* (1999), pp 245-250.
79. N. Mitianoudis and M. Davies. Audio source separation of convolutive mixtures. *IEEE Trans. on Speech and Audio Process* (2002).
80. A. Westner and V. M. Bove. Applying blind source separation and deconvolution to real-world acoustic environments. In *proceeding 106th of Audio Engineering Society* (1999).
81. V. C. Soon, L. Tong, Y. F. Huang, and R. Liu, A wideband blind identification approach to speech acquisition using a microphone array. In *ICASSP'92* (1992), pp. 293–296.
82. S. Kurita, H. Saruwatari, S. Kajita, K. Takeda, and F. Itakura, Evaluation of frequency-domain blind signal separation using directivity pattern under reverberant conditions. In *ICASSP'00* (2000), pp. 3140–3143.
83. F. Asano, S. Ikeda, M. Ogawa, H. Asoh, and N. Kitawaki, Combined approach of array processing and independent component analysis for blind separation of acoustic signals. *IEEE Trans. Speech Audio Process*, vol. **11**(3) 204–215 (2003).
84. N. Mitianoudis and M. E. Davies, Permutation alignment for frequency domain ICA using subspace beamforming methods. In *ICA'04* (2004), pp. 669–676.

85. W. Baumann, B.-U. Köhler, D. Kolossa, and R. Orglmeister, Real time separation of convolutive mixtures. *In ICA '01* (2001), pp. 65–69.
86. F. Tordini and F. Piazza, A semi-blind approach to the separation of real world speech mixtures. *In IJCNN'02*. **2**, (2002) pp. 1293–1298.
87. M. Knaak, M. Kunter, and D. Filbert, “Blind source separation for acoustical machine diagnosis. *In DSP'02* (2002).
88. A. Dapena and C. Serviere, A simplified frequencydomain approach for blind separation of convolutive mixtures. *In ICA '01* (2001), pp. 569–574.
89. C. Mejuto, A. Dapena, and L. Castedo, “Frequencydomain infomax for blind separation of convolutive mixtures. *In ICA '00* (2000), pp. 315–320.
90. I. Kopriva, Željko Devčić, and H. Szu, An adaptive short-time frequency domain algorithm for blind separation of nonstationary convolved mixtures. *In IJCNN'01* (2001) pp. 424–429.
91. S. Dubnov, J. Tabrikain, and M. Arnon-Targan, A method for directionally-disjoint source separation in convolutive environment. *In ICASSP'04*, **V** (2004), 489–492.
92. M. Joho, Blind signal separation of convolutive mixtures: A time-domain joint-diagonalization approach. *In ICA '04* (2004), pp. 578–585.
93. A. Hiroe, Solution of permutation problem in frequency domain ica, using multivariate probability density functions. *In ICA '06* (2006), pp. 601–608.
94. D. Kolossa, B. uwe Khler, M. Conrath, and R. Orgelmeister, Optimal permutation correlation by multiobjective genetic algorithms. *In ICA '01* (2001), pp. 373–378.
95. K. Kamata, X. Hu, and H. Kobataka, A new approach to the permutation problem in frequency domain blind source separation. *In ICA '04* (2004), pp. 849–856.
96. E. Ollila and V. Koivunen, Complex ICA using generalized uncorrelating transform. *Signal Process.* **89** (4), 365–377 (2009).
97. L. De Lathauwer, B. De Moor, On the blind separation of non-circular sources. *In EUSIPCO* (2002)
98. J. Eriksson and V. Koivunen, Complex-valued ICA using second order statistics. *In Machine Learning for Signal Processing, 2004. Proceedings of the 2004 14th IEEE Signal Processing Society Workshop* (2004) pp. 183–192.

99. J.-F. Cardoso and A. Souloumiac, Blind beamforming for non-Gaussian signals. In *IEE Proceedings of Radar and Signal Process.* **140**(6) pp. 362–370 (1993).
100. A. Hyvärinen, J. Karhunen, and E. Oja, Independent Component Analysis, Wiley, New York (2001).
101. H. Li, T. Adalı, Gradient and fixed-point complex ICA algorithms based on kurtosis maximization. In *Proc. IEEE Workshop on Machine Learning for Signal Processing (MLSP)* (2006), pp. 85–90.
102. E. Bingham, A. Hyvärinen, A fast fixed-point algorithm for independent component analysis of complex valued signals. *Int. J. Neural Syst.* **10**,1–8 (2000).
103. J. Anemuller, T. J. Sejnowski, and S. Makeig, Complex independent component analysis of frequency-domain electroencephalographic data. *Neural Netw.* **16**(9) 1311–1323 (2003).
104. M. Novey and T. Adalı, On extending the complex FastICA algorithm to noncircular sources. *IEEE Trans. Signal Process.* **56**(5) pp. 2148–2154 (2008).
105. H. Li and T. Adalı, A class of complex ICA algorithms based on the kurtosis cost function., *IEEE Trans. Neural Netw.*, vol. **19** (3) 408–420 (2008).
106. X.-L. Li and T. Adalı, Complex independent component analysis by entropy bound minimization., *IEEE Tran. Circuits and Systems I: Regular Papers.* **57**(7) 1417–1430 (2010).
107. J.-F. Cardoso and T. Adalı, The maximum likelihood approach to complex ICA. In *ICASSP* (2006).
108. T. Adalı, H. Li, M. Novey, and J.-F. Cardoso, Complex ICA using nonlinear functions. *IEEE Trans. Signal Process.* **56** (9) 4536–4544 (2008).
109. J.-C. Cexus, Analyse des signaux non-stationnaires par Transformation de Huang, Opérateur de Teager-Kaiser, et Transformation de Huang-Teager (THT). Thèse doctorat, l'Université de Rennes 2005.
110. M. U. Altaf, T. Gautama, T. Tanaka, and D. P. Mandic. Rotation invariant Empirical Mode Decomposition. In *IEEE Int. Conf on Acoust. Speech and Signal Proc* (2007).
111. A. Kacha, F. Grenez, J. Schoentgen. Assessment of Disordered Voices Using Empirical Mode Decomposition in the Log-Spectral Domain. *Interspeech* (2012).

112. A. Kacha, F. Grenez, J. Schoentgen. Empirical Mode Decomposition-Based Spectral Acoustic Cues for Disordered Voices Analysis, *Interspeech* (2013), pp. 3632-3636
113. A. Kacha, F. Grenez, J. Schoentgen. Bivariate analysis of disordered connected speech using temporal and spectral acoustic cues. *ICASSP* (2014), 985-989
114. A. Kacha, F. Grenez, J. Schoentgen, Multiband vocal dysperiodicities analysis using empirical mode decomposition in the log-spectral domain, *Biomed Signal Process and Control*, **17**, 11–20(2015).
115. M. Kemiha, A. Kacha, Estimation of spectral envelope and harmonic component in speech using empirical mode decomposition, *ICNCRE* (2013).
116. N. Chatlani, J. Soraghan. DME-Based Filtering (DMEF) of Low-Frequency Noise for Speech Enhancement. *IEEE Trans. Audio, Speech, and lang. Process*, **20**(4), 1158-1166 (2012).
117. L. R. Rabiner. Digital-Formant Synthesizer for Speech-Synthesis Studies. *J. Acoust. Soc. Amer*, **43**(4) 822-828 (1968).
118. M. Kemiha, A. Kacha Empirical mode decomposition for speech deconvolution with application to formant and fundamental frequencies estimation. *SIVA* (2013).
119. L. R. Rabiner, A Comparative Performance Study of Several Pitch Detection Algorithms. *IEEE Trans. Acoust Speech, And Signal Process*. **24**(5), (1976).
120. P. Boersma, D. Weenink, Praat: doing phonetics by computer (Version 4.6.09) [Computer program], <http://www.praat.org/>
121. P. Boersma, Accurate short term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sample sound. *In Poceeding of Institute of Phonetic Sciences* (1993) 17 97–110.
122. M. Plumpe, T. Quatieri, D. Reynolds. Modeling of the glottal flow derivative wave form with application to speaker identification. *IEEE Trans. on Speech and Audio Process* (1999).
123. E. Moore, M. Clements, J. Peifer, L. Weisser. Investigating the role of glottal features in classifying clinical depression. *In Proceeding of the 25th International Conference of the IEEE Engineering in Medicine and Biology Society*, **3**, (2003) 2849-2852.

124. D. Yamada, N. Kitaoka, S. Nakagawa. Speech Recognition Using Features Based on Glottal Sound Source. *Trans. of the Institute of Electrical Engineers of Japan*, **122**(12): 2028- 2034 (2002).
125. T. Drugman, G. Wilfart, A. Moinet, T. Dutoit. (2009). Using a pitch-synchronous residual for hybrid HMM/frame selection speech synthesis. *ICASSP* ((2009) 3793-3796.
126. P. Alku, E. Vilkmán. Estimation of the glottal pulseform based on discrete all-pole modeling. *Third international Conference on Spoken Language Processing* (1994) 1619-1622.
127. Alku, J. Svec, E. Vilkmán, F. Sram. Glottal wave analysis with pitch synchronous iterative adaptive inverse filtering. *Speech Comm.* **11**(2-3): 109-118 (1992).
128. D. Brookes, D. Chan. Speaker characteristics from a glottal airway model using glottal inverse filtering. In *Proceeding of Institute of Acoust*, **15**, (1994). 501-508.
129. B. Bozkurt, B. Doval, C. DAlessandro, T. Dutoit. Zeros of Z-Transform Representation With Application to Source-Filter Separation in Speech. *Signal Process. Lett*, **12**(4) (2005).
130. B. Doval, C. dAlessandro, N. Henrich. The voice source as a causal/anticausal linear filter. In *proceedings ISCA ITRW VOQUAL03* (2003). pp 15-19.
131. T. Drugman, B. Bozkurt, T. Dutoit, Complex Cepstrum-based Decomposition of Speech for Glottal Source Estimation. *Interspeech* (2009), pp 116-119.
132. T. Drugman, B. Bozkurt, T. Dutoit. A comparative study of glottal source estimation techniques. *Computer Speech & Language*. **26**(1), 20-34 (2012).
133. M. Stark, M. Wohlmayr, F. Pernkopf. Source-Filter-Based Single-Channel Speech Separation Using Pitch Information. *IEEE trans on audio, speech and language process*. **19**(2), 242-255 (2011).
134. M.Kemiha, A. Kacha, M.Boudjerda, Estimation de la source glottique par décomposition modale empirique. *XXXe édition des journées d'étude sur la parole (JEP'14)* (2014)
135. M. Kemiha, A. Kacha, Glottal Source Estimation based on Bivariate Empirical Mode Decomposition. In *proceedeng of ICEE* (2015)

136. E.-T. Jaynes, Information theory and statistical mechanics, *Physical review*. **106**(4), 620–630 (1957).
137. M. Kemiha, A. Kacha, Complex blind source separation. *Circuit syst signal process* (2017). Doi 10.1007/s00034-017-0539-0
138. J. An emüler, B. Kollmeier, Amplitude modulation decorrelation for convolutive blind source separation. *In Proceeding of ICA*, (2000). pp 215-220.
139. Dr.ir. Emanu, A.P. Habets Room impulse response Room Impulse Response Generator (2010)
140. B. Mijovic, M. De Vos, I. Gligorijevic, S. Van Huffel, Combining DME with ICA for Extracting Independent Sources from Single Channel and Two-Channel Data. *In proceeding of 32nd Annual International Conference of the IEEE EMBS* (2010).
141. TIMIT database. <https://catalog ldc.upenn.edu/ldc93s1>
142. NOIZEUS database <http://ecs.utdallas.edu/loizou/speech/noizeus/>
143. C. Févotte, R. Gribonval, E. Vincent, BSS EVAL toolbox user guide, IRISA (2005), [http://www.irisa.fr/metiss/bss\\_eval](http://www.irisa.fr/metiss/bss_eval)
144. A. Krause, M. Olson, The Basics of S-PLUS (Statistics and Computing), 4th Edition (2005).