



HAL
open science

Séparation Aveugle de Mélanges Convolutifs de Sources

Hakim Boumaraf

► **To cite this version:**

Hakim Boumaraf. Séparation Aveugle de Mélanges Convolutifs de Sources. Traitement du signal et de l'image [eess.SP]. Université Joseph-Fourier - Grenoble I; Institut National Polytechnique de Grenoble - INPG, 2005. Français. NNT: . tel-00011643

HAL Id: tel-00011643

<https://theses.hal.science/tel-00011643>

Submitted on 17 Feb 2006

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

à ma famille

Table des matières

Table des matières	v
Remerciements	ix
Notations	xi
Abréviations	xiii
Introduction Générale	15
1 La Séparation de Sources	19
1.1 Principe	19
1.2 Applications	19
1.3 Modélisation mathématique et séparation	21
1.3.1 Modèles de mélange de sources	22
1.3.1.1 Une autre écriture du modèle	24
1.3.1.2 Environnement bruité	24
1.3.1.3 Type de mélange : convolutif ou instantané	24
1.3.2 Idée de séparation	25
1.3.3 Hypothèses de base, indépendance et statistiques utilisées	26
1.3.3.1 Hypothèses	26
1.3.3.2 Indépendance	26
1.3.3.3 Statistiques utilisées	27
1.3.4 Ambiguïté de séparation	28
1.3.4.1 Ambiguïté de permutation	28
1.3.4.2 Ambiguïté d'échelle	29
1.3.4.3 Ambiguïté de filtrage	30
1.3.4.4 Ambiguïté complète	30
1.3.4.5 Système global	31
1.4 Conclusion	32
2 Algorithmes de Séparation de Sources	33
2.1 Mélanges instantanés	33
• <u>algorithme de J. Héroult et al. [59]</u> :	34

• <u>l'algorithme JADE de J.-F. Cardoso [30]</u> :	35
• <u>l'algorithme AMUSE de L. Tong et al. [102]</u> :	36
• <u>l'algorithme SOBI de A. Belouchrani [7]</u> :	36
• <u>l'algorithme de D.-T. Pham et J.-F. Cardoso [90]</u> :	38
• <u>l'algorithme TFS de A. Belouchrani et M. G. Amin [9, 10]</u> :	40
• <u>l'algorithme OPCA de K. I. Diamantaras et Th. Papadimitriou [44]</u> :	41
• <u>autres algorithmes</u> :	42
2.2 Mélanges convolutifs	42
• <u>l'algorithme de C. Jutten et al. [61, 100]</u> :	43
• <u>l'algorithme de R. Mukai et al. [78, 79, 80]</u> :	44
• <u>l'algorithme de P. Smaragdis [96]</u> :	44
• <u>l'algorithme SIMO-ICA de T. Takatani et al. [99]</u> :	45
• <u>l'algorithme MSICA de T. Nishikawa et al. [82]</u> :	46
• <u>l'algorithme de W. Wang et al. [104]</u> :	47
• <u>l'algorithme de E. Weinstein et al. [106]</u> :	47
• <u>l'algorithme de J. Anemüller et B. Kollmeier [4]</u> :	48
• <u>l'algorithme de L. Parra et C. Spence [85, 84]</u> :	49
• <u>l'algorithme de H.-C. Wu et J. C. Principe [110]</u> :	50
• <u>l'algorithme de N. Mitianoudis et M. Davis [76]</u> :	50
• <u>l'algorithme de A. Westner et V. M. Bove [108]</u> :	51
• <u>l'algorithme de N. Murata et S. Ikeda [81]</u> :	51
2.3 Conclusion	53
3 Algorithmes pour la Séparation Aveugle de Mélanges Convolutifs de Sources	55
3.1 Mélanges acoustiques	56
3.2 La nonstationnarité	57
3.3 Méthodes de séparation dans le cas non bruité	58
3.3.1 Estimation des matrices spectrales	59
3.3.2 Diagonalisation conjointe adoptée	61
3.3.3 Indétermination	62
3.3.4 Résolution d'ambiguïté basée sur la continuité	64
3.3.5 Résolution d'ambiguïté en utilisant les profils d'énergie	67
3.3.5.1 Estimation des profils par moyenne glissante	70
3.3.5.2 Estimation des profils par la SFD	70
3.4 Méthode de séparation dans le cas bruité	72
3.4.1 Principe de l'approche	73
3.4.2 L'algorithme EM	74
3.4.2.1 Version première de l'algorithme EM	75
3.4.2.2 Version étendue de l'algorithme EM	78
3.4.3 Estimation des signaux sources	80

3.5	Conclusion	81
4	Simulations Numériques et Résultats	83
4.1	Jeux de données	83
4.1.1	Jeux de données de A. G. Westner	83
4.1.2	Jeux de données de l'université McMaster	85
4.2	Indices de performance de la séparation	86
4.3	Évaluation numérique	89
4.3.1	Simulations dans le cas non bruité	89
4.3.1.1	Continuité de la réponse fréquentielle	89
4.3.1.1.1	Discussion	94
4.3.1.2	Profils d'énergie par moyenne globale	95
4.3.1.2.1	Discussion	97
4.3.1.3	Profils d'énergie par moyenne locale et par SFD	97
4.3.1.3.1	Réponse impulsionnelle à 256 retards	98
4.3.1.3.2	Réponse impulsionnelle à 512 retards	104
4.3.1.3.3	Réponse impulsionnelle à 1024 retards	108
4.3.1.3.4	Discussion	112
4.3.2	Comparaison avec d'autres méthodes	113
4.3.2.1	Expériences et mesure de performance	113
4.3.2.2	Évaluation des résultats	114
4.3.3	Simulations dans le cas bruité	116
4.3.3.1	Le bruit	116
4.3.3.1.1	Blanc Gaussien	116
4.3.3.1.2	Autorégressif Gaussien	117
4.3.3.2	Réponse impulsionnelle à 512 retards	117
4.3.3.3	Réponse impulsionnelle à 1024 retards	121
4.3.3.4	Discussion	128
4.4	Conclusion	129
	Conclusion Générale	131
	Bibliographie	cxxxiii
	Annexes	cxli
A	Série de Fourier Discrète	cxliii
A.1	Définition	cxliii
A.2	Série de Fourier Discrète	cxliii
A.2.1	Forme exponentielle de la SFD	cxliii
A.2.2	Forme trigonométrique de la SFD	cxliv
A.2.3	Remarques	cxlv

B	Sous-échantillonnage par interpolation	cxlvii
B.1	Sous-échantillonnage de 50% ($m = 2$)	cxlviii
B.2	Sous-échantillonnage de 25% ($m = 4$)	cxlix
B.3	Loi générale	cxlix
B.4	Calcul détaillé ($m = 4$)	cl
C	Maximum de vraisemblance	cliii
C.1	Propriétés de la fonction de vraisemblance	cliv
C.2	Cas du modèle de mélange $\mathbf{x} = A\mathbf{s}$	cliv
C.3	Cas de K variables Gaussiennes multidimensionnelles	clv
D	Convolution sectionnée	clvii
E	Indice d'Amari	clix
F	Estimation spectrale	clxi
F.1	Choix du meilleur estimateur	clxi
F.2	Estimateur spectral simple	clxii
F.2.1	Biais de l'estimateur simple	clxii
F.2.2	Variance de l'estimateur simple	clxii
F.2.3	Commentaire	clxii
F.3	Estimateur spectral moyenné	clxiii
F.3.1	Biais de l'estimateur moyenné	clxiii
F.3.2	Variance de l'estimateur moyenné	clxiii
F.3.3	Commentaire	clxiv
F.4	Estimateur spectral adouci	clxiv
F.4.1	Biais de l'estimateur adouci	clxiv
F.4.2	Variance de l'estimateur adouci	clxv
F.5	Estimateur spectral modifié	clxv
F.5.1	Biais et variance de l'estimateur modifié	clxvi
G	Séparation de sources vibratoires	clxvii
G.1	Introduction	clxvii
G.2	Vibrations d'une machine tournante	clxviii
G.2.1	Définition et mesure des vibrations	clxviii
G.2.2	Origine des vibrations	clxix
G.3	Analyse et séparation des sources vibratoires	clxix
G.3.1	Modèle convolutif de mélanges vibratoires	clxx
G.3.2	Séparation des mélanges vibratoires	clxx

Remerciements

Ce travail a été effectué au LIS (Laboratoire des Images et des Signaux) et au LMC (Laboratoire de Modélisation et Calcul).

En premier lieu, je tiens à bien remercier Monsieur Georges-Henri COTTET, Directeur du LMC, et Monsieur Jean-Marc CHASSERY, Directeur du LIS, de m'avoir accueilli dans leurs laboratoires. Je tiens particulièrement à remercier mes directeurs de thèse, Monsieur Antoine PHAM et Madame Christine SERVIÈRE pour avoir accepté de diriger mes recherches, de m'encadrer et de m'aider durant ces années. Leurs compétences, leurs conseils et leurs idées ont été dans l'aboutissement de ce travail. Je leur témoigne de ma reconnaissance.

J'exprime ma gratitude à Monsieur Pierre GENTIL pour avoir accepté mon rattachement à son école doctorale EEATS au long des années de ma thèse, également, pour son aide administrative et sa position inoubliable lors de ma transition au LMC. Je n'oublierai pas de remercier également Monsieur Kosai RAOOF, Monsieur Jean-Louis LACOUME, Monsieur Joël LIENARD, Monsieur Patrick WITOMSKI, Madame Jeanne MALBOS et tous les autres ...

Je tiens à remercier aussi tous mes collègues dans les deux laboratoires, en particulier, Francois COMBET, mon collègue de bureau au LIS et mes collègues de bureau au LMC : Elimane KANE, Isidore-Paul AKAMBITA et Robin GIRARD. Je dois remercier les ingénieurs, Jacques LAURENT et Patrice NAVARRO pour l'aide technique et informatique qu'ils m'ont fournie chaque fois qu'il y avait un besoin. Merci à Bruno et aux secrétaires ...

Je remercie profondément et sincèrement ma femme pour son aide, son soutien et son encouragement. Le grand remerciement, je le dois à mes parents, c'est grâce à eux que je suis arrivé là. Je dirai également merci à tous mes amis pour leurs encouragements ; merci à l'ami, que j'ai rencontré au Japon, avec qui je tiens de tout coeur, beaucoup travailler et apprendre, Ali MANSOUR.

Je tiens à remercier les membres du jury :

- Monsieur Yannick DEVILLE, Professeur à l'Université Paul Sabatier, pour m'avoir fait l'honneur d'être rapporteur de ma thèse, pour le travail qu'il a fourni pour corriger ce rapport et pour avoir ainsi contribué à lui donner sa forme actuelle. Qu'il soit assuré ici de ma sincère gratitude.

- Monsieur Jean-François CARDOSO, Directeur de recherche au CNRS, pour m'avoir fait l'honneur d'être rapporteur de ma thèse.
- Monsieur Antoine PHAM, Directeur de recherche au CNRS, pour avoir dirigé ce travail, pour ses conseils scientifiques, ses critiques et ses remarques et surtout pour son aide.
- Madame Christine SERVIÈRE, Chargée de recherche au CNRS, pour avoir co-encadré cette thèse et pour son apport à ce travail.
- Monsieur Christian JUTTEN, Professeur à l'Université Joseph Fourier, pour l'honneur qu'il m'a fait de présider ce jury et pour avoir accepté d'évaluer ce travail. Profonds remerciements pour sa contribution significative à la correction et l'amélioration de mon manuscrit.
- Monsieur Ali MANSOUR, Maître de conférence à l'École Nationale Supérieure des Ingénieurs des Études et Techniques d'Armement, pour l'honneur qu'il m'a fait de participer à ce jury et pour avoir accepté d'examiner ce travail.

Enfin, pour finir, je remercie tout ceux qui ont contribué de loin ou de près à l'achèvement de ce travail au long des années de grands apports et de grandes ouvertures.

Notations

- **Matrices et vecteurs :**

h	: les scalaires sont notés par des minuscules italiques.
\mathbf{h}	: les vecteurs sont notés par des minuscules en gras.
H	: les matrices sont notées par des majuscules italiques.
\mathbf{h}^T	: transposée du vecteur \mathbf{h} .
H^T	: transposée de la matrice H .
H^*	: transposée conjuguée de la matrice H .
h_{ij}	: (ij) -ième élément de la matrice H .
$\text{diag}(H)$: matrice diagonale formée à partir de la diagonale de H .
I_n	: matrice identité de dimension $n \times n$.

- **Signaux :**

$(x * y)(t)$: produit de convolution linéaire des signaux $x(t)$ et $y(t)$.
$(\mathbf{x} * \mathbf{y})(t)$: produit matriciel de convolution linéaire des signaux multidimensionnels $\mathbf{x}(t)$ et $\mathbf{y}(t)$.
$E(\cdot)$: espérance mathématique.
$Cum_r(\cdot)$: cumulants d'ordre r .
$C_{xx}(t)$: autocorrélation du signal $x(t)$.
$C_{xy}(t)$: intercorrélation des signaux $x(t)$ et $y(t)$.
$S_x(f)$: densité spectrale du signal $x(t)$.
$S_{xy}(f)$: densité spectrale d'interaction des signaux $x(t)$ et $y(t)$.
$P_x(f)$: périodogramme de $x(t)$.
$Re(\cdot)$: partie réelle.
$Im(\cdot)$: partie imaginaire.
$ \cdot $: module ou valeur absolue.
$\delta(t)$: distribution de Dirac.
t	: variable du temps.
f	: variable de Fourier (fréquence).

Abréviations

ACP	: Analyse en Composantes Principales.
AMUSE	: Algorithm for Multiple Unknown Signals Extraction.
BSS	: Blind Source Separation.
DSP	: Digital Signal Processing.
dsp	: densité spectrale de puissance.
EM	: Expectation-Maximization.
GSM	: Global System for Mobile.
i.i.d.	: indépendant et identiquement distribué.
JADE	: Joint Approximate Diagonalization of Eigen-matrices.
MIMO	: Multiple-Input Multiple-Output.
MV	: Maximum de Vraisemblance.
RIF/FIR	: filtre à Réponse Impulsionnelle Finie.
RII/IIR	: filtre à Réponse Impulsionnelle Infinie.
SDMA	: Spatial Division Multiple Access.
SdS	: Séparation de Sources.
SFD	: Série de Fourier Discrète.
SLIT	: Système Linéaire et Invariant dans le Temps.
SNR	: Signal to Noise Ratio.
SOBI	: Second Order Blind Identification.
STFT	: Short Time Frequency Transform.
SVD	: Singular Value Decomposition.
TF	: Transformée de Fourier.
TFD	: Transformée de Fourier Discrète.
va.	: variable aléatoire.

Introduction Générale

La discipline du traitement de signal se distingue par la nature de données à traiter dans ce cadre : *signal*. Le signal est le support de l'information et le moyen de communication entre les hommes. Il est, dans la plupart des cas, mesuré par des capteurs et issu de notre environnement physique réel : ondes sonores, vibrations sismiques, images visuelles, ...

Avec l'apparition des ordinateurs vers les années 60, le traitement de signal, dont les premières applications reviennent à la première décennie du 19^{ième} siècle, connaît une nouvelle variante : le traitement numérique du signal (ou **DSP**, pour : *Digital Signal Processing*). Par la suite, le traitement de signal reconnaît un grand développement qui continue jusqu'à l'heure actuelle. Tous les outils mathématiques, algorithmiques calculatoires utilisés en DSP ont conduit à une révolution technologique dans plusieurs domaines d'applications : reconnaissance de parole, rehaussement d'images, télécommunication, ...

Le DSP a pour vocation d'extraire l'information véhiculée par le signal. Ceci s'appuie sur la connaissance soit du signal, soit du système, soit des deux. Les systèmes digitaux sont, de nos jours, très utilisés dans l'enregistrement et l'analyse audio (musique moderne, studios de films vidéo, ...). Ces systèmes de traitement audio voient en leur entrée, selon le cas, un ou plusieurs signaux. Cependant, l'environnement externe impose à ces derniers, une transformation au cours de leur propagation aux DSP et également une superposition avec d'autres signaux de l'environnement. Un cas typique de tels systèmes se trouve dans les organes auditifs de l'être humain. En effet, l'être humain est capable de focaliser, dans le mélange provenant de son environnement, sur une des sources de signal qu'il reçoit à son oreille. Dans le cas d'une faiblesse de cet organe, tel que chez les malentendants, un DSP fait améliorer les performances.

Dans cet objectif, plusieurs travaux ont été menés en traitement de signal et plus précisément en séparation de sources : "blind separation of statics/dynamics acoustically-mixed sounds" [109, 14, 108, 76] (voir également [60, page 446]). Dans le cadre de cette thèse nous étudierons le problème de séparation de mélanges convolutifs de sources ; en particulier, des sources auditives.

Le problème de séparation de sources consiste à retrouver des signaux utiles provenant de différentes sources (par abus de langage on dit : signaux sources), qui ont été filtrées et additionnées (*i.e.* mélangées) en se propageant vers un ensemble de capteurs. Ce problème se rencontre dans plusieurs applications [70, 71, 36, 3] : radio-communication,

séparation des signaux sismiques, monitoring des réacteurs nucléaires, surveillance des aéroports, signaux biomédicaux, rehaussement de la parole, ...

La propagation des signaux à séparer, de leurs sources productrices aux capteurs de mesure, a été initialement modélisée par les premiers fondateurs de cette nouvelle discipline à base d'un produit mathématique simple. Ceci correspond au cas appelé "*instantané*" de la séparation de sources. Plus tard, il s'est avéré que cette modélisation ne convenait pas à toutes les situations rencontrées dans la pratique. C'est pourquoi, des modélisations plus réalistes ont été proposées. L'une de ces modélisation interprète le phénomène de propagation comme une opération de filtrage, c'est-à-dire qu'elle suppose que l'environnement est caractérisé par une fonction mathématique dépendante du temps et réalise une opération plus complexe, qui est un produit de convolution, afin d'engendrer les mélanges dits convolutifs. Ceci correspond au cas plus intéressant de la séparation de sources appelé "*convolutifs*".

L'application qui nous intéresse est la séparation aveugle de mélanges convolutifs de sources non stationnaires (acoustiques).

- Dans le premier chapitre, nous présenterons le problème de séparation de sources en expliquant son principe, en donnant son modèle mathématique et en citant quelques exemples d'application de la séparation de sources.
- Dans le deuxième chapitre, en faisant un état de l'art sur quelques méthodes de séparation de sources, nous allons présenter les principales idées utilisées pour résoudre ce problème ; nous évoquerons, essentiellement, les méthodes basées sur les statistiques d'ordre 2 et dédiées au cas convolutif.
- En deux parties, dans le chapitre suivant, nous présenterons notre contribution à la séparation de sources :
 - en premier, nous traiterons le cas des mélanges convolutifs sans bruit additif, où nous décrirons notre approche et verrons comment nous obtenons la séparation en exploitant la nonstationnarité intrinsèque des signaux sources. Nous nous focaliserons sur le problème d'ambiguïté, résultant de la séparation. Nous proposerons des nouvelles techniques, pour éliminer l'ambiguïté et mettre au point l'approche développée.
 - ensuite, nous nous placerons dans une situation plus complexe où nous considérons la présence d'un bruit additif sur les capteurs. Pour accomplir la séparation dans ce contexte, nous proposerons une méthode du maximum de vraisemblance. À partir de la modélisation mathématique de cette situation, la vraisemblance des paramètres de ce modèle est formulée et un algorithme itératif est utilisé pour

son estimation. L'algorithme utilisé, est basé sur la technique 'EM' (Expectation-Maximization).

- le dernier chapitre aura pour objectif, la présentation des résultats de la simulation avec des jeux de données réelles, ainsi que la mesure de performance, dans les deux cas sans et avec bruit.
- enfin, nous terminerons ce manuscrit par une conclusion générale et des perspectives.

Chapitre 1

La Séparation de Sources

1.1 Principe

Le plus souvent dans notre environnement, plusieurs sources de signal (point de départ) émettent dans cet environnement des signaux¹ qui, en se propageant vers un point d'arrivée², subissent une transformation par le milieu de propagation et s'associent pour composer des mélanges complexes.

Un des exemples concrets où l'on rencontre ce phénomène, est celui qui se passe quotidiennement chez l'être humain quand il reçoit à son oreille, les différents sons (signaux) qui se trouvent dans son environnement. En réalité, l'être humain reçoit un mélange des sources de l'environnement. Il est capable, par ses organes, de séparer la source d'intérêt des mélanges.

Typiquement, on retrouve ce problème en traitement de signal sous le thème de *séparation de sources*. La séparation de sources, dans son principe, a pour objectif de discerner les sources qui se trouvent dans un ensemble de mélanges, sans aucune connaissance *a priori* sauf une hypothèse sur les liens statistiques entre les sources³.

1.2 Applications

Depuis sa naissance jusqu'à aujourd'hui [59], la séparation de sources est l'objet de développements accrus étant donné l'intérêt qu'elle présente dans plusieurs situations et dans différentes applications. Dans ce paragraphe, nous essaierons d'énumérer quelques applications de séparation de sources (pour un aperçu plus exhaustif, voir par exemple [41, pages 177-212],[71, 3] et pour les applications industrielles voir [42]) :

- Dans leur travail [108], Westner et Bove utilisent la séparation de sources dans une application de salles acoustiques. Ils adoptent un critère d'info-

¹par abus de langage, on appelle ces signaux : *sources*.

²en général, ce dernier est un capteur placé géographiquement dans l'environnement.

³cette hypothèse est l'indépendance mutuelle des sources.

max qu'ils trouvent, comparé à d'autres, le mieux adapté à un tel cas de séparation de sources.

- Quant à Mitianoudis [76], il s'en sert pour séparer des signaux audio. Il réalise la séparation dans le plan temps-fréquence et il propose une solution au problème de permutation. Enfin, pour valider son travail, il compare sa méthode à trois autres méthodes en se basant sur un simple critère d'évaluation.
- Mansour [70], avec la séparation de sources, traite le cas des signaux de paroles qui sont intrinsèquement non stationnaires⁴. Il considère un cas simple où les sources se propagent de la même façon (cas de mélanges instantanés⁵). La séparation de sources est effectuée au 2nd ordre et le papier compare trois méthodes de séparation.
- En électrotechnique, la séparation de sources a aussi été utilisée; Capdevielle *et al.* [25], Fabry [46] et Gelle *et al.* [53] l'ont appliquée aux machines tournantes.
- Un autre domaine de séparation de sources est la télécommunication. J. M. F. Xavier [111] adopte une modélisation de SdS⁶ pour les GSM MIMO à architecture SDMA.
- En biomédecine, pour l'ECG par exemple, une étude récente basée sur la SdS a été développée dans [50]. Par ailleurs, Deville [43] présente un panorama des applications biomédicales des méthodes de séparation aveugle de sources des signaux ECG, MEG et EEG. Nous retrouvons dans cet article la présentation d'une étude où la SdS est considérée pour traiter le problème d'analyse du contrôle du cœur par le système nerveux autonome, dont les anomalies jouent un rôle important dans de nombreuses situations physiopathologiques.
- En rehaussement de parole, W. Bobillet *et al.* [17] utilisent également la séparation de sources; ils prennent en compte le bruit (coloré additif) et essaient de le supprimer via cette approche. Dans leur modèle, la propagation est représentée par un filtre RIF et la séparation est réalisée au 2nd ordre avec une diagonalisation conjointe, en procédant par des blocs.
- Un domaine nouveau et prometteur dans lequel la SdS intervient (voir [63]) est l'interface machine/homme.

⁴pour la définition de la nonstationnarité voir [68, pages 78-79].

⁵nous définissons dans le paragraphe suivant, le type de mélanges qu'il peut y avoir en SdS.

⁶toutes les abréviations sont définies dans la page [xiii](#).

- Bentoumi et ses collaborateurs ont appliqué concrètement la SdS pour détecter et classifier les défauts dans les rails de chemin de fer [13].
- Dans une application réelle de radar, Ebihara [45] a implémenté l'algorithme de Belouchrani et Amin [9] pour séparer les signaux radar. Il conclut, à l'aide d'expériences et de simulations, que la séparation de sources accomplit cette tâche avec succès, alors que la méthode de traitement de signal conventionnellement utilisée pour cela échoue.
- En astronomie et en traitement d'images astronomiques, la séparation de sources a été employée par Nuzillard [83].
- Dans son système optique, composé de microprocesseur, d'une électronique de puissance (50W) et d'une antenne travaillant à 10GHz, Smith [97] *et al.* ont implémenté et utilisé une approche adaptative de séparation de sources.
- Un autre domaine où l'on retrouve également la séparation de sources, est la prospection sismique. Le Bihan *et al.* [15] l'utilisent afin de séparer les ondes sismiques provenant des capteurs multicomposantes.

1.3 Modélisation mathématique et séparation

Comme nous l'avons expliqué dans l'introduction, les signaux sources⁷ se propagent dans un environnement (voir la figure 1.1). Ce dernier implique une transformation des signaux qui est supposée linéaire⁸, *i.e.* possédant les caractéristiques d'additivité et d'homogénéité, et aussi invariante dans le temps⁹. Dans cet environnement les sources seront transformées puis superposées, construisant ainsi les mélanges que nous recevons sur les capteurs.



FIG. 1.1 – phénomène de mélanges de signaux

En se plaçant au niveau des capteurs, nous allons observer des mélanges complexes des sources de départ. De plus, si nous ne connaissons rien a priori ni sur les sources d'origine, ni rien sur le canal de propagation¹⁰, comment allons nous discerner les signaux d'origine pour retrouver les composantes des mélanges ?

⁷supposés, statistiquement, mutuellement indépendants.

⁸on se place dans le cas simple de séparation de sources.

⁹une définition complète des SLIT peut être trouvée dans [75].

¹⁰d'où le mot "aveugle" dans BSS (Blind Source Separation).

1.3.1 Modèles de mélange de sources

Afin de modéliser le phénomène de propagation (mélanges)¹¹ des sources, nous allons noter les signaux mélanges : $x_i(t)$ (pour $i = 1, \dots, p$) et les signaux sources : $s_j(t)$ (pour $j = 1, \dots, r$), où t est la variable *temps* indiquant l'évolution temporelle des r signaux sources et des p signaux mélanges. Schématiquement, le phénomène se présente comme suit¹² :

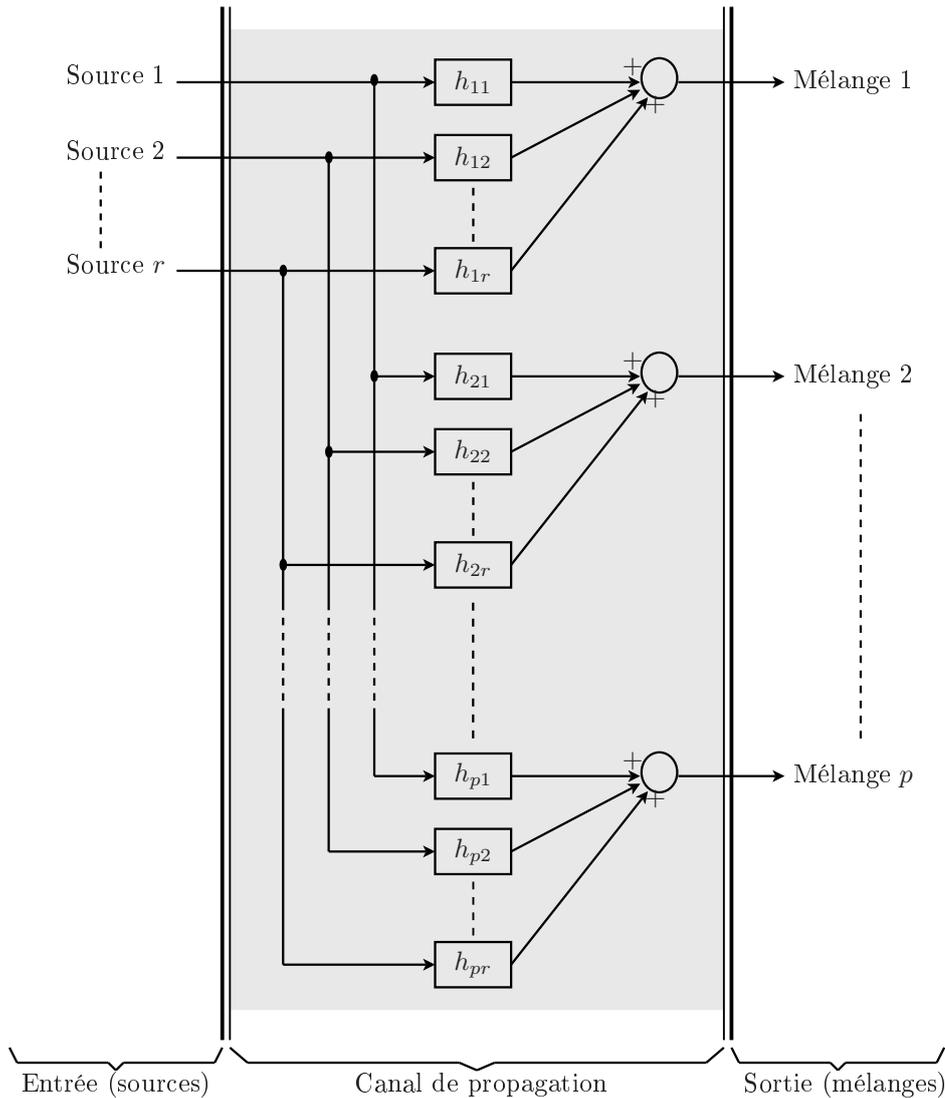


FIG. 1.2 – schéma de propagation et de mélange

On note que les fonctions $h_{ij}(t)$ (t variable du temps) définissent les opérations de transformation réalisées par le canal de propagation, respectivement, sur la source j se propageant vers le capteur i .

¹¹nous appelons le modèle correspondant : *modèle de mélange*.

¹²on observe p mélanges tq : $p \geq r$ ($p > r$ correspond au cas surdéterminé de la SdS [60]).

Ainsi, sous l'hypothèse de linéarité¹³ de l'environnement, la transformation subie par les sources dans le canal de propagation sera définie par l'opération de convolution ; en d'autres termes, cette transformation n'est autre qu'une opération de filtrage linéaire des signaux sources par les filtres à réponse impulsionnelle $h_{ij}(t)$ ¹⁴. Enfin, à chaque capteur nous mesurons un signal mélange qui est égal à la somme des filtrées des sources sur ce capteur (voir la figure 1.2). Plus formellement, nous écrivons le mélange $x_i(t)$ comme suit :

$$\begin{aligned} x_i(t) &= \sum_{j=1}^r (h_{ij} * s_j)(t) \\ &= \sum_{j=1}^r \left(\int_{-\infty}^{+\infty} h_{ij}(\tau) s_j(t - \tau) d\tau \right) \end{aligned} \quad (1.1)$$

Dans le cadre d'un traitement numérique aux ordinateurs¹⁵, nous considérons donc, la version discrète de notre modélisation mathématique, dans laquelle les signaux sont à variable de temps discrète n (échantillons temporels) :

$$\begin{aligned} x_i(n) &= \sum_{j=1}^r (h_{ij} * s_j)(n) \\ &= \sum_{j=1}^r \left(\sum_{k=-\infty}^{+\infty} h_{ij}(k) s_j(n - k) \right) \\ &= \sum_{k=-\infty}^{+\infty} \sum_{j=1}^r h_{ij}(k) s_j(n - k) \end{aligned} \quad (1.2)$$

En faisant un rangement, en vecteurs, des sources et des filtres (respectivement dans \mathbf{s} , $\tilde{\mathbf{h}}_i$), nous aboutissons à la simple écriture vectorielle de l'équation (1.2) :

$$x_i(n) = \sum_{k=-\infty}^{+\infty} \tilde{\mathbf{h}}_i^T(k) \mathbf{s}(n - k) \quad (1.3)$$

En pratique, les filtres $h_{ij}(n)$ sont des séquences finies de taille m . Ainsi :

$$x_i(n) = \sum_{k=0}^{m-1} \tilde{\mathbf{h}}_i^T(k) \mathbf{s}(n - k) = (\tilde{\mathbf{h}}_i^T * \mathbf{s})(n) \quad (1.4)$$

Enfin pour obtenir le modèle de mélange, c'est-à-dire l'écriture formelle qui prend en

¹³dans le cas non linéaire, nous disons uniquement, que les mélanges sont reliés aux sources par une fonction f non linéaire : $\mathbf{x} = f(\mathbf{s})$.

¹⁴pour une définition complète de la convolution et du filtrage voir, par exemple, [98].

¹⁵typiquement, les signaux continus sont soit échantillonnés directement pour un traitement numérique (sous condition de respecter le théorème de Shannon), soit sur-échantillonnés pour une simulation sur ordinateur [98].

compte l'ensemble des signaux mélanges, nous considérons le vecteur des mélanges \mathbf{x} et la matrice des filtres H ¹⁶. Le modèle de mélange sera donné par :

$$\mathbf{x}(n) = \sum_{k=0}^{m-1} H(k)\mathbf{s}(n-k) = (H * \mathbf{s})(n) \quad (1.5)$$

1.3.1.1 Une autre écriture du modèle

Le même modèle de mélange peut être écrit sous une autre forme équivalente à celle donnée dans la section 1.3.1. Comme dans la thèse de Belouchrani [7], en considérant les vecteurs $\mathbf{h}_j(n)$ qui contiennent, dans l'ordre, les éléments de la colonne j se trouvant dans la matrice H et en considérant les signaux sources $s_j(n)$, le modèle de mélange peut être formulé comme suit¹⁷ :

$$\mathbf{x}(n) = \sum_{j=1}^r \sum_{k=0}^{m-1} \mathbf{h}_j(k)s_j(n-k) = \sum_{j=1}^r (s_j * \mathbf{h}_j)(n) \quad (1.6)$$

1.3.1.2 Environnement bruité

En cas de présence de bruit dans l'environnement, un bruit $b_i(n)$ vient se superposer sur chaque capteur i ($i = 1 \dots, p$). En rassemblant ces signaux bruits dans un vecteur $\mathbf{b}(n)$, le modèle mathématique de mélange relatif à cette situation s'écrit comme suit :

$$\mathbf{x}(n) = (H * \mathbf{s})(n) + \mathbf{b}(n) \quad (1.7)$$

Dans ce qui suit, pour des raisons de simplification, nous considérons le cas non bruité de la séparation de sources et celui dont le nombre de sources est égale au nombre de capteurs.

1.3.1.3 Type de mélange : convolutif ou instantané

Le modèle de mélange exprimé dans le domaine temporel par l'équation (1.5), représente le cas général où les filtres de mélange exprimés par la matrice des filtres (ou matrice de mélange) $H(n)$, ont des coefficients non nuls en plusieurs instants n . Dans ce cas, les échantillons des sources se convolueront aux filtres pour donner les signaux qui sont des *mélanges convolutifs de sources*.

Dans certains cas, ce modèle peut être plus simple. En effet, quand les réponses impulsionnelles des filtres de propagation, exprimées par la matrice $H(n)$, sont toujours

¹⁶chaque ligne de cette matrice contient, dans l'ordre des sources, tous les filtres relatifs à {1 capteur/toutes les sources} et chaque colonne contient, dans l'ordre des capteurs, tous les filtres relatifs à {1 source/tous les capteurs}.

¹⁷pour des raisons de clarté, cette forme sera adoptée dans la section 1.3.4.

nulles sauf à l'instant d'indice $n = 0$, les signaux mélanges seront instantanément établis. C'est pourquoi on dit dans ce cas qu'il y a un *mélange instantané des sources*¹⁸. Le modèle de mélange devient plus simplement :

$$\mathbf{x}(n) = H(0)\mathbf{s}(n) \quad (1.8)$$

ou bien sous la deuxième forme (1.6) :

$$\mathbf{x}(n) = \sum_{j=1}^r s_j(n)\mathbf{h}_j(0) \quad (1.9)$$

On note que chaque cas, convolutif ou instantané, modélise des situations bien précises qu'on retrouve dans la pratique.

1.3.2 Idée de séparation

Sous sa forme matricielle (équation (1.5)), le mélange des signaux sources $s_j(n)$ composant les signaux capteurs $x_i(n)$, qu'on appelle aussi *système de transfert*, peut être mis en schémas-blocs (général) comme suit :

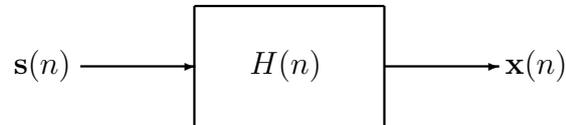


FIG. 1.3 – *système de transfert*

L'idée de séparation résolvant ce problème, consiste à estimer un système inverse¹⁹ à $H(n)$, de sorte que l'application des signaux capteurs à ce dernier, fournisse les signaux sources recherchés²⁰. Le schéma global de séparation de sources peut être, ainsi, tracé comme suit :

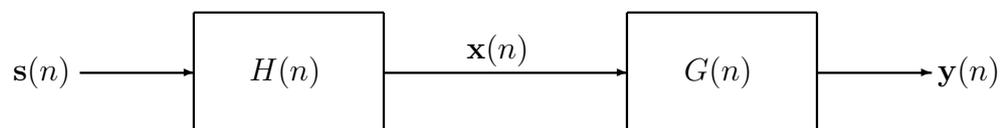


FIG. 1.4 – *schéma de séparation (système global)*

¹⁸par abus de langage, on dit également, mélanges instantanés (tout court) ou modèle instantané et pareil pour le cas convolutif (mélanges convolutifs / modèle convolutif).

¹⁹on dit, également, matrice séparante ou de séparation.

²⁰les estimés des signaux sources sont notés $y_j(n)$.

1.3.3 Hypothèses de base, indépendance et statistiques utilisées

1.3.3.1 Hypothèses

Afin d'accomplir la séparation des signaux, décrite dans les sections précédentes, nous résumons dans ce paragraphe les hypothèses sur lesquelles la séparation de sources est basée :

hypothèses sur les sources :

h1 : presque toutes les procédures de séparation aveugle de sources supposent que, statistiquement, les sources sont mutuellement indépendantes²¹.

h2 : De plus, dans le cas d'un processus i.i.d., au plus une source peut être gaussienne. Comon [35] (voir aussi [23]) démontre ce théorème dans le cas instantané (filtre de mélange constant) en utilisant les travaux de Cramer (1936) et de Darmois (1953). En effet, *minimiser la dépendance, c'est minimiser la corrélation tout en maximisant la non gaussianité* [27].

hypothèses sur l'environnement de propagation et les mélanges :

h3 : la plupart des méthodes de séparation de sources traitent le cas linéaire d'environnement de propagation et, ainsi, de mélanges.

h4 : en général, on suppose que le nombre de sources r est égal au nombre de capteurs p . Certaines méthodes traitent le cas particulier $p < r$.

1.3.3.2 Indépendance

Comme nous l'avons mentionné dans le précédent paragraphe et aux sections précédentes, la plupart des procédures de séparation aveugle de sources suppose l'hypothèse **h1**. Pour cela, différents critères de dépendance peuvent être trouvés dans la littérature (voir par exemple [41, pages 17-42], [68, 71, 26]). Nous résumons, dans ce qui suit, ces concepts de mesure de dépendance statistique :

- **moments et cumulants** : issus de la deuxième fonction caractéristique, notée : $\Psi_{\mathbf{u}}(\mathbf{v})$ ²², les cumulants (d'ordre q) de la variable aléatoire \mathbf{u} sont définis [68, 72, chapitre 1] par :

$$Cum_q(\mathbf{u}) = (-j)^q \left. \frac{d^q \Psi_{\mathbf{u}}(\mathbf{v})}{d\mathbf{v}^q} \right|_{\mathbf{v}=0} \quad (1.10)$$

On montre (voir par exemple [41, pages 20-21]) que les cumulants sont adaptés à 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 zéro²³.

²¹par définition, deux variables aléatoires sont indépendantes lorsque leur densité de probabilité est égale au produit de leurs densités marginales.

²²par transformation de Fourier de la distribution de la variable aléatoire \mathbf{u} , nous définissons la première fonction caractéristique $\Phi_{\mathbf{u}}(\mathbf{v}) = E[\exp(i\mathbf{u}\mathbf{v})]$, et la seconde par $\Psi_{\mathbf{u}}(\mathbf{v}) = \ln(\Phi_{\mathbf{u}}(\mathbf{v}))$.

²³en pratique, on se contente des ordres faibles des cumulants (2 à 4 en général).

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. C'est pourquoi, d'autres mesures de dépendance ont été développées.

- **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 la quantité :

$$\delta(p_u, p_v) = \int \log \frac{p_u(x)}{p_v(x)} p_u(x) dx \quad (1.11)$$

En utilisant les propriétés de la fonction log, nous démontrons que $\delta(p_u, p_v) \geq 0$ et 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.

- **information mutuelle** : ce critère calcule l'information mutuelle entre un vecteur et ces composantes (scalaires) aléatoires. Soit le vecteur \mathbf{u} de k composante u_i ($i = 1, \dots, k$),

$$i(p_{\mathbf{u}}) = \int p_{\mathbf{u}}(\mathbf{v}) \log \frac{p_{\mathbf{u}}(\mathbf{v})}{\prod_{i=1}^k p_{u_i}(v_i)} d\mathbf{v} \quad (1.12)$$

est l'information mutuelle qui mesure, la divergence de Kullback-Leibler entre $p_{\mathbf{u}}(\mathbf{v})$ et $\prod_{i=1}^k p_{u_i}(v_i)$, c'est-à-dire l'indépendance des u_i ($i(p_{\mathbf{u}}) = 0$).

1.3.3.3 Statistiques utilisées

Nous avons montré dans la sous section précédente que les cumulants peuvent être utilisés comme mesure d'indépendance dans la procédure de séparation de sources. Nous allons, dans cette sous-section, voir quel est l'ordre suffisant des statistiques, pouvant répondre à cette mesure²⁴ :

- **statistiques d'ordre 2** : partant de la décomposition en SVD, de la matrice des filtres de mélange H ($H = U\Delta^{1/2}V$: Δ est une matrice diagonale, U , V sont deux matrices unitaires), nous pouvons montrer (avec l'hypothèse de sources à puissance unité) que la matrice de covariance des mélanges ($\mathbf{x} = H\mathbf{s}$)²⁵ s'écrit :

$$\Gamma = E(\mathbf{x}\mathbf{x}^*) = U\Delta U^* \quad (1.13)$$

Il s'avère²⁶ ainsi, qu'il est impossible d'accomplir la séparation à l'ordre deux sans hypothèses supplémentaires. Cela est dû au fait que la matrice Γ est indépendante de la matrice V , qui est nécessaire pour retrouver l'inverse de H .

- **statistiques d'ordre 3** : ces statistiques ne sont généralement pas utilisées car elles sont nulles dans le cas des signaux à densité symétrique.

²⁴pour plus de détails voir [71].

²⁵dans ce qui suit, nous omettons la notation de la variable (n).

²⁶cette démonstration est aussi intitulée dans la littérature : insuffisance de l'ACP.

- **statistiques d'ordre 4** : à l'opposé des statistiques d'ordre deux qui sont insuffisantes pour réaliser la séparation, les cumulants d'ordre quatre permettent de retrouver²⁷ les sources recherchées dans les mélanges [71, 31].

1.3.4 Ambiguïté de séparation

Séparer aveuglement les sources contenues dans leurs mélanges est entaché d'ambiguïté. En effet, la séparation peut être obtenue avec une infinité de solutions [101, 21]. Nous allons, dans les sous-sections suivantes, montrer l'ambiguïté du modèle de séparation. Nous montrons par ailleurs pour chaque cas, instantané ou convolutif, le comportement de l'ambiguïté et son incidence sur la reconstruction des sources.

1.3.4.1 Ambiguïté de permutation

L'estimation de la matrice inverse des filtres n'est pas unique. Ceci peut être vu sur le modèle de mélange (cas instantané) en permutant par exemple deux colonnes (*e.g.* d'indice 1 et 2) dans la matrice de mélange :

$$\begin{aligned} \mathbf{x} &= \sum_{j=1}^p s_j \mathbf{h}_j \\ &= s_1 \mathbf{h}_1 + s_2 \mathbf{h}_2 + \cdots + s_p \mathbf{h}_p \end{aligned} \quad (1.14)$$

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

$$\begin{aligned} \mathbf{x} &= s_1 \mathbf{h}_1 + s_2 \mathbf{h}_2 + \cdots + s_p \mathbf{h}_p \\ &= s_2 \mathbf{h}_2 + s_1 \mathbf{h}_1 + \cdots + s_p \mathbf{h}_p \\ &= \begin{bmatrix} h_{12}s_2 \\ \vdots \\ h_{p2}s_2 \end{bmatrix} + \begin{bmatrix} h_{11}s_1 \\ \vdots \\ h_{p1}s_1 \end{bmatrix} + \cdots + \begin{bmatrix} h_{1p}s_p \\ \vdots \\ h_{pp}s_p \end{bmatrix} \\ &= \begin{bmatrix} h_{12} & h_{11} & \cdots & h_{1p} \\ h_{22} & h_{21} & \cdots & h_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ h_{p2} & h_{p1} & \cdots & h_{pp} \end{bmatrix} \begin{bmatrix} s_2 \\ s_1 \\ \vdots \\ s_p \end{bmatrix} \end{aligned} \quad (1.15)$$

Nous constatons, par rapport à la matrice de mélange initiale et par rapport au vecteur sources initial, qu'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*²⁸ est modélisée par une matrice de permutation

²⁷les signaux Gaussiens ont les cumulants d'ordre ≥ 3 nuls, c'est pourquoi on suppose en plus, \mathbf{h}_2 .

²⁸voir [21] et [55, pages 109-110] pour la définition de la matrice de permutation.

P qui contient seulement un élément égal à un par ligne et par colonne.

En notant la nouvelle matrice de mélange : \tilde{H} et le nouveau vecteur sources $\tilde{\mathbf{s}}$, le vecteur mélange s'écrira :

$$\begin{aligned}
 \mathbf{x} &= \tilde{H}\tilde{\mathbf{s}} \\
 &= HP\tilde{\mathbf{s}} \\
 &= HPP^{-1}\mathbf{s} \\
 &= H\mathbf{I}\mathbf{s} \\
 &= H\mathbf{s}
 \end{aligned} \tag{1.16}$$

On conclut que 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.

1.3.4.2 Ambiguïté d'échelle

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

$$\begin{aligned}
 \mathbf{x} &= s_1\mathbf{h}_1 + s_2\mathbf{h}_2 + \dots + s_p\mathbf{h}_p \\
 &= \left(\frac{1}{\alpha_1}s_1\right)(\alpha_1\mathbf{h}_1) + \left(\frac{1}{\alpha_2}s_2\right)(\alpha_2\mathbf{h}_2) + \dots + \left(\frac{1}{\alpha_p}s_p\right)(\alpha_p\mathbf{h}_p) \\
 &= \begin{bmatrix} \alpha_1 h_{11} & \alpha_2 h_{12} & \dots & \alpha_p h_{1p} \\ \alpha_1 h_{21} & \alpha_2 h_{22} & \dots & \alpha_p h_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_1 h_{p1} & \alpha_2 h_{p2} & \dots & \alpha_p h_{pp} \end{bmatrix} \begin{bmatrix} \frac{s_1}{\alpha_1} \\ \frac{s_2}{\alpha_2} \\ \vdots \\ \frac{s_p}{\alpha_p} \end{bmatrix}
 \end{aligned} \tag{1.17}$$

En factorisant ces scalaires dans une matrice, nous obtenons une matrice diagonale qu'on note D . En notant cette fois la nouvelle matrice de mélange \tilde{H} et le nouveau vecteur sources $\tilde{\mathbf{s}}$,

$$\begin{aligned}
 \mathbf{x} &= \tilde{H}\tilde{\mathbf{s}} \\
 &= HD\tilde{\mathbf{s}} \\
 &= HDD^{-1}\mathbf{s} \\
 &= H\mathbf{I}\mathbf{s} \\
 &= H\mathbf{s}
 \end{aligned} \tag{1.18}$$

Nous concluons que la multiplication des colonnes de la matrice de mélange par un scalaire ne change pas les mélanges, mais montre que l'amplitude des sources est indéterminée.

1.3.4.3 Ambiguïté de filtrage

Le plus souvent, la séparation de mélanges convolutifs de sources est ramenée à un ensemble de séparations instantanées. Ceci est réalisé par le passage dans le domaine de Fourier, où nous obtenons un modèle de mélange instantané à chaque fréquence. Comme nous l'avons vu, des ambiguïtés peuvent ainsi se produire à chaque fréquence ce qui implique un filtrage des signaux dans le temps ; Ceci bien évidemment sans remettre en cause le procédé de séparation de sources :

$$\begin{aligned} \mathbf{x} = H * \mathbf{s} &\xrightarrow{\mathcal{TF}} \mathbf{x}(f) = \left(H(f)F(f) \right) \left(F(f)^{-1} \mathbf{s}(f) \right) \\ &\quad \downarrow \mathcal{TFI} \\ \mathbf{x} &= (H * F) * (F^{-1} * \mathbf{s}) = \tilde{\tilde{H}} * \tilde{\tilde{\mathbf{s}}} \end{aligned} \quad (1.19)$$

En conséquence, la séparation des mélanges convolutifs peut être accomplie à un filtrage près²⁹ sans remettre en cause la démarche de séparation. Il faut par ailleurs noter que cette ambiguïté est la plus complexe et présente une grande gêne à la reconstruction des sources.

1.3.4.4 Ambiguïté complète

Dans le cas où les ambiguïtés s'associent, la solution que nous obtenons par la séparation de sources (dans le cas instantané) s'écrira de la forme :

$$\begin{aligned} \mathbf{x} &= (HDP)(P^{-1}D^{-1}\mathbf{s}) \\ &= (HDP)\mathbf{y} \\ &= G^{-1}\mathbf{y} \end{aligned} \quad (1.20)$$

où $G^{-1} = HDP$ et \mathbf{y} sont les nouvelles matrices de mélange et vecteur sources, qui forment les signaux mélanges aux capteurs. Ainsi, la séparation de sources, en partant des signaux mélanges, va rechercher à séparer des signaux sources contenus dans ces mélanges par l'estimation de la matrice G (système inverse du système de mélange) qui permet de calculer le vecteur des sources séparées (estimées), tels que :

$$\begin{cases} G = \hat{H}^{-1} = (HDP)^{-1} = (P^{-1}D^{-1}H^{-1}) \\ \text{et} \\ \mathbf{y} = \hat{\mathbf{s}} = G\mathbf{x} = (P^{-1}D^{-1}\mathbf{s}) \end{cases} \quad (1.21)$$

²⁹ $F(f)$ désigne le filtre d'indétermination.

Nous concluons que la séparation aveugle de sources, en plus des erreurs d'estimation, ne peut retrouver les signaux sources dans les mélanges linéaires, qu'à une permutation près, à un facteur d'échelle près et/ou à un filtrage près (cas convolutif).

Dans le cas favorable où les matrices P et D sont égales à la matrice identité, les sources estimées s'écriront :

$$\begin{aligned} \mathbf{y} &= (P^{-1}D^{-1}\mathbf{s}) \\ &= \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} s_1 + \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix} s_2 + \cdots + \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix} s_p \end{aligned} \quad (1.22)$$

Finalement, nous pouvons dire que les sources seront estimées à une possible permutation près et/ou un possible facteur d'échelle près³⁰. Dans le cas convolutif, toutes les indéterminations peuvent être modélisées par une opération complexe de filtrage des signaux. On dit dans ce cas, que les sources seront estimées à un filtrage près.

1.3.4.5 Système global

Nous rappelons que le système global, qu'on note C , est composé du système de mélange en cascade avec le système de séparation. Il permet d'obtenir les signaux séparés directement, à partir des signaux d'origine ($\mathbf{y} = C\mathbf{s}$). En prenant en compte l'ambiguïté intrinsèque de la SdS, ce système s'écrit :

$$\begin{aligned} C &= GH \\ &= (P^{-1}D^{-1}H^{-1})H \end{aligned} \quad (1.23)$$

Nous constatons que le système global sera, en plus des erreurs relatives à l'estimation, ambigu car les lignes peuvent être permutées et/ou multipliées, chacune, par un scalaire $1/\alpha_i$. En effet, si on note U l'estimée du filtre identité (ses éléments sont notés u_{ij}), le

³⁰dans le cas de sources complexes, un terme de phase définit les signaux ce qui engendre, en plus, une séparation à un déphasage (phase-shift) près [31].

filtre global estimé par la SdS sera de la forme :

$$\begin{aligned}
C &= P^{-1}D^{-1}U \\
&= P^{-1} \begin{bmatrix} 1/\alpha_1 & 0 & \dots & 0 \\ 0 & 1/\alpha_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1/\alpha_p \end{bmatrix} U \\
&= \begin{bmatrix} 0 & u_{22}/\alpha_2 & \dots & 0 \\ u_{11}/\alpha_1 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & u_{pp}/\alpha_p \end{bmatrix}
\end{aligned} \tag{1.24}$$

où la permutation dans cet exemple a été supposée parvenue dans les lignes d'indices 1 et 2, *i.e.* :

$$P^{-1} = \begin{bmatrix} 0 & 1 & \dots & 0 \\ 1 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix} \tag{1.25}$$

Si H, G, C sont respectivement les matrices de mélange, de séparation et d'estimation directe des sources (matrice globale), et si P, D sont respectivement une matrice de permutation et une matrice diagonale, alors, la multiplication à droite de la matrice de séparation, par P et/ou D ne remet pas en cause le processus de mélange, *i.e.* toutes les matrices de la forme HDP sont des matrices de mélange, la multiplication à gauche de la matrice de séparation, par P^{-1} et/ou D^{-1} ne remet pas en cause la séparation, *i.e.* toutes les matrices de la forme $P^{-1}D^{-1}\hat{H}^{-1}$ sont des matrices de séparation. De plus, dans le cas convolutif, le filtrage des signaux (sources et réponses impulsionnelles) ne remet pas en cause la procédure de séparation. Enfin, toutes les matrices de la forme $C = F^{-1} * U$ sont des matrices globales.

1.4 Conclusion

Dans ce chapitre nous avons donné une vue générale sur le problème de la séparation de sources : quelques exemples d'application, le modèle mathématique, les types de mélanges des sources, les hypothèses et statistiques utilisées en SdS. Nous retiendrons que la solution fournie par les méthodes de séparation aveugle n'est pas unique mais les sources séparées sont estimées à une permutation et à un facteur d'échelle (instantané) ou à un filtrage près (convolutif).

Chapitre 2

Algorithmes de Séparation de Sources

Après avoir introduit le problème de la séparation de sources dans le premier chapitre, nous allons présenter, dans ce chapitre, le principe de quelques algorithmes de séparation de sources. Nous trouvons dans la littérature différentes classifications d'algorithmes : selon le type de mélanges [71], selon le critère utilisé [60] et selon la nature du traitement réalisé [7]. La nôtre ne se prétend pas être exhaustive, mais essaie de donner, en deux sections, les idées de quelques algorithmes trouvés dans la littérature. Dans la première section, nous présentons les algorithmes qui traitent le cas du mélange instantané. Dans la seconde section, nous focalisons sur les algorithmes dédiés au cas convolutif en général et au cas de signaux acoustiques, en particulier. Cependant, le lecteur trouvera une classification exhaustive dans les références suivantes : [71, 1] et [60].

Pour chaque type de mélange, les algorithmes sont présentés dans un cadre général, en fonction de l'ordre de leurs statistiques et de leur critère, en fonction de la structure non linéaire qu'ils utilisent ou bien de l'hypothèse de mélange linéaire des signaux sources ou en fonction de la nature de la séparation de sources qu'ils réalisent soit dans le domaine de Fourier, soit dans le temps.

2.1 Mélanges instantanés

En 1985, Héroult, Jutten et Ans publiaient le premier algorithme de séparation de sources. Ce travail qui a inauguré la recherche dans le domaine de la séparation de sources, traite le cas du mélange instantané. Dès lors, beaucoup d'articles ont été publiés sur la séparation de sources, notamment sur les mélanges instantanés. Nous présentons, en premier, l'algorithme initial de la SdS [59]. Celui-ci modélise la structure de séparation sous la forme d'un réseau de neurones bouclé. Quant au reste des algorithmes de cette section, ils traitent des mélanges supposés linéaires. Les algorithmes de Héroult *et al.* [59] et de Cardoso [30] et de nombreux autres utilisent les statistiques d'ordre supérieur à deux dans la conception des contraintes d'indépendance des sources. Cependant, d'autres algorithmes [102, 7, 90, 9, 10, 44] peuvent être fondés sur les statistiques d'ordre deux moyennant des informations a priori supplémentaires. Dans ces algorithmes la seule

information disponible dans les observations est interprétée soit par le calcul des matrices de corrélation, soit par des matrices de covariance ou soit par les matrices spectrales. L'algorithme de Belouchrani et Amin [9, 10] procède dans le plan temps-fréquence tandis que le reste des algorithmes procèdent dans le temps. Certains de ces algorithmes réalisent une diagonalisation conjointe d'un ensemble de matrices. Ceux de Cardoso et de Belouchrani sont basés sur la technique Jacobienne et celui de Pham [90] est issu de la technique du maximum de vraisemblance :

- *algorithme de J. Héroult et al. [59]* : l'architecture de calcul neuromimétique utilisée pour l'analyse des messages composites neurobiologiques, fut à la base du développement de la séparation de sources. Dans leur article [59], Héroult *et al.* formulent le problème de séparation de sources, qu'ils définissent comme suit : "*supposons que, pour l'analyse d'un système complexe, nous disposions d'un ensemble de n voies de mesure. Chaque voie, issue d'un capteur spécifique, fournit à chaque instant t un signal image d'une certaine combinaison d'un nombre limité de variables internes du système. En supposant que ces variables soient des primitives indépendantes, ou simplement non corrélées, est-il possible de les isoler par une méthode d'analyse statistique, c'est-à-dire sans connaissance a priori sur ces variables ni sur leur poids relatif dans chaque voie de mesure ?*".

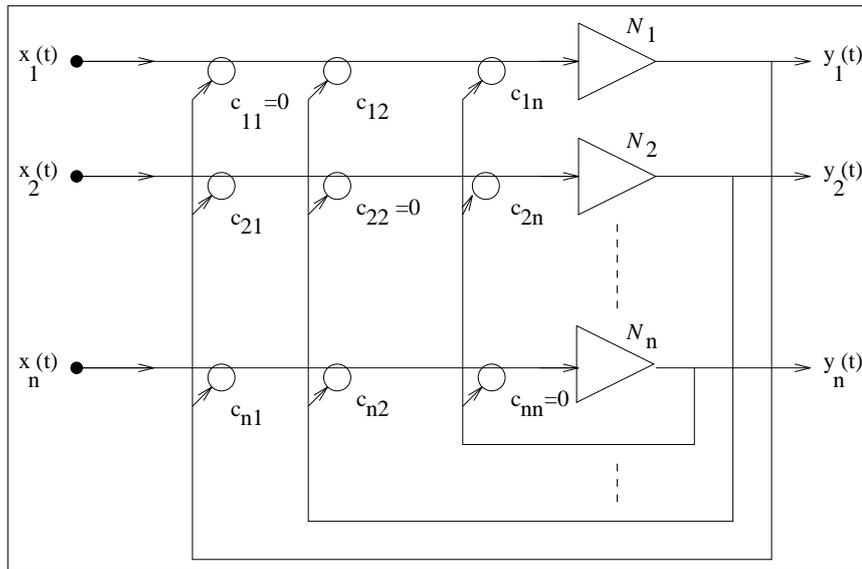


FIG. 2.1 – réseau neuromimétique

C'était la première définition de la SdS. À la base de cette définition et ses hypothèses, le problème traité a été résolu dans le cadre d'observation de mélanges linéaires instantanés, temporels et non bruités. Le principe de résolution se résume comme suit :

1. le vecteur des mélanges $\mathbf{x}(t)$ appliqué à l'entrée de l'architecture (à gauche) étant lié au vecteur des sources statistiquement indépendantes $\mathbf{s}(t)$ par la transformation linéaire $\mathbf{x}(t) = H\mathbf{s}(t)$, l'estimation des signaux sources est donnée par la sortie du réseau neuromimétique (à droite) $\mathbf{y}(t)$, où chacune de ses composantes y_i , rela-

tive à l'opérateur N_i est estimée par la soustraction des contributions, des autres composantes de sortie, à l'entrée du même indice i , *i.e.* :

$$y_i = x_i - \sum_{j=1}^n c_{ij} \cdot y_j \quad (2.1)$$

sous la contrainte $c_{ii} = 0$

2. le problème se réduit désormais à estimer les paramètres c_{ij} de l'architecture.
3. Pour cela, deux fonctions non linéaires différentes f, g ont été utilisées pour approcher la contrainte d'indépendance¹ (mieux qu'à l'ordre 2) des composantes y_j et pour estimer ces coefficients via une procédure adaptative :

$$c_{ij}(t+1) = c_{ij}(t) + \mu_{ij} \phi_{ij}(C(t), \mathbf{y}), \quad i \neq j, i \leq n, j \leq n. \quad (2.2)$$

où $\phi_{ij}(C(t), \mathbf{y}) = f(y_i)g(y_j)$, $C(t)$ est la matrice des coefficients c_{ij} et μ est un scalaire appelé pas d'adaptation.

À noter que les auteurs ont proposé d'améliorer la règle d'adaptation en introduisant le signe de la dérivée de ϕ_{ij} et les cumulants. Ainsi l'applicabilité est généralisée à toutes les sources indépendamment de leur loi de probabilité :

$$c_{ij}(t+1) = c_{ij}(t) + \mu_{ij} \text{signe} \left[\frac{\partial \text{Cum}(y_i^3 y_j)}{\partial c_{ij}(t)} \right] \text{Cum}(y_i^3 y_j)(t) \quad (2.3)$$

• *l'algorithme JADE de J.-F. Cardoso [30]* : L'algorithme JADE commence par une étape de blanchiment dans laquelle les observations blanchies sont à matrice spectrale diagonale. Cette étape est basée sur l'utilisation des statistiques d'ordre deux. Elle rend les observations indépendantes à cet ordre. Puis une deuxième étape dite de rotation va consister à rendre les observations indépendantes à l'ordre quatre via un critère à base des cumulants d'ordre quatre. L'auteur résume son algorithme de la sorte :

1. constituer la matrice de covariance R_x et calculer une matrice de blanchiment W ;
2. constituer les cumulants d'ordre quatre $\text{Cum}_4(\mathbf{z})$ (noté aussi Q_z), du processus blanchi $\mathbf{z} = W\mathbf{x} = U\mathbf{s} + W\mathbf{n}$, (U : matrice unitaire, W : matrice de blanchiment et \mathbf{n} : bruit blanc), et calculer l'ensemble de ses n plus significatives (ordre croissant en valeur absolu) valeurs et matrices propres $\{\lambda_r, M_r\}$; $1 \leq r \leq n$, (n est le nombre des sources).

En effet, $Q_z = \sum_{p=1}^n k_p M_p M_p^*$ (k_p est le kurtosis de la source s_p , M est une matrice hermitienne). Par analogie avec la notion de décomposition en valeurs propres, la

¹l'idée derrière cela, est que si nous pouvons retrouver deux fonctions impaires et non linéaires f et g , tel que $f(y_i)$ et $g(y_j)$ ($i \neq j$) soient décorrélées, alors les composantes seront indépendantes au sens de la minimisation d'intercumulants d'ordre 4.

décomposition des cumulants est constituée de "matrices propres" M_r et valeurs propres λ_r ;

3. diagonaliser conjointement l'ensemble des matrices $\lambda_r M_r$ par une matrice unitaire \hat{U} , où la diagonalisation a été définie par la maximisation du critère suivant :

$$C(U, \mathcal{N}) = \sum_{r=1}^n |\text{diag}(U^* N_r U)|^2 \quad (2.4)$$

où $|\text{diag}(\cdot)|$ est la norme du vecteur composé des éléments diagonaux de la matrice argument et $\mathcal{N} = \{N_r | 1 \leq r \leq n\}$ est l'ensemble à diagonaliser ;

4. estimer la matrice de séparation $A = W \# U$

où le symbole $\#$ dénote la matrice pseudo-inverse de Moore-Penrose.

L'auteur utilise la technique de Jacobie étendue pour maximiser son critère de diagonalisation conjointe.

• l'algorithme AMUSE de L. Tong et al. [102] : Dans cet article, les auteurs supposent que les sources sont colorées et de spectres différents. Ils formulent mathématiquement le problème de SdS en tenant compte de la présence du bruit additif $\mathbf{n}(t)$. Ils proposent un algorithme simple pour l'identification, dans lequel ils se basent sur la matrice de covariance des sources. AMUSE se résume ainsi :

1. estimer la matrice de covariance des observations $R_{\mathbf{x}}(\tau) = E \{\mathbf{x}(t)\mathbf{x}(t - \tau)\}$;
2. calculer la SVD de cette matrice : $R_{\mathbf{x}} = U\Phi U^T = U_s \Psi^2 U_s^T + \sigma^2 I$, où $U = U_s U_n$ est unitaire, $\Phi = \text{diag}(\psi_1 + \sigma^2, \psi_2 + \sigma^2, \dots, \psi_m + \sigma^2, \sigma^2, \dots, \sigma^2)$, $R_{\mathbf{n}} = \sigma^2 I$ et $\Psi = \text{diag}(\psi_1, \psi_2, \dots, \psi_m)$;
3. faire une transformation orthogonale : $\mathbf{y} = C\mathbf{x} = C(As + \mathbf{n})$, où A est la matrice de mélange, $C = \text{diag}\left(\frac{1}{\psi_1}, \frac{1}{\psi_2}, \dots, \frac{1}{\psi_m}\right)$;
4. définir le nouveau système : $\mathbf{y} = B\mathbf{s} + W$, où $B = CA$ et $W = C\mathbf{n}$ (\mathbf{n} : vecteur bruit additif) ;
5. choisir un retard τ tel que : $\frac{R_{\mathbf{y}}(\tau) + R_{\mathbf{y}}(\tau)^T}{2}$ ait des valeurs propres distinctes ; V est la matrice des vecteurs propres de cette décomposition en valeurs et vecteurs propres ;
6. estimer la matrice de séparation : $\hat{A} = U_s \Phi V$;
7. estimer les signaux sources : $\hat{\mathbf{s}} = V^T \mathbf{y}$.

• l'algorithme SOBI de A. Belouchrani [7] : dans le même esprit (sources colorées de spectres différents), à base des statistiques d'ordre deux, Belouchrani a développé, dans le cadre de sa thèse, une méthode de séparation de sources instantanément mélangées et bruitées². Pour cela, il fait appel à la propriété d'intercorrélacion des signaux sources qui

²le bruit considéré est supposé indépendant des sources et spatialement blanc (n'intervient pas dans la fonction de corrélation de la partie signal)

est interprétée par la matrice d'intercovariance à plusieurs retards. Ce travail est apparu dans l'article [8].

Nous avons montré dans la section 1.3.3.3 qu'il est impossible d'accomplir la séparation aveugle de sources via les statistiques d'ordre deux. Pourquoi SOBI (ou AMUSE) arrive-elle à identifier les sources ? Pour répondre à cette question nous allons exposer le principe de la méthode :

Sous l'hypothèse de signaux sources centrés et décorrélés, à spectres normalisés différents (sources différentes) et dans le cadre d'un processus d'entrée ergodique³ et stationnaire multivarié, la matrice de covariance est égale à :

$$E[\mathbf{s}(t + \tau)\mathbf{s}(t)^*] = \text{diag}[\rho_1(\tau), \rho_2(\tau), \dots, \rho_n(\tau)] \quad (2.5)$$

où $\rho_i(\tau) = E[s_i(t + \tau)s_i^*(t)]$ est l'autocovariance de s_i et $\text{diag}[\cdot]$ est la matrice diagonale formée par les éléments de son argument.

La séparation passe par une étape de blanchiment permettant de transformer le système de transfert d'un nouveau système \mathbf{z} garantissant :

1. le filtre de mélange est inversible (matrice de mélange unitaire) :

$$\mathbf{z} = W\mathbf{x} = W[\mathbf{y}(t) + \mathbf{n}(t)] \quad (2.6)$$

où $\mathbf{y}(t) = A\mathbf{s}(t)$, A est la matrice de mélange.

En effet :

Sous l'hypothèse de sources à puissance unité ($R_s(0) = I$)⁴ et en absence du bruit, l'énergie de la partie signal des mélanges s'écrit :

$$R_y(0) = E[\mathbf{y}(t)\mathbf{y}(t)^*] = AE[\mathbf{s}(t)\mathbf{s}(t)^*]A^* = AA^* \quad (2.7)$$

Ainsi, l'application de la matrice de blanchiment au vecteur \mathbf{y} donne :

$$E[W\mathbf{y}(t)\mathbf{y}(t)^*W^*] = WR_y(0)W^* = WAA^*W^* = I = UU^* \quad (2.8)$$

où $U = WA$ est unitaire ; Ceci implique :

$$A = W\#U \quad (2.9)$$

où le symbole $\#$ dénote la matrice pseudo-inverse de Moore-Penrose.

Cela nous permet enfin d'écrire l'équation (2.6) (blanchiment des observations) comme suit :

$$\mathbf{z} = WAs(t) + W\mathbf{n}(t) = \boxed{U\mathbf{s}(t) + W\mathbf{n}} \quad (2.10)$$

³voir [75, pages 82-85] pour la définition de l'ergodisme.

⁴vu l'indétermination d'échelle, cette convention ou standardisation n'engendre pas la perte de généralité car la rang de puissance des sources est relevé par la colonne correspondante de la matrice de mélange

où W peut être estimée à partir de la matrice de covariance de la partie signal (\mathbf{y}) du processus initial et la connaissance de la forme de la covariance du bruit⁵.

2. la matrice de covariance (du nouveau système) est diagonalisable (à une ambiguïté près) par U , certifiant son existence :

$$R_{\mathbf{z}} = WR_{\mathbf{x}}W^* = UR_{\mathbf{s}}U^* = \boxed{U \text{diag}[\rho_1(\tau), \rho_2(\rho), \dots, \rho_n(\tau)]U^*} \quad (2.11)$$

L'auteur montre la difficulté de retrouver la matrice U à partir de la diagonalisation de la covariance, du processus blanchi, à un retard donné. En effet, cela revient à l'ambiguïté intrinsèque de la SVD que nous résumons comme suit :

- a) une matrice n-carrée est diagonalisable si elle possède n vecteurs propres linéairement indépendants ;
- b) des vecteurs propres associés à des valeurs propres distinctes sont linéairement indépendants ;
- c) il existe une infinité de vecteurs propres associés à chaque valeur propre

Il propose une situation plus favorable qui revient à diagonaliser conjointement plusieurs matrices de covariance, à plusieurs retards. Cela augmente significativement la robustesse de la séparation.

Après l'estimation de la matrice U il sera possible d'estimer les sources et les filtres de mélange :

$$\hat{\mathbf{s}}(t) = \hat{U}^* \hat{W} \mathbf{x}(t) \quad (2.12)$$

$$\hat{A} = \hat{W} \# \hat{U} \quad (2.13)$$

On conclut que l'exploitation de l'information contenue dans les matrices de covariance, en plusieurs retards, rend la séparation statistiquement plus efficace en faisant une estimation à base d'un large ensemble de statistiques. Au quel cas les statistiques d'ordre deux réalisent la séparation. L'utilisation des informations statistiques, disponibles sur les sources, à plusieurs retards rend la SdS possible via les statistiques d'ordre deux.

• *l'algorithme de D.-T. Pham et J.-F Cardoso [90]* : dans ce travail les auteurs ont développé un algorithme simple et efficace pour la séparation des signaux sources non stationnaires à amplitude variable ; par exemple pour la séparation des signaux de parole. L'algorithme optimise mathématiquement, "on-line" ou bien "off-line", deux fonctions de coût selon une approche fondée sur le maximum de vraisemblance.

Sous l'hypothèse de vecteurs aléatoires Gaussiens, pour le développement de leurs algorithmes, les auteurs partent du maximum de vraisemblance qui est défini dans ce

⁵la méthode n'exige pas que le bruit soit spatialement blanc ni d'une forme particulière.

contexte par :

$$C_{ML} = \frac{1}{T} \sum_{t=1}^T \frac{1}{2} \text{tr}[\Sigma^{-2}(t)A^{-1}X(t)X(t)^T A^{-T}] + \frac{1}{2} \log \det[2\pi\Sigma^2(t)] + \log |\det A|, \quad (2.14)$$

où A est la matrice de mélange et Σ est la matrice de covariance des sources.

En considérant les T échantillons partagés en L "blocs gaussiens", avec des matrices de covariance R_l différentes (modèle du non stationnarité), ils développent leur calcul pour montrer que le minimum du critère de maximum de vraisemblance dans ce contexte est :

$$C_{ML}^* = \sum_{l=1}^L w_l \text{off}(A^{-1}\hat{R}_l A^{-T}) + \text{Constante}, \quad (2.15)$$

où $\text{off}(\cdot)$ mesure la déviation à la diagonalité : on peut démontrer que $\text{off}(\cdot) = D\{.\}$ et $D\{.\}$ = divergence de Kullback-Leibler entre deux densités de probabilité. Celui-ci constitue une fonction de coût qui est un critère de diagonalisation conjointe mesurant la séparabilité. Dans cette équation, w_l est la proportion entre les échantillons de données dans le l -ième sous bloc et le nombre de blocs T .

La deuxième fonction de coût proposée est fondée sur l'information mutuelle gaussienne (sources gaussiennes)⁶. Ils remplacent la covariance par un estimateur de noyau non paramétrique ; ainsi le critère est donné par :

$$C_{MI} = \frac{1}{L} \sum_{l=1}^L \text{off}[B\hat{R}(lT/L)B^T] \quad (2.16)$$

À partir de cela, sont issus les algorithmes suivants :

a. "*algorithme off-line*" : dans cet algorithme le critère du maximum de vraisemblance (2.15) est minimisé par l'algorithme de diagonalisation conjointe proposé par D.-T. Pham [89].

b. "*algorithmes on-line*" :

i. *gradient stochastique simple* : la matrice de séparation est itérativement adaptée selon :

$$\hat{B}(t+1) = \hat{B}(t) - \lambda G(t)\hat{B}(t) \quad (2.17)$$

avec $G(t) = \hat{\Sigma}^{-2}\hat{\mathbf{S}}\hat{\mathbf{S}}^T - I$ et λ est une petite constante positive.

Les éléments de $\hat{\Sigma}^2$, $\hat{\sigma}_k^2$, sont calculés par :

$$\hat{\sigma}_k^2(t) = \hat{\sigma}_k^2(t-1) + \rho[\hat{s}_k^2(t) - \hat{\sigma}_k^2(t-1)], \quad (2.18)$$

ρ est une petite constante positive de l'étape d'apprentissage.

⁶les auteurs mentionnent que leurs algorithmes sont applicables même dans le cas de sources non-stationnaires non-Gaussiennes.

ii. *technique quasi-Newton* : elle se résume ainsi :

1. calculer $\hat{\mathbf{s}} = B(t-1)\mathbf{x}$, mettre à jour $\hat{\sigma}_i^2(t)$ selon (2.18) et $\hat{\omega}_{ij}(t)$ selon :

$$\hat{\omega}_{ij}(t) = \hat{\omega}_{ij}(t-1) + \lambda[\hat{\sigma}_j^2(t)/\hat{\sigma}_i^2(t) - \hat{\omega}_{ij}(t-1)] \quad (2.19)$$

2. mettre à jour $\hat{B}(t)$ ($\hat{B}(t) = \hat{B}(t-1) - \lambda H(t)\hat{B}(t-1)$) où la diagonale de H est mise à zéro et son antidiagonale est la solution de :

$$\begin{bmatrix} h_{ij} \\ h_{ji} \end{bmatrix} = \begin{bmatrix} \hat{\omega}_{ij} & 1 \\ 1 & \hat{\omega}_{ji} \end{bmatrix}^{-1} \begin{bmatrix} \hat{s}_i \hat{s}_j / \hat{\sigma}_i^2 \\ \hat{s}_j \hat{s}_i / \hat{\sigma}_j^2 \end{bmatrix} \quad (2.20)$$

iii. une version "on-line" de l'approche "bloc-Gaussien" est aussi proposée.

En résumé, les auteurs décrivent, plus précisément, leurs algorithmes comme étant applicables aux sources indépendantes à "lente" modulation d'amplitude.

• *l'algorithme TFS de A. Belouchrani et M. G. Amin [9, 10]* : un algorithme pour les sources non stationnaires a été proposé dans ce papier. Il utilise aussi les statistiques d'ordre deux, mais à la différence de SOBI, il est basé sur l'exploration de l'information statistique concernant les sources, dans le plan temps-fréquence. L'algorithme sépare aussi des sources gaussiennes. Ce travail a été réétudié récemment par C. Févotte [49], A. Yeredor [112], où une extension de cet algorithme a été proposée.

La séparation s'obtient par la diagonalisation conjointe d'un ensemble de matrices de distribution spatiale temps-fréquence qui informent sur le lien statistique spatial entre les sources. L'algorithme se résume comme suit :

1. estimer la matrice d'autocorrélation $\hat{\mathbf{R}}$ des observations à partir de T échantillons de données. Mettre cette matrice, par diagonalisation, sous la forme :

$$\hat{\mathbf{R}} = H\Delta H^* \quad (2.21)$$

où $\lambda_1, \lambda_2, \dots, \lambda_n$ sont les valeurs propres de $\hat{\mathbf{R}}$ et $H = [h_1, h_2, \dots, h_n]$ (n : nombre des sources) est la matrice des vecteurs propres de $\hat{\mathbf{R}}$.

2. étant donné :

$$\hat{\mathbf{R}} = H\Delta H^* = H_s H_n \Delta H_s^* H_n^* = H_s \tilde{\Delta} H_s^* + \rho^2 I \quad (2.22)$$

où H_s est la matrice des valeurs propres, issue de la SVD de R_s , H_n est la matrice des valeurs propres, issue de la SVD de R_n et $\Delta = \text{diag}[\lambda_1^2 + \rho^2, \lambda_2^2 + \rho^2, \dots, \lambda_n^2 + \rho^2, \rho^2, \dots, \rho^2]$ ($\text{diag}[\cdot]$ est la matrice diagonale formée par les éléments de son argument). La variance (puissance) du bruit (supposé blanc) est estimée en moyennant

les $m - n$ petites valeurs propres dans Δ :

$$\hat{\rho}^2 = \frac{1}{m - n} \sum_{i=n+1}^m \lambda_i \quad (2.23)$$

Ainsi, étant donné :

$$\begin{cases} \hat{R} = H\Delta H^*; \\ AA^* = \hat{R} - \rho^2 I; \\ WAA^*W^* = I \end{cases} \quad (2.24)$$

l'estimé de la matrice de blanchiment peut être choisi :

$$\hat{W} = \hat{\Delta}\hat{H}^* = \text{diag}[\lambda_1^2 - \rho^2, \lambda_2^2 - \rho^2, \dots, \lambda_n^2 - \rho^2]^{-\frac{1}{2}} [h_1, \dots, h_n]^* \quad (2.25)$$

3. diagonaliser conjointement les matrices de l'STFD de $z(t)$ calculées à un point fixe (t_i, f_i) , $i = 1, \dots, K$ du plan temps-fréquence correspondant aux autotermes du signal. La diagonalisation conjointe de l'ensemble de matrices de "distribution spatiale temps-fréquence" (ou covariance) est résolue comme dans [8] par la technique de Jacobi en minimisant les termes hors diagonale.
4. estimer $\hat{\mathbf{s}}(t) = \hat{U}^* \hat{W} \mathbf{x}(t)$ et $\hat{A} = \hat{W} \# \hat{U}$.

L'analyse de la performance de l'algorithme montre qu'il est plus robuste que SOBI.

• l'algorithme OPCA de K. I. Diamantaras et Th. Papadimitriou [44] : l'OPCA est une version étendue de la technique classique d'analyse en composante principale (voir définition) les auteurs ont développé un algorithme qui utilise les statistiques d'ordre deux. Comme la PCA, qui correspond à la décomposition en valeurs propres d'une seule matrice de covariance, selon les auteurs, l'OPCA correspond à la décomposition d'une paire de matrices de variance-covariance en valeurs propres. Ainsi, l'objectif de cette technique est de maximiser la fonction de coût donnant le rapport entre une paire de signaux n -dimensionnel \mathbf{u} et \mathbf{v} :

$$J_{OPCA}(\mathbf{w}) = \frac{E(\mathbf{w}^T \mathbf{u})^2}{E(\mathbf{w}^T \mathbf{v})^2} = \frac{\mathbf{w}^T R_u \mathbf{w}}{\mathbf{w}^T R_v \mathbf{w}} \quad (2.26)$$

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

Définition 2.1 ([60, page 126]) *la PCA d'un vecteur aléatoire $\mathbf{x}(k) \in \mathbb{R}^m$ est obtenu en retrouvant un système de coordonnées réalisant une opération d'orthogonalisation de sorte que les composantes de \mathbf{x} dans ce système deviennent décorréélées, orthogonales et ordonnées en énergie croissante.*

Afin d'appliquer l'OPCA, l'idée des auteurs est de considérer un nouveau signal (y)

issu du filtrage des observations par un filtre $\mathbf{h} = [h_0, \dots, h_M]$:

$$\mathbf{y}(k) = \sum_{m=0}^M h_m \mathbf{x}(k - l_m) \quad (2.27)$$

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

$$R_y(0)A^{-T} = R_x(0)A^{-T}D \quad (2.28)$$

où $D = \sum_{p,q=0}^M h_p h_q R_s(l_p - l_q)$ (l_p, l_q sont deux retards).

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

Il se pose après le problème d'initialisation de \mathbf{h} et le critère qui détermine optimalement les valeurs de ces retards. Les auteurs proposent un filtre initial de la forme $\mathbf{h} = [h_0, h_1] = [1, \alpha]$. Ils proposent aussi un processus itératif pour calculer α en utilisant un critère $J(\alpha)$ fondé sur la distribution des valeurs propres.

- autres algorithmes : plusieurs autres algorithmes basés sur différents critères, existent aussi dans la littérature (voir [71, 1]). Nous pouvons mentionner par exemple : approches basées sur le maximum de vraisemblance [91, 28, 11], approches basées sur l'information mutuelle Gaussienne [88], approche basée sur la déflation [62], approche géométrique [73], approches basées sur le maximum d'entropie [6].

2.2 Mélanges convolutifs

Le problème de séparation de mélanges convolutifs de sources intervient dans beaucoup de domaines d'applications. Les premiers travaux modélisant cette situation de mélanges ont été initiés au début des années 90 ; selon ce que l'on trouve dans la littérature [71, 69, 24] le premier travail qui traite le mélange convolutif est celui de C. Jutten *et al.* [61, 100]. Le mélange convolutif, relativement au mélange instantané, a été moins étudié mais il existe néanmoins beaucoup d'algorithmes pour ce type de mélange qui convient à une variété d'applications physiques.

M. Babaie-Zadeh classe dans sa thèse [5] qui contient un excellent état de l'art, les algorithmes convolutifs selon le domaine de traitement et les statistiques utilisées en : algorithmes temporels à statistiques d'ordre supérieur, algorithmes temporels à statistiques d'ordre deux et enfin des algorithmes fréquentiels transformant le convolutif en un en-

semble d'instantané^{7,8}. Nous adoptons ce même ordre de classification pour citer quelques algorithmes pour la séparation de mélanges convolutifs : les algorithmes [78, 79, 80, 96] sont basés sur un traitement dans le domaine fréquentiel. Celui-ci utilise des statistiques d'ordre supérieur. Les algorithmes dans [104, 85, 84, 110, 76, 81] adoptent une approche temps-fréquence, par contre celui de Weinstein [106] sépare les sources en agissant sur l'adaptation du filtre inverse en fréquence. Les algorithmes [61, 100, 99, 4] procèdent, quant à eux, dans le temps. Nous retrouvons différents critères d'indépendance dans ces travaux, certains sont basés sur le gradient [110, 108, 85, 84] et d'autres sur le maximum de vraisemblance [76]. Quelques articles considèrent la propriété de nonstationnarité des signaux [104, 76, 85, 84]. Cependant, peu de travaux dans le domaine fréquentiel proposent des idées efficaces pour résoudre le problème de l'ambiguïté de permutation dans le cas convolutif [81, 85, 84, 110, 4].

• *l'algorithme de C. Jutten et al.* [61, 100] : dans ce travail les auteurs proposent une extension et une généralisation de leur méthode déjà développée pour le mélange instantané. Le principe de cette méthode qui considère deux sources se résume comme suit :

1. avec le même principe, l'estimée de chaque source i est calculée en fonction des contributions des autres selon :

$$\hat{s}_i(t) = e_i(t) - \sum_{j=1}^p \sum_{k=0}^M c_{ij}(k) \hat{s}_j(t-k) \quad tq : i \neq j \quad (2.29)$$

sous la contrainte : $c_{ii} = 0$

2. par analogie au cas du mélange instantané, la règle itérative d'ajustement des coefficients c_{ij} peut être :

i.

$$c_{ij}(t+1, k) = c_{ij}(t, k) + \alpha f(\hat{s}_i(t)) g(\hat{s}_j(t-k)) \quad (2.30)$$

dans le cas où le critère d'indépendance est celui de l'annulation des moments croisés d'ordre impaire : $E[f(s_i)g(s_j)]$

ii.

$$c_{ij}(t+1, k) = c_{ij}(t, k) - \alpha \text{signe} \left(\frac{\partial \text{Cum}(\hat{s}_i(t)^2 \hat{s}_j(t-k)^2)}{\partial c_{ij}(t, k)} \right) \text{Cum}(\hat{s}_i(t)^2 \hat{s}_j(t-k)^2) \quad (2.31)$$

si le critère d'indépendance est celui de l'annulation de l'intercumulant symétrique : $\text{Cum}(s_i(t)^2 s_j(t-k)^2)$

⁷pour cette classe C. Févotte présente dans sa thèse [48] une généralisation des algorithmes de l'instantané au cas convolutif

⁸le problème majeur de cette approche, qui semble simple, est l'ambiguïté de permutation locale à chaque canal de fréquence lors de la reconstruction des sources

iii.

$$c_{ij}(t+1, k) = c_{ij}(t, k) - \alpha \operatorname{signe} \left(\frac{\partial \operatorname{Cum}(\hat{s}_i(t)\hat{s}_j(t-k)^3)}{\partial c_{ij}(t, k)} \right) \operatorname{Cum}(\hat{s}_i(t)\hat{s}_j(t-k)^3) \quad (2.32)$$

si le critère d'indépendance est celui de l'annulation de l'intercumulant dissymétrique : $\operatorname{Cum}(s_i(t)s_j(t-k)^3)$

• *l'algorithme de R. Mukai et al. [78, 79, 80]* : Mukai *et al.* ont étudié le problème de séparation et déréverbération de mélanges convolutifs de sources. Ils tiennent compte des propriétés acoustique des signaux, ils traitent le problème en fréquence (ensemble d'ICA fréquentiel). Leur procédure consiste en :

1. une TF court terme :

$$\mathbf{x}(\omega, m) = H(\omega)\mathbf{s}(\omega, m), \quad \mathbf{y}(\omega, m) = W(\omega)\mathbf{x}(\omega, m) \quad (2.33)$$

où H , W sont les filtres de mélange et de séparation ; m est l'index du bloc de TF.

2. W est calculé en utilisant :

$$W_{i+1} = W_i + \eta [\operatorname{diag}(\langle \Phi(\mathbf{y})\mathbf{y}^H \rangle) - \langle \Phi(\mathbf{y})\mathbf{y}^H \rangle] W_i \quad (2.34)$$

où Φ est une fonction non linéaire définie par :

$$\Phi(\mathbf{y}) = \frac{1}{1 + e^{-\operatorname{Re}(\mathbf{y})}} + j \frac{1}{1 + e^{-\operatorname{Im}(\mathbf{y})}} \quad (2.35)$$

3. les auteurs ont donné beaucoup d'importance à l'évaluation de la performance du système de séparation. Pour ce, ils analysent les parts de la propagation directe de chaque source et sa contribution dans l'autre capteur. Les signaux sont calculés dans le domaine temporel par un produit de convolution, après le calcul de la réponse impulsionnelle du filtre de séparation par une TF inverse. De plus la réduction de réverbération est aussi analysée.

• *l'algorithme de P. Smaragdis [96]* : En passant dans le domaine temporel en l'instantané dans le domaine fréquentiel, Smaragdis construit aussi un algorithme qui sépare les sources des mélanges convolutifs, fréquence par fréquence :

1. par une transformée de Fourier court terme, le modèle adopté est :

$$\mathbf{x}_f(t) = A_f \mathbf{s}_f(t) \quad (2.36)$$

où t est la variable relative au bloc de la TF et f est la variable de Fourier.

2. afin de minimiser la dépendance, l'auteur a choisi un critère qui existe dans la

littérature pour le cas du mélange instantané. Il est défini par l'équation :

$$\Delta W \propto [I - f(\mathbf{y}(t))\mathbf{y}^H(t)] W \quad (2.37)$$

où W est le filtre de séparation recherché, \mathbf{y} est l'estimée des sources et $f(\cdot)$ est une fonction non linéaire.

3. le problème de permutation locale aux canaux de fréquence est résolu en imposant deux contraintes. La première est relative à la norme du filtre :

$$W_f^{norm} = W_f^{orig} |W_f^{orig}|^{-\frac{1}{N}} \quad (2.38)$$

où W_f^{norm} , W_f^{orig} sont respectivement le filtre normalisé et original à la fréquence f et N est le nombre des sources.

Et la deuxième est relative à la continuité du filtre en fréquence :

$$\Delta_a W_{f+1} = \Delta_e W_{f+1} + k \Delta W_f \quad (2.39)$$

où $\Delta_a W_{f+1}$ et $\Delta_e W$ sont les facteurs (respectivement appliqués et estimés) de mise à jour et $0 < k < 1$ est une constante.

- *l'algorithme SIMO-ICA de T. Takatani et al. [99]* : les auteurs proposent dans ce travail un nouvel algorithme pour la séparation de signaux acoustiques dans un système à une entrée et plusieurs sorties. L'algorithme est composé d'un ensemble d'analyse en composantes principales (voir définition ci-dessous) et d'un contrôleur de fidélité au système de séparation. L'expérimentation de cet algorithme montre une bonne qualité des signaux audio séparés.

Définition 2.2 ([34, page 7]) *l'ICA d'un vecteur aléatoire $\mathbf{x}(k) \in \mathbb{R}^m$ est obtenu en retrouvant une matrice séparatrice G , $n \times m$ ($m \geq n$) de rang plein, tel que le vecteur signal de sortie $\mathbf{y}(k) = [y_1(k), \dots, y_n(k)]$, défini par :*

$$\mathbf{y}(k) = G\mathbf{x}(k), \quad (2.40)$$

contiendra les estimées des composantes source $\mathbf{s}(k) \in \mathbb{R}^n$, les plus indépendantes possibles selon une évaluation par une fonction de coût comme la divergence de Kullback-Leibler.

En étudiant l'ICA standard, les auteurs ont constaté que celle-ci ne pouvait pas sauvegarder la qualité des signaux audio. C'est pourquoi ils ont proposé une extension d'ICA dans leur algorithme qui procède comme suit :

1. calculer les signaux séparés du l -ième ICA (dans le SIMO-ICA) :

$$\mathbf{y}_{ICA}(t) = [y_k^{(l)}(t)]_{k1} = \sum_{n=0}^{D-1} W_{ICA}(n) \mathbf{x}(t-n) \quad (2.41)$$

où D est la longueur du filtre de séparation W .

2. l'efficacité du système ("contrôleur de fidélité") sera évaluée par une nouvelle fonction coût à minimiser :

$$\left\| \sum_{l=1}^L \mathbf{y}_{ICA}(t) - \mathbf{x}(t - D/2) \right\|^2 \quad (2.42)$$

où $\|\cdot\|$ désigne la norme euclidienne.

Avec ces deux étapes, la séparation est achevée en sauvegardant la qualité sonore des sources.

• *l'algorithme MSICA de T. Nishikawa et al. [82]* : dans cet algorithme deux approches ICA, temporelles et fréquentielles, sont combinées afin de construire une procédure de séparation plus robuste. L'algorithme procède en deux étapes :

- i. l'FDICA est appliquée dans cette étape où la séquence de sortie \mathbf{z} est calculée par le filtre de séparation optimal issu du critère :

$$W_{i+1}^{(l)}(f) = W_i^{(l)}(f) + \alpha [\text{diag}(\langle \Phi(\mathbf{z}(f, t)) \mathbf{z}^*(f, t) \rangle_t) - \langle \Phi(\mathbf{z}(f, t)) \mathbf{z}^*(f, t) \rangle_t] W_i^{(f)}(f) \quad (2.43)$$

où Φ est une fonction non linéaire.

- ii. l'TDICA est appliquée dans cette étape pour une optimisation supérieure. L'algorithme cherche à nouveau un filtre w . Pour ce, trois variantes d'TDICA sont proposées :

1. le filtre est obtenu en minimisant le critère :

$$Q(w^{(m)}(\tau)) = \frac{1}{2B} \sum_{b=1}^B \left\{ \log \left(\frac{\det \text{diag} \langle \mathbf{y}(t) \mathbf{y}(t)^T \rangle_t^{(b)}}{\det \langle \mathbf{y}(t) \mathbf{y}(t)^T \rangle_t^{(b)}} \right) \right\} \quad (2.44)$$

où B est le nombre de blocs.

2. le critère proposé cette fois est :

$$KLD(w^{(m)}(\tau)) = \int p(\mathbf{y}(t)) \log \frac{p(\mathbf{y}(t))}{\prod_{l=1}^L \prod_{t=0}^{T-1} p(y_l(t))} d\mathbf{y}(t) \quad (2.45)$$

3. en dernier, c'est le critère suivant :

$$w_{i+1}^{(n)}(\tau) = w_i^{(n)}(\tau) + \alpha \sum_{d=0}^{Q-1} [\text{diag}(\langle \Phi(\mathbf{y}(t)) \mathbf{y}^T(t - \tau + d) \rangle_t) - \langle \Phi(\mathbf{y}(t)) \mathbf{y}^T(t - \tau + d) \rangle_t] w_i^{(n)}(d) \quad (2.46)$$

• l'algorithme de W. Wang et al. [104] : basé sur la diagonalisation conjointe et exploitant la nonstationnarité des signaux sources, l'algorithme proposé par les auteurs utilise les statistiques d'ordre deux pour accomplir la séparation. Grâce à la non stationnarité les signaux sont découpés en blocs où une TF est calculée sur chacun d'eux. Le problème est ramené ainsi, à un ensemble d'ICA fréquentiel à résoudre par diagonalisation conjointe. Le principe est le suivant :

1. construire les matrices de covariance à diagonaliser conjointement selon :

$$R_Y(\omega, k) = W(\omega)[R_X(\omega, k) - R_V(\omega, k)]W^*(\omega) \quad (2.47)$$

où W est le filtre de séparation et V est le bruit additif.

2. pour la diagonalisation conjointe, utiliser le critère :

$$J(W) = \arg \min_W \sum_{\omega=1}^T \sum_{k=1}^K \{J_M(W)(\omega, k) + \lambda J_C(W)(\omega, k)\} \quad (2.48)$$

où λ est un facteur de blanchiment et :

$$J_M(W)(\omega, k) = \|R_Y(\omega, k) - \text{diag}[R_Y(\omega, k)]\|^2 \quad (2.49)$$

$$J_C(W)(\omega, k) = \|\text{diag}[W(\omega) - I]\|^2 \quad (2.50)$$

3. pour trouver une solution à (2.48) au sens du moindre carré, les auteurs proposent le gradient stochastique :

$$W^{(l+1)}(\omega) = W^{(l)}(\omega) - \mu(\omega) \frac{\partial J}{\partial W^{(l)*}(\omega)} \quad (2.51)$$

Pour résoudre le problème de permutation, les auteurs suggèrent plusieurs méthodes qui existent déjà dans la littérature.

• l'algorithme de E. Weinstein et al. [106] : Ce travail s'intéresse aux situations physiques réelles, tel que l'environnement acoustique. En modélisant l'environnement par un système LIT qu'on lui considère deux entrées et deux sorties, les auteurs proposent un algorithme basé sur les statistiques d'ordre deux. Ils supposent que les sources sont décorréliées et imposent, pour réaliser la séparation, des contraintes appropriées sur le système de reconstruction des sources. Pour cela ils utilisent le principe du moindre carré

de Widrow.

Le principe de leurs approches est le suivant :

1. dans l'objectif de reconstruire les sources s_1, s_2 à un filtre près, respectivement F_1, F_2 , ils ajustent les composantes du filtre inverse G . Pour ce, ils supposent sans perte de généralité que $H_{11} = H_{22} = 1$ ⁹, et selon les deux cas de permutation, ajustent les éléments du filtre afin d'avoir une diagonale (ou antidiagonale) nulle dans le filtre global $H(\omega)G(\omega)$.
2. pour aboutir à cela ils proposent un critère de décorrélation en fonction des densités de puissance des observations :

$$P_{y_1 y_2}(\omega) - G_{12}(\omega)P_{y_2 y_2}(\omega) - G_{21}^*(\omega)P_{y_1 y_1}(\omega) + G_{12}G_{21}^*(\omega)P_{y_2 y_1}(\omega) = 0 \quad (2.52)$$

3. étant donnée l'infinité de solutions à cette équation, les auteurs proposent une solution qui améliore la solution du moindre carré de Widrow :

$$G_{12}(\omega) = \frac{P_{y_1 y_2}(\omega) - G_{21}^*(\omega)P_{y_1 y_1}(\omega)}{P_{y_2 y_2}(\omega) - G_{21}^*(\omega)P_{y_2 y_1}(\omega)} \quad (2.53)$$

$$G_{21}(\omega) = \frac{P_{y_2 y_1}(\omega) - G_{12}^*(\omega)P_{y_2 y_2}(\omega)}{P_{y_1 y_1}(\omega) - G_{12}^*(\omega)P_{y_1 y_2}(\omega)} \quad (2.54)$$

4. enfin, dans l'implémentation de leur algorithme (version fréquentielle), la densité spectrale de puissance est estimée par le périodogramme et les équations (2.53) sont mises dans un processus itératif. Dans le domaine temporel, l'implémentation de l'algorithme est basée sur la transformée inverse de Fourier.

• l'algorithme de J. Anemüller et B. Kollmeier [4] : Anemüller *et al.* ont étudié la séparation des mélanges convolutifs de signaux de parole. L'algorithme qu'ils ont proposé pour cela est dérivé du principe de décorrélation par modulation d'amplitude. Cet algorithme rassemble l'information nécessaire à la séparation, simultanément de plusieurs fréquences. À partir de l'analyse d'un spectrogramme de parole, les auteurs déduisent que la production de la voix humaine suit naturellement le changement d'amplitude en fonction de la variable temps du spectrogramme dans différentes fréquences. C'est pourquoi, ils appellent cette propriété "corrélacion de modulación d'amplitude" (AMCor). L'algorithme se résume en :

1. une matrice d'AMCor est calculée en considérant deux séries temporelles de deux sources différentes en deux fréquences :

$$C^{xy} = [c_{\alpha\beta}^{xy}] = E \{ \mathbf{x}_\alpha(t), \mathbf{x}_\beta(t) \} \quad (2.55)$$

⁹une paramétrisation similaire a été adoptée dans [95].

2. la séparation consiste à minimiser l'AMCor en utilisant le critère :

$$H\left(\{\hat{W}_\alpha\}\right) = \sum_{i,j \neq i} \text{tr}\left([C^{ij}]^T C^{ij}\right) \quad (2.56)$$

• l'algorithme de L. Parra et C. Spence [85, 84] : Parra s'est intéressé au problème des signaux acoustiques simultanément enregistrés dans un environnement réverbérant. Pour les identifier, il a proposé une méthode de séparation aveugle de mélanges convolutifs, basée sur la non stationnarité des signaux. En effet, cette propriété permet de construire un ensemble suffisant d'informations conduisant à la séparation. Il reconstruit les sources par un filtre inverse FIR.

Comme la majorité des algorithmes de séparation de mélanges convolutifs, l'algorithme de Parra transforme le problème dans le domaine de Fourier en un ensemble de séparations effectuées fréquence par fréquence. L'algorithme procède comme suit :

1. pour des considérations pratiques, le passage au domaine de Fourier est réalisé par une TFD et le modèle de mélange est exprimé par :

$$\mathbf{x}(\omega, t) = A(\omega)\mathbf{s}(\omega, t) + \mathbf{n}(\omega, t) \quad (2.57)$$

où t est l'index des blocs et ω est l'index de la variable de Fourier.

2. vu l'hypothèse de nonstationnarité des sources¹⁰ la matrice de covariance est exprimée en fonction de la variable index-bloc t :

$$R_x(\omega, t) = A(\omega)\Lambda_s(\omega, t)A^H + \Lambda_n(\omega, t) \quad (2.58)$$

En pratique la matrice de covariance des mélanges est estimée par le périodogramme :

$$R_x(\omega, t) = \frac{1}{N} \sum_{n=0}^{N-1} \mathbf{x}(\omega, t + nT)\mathbf{x}^H(\omega, t + nT) \quad (2.59)$$

3. vu l'hypothèse d'indépendance qui implique la décorrélation, les matrices $\Lambda_s(\omega, t)$ et $\Lambda_n(\omega, t)$ sont diagonales. Ainsi le problème se ramène à une diagonalisation conjointe des matrices de même fréquence :

$$\Lambda_s(\omega, t) = W(\omega) (R_x(\omega, t) - \Lambda_n(\omega, t)) W^H(\omega) \quad (2.60)$$

où W est le filtre inverse issu de la diagonalisation conjointe.

4. le critère de diagonalisation utilisé est adaptatif et utilise la méthode des moindres carrés. Enfin, pour résoudre le problème de permutation, les auteurs imposent la contrainte de continuité ("lissage") dans le domaine des fréquences.

¹⁰étant donnée la transformation linéaire $\mathbf{x} = A\mathbf{s}$, la propriété de nonstationnarité est propagée aux signaux mélanges.

• *l'algorithme de H.-C. Wu et J. C. Principe [110]* : avec le même principe, les auteurs résolvent le problème du mélange convolutif dans le domaine de Fourier fréquence par fréquence. Ils diagonalisent conjointement les matrices de covariance des mélanges de même fréquence. Pour ce, ils utilisent leur nouvel algorithme de diagonalisation conjointe. La résolution se résume en :

1. initialiser l'algorithme par une valeur initiale des filtres de séparation $W(\omega)$.
2. calculer les matrices de covariance $\Sigma_Y(\omega, t_i)$:

$$\Sigma_Y(\omega, t_i) = W(\omega)\Sigma(\omega, t_i)W^H(\omega) \quad (2.61)$$

3. calculer la nouvelle adaptation de W :

$$W^{new}(\omega) = W^{old}(\omega) + \Delta W(\omega) \quad (2.62)$$

où :

$$\Delta W(\omega) = \frac{\partial J(\omega)}{\partial W(\omega)} \quad (2.63)$$

$$J(\omega) = \sum_{t_i} \left\{ \sum_{j=1}^N \log[\sigma_{Y_{jj}}(\omega)] - \log[|\det[\Sigma_Y(\omega, t_i)]|] \right\} \quad (2.64)$$

4. construire $W(t)$ et itérer jusqu'à la convergence.

Pour résoudre le problème de permutation, les auteurs imposent des contraintes sur la longueur du filtre de séparation et une troncature en blocs apodisée.

• *l'algorithme de N. Mitianoudis et M. Davis [76]* : ce travail concerne le problème de la séparation des sources audio mélangées dans un contexte réel. Les auteurs adoptent le principe de l'ICA dans le domaine fréquentiel, et ils proposent un algorithme basé sur le passage dans le domaine STFT. Ils proposent également une solution au problème de permutation persistant à chaque fréquence.

1. le filtre inverse est estimé en maximisant la log-vraisemblance :

$$\log p(W_f | \mathbf{x}(f)) \propto \log p(W_f) + E\{\log p(\mathbf{u}(f))\} + \log \det W_f \quad (2.65)$$

2. deux contraintes sont imposées pour résoudre le problème de permutation :

- i. une contrainte de couplage fréquentiel du canal de propagation :

$$p(W_f | W_{f-1}) \propto \exp\left(-\frac{1}{2\sigma^2} \|W_f - W_{f-1}\|_F\right) \quad (2.66)$$

- ii. une contrainte de couplage fréquentiel des sources :

$$p(u_k(f, t)) \propto \beta_k(t)^{-1} \exp(-h(u_k(f, t)/\beta_k(t))) \quad (2.67)$$

où β est un paramètre de non stationnarité relatif au bloc t et $h(\cdot)$ définit la forme de la densité des sources u .

- de cette nouvelle proposition, les auteurs suggèrent un complément à la méthode de Smaragdis [96] pour renforcer sa contrainte d'anti-permutation qu'ils qualifient d'insuffisante. Le nouvel estimateur est :

$$\Delta W_f = \eta(I - \beta_k(t)^{-1} E g(\mathbf{u}(f, t)) \mathbf{u}^H(f, t)) W_f \quad (2.68)$$

où $g(u) = u/|u|$ et $\beta_k(t) = \frac{1}{N} \sum_f |u_k(f, t)|$

- enfin, les auteurs fournissent pour leur algorithme une implémentation qui améliore sa rapidité et sa stabilité.

• *l'algorithme de A. Westner et V. M. Bove [108]* : Westner a travaillé sur l'application d'algorithmes de séparation de sources aux vrais environnements de mélange des sources. Il a effectué des expériences réelles dans des studios d'enregistrement acoustique. Les signaux enregistrés sont ensuite appliqués aux algorithmes de la SdS. Selon son expérience, Westner conclut que les algorithmes basés sur l'infomax sont les mieux adaptés à la séparation des sons acoustiquement mélangés.

- il utilise le critère du gradient stochastique pour l'identification du filtre de séparation W :

$$\Delta W \propto [W^T]^{-1} + g(\mathbf{u}(t)) x(t)^T \quad (2.69)$$

- il utilise pour la séparation une version modifiée de l'algorithme de Smaragdis [96] dans laquelle, précède une étape de prétraitement du bruit.

• *l'algorithme de N. Murata et S. Ikeda [81]* : Par ce travail, les auteurs ont proposé un algorithme en ligne pour la séparation de sources qui sont enregistrées dans un environnement réel. En particulier ils considèrent la situation où des personnes parlent en se déplaçant dans une salle (cas dynamique de la séparation de sources). Ils présentent une extension de leurs algorithmes dans une nouvelle version, dont l'idée de base est de transposer les signaux mélangés dans le domaine temps-fréquence. L'algorithme proposé se résume en plusieurs étapes :

- en se basant sur la propriété de la voix humaine qui est stationnaire pour une durée inférieure à 10 msec et devient non stationnaire à partir d'une durée de 100 msec, les auteurs appliquent une transformée de Fourier aux signaux mélangés. Ils obtiennent le spectrogramme suivant :

$$\hat{\mathbf{x}}(\omega, t_s) = \sum_t \exp^{-j\omega t} \mathbf{x}(t) w(t - t_s) = \hat{A}(\omega) \hat{\mathbf{s}}(\omega, t_s) \quad (2.70)$$

où w est une fenêtre d'appodisation (de Hamming).

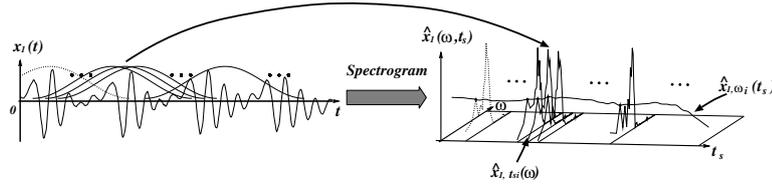


FIG. 2.2 – spectrogramme

2. pour l'extraction de composantes indépendantes, ils utilisent un réseau de neurones dans lequel la sortie est définie par :

$$\hat{\mathbf{u}}(\omega, t_s) = \hat{\mathbf{x}}(\omega, t_s) - B(\omega, t_s)\hat{\mathbf{u}}(\omega, t_s) \quad (2.71)$$

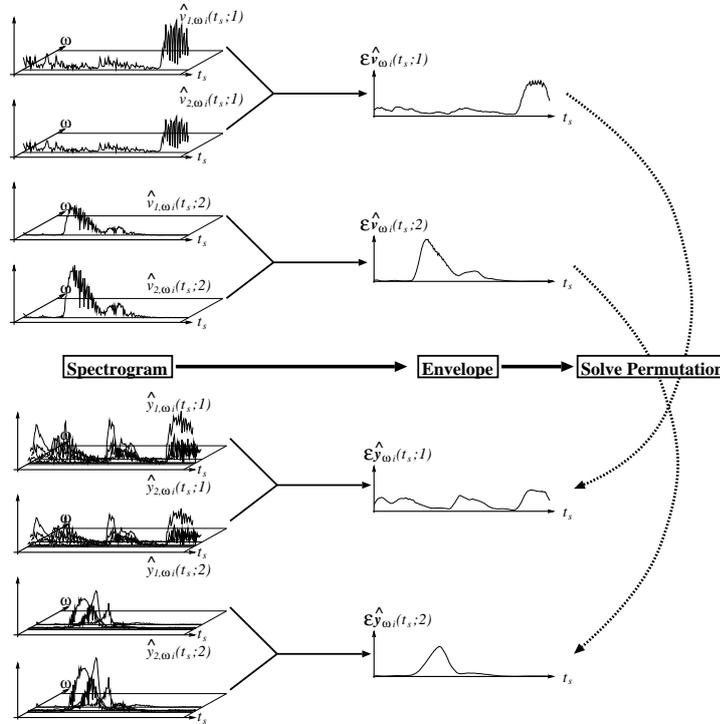


FIG. 2.3 – reconstruction du spectrogramme

3. les sources sont estimées par l'équation :

$$\hat{\mathbf{v}}_\omega(t_s; i) = (B(\omega, t_s) + I)(0, \dots, \hat{u}_i(\omega, t_s), \dots, 0)^T \quad (2.72)$$

où B est déterminée par :

$$B(\omega, t_s + \Delta T) = B(\omega, t_s) - \eta(B(\omega, t_s) + I)(\text{diag}(\phi(\mathbf{z})\mathbf{z}^*)) \quad (2.73)$$

où $\mathbf{z} = \hat{\mathbf{u}}(\omega, t_s)$ et $\phi(z) = \tanh(\text{Re}(z)) + i \tanh(\text{Im}(z))$

4. les auteurs proposent de résoudre le problème de permutations fréquentielles en se

référant à la structure temporelle des signaux. Ils définissent l'enveloppe :

$$\xi_{\hat{\mathbf{v}}_\omega}(t_s; i) = \frac{1}{2M} \sum_{t'_s=t_s-M}^{t_s+M} |\hat{\mathbf{v}}_\omega(t'_s; i)|, \quad (2.74)$$

ils cherchent ainsi la permutation $\sigma_\omega(i)$ qui maximise la corrélation entre $\xi_{\hat{\mathbf{v}}_\omega}(t_s; \sigma_\omega(i))$ et $\xi_{\hat{\mathbf{y}}_\omega}(t_s; i) = \xi_{\sum_{\omega'} \hat{\mathbf{v}}_{\omega'}(t_s; \sigma_{\omega'}(i))}$

2.3 Conclusion

Dans ce chapitre nous avons revu les algorithmes de séparation de sources trouvés dans la littérature (cas de mélanges linéaires instantanés et convolutifs). Fondamentalement, la littérature est riche d'algorithmes de séparation de sources de mélanges instantanés où les critères de dépendance sont variés et les résultats de séparation sont de très bonne qualité. Cependant, ces algorithmes n'ont pas la capacité de séparer les signaux rencontrés dans la majorité des situations physiques réelles, qui correspondent au cas convolutif.

La plupart des algorithmes de séparation de mélanges convolutifs de sources transforme le problème en un ensemble de séparations instantanées dans le domaine de fréquence où les mêmes critères de dépendance sont réutilisés. La performance de ces algorithmes dépend du critère de séparation ainsi que de la résolution du problème crucial d'ambiguïté ! Peu d'algorithmes du cas convolutif présentent une bonne qualité de séparation dans le cas de signaux réels, notamment dans le cas d'enregistrements effectués dans un environnement acoustique réel. En conséquence, nous développons un algorithme de séparation de mélanges convolutifs de sources dédié aux signaux acoustiques (audio, parole, ...). Nous nous focalisons en particulier sur la bonne résolution du problème local de permutation qui persiste à chaque canal de fréquence.

Chapitre 3

Algorithmes pour la Séparation Aveugle de Mélanges Convolutifs de Sources

Les systèmes digitaux de traitement du signal sont de nos jours très utilisés dans l'enregistrement, le traitement et la perception audio. Dans le cadre de cette thèse, nous nous sommes intéressés au développement des algorithmes pour la séparation de mélanges convolutifs, entre autres, de sources audio. La motivation pour cela est l'applicabilité de ces développements sur des signaux et des réponses impulsionnelles réelles fournies par le partenaire McMaster University dans le cadre du projet BLISS [16], auquel nous avons participé, pour la validation d'algorithmes dédiés aux systèmes digitaux d'aide aux malentendants¹ [18, 103]. Nous nous sommes fixés dans ce projet, l'objectif d'explorer les aspects d'utilisation des techniques de séparation de sources pour l'aide aux malentendants, en commençant par des situations contrôlées où les sources et les capteurs sont immobiles sans présence de bruit, puis dans une deuxième étape, à des situations moins contrôlées (personnes qui parlent en se déplaçant dans un environnement réel).

Biologiquement, le fonctionnement et la capacité d'un système auditif humain normal à réaliser la séparation auditive de sources (ou l'analyse de scènes auditives) ne sont pas encore totalement compris [103]! En effet, l'être humain est capable, dans un environnement acoustique réel, de focaliser son attention sur une source sonore brouillée dans un mélange de sons. Malheureusement, pour les malentendants, cette capacité de focalisation devient sévèrement limitée. Afin de compenser cette faiblesse, des circuits à base de DSP et microphones sont développés pour être placés sur l'oreille et amplifier les signaux, présentant ainsi une aide aux malentendants. Le problème est que cette amplification ne les aide pas à comprendre la parole et interpréter l'information dans un milieu bruité ou réverbérant.

Notre challenge est de fournir et valider des algorithmes (dont l'application principale

¹par abus de langage on dit, tout court, *aide aux malentendants* [94, page 8], en omettant "système digital".

est l'aide aux malentendants) qui seront capables de reconnaître et d'isoler les différents sons qui se trouvent dans un environnement acoustique réel contenant plusieurs sources sonores.

Les algorithmes de séparation de sources acoustiques doivent permettre aux systèmes d'acquisition acoustique l'accès aux différents objets audio qui se trouvent dans l'environnement acoustique. Cela doterait le système de certaines caractéristiques comme par exemple : choix de la façon dont ces signaux doivent être enregistrés et traités, analyse de l'environnement acoustique (nombre et localisation de sources, réverbération, ...), perception auditive, ...; Westner [109] a regroupé un ensemble d'application pour les algorithmes de séparation aveugle de sources acoustiques :

- studio de musique ;
- studio de films ;
- télé-présence ;
- traitement de l'audio mobile.

Ce chapitre est consacré à la description des algorithmes développés et la résolution de l'ambiguïté de permutation effectuée dans le domaine de fréquence.

3.1 Mélanges acoustiques

Dans un environnement acoustique réel, les *objets*² émettent des sons qui se dispersent dans tous les sens, en conséquence :

- un malentendant dans une salle acoustique reçoit une scène auditive complexe, à cause des différentes localisations géographiques des objets et de leur contenu fréquentiel et temporel varié.
- un autre facteur rend la situation plus complexe : c'est la réverbération de la salle dont les échos amplifient l'altération des signaux et embrouillent la localisation et la séparation des sources.
- de plus, les sons arrivent à l'oreille dans tous les sens (voir Fig. 3.1), chevauchés en temps et en fréquence.

Tous ces effets acoustiques rendent la séparation dans ce contexte complexe *très difficile*. En effet, le système d'aide au malentendant (à chaque microphone) ne capte pas que les copies des sources (à un retard de propagation près), mais aussi plusieurs copies (modifiées) réfléchies des autres sources. Le niveau de distorsion des signaux captés dépend des caractéristiques acoustiques de réverbération (voir définition ci-dessous) de l'environnement.

²chaque objet acoustique correspond à une source de signal

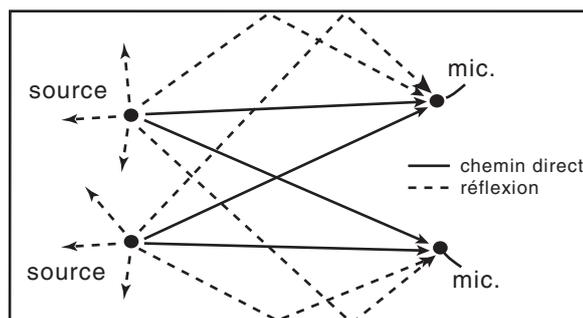


FIG. 3.1 – chemins acoustiques de 2 sources dans une salle réelle

Définition 3.1 ([74, pages 191,560]) *Un écho est une onde qui a été réfléchiée ou, en d'autres termes, retournée avec une amplitude et un retard suffisants pour qu'elle soit perçue comme une onde distincte. La réverbération sonore est la persistance à un point donné des sons réfléchis (échos) par un mur ou une surface d'un environnement acoustique, après la réception directe du son source.*

La mission de l'aide aux malentendants sera, en conséquence, de débrouiller l'information acoustique afin de déterminer "qui" parle et "d'où" (*i.e.* déterminer tous les flux sonores qui existent dans l'environnement). Ainsi, le malentendant compensera sa difficulté et pourra entendre le son qui l'intéresse sans la perturbation de tout le fond acoustique présent dans l'environnement.

Afin de répondre à cet objectif, nous utiliserons les techniques de séparation de sources pour développer un algorithme qui constituera le cœur d'un DSP d'aide aux malentendants.

3.2 La nonstationnarité

Comme il a été vu dans le Chapitre 2, la résolution du problème de séparation de sources dans le cas du mélange convolutif s'effectue souvent par des approches qui transforment le mélange convolutif en un ensemble de mélanges instantanés. Ces approches transposent le problème dans le domaine de Fourier par une TFD. En conséquence, il est possible d'utiliser les méthodes des mélanges instantanés à chaque canal de fréquence puis de reconstruire les sources en ayant l'ensemble des solutions fréquentielles. Le problème semble, ainsi, simplifié. Cependant, la mise en œuvre de cette approche se heurte à certains problèmes.

En effet, en plus du problème d'ambiguïté d'échelle et de permutation intrinsèque à la séparation de sources, la plupart des méthodes de séparation reposent sur l'hypothèse de non gaussianité des sources. Or, dans le cas général, il est bien connu que la TFD produit des signaux qui sont asymptotiquement gaussiens. Capdevielle a étudié dans sa thèse [24, chapitre 2] la distance à la gaussianité des signaux, après transformée de Fourier

et a proposé un test de mesure pour la validation d'applicabilité des algorithmes dans le domaine fréquentiel. Par ailleurs, Pham et Cardoso [90] ont proposé une alternative qui permet de réaliser la séparation de sources tout en ignorant leur non gaussianité. Cette méthode exploite la *nonstationnarité* des sources, comme dans le cas des sources audio par exemple, pour construire un ensemble de statistiques différentes permettant d'accomplir la séparation de sources qu'elles soient gaussiennes ou non.

Dans le développement de notre algorithme nous avons adopté cette idée de nonstationnarité des sources pour réaliser la séparation de signaux audio dans le domaine de fréquence.

3.3 Méthodes de séparation dans le cas non bruité

Afin de modéliser la situation où plusieurs signaux se propagent à partir de leurs sources productrices et arrivent à un ensemble de capteurs de mesure³, nous nous plaçons dans le contexte convolutif représenté par le modèle :

$$\mathbf{x}(n) = (H * \mathbf{s})(n) \quad (3.1)$$

où H est la matrice des filtres modélisant l'environnement acoustique de propagation, \mathbf{x} est le vecteur des signaux (mélanges) reçus par le système d'aide, et \mathbf{s} est le vecteur des signaux originaux qui se sont propagés dans l'environnement et sont modifiés par les filtres ; n représente la variable discrète du temps⁴.

Pour répondre à l'objectif visant la résolution de ce problème, nous devons identifier l'environnement acoustique de propagation (*i.e.* la matrice des filtres inverses $G(n)$) et reconstruire les sources. *i.e.*

$$\mathbf{y}(n) = (G * \mathbf{x})(n) \quad (3.2)$$

où $\mathbf{y}(n)$ est le vecteur des estimés des signaux sources.

Ainsi, le problème consiste à chercher le filtre de séparation G de sorte que les sources estimées soient aussi indépendantes que possible.

Afin de simplifier le problème nous ramenons le produit de convolution matriciel à un produit matriciel simple par l'application d'une TFD. Le modèle de séparation devient :

$$\mathbf{y}(f) = G(f)\mathbf{x}(f) \quad (3.3)$$

où f est la variable de Fourier discrète, et $\mathbf{y}(f)$, $G(f)$, $\mathbf{x}(f)$ sont, respectivement, les TFD de $\mathbf{y}(n)$, $G(n)$, $\mathbf{x}(n)$.

En adoptant une approche au second ordre, l'idée de séparation consiste à rendre les

³plusieurs cas réels correspondants à cette situation peuvent être énumérés. Nous en citons le problème d'aide aux malentendants par exemple.

⁴tous les signaux sont des séquences numériques d'échantillons

sources décorréliées. Nous construisons ainsi leur matrice spectrale :

$$\begin{aligned} (3.3) \implies S_y(f) &= E((G(f)\mathbf{x}(f))(G(f)\mathbf{x}(f))^*) \\ &= G(f)S_x(f)G^*(f) \end{aligned} \quad (3.4)$$

La décorrélation des sources estimées revient, en conséquence, à rendre la matrice $S_y(f)$ diagonale.

Afin de mettre en œuvre ce principe en surmontant le problème d'unicité de la solution⁵, nous exploitons la nonstationnarité des sources en considérant la matrice spectrale des sources qui est variable en temps (selon la division en blocs de stationnarité) et aussi en fréquence. De cette manière, nous construisons un ensemble de statistiques qui nous informe sur l'évolution du contenu fréquentiel des signaux dans le temps et qui nous permet aussi de séparer les sources. Le spectre d'un bloc temporel autour de l'instant n s'écrit :

$$S_y(n, f) = G(f)S_x(n, f)G^*(f) \quad (3.5)$$

Si les signaux \mathbf{y} sont indépendants, l'échange d'énergie entre les sources doit être négligeable (faible ou nul). Ainsi, l'ensemble des matrices spectrales $S_y(n, f)$ de tous les blocs n à une fréquence f , doit être le plus diagonale possible. Pour accomplir cela nous réalisons une diagonalisation simultanée de ces matrices spectrales, par une matrice qui rend l'ensemble le plus possible diagonal. Cette matrice "diagonalisante" représente la matrice de séparation à la fréquence f . Nous discuterons dans la section 3.3.2 le critère et le principe de la diagonalisation conjointe adoptée.

3.3.1 Estimation des matrices spectrales

La première étape de calcul dans notre algorithme est l'estimation de la densité spectrale des signaux mélanges. Dans la littérature nous retrouvons plusieurs méthodes d'estimation spectrale (voir par exemple [47] et l'Annexe F) ; Comme nous l'avons expliqué dans la section précédente, nous considérons que les signaux sont à contenu spectral variable dans le temps, d'où le choix d'une méthode qui procède sur des blocs (chevauchés ou non) afin de décrire la distribution de l'énergie en fonction du temps et de la fréquence [75, pages 64-68], d'une part ; Et d'autre part, afin d'estimer la dsp de l'ensemble par l'association des dsp des blocs [75, page 106], selon la propriété de nonstationnarité des signaux. Cependant, le choix de la taille de la fenêtre⁶ de TFD glissante est un paramètre essentiel dans l'estimation spectrale [75, page 65].

En effet, nous calculons en premier le périodogramme des blocs selon la formule

⁵Belouchrani *et al.* ont discuté de cela dans leur article [8] et ont montré qu'un contexte de diagonalisation conjointe de plusieurs matrices est plus favorable.

⁶pour affaiblir l'effet de bord de la troncature, il existe un large choix de la forme de cette fenêtre (voir l'Annexe F).

suivante :

$$P_x(k, f) = \frac{1}{N} \left[\sum_{n=k+1}^{k+N} \mathbf{x}(n) e^{-j2\pi \frac{nf}{N}} \right] \left[\sum_{n=k+1}^{k+N} \mathbf{x}(n) e^{-j2\pi \frac{nf}{N}} \right]^* \quad (3.6)$$

où N qui est la taille du bloc à choisir est égale à une puissance de 2 et k est l'indice représentant le k -ième bloc.

Afin de réduire l'effet de la variance du spectre nous effectuons ensuite une opération de lissage par des moyennes locales de $2m + 1$ fréquences (consécutives). L'estimée du spectre, dit : "*estimateur spectral lissé*", est donnée par :

$$\hat{S}_x(k, f) = \frac{1}{2m + 1} \sum_{l=f-m}^{f+m} P_x(k, l \bmod N) \quad (3.7)$$

où \bmod est l'opération mathématique "modulo" exprimant la circularité de la TFD utilisée pour la transposition des signaux dans le domaine de Fourier.

Une autre variante d'estimateur est utilisée pour améliorer les propriétés statistiques du spectre, mais aussi pour réduire l'effet de bord en pré-multipliant les blocs par une fenêtre d'apodisation. La fenêtre utilisée est une forme d'Hanning définie par :

$$w(n) = \begin{cases} \sqrt{2/(3N)} [1 - \cos(2\pi n/N + \pi/N)] & : 0 \leq n \leq (N - 1) \\ 0 & : \text{ailleurs} \end{cases} \quad (3.8)$$

Ainsi, le périodogramme (d'un point τ) est calculé comme suit :

$$P_x(\tau, f) = \left[\sum_{n=\tau+1}^{\tau+N} w(n - \tau) \mathbf{x}(n) e^{-j2\pi \frac{nf}{N}} \right] \left[\sum_{n=\tau+1}^{\tau+N} w(n - \tau) \mathbf{x}(n) e^{-j2\pi \frac{nf}{N}} \right]^* \quad (3.9)$$

Ensuite, sous l'hypothèse de signaux non stationnaires, nous calculons la moyenne mobile des périodogrammes sur m points (blocs) : τ_1, \dots, τ_m . L'estimée du spectre au point temporel $(\tau_k + \tau_{mk} + N - 1)/2$ (appelée : "*estimateur spectral moyenné*") est donnée par :

$$\hat{S}_x \left(\frac{\tau_k + \tau_{mk} + N - 1}{2}, f \right) = \frac{1}{m} \sum_{k=1}^m P_x(\tau_k, f) \quad (3.10)$$

La discrétisation en fréquence a été faite avec un pas égal à $1/N$ ($f = k/N$, $k = 0, \dots, N - 1$), où N est une puissance de 2 pour bénéficier des avantages de l'algorithme de la TFR. La résolution fréquentielle est ainsi déterminée par la longueur de la fenêtre N et la résolution temporelle par $m\delta$ (où $\delta = \tau_i - \tau_{i-1}$). Prendre $\delta \gg 1$ permet de réduire le coût de calcul, mais dégrade significativement l'estimateur. En conséquence, nous avons été amené à choisir $m\delta^7$ en faisant un compromis entre la résolution fréquentielle et la résolution temporelle. En effet, cette méthode est plus flexible pour ajuster la résolution

⁷ δ peut être une fraction de N sans qu'il ait une dégradation significative.

et aide davantage à réduire le biais.

3.3.2 Diagonalisation conjointe adoptée

Le principe de notre méthode de séparation de mélanges convolutifs de sources, qui revient finalement à diagonaliser simultanément les matrices spectrales de même fréquence et de tous les blocs ($S_i : i = 1, 2, \dots, b$; b est le nombre de blocs), implique la recherche d'une matrice G de sorte que la multiplication de chacune des matrices spectrales de l'ensemble, à gauche par cette matrice et à droite par sa transposée conjuguée résulte une nouvelle matrice (GS_iG^*) qui soit le plus diagonale possible. Ce processus est appelé *diagonalisation conjointe* et la matrice qui en résulte est notée : G .

Historiquement, le besoin de cette procédure est apparu en Mécanique Quantique vers les années 40 ; Le premier algorithme de diagonalisation conjointe [54] repose sur un critère utilisant la technique de *Jacobi-Like* [105]. Dans cet algorithme initial, le principe (critère) de la diagonalisation conjointe⁸ a été défini par les auteurs comme suit :

Définition 3.2 Soit un ensemble $\mathcal{M} = \{M_1, M_2, \dots, M_K\}$ de K matrices de taille $n \times n$. L'opération de "diagonalisation conjointe" consiste à retrouver la matrice V qui permet de minimiser la fonction suivante :

$$\mathcal{C}(\mathcal{M}, V) \stackrel{\text{def}}{=} \sum_{k=1}^K \text{off}(V^* M_k V) \quad (3.11)$$

où l'"off" d'une matrice D , $n \times n$ est l'opération définie sur les éléments anti-diagonaux de cette matrice par $\text{off}(D) \stackrel{\text{def}}{=} \sum_{i \neq j} |d_{ij}|^2$.

Ensuite, d'autres algorithmes ont été développés pour la diagonalisation conjointe (voir [22]). En séparation de sources et en analyse en composantes principales, cette idée a conduit à des développements performants [32, 8, 89].

Quant à la diagonalisation conjointe, dans notre méthode de séparation, nous adoptons un critère lié à l'information mutuelle Gaussienne et basé sur le maximum de vraisemblance. L'expérimentation de celui-ci dans le cas instantané a montré qu'il est très performant ! Ce critère maximise la mesure de diagonalité de l'ensemble des matrices $G(f)\hat{S}_x(t_i, f)G^*(f)$, ($i = 1, \dots, K$), qui est définie comme suit :

$$\frac{1}{2} \left\{ \log \det \text{diag}[G(f)\hat{S}_x(t, f)G^*(f)] - \log \det[G(f)\hat{S}_x(t, f)G^*(f)] \right\} \quad (3.12)$$

où diag désigne l'opérateur qui transforme une matrice en une autre matrice diagonale de mêmes termes diagonaux.

⁸on note que cette définition n'est pas unique, d'ailleurs le critère que nous avons adopté diffère de celui-ci. Dans notre critère, $\text{off}(D)$ est donné par l'équation (3.12)

Or le dernier terme dans (3.12) étant égal à $2 \log |\det G(f)| + \log \det \hat{S}_x(t, f)$, où le terme $\log \det \hat{S}_x(t, f)$ est constant et donc peut être ignoré. Alors, le critère de diagonalisation de l'ensemble des matrices spectrales de la même fréquence f (et correspondante chacune à un bloc de stationnarité) devient :

$$\sum_{t=1}^b \left\{ \frac{1}{2} \log \det \text{diag}[G(f)\hat{S}_x(t, f)G^*(f)] - \log |\det G(f)| \right\} \quad (3.13)$$

La minimisation de ce critère est effectuée en faisant, itérativement, des transformations Jacobiennes sur une paire de lignes de la matrice G comme suit :

$$\begin{bmatrix} G_i \\ G_j \end{bmatrix} \leftarrow T_{ij} \begin{bmatrix} G_i \\ G_j \end{bmatrix} \quad (3.14)$$

où G_i , G_j sont respectivement, la i -ième et la j -ième ligne de la matrice G ; T_{ij} est une matrice 2×2 non singulière choisie de sorte que le critère décroisse progressivement de manière suffisante. Ensuite, le processus se réitère sur toutes les paires de lignes jusqu'à la convergence. La matrice T_{ij} qui convient à cette démarche est donnée par :

$$T_{ij} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \frac{2}{1 + g_{ij}g_{ji} - g_{ij}^*g_{ji}^* + \sqrt{(1 + g_{ij}g_{ji} - g_{ij}^*g_{ji}^*)^2 - 4g_{ij}g_{ji} - g_{ji}^*g_{ij}^*}} \begin{bmatrix} 0 & g_{ij} \\ g_{ji} & 0 \end{bmatrix} \quad (3.15)$$

où g_{ij} et g_{ji} sont les solutions de :

$$\begin{bmatrix} w_{ij} & 1 \\ 1 & w_{ji} \end{bmatrix} \begin{bmatrix} g_{ij} \\ g_{ji}^* \end{bmatrix} = \begin{bmatrix} h_{ij} \\ h_{ji}^* \end{bmatrix}, \quad (3.16)$$

h_{ij} et w_{ij} étant donnés par :

$$h_{ij} = \frac{1}{b} \sum_{l=1}^b \frac{(G\hat{S}_{x_l}G^*)_{ij}}{(G\hat{S}_{x_l}G^*)_{ii}}, \quad w_{ij} = \frac{1}{b} \sum_{l=1}^b \frac{(G\hat{S}_{x_l}G^*)_{jj}}{(G\hat{S}_{x_l}G^*)_{ii}} \quad (3.17)$$

Une implémentation rapide en code Matlab et en C est disponible à l'adresse [87]. Nous avons choisi cet algorithme (fonction Matlab) pour l'étape de diagonalisation conjointe de notre méthode de séparation.

3.3.3 Indétermination

Notre approche spectrale, qui présente l'avantage de ramener le problème de séparation de mélanges convolutifs à un ensemble de diagonalisations conjointes simples et résolubles indépendamment, fréquence par fréquence, est cependant entachée d'ambiguïté.

En effet, la diagonalisation conjointe est sensible à l'ambiguïté intrinsèque de la séparation de sources, expliquée dans la section 1.3.4. Cela provient du fait que lorsque l'on multiplie une matrice diagonale par une autre matrice diagonale (ou par une matrice de permutation), à gauche et sa transposée à droite, elle reste diagonale ; *i.e.* si $G(f)$ est la matrice qui diagonalise conjointement les matrices spectrales à la fréquence f , $D(f)$ et $P(f)$ sont respectivement une matrice diagonale et une matrice de permutation, alors :

$$\log \det \text{diag}[P(f)D(f)G(f)\hat{S}_x(t, f)G^*(f)D(f)P^T(f)] - \log \det[P(f)D(f)G(f)\hat{S}_x(t, f)G^*(f)D(f)P^T(f)] = \log \det \text{diag}[G(f)\hat{S}_x(t, f)G^*(f)] - \log \det[G(f)\hat{S}_x(t, f)G^*(f)] \quad (3.18)$$

En effet, en utilisant⁹ les propriétés $\det(PAP^T) = \det(A)$, $\det \text{diag}(PAP^T) = \det \text{diag}(A)$ et $\text{diag}(DAD) = D \text{diag}(A)D$ et $\det(AB) = \det(A)\det(B)$:

$$\begin{aligned} & \log \det \text{diag}[PDG\hat{S}_xG^*DP^T] - \log \det[PDG\hat{S}_xG^*DP^T] = \\ & \quad \log \det \text{diag}[DG\hat{S}_xG^*D] - \log \det[DG\hat{S}_xG^*D] \\ & = \log \det[D(\text{diag}[G\hat{S}_xG^*]D)] - \log \det[D] \det[G\hat{S}_xG^*D] \\ & = \log \det[D] + \log \det[\text{diag}[G\hat{S}_xG^*]D] - \log \det[D] - \log \det[G\hat{S}_xG^*D] \\ & = \log \det[\text{diag}[G\hat{S}_xG^*]D] - \log \det[G\hat{S}_xG^*D] \\ & = \log \det \text{diag}[G\hat{S}_xG^*] + \log \det[D] - \log \det[G\hat{S}_xG^*] - \log \det[D] \\ & = \log \det \text{diag}[G\hat{S}_xG^*] - \log \det[G\hat{S}_xG^*] \end{aligned}$$

On en déduit que si $G(f_i)$ est la solution fournie par notre algorithme à la fréquence f_i , alors $P(f_i)D(f_i)G(f_i)$ est une solution qui est aussi valable en cette fréquence. Finalement, cette indétermination de la solution n'est pas due au fait de travailler fréquence par fréquence, mais au fait que la maximisation de la diagonalité des matrices spectrales est intrinsèquement ambiguë : $\mathcal{C}(P(f)D(f)G(f)) = \mathcal{C}(G(f))$.

Par ailleurs, la séparation fréquence par fréquence est très importante, car l'application des procédures de séparation dans un environnement acoustique implique que les réponses impulsionnelles des filtres de mélange utilisés sont très longues, ce qui nécessite, pour une approche temporelle, la minimisation d'un critère dépendant d'un très grand nombre de paramètres, à l'opposé de l'approche fréquentielle qui revient à minimiser un critère plus simple.

L'ambiguïté de séparation de sources est un problème difficile, en particulier dans le cas d'un environnement acoustique [109, 80, 78, 79]. Nous trouvons dans la littérature des idées qui sont proposées afin de surmonter ce problème, par exemple, on peut contraindre le filtre de séparation à avoir un support FIR en temps, étant donnée que la permutation induit un filtre IIR. Cependant, cette idée ne convient pas aux environnements réels qui sont caractérisés par une longue réponse impulsionnelle, ce qui implique que le filtre inverse de séparation le sera également [110]. D'autres idées consistent par exemple à

⁹pour la clarté de la démonstration, nous omettons les variables (f) et (t, f).

introduire un couplage entre les solutions dans le domaine fréquentiel [96, 4], ou bien à utiliser l'hypothèse de continuité de la fonction de transfert de l'environnement de propagation [110, 96].

Dans notre approche [92, 93, 20], nous proposons en premier d'exploiter la continuité du filtre de séparation dans le domaine fréquentiel $G(f)$ en fonction de la fréquence f . Cette idée a l'avantage d'être uniquement relative à cette hypothèse et son coût de calcul est modeste. Cependant, nous avons constaté que cette approche a un faible effet dans les contextes réels d'environnements acoustiques, d'où une seconde proposition qui exploite directement la caractéristique intrinsèque de la variation d'énergie des sources dans le temps. Cette approche vient en complément de la première et permet en même temps de compenser le défaut des fausses détections.

Nous discuterons en détails ces approches d'élimination d'ambiguïté dans les sections suivantes.

3.3.4 Résolution d'ambiguïté basée sur la continuité

Étant donnée que la permutation tend à créer des sauts entre deux fréquences f_{l-1} et f_l , nous imposons que $G(f)$ soit continue en fonction de f [92], afin d'éliminer cette ambiguïté. Sachant qu'en pratique, le passage dans le domaine de Fourier se fait par TFD, *i.e.* $G(f)$ est calculée sur une grille finie de fréquences $f_0 < \dots < f_L$, il est possible de détecter si un saut a eu lieu dans l'intervalle $[f_{l-1}, f_l]$ en comparant $G(f_{l-1})$ avec $G(f_l)$. Pour effectuer cette comparaison, nous considérons le produit $G(f_{l-1})G^{-1}(f_l)$ que l'on note $R(f_{l-1}, f_l)$.

En supposant que le critère de diagonalisation conjointe fournit la solution $G(f)$ à une matrice de permutation près $P(f)$ et une matrice diagonale près $D(f)$, $R(f_{l-1}, f_l)$ s'écrit comme suit :

$$\begin{aligned}
 R(f_{l-1}, f_l) &= G(f_{l-1})G^{-1}(f_l) \\
 &= [P(f_{l-1})D(f_{l-1})\hat{H}^{-1}(f_{l-1})][P(f_l)D(f_l)\hat{H}^{-1}(f_l)]^{-1} \\
 &= P(f_{l-1})[D(f_{l-1})\hat{H}^{-1}(f_{l-1})\hat{H}(f_l)D^{-1}(f_l)]P^{-1}(f_l) \\
 &= P(f_{l-1})P^{-1}(f_l)\{P(f_l)[D(f_{l-1})\hat{H}^{-1}(f_{l-1})\hat{H}(f_l)D^{-1}(f_l)]P^{-1}(f_l)\}
 \end{aligned} \tag{3.19}$$

Comme la fonction $\hat{H}(f)$ est continue, le produit $\hat{H}(f_{l-1})\hat{H}^{-1}(f_l)$ est presque une matrice identité (f_l est supposée assez proche de f_{l-1}); donc le produit qui est entre crochets $[\cdot]$ est presque une matrice diagonale. D'autre part, Le produit entre accolades $\{\cdot\}$ l'est aussi, car multiplier, à gauche et à droite, une matrice par $P(f_l)$ et $P^{-1}(f_l)$ revient à permuter les lignes et les colonnes de cette matrice (suivant une même permutation). Ainsi :

$$R(f_{l-1}, f_l) = P(f_{l-1})P^{-1}(f_l)\Delta(f_{l-1}, f_l) \tag{3.20}$$

où $\Delta(f_{l-1}, f_l)$ est une matrice presque diagonale.

Cette approche est mieux qu'une comparaison directe entre $G(f_{l-1})$ et $G(f_l)$, car une matrice presque diagonale est facile à reconnaître. En effet, par cette approche, nous pouvons supprimer un saut survenu entre f_{l-1} et f_l en cherchant une matrice de permutation \mathcal{P} telle que la matrice $\mathcal{P}R(f_{l-1}, f_l)$ soit, le plus possible, diagonale, selon un critère que nous allons définir. Si la matrice trouvée, notée $\mathcal{P}(f_{l-1}, f_l)$, est la matrice identité, nous déduisons qu'il n'y a pas eu un saut de permutation entre f_{l-1} et f_l et donc on ne fait rien. Sinon, nous déduisons qu'il y a eu un saut entre f_{l-1} et f_l , et ainsi nous pouvons le supprimer en prémultipliant $G(f_{l-1})$ par $\mathcal{P}(f_{l-1}, f_l)$.

A ce stade, il reste à définir un critère pour mesurer la diagonalité de $\mathcal{P}R(f_{l-1}, f_l)$. Comme les matrices $G(f_l)$ sont intrinsèquement ambiguës au niveau de l'échelle, il est préférable d'utiliser un critère qui soit invariant par rapport au changement d'échelle, *i.e.* au changement issu d'une multiplication par une matrice diagonale D . Un critère qui possède cette propriété est :

$$d(A) = d(DA) = \min_{(i_1, \dots, i_K) \neq (1, \dots, K)} \frac{|a_{i_1 1}| \dots |a_{i_K K}|}{|a_{11}| \dots |a_{KK}|} \quad (3.21)$$

où a_{ij} désigne le terme général de la matrice A (d'ordre K) et le minimum porte sur toutes les permutations (i_1, \dots, i_K) de $1, \dots, K$, sauf la permutation identité.

En adoptant le précédent critère, notre procédure d'élimination d'ambiguïté, basée sur la continuité, se résume comme suit :

- (i) calculer la matrice $R(f_{l-1}, f_l) = G(f_{l-1})G^{-1}(f_l)$, $l = 1, \dots, L$.
- (ii) pour chaque $l = 2, \dots, L$, chercher la permutation (i_1, \dots, i_K) de telle sorte que le produit $\prod_{j=1}^K |R(f_{l-1}, f_l)_{i_j j}|$ soit le plus grand parmi toute autre permutation. $(R(\cdot))_{i_j j}$ représente l'élément d'indice (i_j, j) de la matrice $R(\cdot)$.
- (iii) si (i_1, \dots, i_K) est la permutation identité, ne rien faire. Sinon permuter les lignes $1, \dots, K$, de $G(f_1), \dots, G(f_L)$, par les lignes i_1, \dots, i_K .

Cependant, d'une part, l'étape (ii) de la procédure peut être très coûteuse en temps de calcul pour K grand, car il faut chercher toutes les permutations et d'autre part, nous rappelons que $R(f_{l-1}, f_l)$ doit être proche d'une matrice diagonale permutée. Alors, en définissant les indices i_1, \dots, i_K par :

$$i_j = \arg \max_{i=1, \dots, K} |R(f_{l-1}, f_l)_{ij}| \quad (3.22)$$

il est fortement possible que les i_1, \dots, i_K soient différents et forment une permutation de $1, \dots, K$ qui définit précisément la permutation recherchée. Sinon, on doit en principe réexaminer toutes les permutations, mais ce cas est exceptionnel car il signifie que la

matrice $R(f_{l-1}, f_l)$ n'est pas proche d'une matrice diagonale permutée ; La permutation dans ce cas est de toute façon mal déterminée.

Afin de réduire le coût de calcul, nous avons développé une routine sous-optimale comme suit :

- (1) on cherche la permutation i_1 qui maximise $|R(f_{l-1}, f_l)_{i_1}|$.
- (2) pour $j = 2, \dots, K$, on cherche d'abord

$$i_j^* = \arg \max_{i=1, \dots, K} |R(f_{l-1}, f_l)_{ij}|$$

si $i_j^* \notin \{i_1, \dots, i_{j-1}\}$ on prend $i_j = i_j^*$; sinon c'est que $i_j^* = i_k$ pour un certain $k < j$. On cherche alors

$$i'_j = \arg \max_{i_j^* \notin \{i_1, \dots, i_{j-1}\}} |R(f_{l-1}, f_l)_{ij}|, \quad i'_k = \arg \max_{i_j^* \notin \{i_1, \dots, i_{j-1}\}} |R(f_{l-1}, f_l)_{ik}|.$$

Puis on compare

$$|R(f_{l-1}, f_l)_{i_k k} R(f_{l-1}, f_l)_{i'_j j}| \text{ et } |R(f_{l-1}, f_l)_{i'_k k} R(f_{l-1}, f_l)_{i_k j}|$$

si le premier produit est plus grand, on prend $i_j = i'_j$, sinon on remplace i_k par i'_k et on prend $i_j = i_k$.

Via cette procédure nous avons pu réduire l'ambiguïté de permutation variable (selon les fréquences) à une permutation fixe car la fonction de permutation ne peut être continue que si elle est constante. Toutefois, cette procédure ne permet pas d'enlever l'ambiguïté d'échelle car on peut toujours prémultiplier $G(f)$ par une matrice diagonale $D(f)$ continue en f et obtenir une fonction matricielle continue.

Une autre procédure plus simple et presque aussi performante, relativement au fonctionnement de notre algorithme de diagonalisation conjointe, est celle qui au lieu de diagonaliser conjointement les matrices $\hat{S}_x(t, f_l)$, diagonalise conjointement les matrices $G(f_{l-1})\hat{S}_x(t, f_l)G^*(f_{l-1})$; où $G(f_{l-1})$ est la solution du précédent problème de diagonalisation conjointe de $\hat{S}_x(f_{l-1})$.

En effet, si $G(f)$ est continue ($G(f_{l-1})$ proche de $G(f_l)$), $G(f_{l-1})\hat{S}_x(t, f_l)G(f_{l-1})$ doivent rester proches d'une matrice diagonale et la solution à leur problème de diagonalisation conjointe est, ainsi, proche de la matrice identité. Or, la procédure de diagonalisation conjointe opère par transformations successives des matrices à diagonaliser en les pré- et post- multipliant par une matrice appropriée. Celle-ci est choisie chaque fois, parmi deux matrices qui se diffèrent uniquement par une permutation, comme celle qui est la plus proche de l'identité. Donc, le rapport entre les solutions aux problèmes de la diagonalisation conjointe en deux fréquences successives est une matrice proche de l'identité.

De plus, en initialisant l'algorithme de diagonalisation conjointe à chaque fréquence par la solution de la précédente fréquence (*i.e.* à la fréquence f_l , on diagonalise $G(f_{l-1})\hat{S}_x(t, f_l)G(f_{l-1})$), la procédure fournit plus rapidement la solution car l'algorithme est initialisé à une valeur proche de la solution. Le rapport $G(f_{l-1})G^{-1}(f_l)$ sera proche de l'identité, sauf si $G(f)$ varie rapidement en fonction de f , auquel cas la permutation est de toutes façons mal détectée.

Cette procédure peut être implémentée comme une sous-routine complémentaire à l'algorithme de diagonalisation conjointe, en conjonction à une autre procédure d'élimination d'ambiguïté telle que la première procédure présentée dans ce paragraphe.

3.3.5 Résolution d'ambiguïté en utilisant les profils d'énergie

Avec la précédente procédure d'élimination d'ambiguïté, qui est basée sur la continuité de la fonction de transfert (de mélange) acoustique $H(f)$ (de la même façon on impose que la réponse fréquentielle de séparation $G(f)$ soit continue), nous avons pu éliminer la majorité des permutations fréquentielles. Cependant, le désavantage de cette méthode, en plus de sa puissance limitée à détecter les permutations dans le contexte de salles réverbérantes, est dû au fait qu'une fausse détection à une fréquence donnée peut causer une fausse permutation de tout un bloc de fréquences.

Afin de surmonter ce défaut et d'éliminer les permutations qui peuvent subsister après le processus de la première méthode, nous proposons une nouvelle méthode [93] qui procède à la sortie de l'algorithme de diagonalisation conjointe couplé avec la contrainte de continuité du filtre de séparation (*i.e.* initialisé à chaque fréquence par la solution de la fréquence précédente), élaborant donc une deuxième étape qui est un complément et un renforcement à l'élimination d'ambiguïté de notre algorithme de séparation de sources.

Notre idée part de la caractéristique intrinsèque de nonstationnarité des sources et de la variation de leur énergie dans le temps. En effet, la transposition des signaux dans le domaine de Fourier montre qu'ils ont une forme caractérisée par des zones de faible énergie entre lesquelles se trouvent des zones d'énergie importantes. Ainsi, le principe de notre méthode consiste à admettre que l'énergie dans le domaine des fréquences varie de manière proportionnelle à sa variation dans le temps ; *e.g.* les zones de faible énergie correspondent aux instants de silences de la parole. Donc, pour une source s_i , l'évolution de son spectre à court terme $S_{s_i}(t, f)$ comme une fonction de t sera supposée similaire au long de toutes les fréquences. En conséquence, les zones de silence à l'instant t fourniront les $S_s(t, f)$ minima et, de la même façon, un bloc temporel t de parole produira (sur tout l'horizon des fréquences) un $S_s(t, f)$ maxima. Par ailleurs, les sources s_i sont supposées à spectres $S_{s_i}(t, f)$ assez différents.

Pour mettre en œuvre cette procédure, nous construisons, à partir des sources estimées après diagonalisation conjointe où certaines permutations peuvent subsister, des "profils" qui représentent la distribution globale d'énergie.

La distribution d'énergie (ou profil d'énergie) d'une source s_i , notée $\epsilon_{s_i}(t)$, est une fonction qui ne dépend que des instants t (*i.e.* du bloc temporel d'indice t) et à chaque instant elle est égale à une quantité énergétique $\bar{\epsilon}_{s_i}(t)$. En accord avec notre hypothèse de variation d'énergie, $\bar{\epsilon}_{s_i}(t)$ qui représente l'énergie globale du bloc t de la source i , est constante (*i.e.* la variation énergétique supposée similaire au long de toutes les fréquences) et estimée comme étant la différence entre la moyenne des énergies fréquentielles logarithmiques du même bloc et la moyenne globale de l'énergie logarithmique fréquentielle de la source.

En pratique, on passe par les étapes suivantes :

- (i) calculer le logarithme des énergies fréquentielles (spectre) des sources (estimées après l'étape de diagonalisation), $\epsilon_{s_i}(t, f)$:

$$\epsilon_{s_i}(t, f) = \log \left(\hat{S}_{s_i}(t, f) \right)$$

- (ii) calculer l'énergie logarithmique moyenne à chaque fréquence f_l :

$$\bar{\epsilon}_{s_i}(t, f_l) = \frac{1}{b} \sum_{t=1}^b \epsilon_{s_i}(t, f_l) \quad : \quad b = \text{nombre de blocs}$$

- (iii) centrer par rapport à $\bar{\epsilon}_{s_i}(t, f_l)$:

$$\epsilon'_{s_i}(t, f_l) = \epsilon_{s_i}(t, f_l) - \bar{\epsilon}_{s_i}(t, f_l)$$

- (iv) calculer l'énergie (logarithmique fréquentielle) de chaque bloc t :

$$\bar{\epsilon}_{s_i}(t) = \frac{1}{L} \sum_{f=0}^{L-1} \epsilon'_{s_i}(t, f)$$

Les étapes (i) et (iii) reviennent à représenter la distribution d'énergie des sources par une variante centrée, tout en éliminant l'ambiguïté d'échelle de la solution. En effet, si l'algorithme de diagonalisation conjointe produit un facteur d'échelle $d(f)$ à la fréquence f , les quantités énergétiques correspondantes vont changer avec les mêmes facteurs, mais le centrage de leur logarithme éliminera ce facteur comme s'il s'agissait de centrer l'énergie non ambiguë. Soit par exemple le centrage de l'énergie correspondante au bloc

$t = \alpha$ et à la fréquence f :

$$\begin{aligned}
& \log(d(f)\epsilon_{s_i}(\alpha, f)) - \frac{1}{b} \sum_{t=1}^b \log(d(f)\epsilon_{s_i}(t, f)) = \\
& \quad \log(d(f)) + \log(\epsilon_{s_i}(\alpha, f)) - \frac{1}{b} \sum_{t=1}^b [\log(d(f)) + \log(\epsilon_{s_i}(t, f))] \\
& = \log(d(f)) + \log(\epsilon_{s_i}(\alpha, f)) - \frac{1}{b} \sum_{t=1}^b \log(d(f)) - \frac{1}{b} \sum_{t=1}^b \log(\epsilon_{s_i}(t, f)) \quad (3.23) \\
& = \log(d(f)) + \log(\epsilon_{s_i}(\alpha, f)) - \log(d(f)) - \frac{1}{b} \sum_{t=1}^b \log(\epsilon_{s_i}(t, f)) \\
& = \log(\epsilon_{s_i}(\alpha, f)) - \frac{1}{b} \sum_{t=1}^b \log(\epsilon_{s_i}(t, f))
\end{aligned}$$

Ceci démontre l'intérêt mathématique¹⁰ de travailler avec le log de l'énergie, qui revient à circonvier l'ambiguïté d'échelle de la séparation.

Après avoir caractérisé chaque source par son profil d'énergie, la procédure détecte et compense les permutations éventuelles. À chaque fréquence, les profils sont utilisés pour comparer l'énergie de chaque source séparée au profil de référence. La procédure est itérative car elle recalcule à chaque pas d'itération, un nouveau profil de référence en tenant compte de l'affectation des sources et des corrections de permutations effectuées à la fin de l'itération précédente. Le profil initial est estimé comme étant la moyenne standard de l'énergie sur toutes les fréquences.

Quant au critère d'affectation, il est basé sur le calcul de la distance quadratique fréquentielle globale entre les sources et leurs profils de comparaison. Cette distance est calculée à toutes les fréquences et l'affectation est examinée fréquence par fréquence en faisant une comparaison entre les distances correspondantes à toutes les permutations possibles des sources (à la fréquence f). La décision d'affectation au profil de comparaison est retenue quand la distance de cette affectation est inférieure à tous les autres cas (issus des autres permutations). Ce critère est défini comme suit :

$$\mathfrak{d}(f) = \min_{\pi \in (\pi_1, \dots, \pi_b)} \sum_{i=1}^r \sum_{t=1}^b \left(\pi^*[\epsilon_{s_i}(t, f)] - \epsilon_{s_i}(t) \right)^2 \quad (3.24)$$

où r est le nombre des sources, b est le nombre des blocs et $\pi^*[\epsilon_{s_i}(t, f)]$ est la permutation (du log de l'énergie de s_i) qui permet que la source soit la plus proche possible de son profil.

¹⁰l'autre intérêt de considérer le log de l'énergie est l'analogie psychologique avec la réception des sons par l'oreille, qui est en log de la puissance des sons. Ceci est bel et bien pour se protéger des fortes puissances !

3.3.5.1 Estimation des profils par moyenne glissante

L'application de la méthode des profils à l'élimination des permutations, montre qu'elle est beaucoup plus performante par rapport à la méthode de continuité. Cependant, les contextes difficiles des salles acoustiques réverbérantes nécessitent une très grande puissance à détecter les permutations aberrantes. C'est pourquoi, nous proposons une amélioration de l'estimation des profils de référence [20].

En effet, l'approximation de la dynamique énergétique d'une source sur un bloc temporel par une constante à toutes les fréquences est grossière dans certains cas, car l'occurrence de plusieurs permutations dans le même bloc peut biaiser l'estimation (par moyenne d'ensemble) du profil et le rend non représentatif à la source. En conséquence, nous allons accepter une faible variation du profil dans les nouvelles méthodes d'estimation :

Une première proposition consiste à améliorer cette estimation du profil par une moyenne glissante (locale) au lieu de la moyenne globale de toutes les fréquences. Cela permettrait d'affiner la caractérisation de la source par son profil en ayant cette fois-ci un profil temps-fréquence plus représentatif, aidant à déceler les permutations difficiles à détecter.

Le profil est calculé, à partir des sources issues de l'étape de diagonalisation conjointe, avec les mêmes étapes décrites dans la section précédente, où l'étape (iv) est remplacée par le calcul de la nouvelle estimation du profil à chaque fréquence f , qui est égale à la moyenne des $(\tau + 1)$ valeurs spectrales centrées sur f . *i.e.*

(iv) calculer l'énergie (logarithmique fréquentielle) de chaque (t, f_l) :

$$\bar{\epsilon}_{s_i}(t, f_l) = \frac{1}{\tau + 1} \sum_{f=f_l-\frac{\tau}{2}}^{f=f_l+\frac{\tau}{2}} \epsilon'_{s_i}(t, f)$$

L'estimation du profil est ajustable en fonction de la taille τ , de la fenêtre spectrale glissante. Par ailleurs, le pas de glissement est toujours fixe et est égale à un.

3.3.5.2 Estimation des profils par la SFD

Avec la nouvelle estimation du profil par moyenne glissante, l'élimination des permutations est encore meilleur, surtout dans les environnements acoustiques difficiles. Cependant, nous avons constaté que la performance de la méthode dépend sensiblement du réglage de la taille de la fenêtre de moyennage. De plus, le réglage optimal est difficile à obtenir, car il faut faire un bon compromis entre le lissage et la variance pour bien caractériser une source par son profil, sans l'influence des permutations qui modifient le spectre aux fréquences d'occurrence.

Dans cette optique nous proposons une deuxième méthode pour l'estimation des profils d'énergie par lissage et où le paramétrage de cette nouvelle méthode est plus facile à régler. Il s'agit de l'utilisation de la Série de Fourier Discrète comme technique d'un lissage contrôlé fréquentiellement.

En effet, étant donné la nature discrète des signaux qu'on traite en pratique, le passage au domaine de Fourier s'effectue par FFT et le spectre résultant est aussi discret. D'autre part, vu la nature réelle des signaux et les propriétés de circularité et de symétrie de la TFD, une Série de Fourier Discrète (ou échantillonnée) est applicable sur le spectre de nos signaux. Cette SFD va permettre de trouver les signaux composites à fréquences pures du signal discret. *i.e.* si $x(n)$ est un signal discret périodique de période T_0 et dont la période d'échantillonnage est $T_e = T_0/N$, alors sa SFD est définie par l'équation suivante :

$$x(nT_e) = x(n) = \sum_{k=0}^{N-1} C_k e^{j2\pi k f_0(nT_e)}, \quad (3.25)$$

où les C_k sont les N coefficients de la série de Fourier discrète qui sont donnés par :

$$C_k = \frac{1}{N} \sum_{n=0}^{N-1} x(n) e^{-j\frac{2\pi}{N} kn} \quad (3.26)$$

Par ailleurs, il est bien connu que les hautes fréquences d'un signal dans le domaine de Fourier correspondent aux fortes dynamiques dans le domaine temporel ; Ainsi, il est possible de réaliser un lissage en utilisant l'outil de Fourier qui contrôle la dynamique du signal par l'élimination des sinusoïdes à hautes fréquences avant de faire une SFD inverse. De cette manière, nous aboutissons à un lissage naturel, homogène et plus facile à contrôler en fonction des hautes fréquences filtrées.

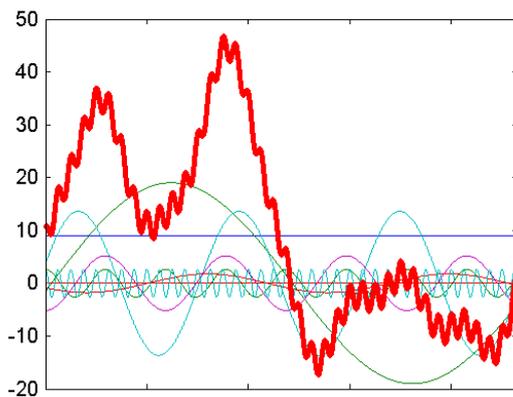


FIG. 3.2 – SFD d'un signal discret

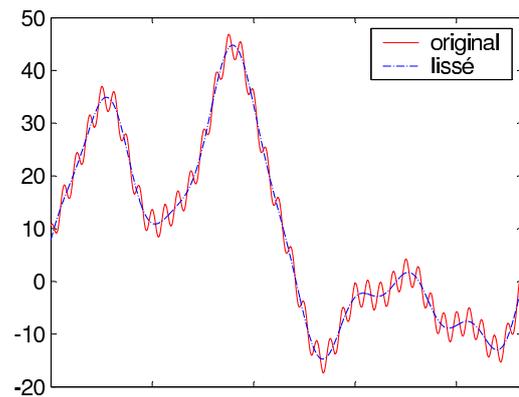


FIG. 3.3 – lissage par SFD

L'intégration de cette technique dans le calcul du profil consiste à changer l'étape (iv) de l'algorithme d'élimination de permutation comme suit :

(iv) lissage par SFD :

a) calculer l'SFD sur chaque bloc :

$$\begin{aligned}\varepsilon(t, f) &= \mathcal{SFD} \left[\epsilon'_{s_i}(t, f) \right] \\ &= \sum_{k=0}^{L-1} C_k e^{j \frac{2\pi}{L} k f} \\ &= \sum_{k=0}^{L-1} \varepsilon_k(t, f)\end{aligned}$$

où C_k est donné par l'équation (3.26)

b) contrôler le lissage par le pourcentage \aleph du filtrage HF :

$$L \longleftarrow \left(1 - \frac{\aleph}{100} \right) L$$

c) calculer l'SFD inverse :

$$\tilde{\varepsilon}_{s_i}(t, f) = \mathcal{SFD}^{-1} \left[\varepsilon(t, f) \right] = \sum_{k=0}^{L-1} \varepsilon_k(t, f)$$

Nous rappelons que le calcul des C_k est simplifié en tenant compte de la propriété de symétrie du spectre. Il est aussi possible d'utiliser la forme trigonométrique de la SFD pour une décomposition plus simple en cosinusoïdes (voir l'Annexe A).

3.4 Méthode de séparation dans le cas bruité

Toujours dans l'objectif de développer des nouvelles méthodes de séparation de mélanges convolutifs de sources, pouvant servir dans beaucoup de situations pratiques comme l'aide aux malentendants par exemple, notre challenge continue mais cette fois avec une situation plus complexe. Il s'agit de traiter le cas bruité de la séparation de mélanges convolutifs de sources. Dans la littérature peu d'articles traitent ce cas, en particulier pour des SNR élevés¹¹ ; *e.g.* [29].

Cette section est consacrée à la description de notre méthode développée pour traiter ce contexte difficile de séparation de sources, modélisé par l'équation suivante :

$$\mathbf{x}(n) = (H * \mathbf{s})(n) + \mathbf{b}(n) \quad (3.27)$$

où les vecteurs $\mathbf{x}(n)$, $\mathbf{s}(n)$, $\mathbf{b}(n)$ représentent respectivement : les mélanges, les sources et les signaux des bruits additifs aux capteurs. Notre problème revient à estimer le vecteur

¹¹valeurs en db ($-10 \log \left[\frac{S_s(t)}{S_n(t)} \right]$)

des sources¹², seulement en connaissant les mélanges et à une hypothèse près.

Pour atteindre notre objectif nous avons utilisé la technique dite "EM" (pour les initiales de "Expectation-Maximization") [39], [7, pages 44-46]. Cette approche est basée sur la maximisation itérative de la log-vraisemblance¹³. En séparation de sources, cette approche a été initialement employée par Belouchrani [7, pages 44-46] dans son algorithme MLS, mais ce développement ne s'appliquait qu'à la séparation des mélanges instantanés [11, 12]. Récemment, cette méthode a été reprise dans le travail de Cardoso *et al.* [38] et le travail de Cardoso et Pham [29] où elle a été réétudiée et optimisée pour l'ICA.

Dans le cadre de cette thèse [19], nous avons étudié l'extension de l'algorithme EM pour le cas convolutif. Nous discuterons dans la section suivante le principe de cette algorithme pour la séparation de mélanges convolutifs de sources et nous présenterons les résultats de simulations relatives au cas bruité avec des SNR élevés dans le chapitre 4.

3.4.1 Principe de l'approche

En partant de notre modèle de mélange (3.27), une transposition dans le domaine de Fourier transformera le produit de convolution en un produit simple. Ainsi, nous nous trouvons avec un modèle plus simple qui s'écrit sous la forme :

$$\mathbf{x}(f) = H(f)\mathbf{s}(f) + \mathbf{b}(f) \quad (3.28)$$

où $\mathbf{x}(f)$ ¹⁴, $\mathbf{s}(f)$ et $\mathbf{b}(f)$ sont respectivement les transformées de Fourier de $\mathbf{x}(n)$, $\mathbf{s}(n)$ et $\mathbf{b}(n)$.

À partir de l'équation (3.28) nous pouvons calculer la matrice spectrale (ou par équivalence : la matrice de covariance) du vecteur mélanges, qui se présente comme suit :

$$S_x(f) = H(f)S_s(f)H^*(f) + S_b(f) \quad (3.29)$$

où $H^*(f)$ ¹⁵ est la matrice transposée conjuguée de $H(f)$.

En pratique, nous calculons cette matrice (selon la nature non stationnaire des signaux et selon l'hypothèse d'indépendance) sur un bloc temporel (section de signal, dont la durée est finie) : *analyse court terme*. À chaque bloc i (et chaque fréquence f) nous disposons ainsi d'une matrice spectrale $R_i = S_x(n_i, f)$ où :

$$S_x(n_i, f) = H(f)S_s(n_i, f)H^*(f) + S_b(n_i, f) \quad (3.30)$$

Ayant cet ensemble de données, le problème revient à estimer les sources ($\mathbf{s}(f)$) à

¹²on note le vecteur des estimées des sources par $\mathbf{y}(n)$, ($\mathbf{y}(n) = \hat{\mathbf{s}}(n)$).

¹³la méthode du maximum de vraisemblance est une méthode d'estimation (voir Annexe C).

¹⁴ f est la variable discrète de Fourier.

¹⁵cette matrice est aussi appelée la transposée hermitienne de $H(f)$.

partir de la formulation de leur modèle mathématique de mélange. Pour cela, nous devons résoudre le problème statistique dont le modèle à considérer (en une fréquence) est : $\mathbf{x}_i = A\mathbf{s}_i + \mathbf{e}_i$, où \mathbf{s}_i , \mathbf{e}_i sont indépendants, gaussiens (de dimension m et K , avec $m \leq K$), de moyennes nulles et respectivement de matrices de covariance D_i et N_i . La matrice A est de taille $K \times m$. De plus si, pour chaque i ($i = 1, \dots, r$), nous disposons d'un échantillon x_i (de taille proportionnelle à w_i), qui nous permet d'estimer sa matrice de covariance R_i ; Alors, le problème revient à estimer A, D_1, \dots, D_r et N_1, \dots, N_r afin d'aboutir à l'estimée de \mathbf{s}_i .

Par ailleurs, nous savons¹⁶ que la fonction log vraisemblance (normalisée) d'un ensemble de matrices de covariance $\Gamma_1, \dots, \Gamma_r$ ($\Gamma_i = AD_iA^* + N_i$ est la matrice de covariance issue du modèle : $\mathbf{x}_i = A\mathbf{s}_i + \mathbf{e}_i$) où leurs matrices accessibles (estimées ou réalisations) correspondantes sont R_1, \dots, R_r , est :

$$C - \frac{1}{2} \sum_{i=1}^r w_i \{ \log \det \Gamma_i + \text{tr}(\Gamma_i^{-1} R_i) \} \quad (3.31)$$

où C est une constante, w_i est un coefficient relatif à la taille¹⁷ des données (nombre d'échantillons) correspondantes à $\Gamma_1, \dots, \Gamma_r$ et tr désigne la trace d'une matrice.

Ainsi, la méthode du maximum de vraisemblance pour estimer l'ensemble A, D_1, \dots, D_r et N_1, \dots, N_r à chaque fréquence f , conduit à minimiser¹⁸ :

$$\sum_{i=1}^r w_i \{ \log \det(AD_iA^* + N_i) + \text{tr}[(AD_iA^* + N_i)^{-1} R_i] \} \quad (3.32)$$

Ce critère peut être minimisé par plusieurs méthodes qui sont coûteuses et lourdes ! C'est pourquoi, nous avons opté pour le choix de l'algorithme EM qui est plus simple et plus rapide.

3.4.2 L'algorithme EM

Expectation-Maximization est une méthode itérative d'estimation : elle fournit un schéma itératif qui permet d'obtenir plus simplement le maximum de vraisemblance en remplaçant le problème difficile d'estimation par une séquence de problèmes simples. Par l'extension de l'ensemble des données observées ("*données incomplètes*") avec certaines variables additionnelles ("*données manquantes ou complémentaires*"), l'estimation est accomplie simplement en deux étapes itératives¹⁹. En effet, traiter les données observées (incomplètes) \mathbf{x} comme une fonction d'un ensemble de données non observées (manquantes) \mathbf{s} , va permettre de construire un large ensemble de données (complètes) et de

¹⁶voir l'Annexe C pour la démonstration de ce résultat.

¹⁷lorsque tous les blocs ont la même taille, $w_i = 1, \quad \forall i$

¹⁸à noter que nous travaillons avec l'opposé de la fonction log-vraisemblance, c'est pourquoi la maximisation est transformée en minimisation.

¹⁹l'arrêt de l'itération est conditionné par la convergence de l'algorithme.

traiter le maximum de vraisemblance en fonction de la densité de probabilité $f(s, \theta)$ (θ est un paramètre), *e.g.* $\mathbf{x} = A\mathbf{s} + \mathbf{e}$ (voir la section C.2 de l'Annexe C). Les deux étapes qui donnent à l'algorithme son nom sont :

- a) Estimer les statistiques complémentaires permettant de calculer la valeur actuelle des paramètres ;
- b) Maximiser la fonction de vraisemblance associée aux statistiques estimées.

L'organigramme de l'algorithme EM se présente comme suit :

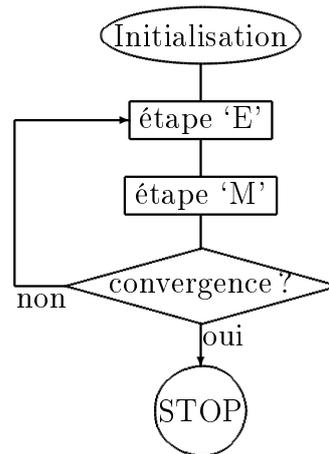


FIG. 3.4 – organigramme de l'algorithme

La théorie montre que l'algorithme EM a la propriété d'avoir une bonne monotonie qui assure la croissance itérative de la vraisemblance jusqu'à un point critique. En effet, à partir d'une valeur initiale, la maximisation augmente d'une itération à l'autre jusqu'à la convergence vers l'estimateur de la vraisemblance [7, pages 44-46]. Parmi les premiers articles sur l'algorithme EM, on trouve dans la littérature le papier de Hartley [58]. Par ailleurs, la formulation de l'algorithme ainsi que la démonstration de sa convergence sont apparues dans l'article de Dempster *et al.* [40].

Dans les sous sections suivantes nous décrivons les étapes de la procédure EM utilisée pour achever la séparation de mélanges convolutifs dans un environnement acoustique bruité. Le développement de cette procédure est passé par deux versions où de grandes améliorations ont été apportées.

3.4.2.1 Version première de l'algorithme EM

Dans l'objectif d'estimer les paramètres A , D_1, \dots, D_r et N_1, \dots, N_r ²⁰ en utilisant la méthode Expectation-Maximization et ainsi de solutionner le problème de séparation de mélanges convolutifs bruités, nous partons de l'ensemble (incomplet) de nos données observées \mathbf{x}_i pour élargir l'ensemble par d'autres données (complémentaires) nécessaires

²⁰pour simplifier la notation nous allons omettre dans ce qui suit la variable de Fourier f ; le temps (blocs) sera indiqué par l'indice i , et r représente le nombre des blocs.

au procédé de notre algorithme. À cette fin, nous considérons le modèle (complet²¹) où les vecteurs \mathbf{s}_i sont supposés observables. En conséquence, nous aurons accès à l'ensemble des matrices de covariance R_i^s de \mathbf{s}_i et d'intercovariance R_i^{xs} entre \mathbf{x}_i et \mathbf{s}_i . La fonction log-vraisemblance du modèle complet est ainsi proportionnelle à :

$$- \sum_{i=1}^r \{ \log \det D_i + \text{tr}(D_i^{-1} R_i^s) + \log \det N_i + \text{tr}[N_i^{-1}(R_i - 2R_i^{xs} A^* + AR_i^s A^*)] \} \quad (3.33)$$

Dans l'étape 'E', ayant formulé notre modèle complet, l'algorithme estime (conditionnellement à \mathbf{x}_i) les matrices R_i^s et R_i^{sx} (notées respectivement \hat{R}_i^s et \hat{R}_i^{sx}). Ceci est réalisé par la régression de \mathbf{s}_i par rapport à \mathbf{x}_i ce qui permet d'écrire \mathbf{s}_i sous la forme $D_i A^* (AD_i A^* + N_i)^{-1} \mathbf{x}_i$ plus un résidu²² décorrélé de \mathbf{x}_i . Ainsi :

$$\begin{aligned} \hat{R}_i^{xs} &= R_i (AD_i A^* + N_i)^{-1} AD_i \\ &= R_i N_i^{-1} A (D_i^{-1} + A^* N_i^{-1} A)^{-1} \end{aligned} \quad (3.34)$$

$$\begin{aligned} \hat{R}_i^s &= D_i A^* (AD_i A^* + N_i)^{-1} R_i (AD_i A^* + N_i)^{-1} AD_i + \\ &\quad D_i - D_i A^* (AD_i A^* + N_i)^{-1} AD_i \\ &= (D_i^{-1} + A^* N_i^{-1} A)^{-1} A^* N_i^{-1} R_i N_i^{-1} A (D_i^{-1} + A^* N_i A)^{-1} + \\ &\quad (D_i^{-1} + A^* N_i^{-1} A)^{-1} \end{aligned} \quad (3.35)$$

L'étape 'M' consiste à minimiser le log-vraisemblance (avec R_i^{sx} et R_i^s remplacées par \hat{R}_i^{sx} et \hat{R}_i^s)²³. Il en résulte que les (nouvelles) quantités

$$D_i^{\text{new}} = \text{diag} \hat{R}_i^s \quad (3.36)$$

et A^{new} et N_i^{new} minimisent

$$\sum_{i=1}^r w_i \{ \log \det N_i + \text{tr}[N_i^{-1}(R_i - 2\hat{R}_i^{xs} A^* + A\hat{R}_i^s A^*)] \} \quad (3.37)$$

La minimisation de (3.37) n'est, par contre, pas facile. Mais dans le cas où $N_i \equiv N$, elle se simplifie. L'expression (3.37) se réduit ainsi à

$$\left(\sum_{i=1}^r \right) \log \det N + \text{tr} \left\{ N^{-1} \left[\left(\sum_{i=1}^r w_i R_i \right) - 2 \left(\sum_{i=1}^r w_i \hat{R}_i^{xs} \right) A^* + A \left(\sum_{i=1}^r w_i \hat{R}_i^s \right) A^* \right] \right\} \quad (3.38)$$

²¹on dit modèle complet car il permet d'associer (compléter) les données manquantes aux données observées (incomplètes).

²²sa covariance est donnée par $D_i - D_i A^* (AD_i A^* + N_i)^{-1} AD_i$.

²³ceci engendre une constante résiduelle dans la log-vraisemblance ; La minimisation est reportée sur d'autres termes.

dans laquelle le dernier terme peut être mis sous la forme

$$\text{tr} \left\{ N^{-1} \left[A - \left(\sum_{i=1}^r w_i \hat{R}_i^{xs} \right) \left(\sum_{i=1}^r w_i \hat{R}_i^s \right)^{-1} \right] \left(\sum_{i=1}^r w_i \hat{R}_i^s \right) \times \right. \\ \left. \left[A^* - \left(\sum_{i=1}^r w_i \hat{R}_i^s \right)^{-1} \left(\sum_{i=1}^r w_i \hat{R}_i^{sx} \right) \right] + \right. \\ \left. N^{-1} \left[\sum_{i=1}^r w_i R_i - \left(\sum_{i=1}^r w_i \hat{R}_i^{xs} \right) \left(\sum_{i=1}^r w_i \hat{R}_i^s \right)^{-1} \left(\sum_{i=1}^r w_i \hat{R}_i^{sx} \right) \right] \right\} \quad (3.39)$$

R_i^{sx} étant la transposée conjuguée de R_i^{xs} , alors

$$A^{\text{new}} = \left(\sum_{i=1}^r w_i \hat{R}_i^{xs} \right) \left(\sum_{i=1}^r w_i \hat{R}_i^s \right)^{-1} \quad (3.40)$$

$$N^{\text{new}} = \text{diag} \left[\sum_{i=1}^r w_i R_i - \left(\sum_{i=1}^r w_i \hat{R}_i^{xs} \right) \left(\sum_{i=1}^r w_i \hat{R}_i^s \right)^{-1} \left(\sum_{i=1}^r w_i \hat{R}_i^{sx} \right) \right] / \sum_{i=1}^r w_i \quad (3.41)$$

où $\text{diag}(\cdot)$ dénote l'opérateur qui construit une matrice diagonale à partir des éléments de sa matrice en mettant les éléments hors diagonale à zéro.

Dans le cas où les matrices N_i sont distinctes, on peut minimiser (3.37) par rapport à N_i , A gardée fixe ; Ce qui donne

$$N_i = \text{diag}(R_i - A \hat{R}_i^{sx} - \hat{R}_i^{xs} A^* + A \hat{R}_i^s A^*) = \text{diag} \left([I - A] \begin{bmatrix} R_i & \hat{R}_i^{xs} \\ \hat{R}_i^{sx} & \hat{R}_i^s \end{bmatrix} \begin{bmatrix} I \\ -A^* \end{bmatrix} \right) \quad (3.42)$$

La matrice du membre à gauche de l'équation précédente est toujours semi-définie positive car $\begin{bmatrix} R_i & \hat{R}_i^{xs} \\ \hat{R}_i^{sx} & \hat{R}_i^s \end{bmatrix}$ l'est (composée des estimées de matrices semi-définies positives). Ainsi

$$A^{\text{new}} = \arg \min_A \sum_{i=1}^r \log \det(R_i - A \hat{R}_i^{sx} - \hat{R}_i^{xs} A^* + A \hat{R}_i^s A^*) \quad (3.43)$$

$$N^{\text{new}} = \text{diag}(R_i - A^{\text{new}} \hat{R}_i^{sx} - \hat{R}_i^{xs} A^{*\text{new}} + A \hat{R}_i^s A^{*\text{new}}) \quad (3.44)$$

Cependant, la minimisation de A^{new} requiert peut être une minimisation complexe. Une solution possible est donc de minimiser (3.37) en premier par rapport à A (en prenant N_i fixe), puis par rapport à N_i tout en gardant A fixe (égale à la valeur trouvée). Ceci conduit à prendre

$$A^{\text{new}} = \arg \min_A \sum_{i=1}^r \text{tr}[N_i^{-1}(R_i - 2A \hat{R}_i^{sx} + A \hat{R}_i^s A^*)] \quad (3.45)$$

(dont le calcul direct est possible car la fonction à minimiser est quadratique), puis calculer N_i comme avant. Cette procédure est toutefois acceptable car dans l'algorithme EM, on n'a pas vraiment besoin de minimiser (3.37) à l'étape 'M', mais seulement de le faire décroître.

3.4.2.2 Version étendue de l'algorithme EM

Dans cette version on part de la relation $\mathbf{y}_i = B^{-1}\mathbf{s}_i$, où les vecteurs aléatoires \mathbf{y}_i sont supposés observables, B est une matrice inconnue. Cette formulation conduit aux paramètres A , B , D_i et N_i , cependant, la log-vraisemblance dans le modèle original (incomplet) ne dépend pas de B . C'est pourquoi il y a une redondance de paramètres, mais ce n'est pas un problème. En effet, d'ores et déjà, la redondance existe dans l'ensemble de paramètres A et D_i car en post-multipliant A par une matrice diagonale et en pré-multipliant D_i par son inverse, la log-vraisemblance du modèle incomplet reste inchangée. Par commodité, nous mettons $C = AB$ et nous travaillons avec les paramètres C , B , D_i et N_i . Ainsi, la redondance apparaît dans le fait que la fonction log-vraisemblance (issue du modèle incomplet) ne dépend pas des paramètres B et C , mais seulement par leur produit CB^{-1} .

Parce que le vecteur \mathbf{y}_i a comme matrice de covariance $B^{-1}D_i(B^{-1})^*$ et $\mathbf{x}_i = C\mathbf{y}_i + \mathbf{e}_i$, il est facile à voir que la log-vraisemblance du modèle complet est proportionnelle à

$$-\sum_{i=1}^r w_i \{ \log \det D_i + \text{tr}(D_i^{-1} B R_i^y B^*) - 2 \log \det B + \log \det N_i + \text{tr}[N_i^{-1}(R_i - 2R_i^{xy} C^* + C R_i^y C^*)] \} \quad (3.46)$$

où R_i^y et R_i^{xy} sont respectivement la matrice de covariance de \mathbf{y}_i et la matrice d'intercovariance entre \mathbf{x}_i et \mathbf{y}_i .

L'étape 'E' de l'algorithme EM va consister à remplacer les matrices non observées R_i^y et R_i^{xy} par leurs espérances conditionnelles sachant \mathbf{x}_i , notées \hat{R}_i^y et \hat{R}_i^{xy} . De $\mathbf{y}_i = B^{-1}\mathbf{s}_i$ nous obtenons

$$\hat{R}_i^y = B^{-1} \hat{R}_i^s (B^{-1})^*, \quad \hat{R}_i^{xy} = \hat{R}_i^{xs} (B^{-1})^* \quad (3.47)$$

L'étape 'M' de l'algorithme, quant à elle, va consister à maximiser la log-vraisemblance avec R_i^y et R_i^{xy} remplacées par \hat{R}_i^y et \hat{R}_i^{xy} et maintenues fixes. Par ailleurs, la maximisation par rapport à B et D_i est complètement découplée de celle par rapport à C et N_i . Nous obtenons

$$B^{\text{new}} = \arg \min_{\tilde{B}} \sum_{i=1}^r w_i [\log \det \text{diag}(\tilde{B} \hat{R}_i^y \tilde{B}^*) - 2 \log \det \tilde{B}] \quad (3.48)$$

$$D_i^{\text{new}} = \text{diag}[B^{\text{new}} \hat{R}_i^y (B^{\text{new}})^*] \quad (3.49)$$

Quant aux nouvelles valeurs de C et N_i , nous obtenons (par les mêmes calculs) dans

le cas $N \equiv N_i$

$$C^{\text{new}} = \left(\sum_{i=1}^r w_i \hat{R}_i^{xy} \right) \left(\sum_{i=1}^r w_i \hat{R}_i^y \right)^{-1} \quad (3.50)$$

$$N_i^{\text{new}} = \text{diag} \left[\sum_i w_i R_i - \left(\sum_{i=1}^r w_i \hat{R}_i^{xy} \right) \left(\sum_{i=1}^r w_i \hat{R}_i^y \right)^{-1} \left(\sum_{i=1}^r w_i \hat{R}_i^{yx} \right) \right] / \sum_{i=1}^r w_i \quad (3.51)$$

où \hat{R}_i^{yx} est la transposée conjuguée de \hat{R}_i^{xy} . En prenant en compte (3.47), les formules précédentes se réduisent à

$$B^{\text{new}} B^{-1} = \arg \min_T \sum_{i=1}^r w_i [\log \det \text{diag}(T \hat{R}_i^s T^*) - 2 \log \det T] \quad (3.52)$$

$$D_i^{\text{new}} = \text{diag}[(B^{\text{new}} B^{-1}) \hat{R}_i^s (B^{\text{new}} B^{-1})^*] \quad (3.53)$$

$$C^{\text{new}} = \left(\sum_{i=1}^r w_i \hat{R}_i^{xs} \right) \left(\sum_{i=1}^r w_i \hat{R}_i^s \right)^{-1} B \quad (3.54)$$

$$N_i^{\text{new}} = \text{diag} \left[\sum_i w_i R_i - \left(\sum_{i=1}^r w_i \hat{R}_i^{xs} \right) \left(\sum_{i=1}^r w_i \hat{R}_i^s \right)^{-1} \left(\sum_{i=1}^r w_i \hat{R}_i^{sx} \right) \right] / \sum_{i=1}^r w_i \quad (3.55)$$

La matrice $B^{\text{new}} B^{-1}$ n'est autre que la matrice qui diagonalise, approximativement, conjointement les matrices R_i^s . Elle peut être calculée via l'algorithme de diagonalisation conjointe, utilisé comme dans le cas non bruité. Une itération de cet algorithme peut être suffisante étant donné que dans l'algorithme EM il n'est pas vraiment nécessaire de maximiser l'espérance conditionnelle de la log-vraisemblance, mais de l'agrandir. L'algorithme de diagonalisation conjointe peut converger rapidement. En effet, il s'applique aux matrices R_i^s issues de l'étape précédente et qui doivent être déjà presque conjointement diagonalisées. Ainsi, nous pouvons accepter que la solution est quasiment atteinte seulement en une itération.

La nouvelle valeur de N_i est exactement égale à celle de la première version de cette algorithme. Quant à la valeur de A , comme elle est égale à CB^{-1} , on obtient pour sa nouvelle valeur

$$A^{\text{new}} = \left(\sum_{i=1}^r w_i \hat{R}_i^{xs} \right) \left(\sum_{i=1}^r w_i \hat{R}_i^s \right)^{-1} (B^{\text{new}} B^{-1})^{-1} \quad (3.56)$$

Cette version de l'algorithme EM peut être vue comme une extension de la première version. Elle procède comme suit : on calcule les nouvelles valeurs de A et N_i comme dans la première version, mais au lieu de prendre comme nouvelles valeurs de D_i la matrice $\text{diag}[\hat{R}_i^s]$, on effectue une diagonalisation conjointe des matrices \hat{R}_i^s et on prend

les éléments des matrices diagonalisées comme celles des D_i ; la matrice diagonalisante est enfin inversée et post-multipliée par la nouvelle valeur de A (de la première version) pour obtenir la nouvelle valeur de A dans cette version.

Quant à la redondance des paramètres dans cette version, nous allons montrer qu'elle est bénéfique. En effet, l'algorithme étendu maximise

$$- \sum_{i=1}^r w_i \{ \log \det D_i + \text{tr}(D_i^{-1} B \hat{R}_i^y B^*) - 2 \log \det B + \log \det N_i + \text{tr}[N_i^{-1} (R_i - 2 \hat{R}_i^{xy} C^* + C \hat{R}_i^y C^*)] \} \quad (3.57)$$

qui est l'espérance mathématique (conditionnelle) de la log-vraisemblance intégrale. Si nous commençons la maximisation dans l'étape 'M' par la matrice B normalisée à l'identité et si nous contraignons cette matrice à rester à l'identité, alors la maximisation est exactement la même que celle de la première version (avec $A = C$).

Ainsi, la croissance de la log-vraisemblance dans la version étendue est supérieure ou égale à celle de la première version. De plus, étant donnée que cette croissance est toujours une borne inférieure à la maximisation actuelle de la log-vraisemblance, il va de soi que la maximisation de la log-vraisemblance dans la version étendue ne peut croître moins que dans la première version.

3.4.3 Estimation des signaux sources

Une fois que les paramètres D_i , N et A sont estimés, les signaux sources peuvent être estimés à partir des observations selon l'équation suivante :

$$\hat{\mathbf{s}}_i = D_i A^* (A D_i A^* + N)^{-1} \mathbf{x}_i = (D_i^{-1} + A^* N^{-1} A)^{-1} A^* N^{-1} \mathbf{x}_i \quad (3.58)$$

À noter que nos équations sont fonctions des deux variables temps n et fréquence f (voir la note 20). Ainsi, le vecteur temporel des signaux sources $\mathbf{s}(n)$ est relié aux observations $\mathbf{x}(n)$ selon cette équation par le filtre dont la réponse fréquentielle est :

$$\begin{aligned} G(f, n) &= D(f, n) A^*(f) [A(f) D(f, n) A^*(f) + N(f)]^{-1} \\ &= [D^{-1}(f, n) + A^*(f) N^{-1} A(f)] A^*(f) N^{-1}(f) \end{aligned} \quad (3.59)$$

Par ailleurs l'estimation des paramètres D_i , N et A est réalisée par l'introduction des matrices spectrales (temps-fréquence) des observations dans l'algorithme EM. Cependant, pour des raisons de qualité d'estimation, ces matrices spectrales sont estimées sur de fenêtres glissantes et avec recouvrement des signaux observations.

En disposant des signaux observations $\mathbf{x}(n)$ et de la réponse impulsionnelle $g(n)$, issue de la transformée de Fourier inverse de $G(f)$, la manière intuitive qui permet de calculer les échantillons de $\mathbf{s}(n)$ est de faire un produit de convolution $(\mathbf{x} * g)(n)$. En considérant

la situation pratique de nos calculs et de nos signaux, à savoir la longueur des mélanges $\mathbf{x}(n)$ par rapport à celle de $g_i(n)$ en plus de la disposition de ces dernières en fonction des blocs i , il était judicieux d'adopter une méthode appropriée qui permet de convoluer séparément chaque section et où le signal final est obtenu en combinant les résultats partiels. Cette méthode est une variante appropriée des méthodes qui consistent à diviser le signal à convoluer en plusieurs sections d'où l'appellation "*de convolution sectionnée*" (voir Annexe D).

Avec cette méthode chaque réponse impulsionnelle $g_i(n)$ sera convoluée avec une section $\mathbf{x}_i(n)$ appropriée du signal mélange, correspondant à la section i . La convolution est calculée comme suit :

$$\mathbf{s}_i(n) = \sum_{l=-L+1}^L g_i(l)\mathbf{x}_i(n-l) \quad (3.60)$$

où $2L$ est le support du filtre $G_i(f)$ ($g_i(n) = 0, \forall n \notin [1-L, L]$). Chaque $\mathbf{x}_i(n)$ correspondant à la section i est défini de la manière suivante :

$$\mathbf{x}_i(n) = \begin{cases} \mathbf{x}(n) & \text{si : } (i-L) \leq n \leq (i+L-1+\mathfrak{L}), \\ 0 & \text{ailleurs.} \end{cases} \quad (3.61)$$

où \mathfrak{L} correspond au pas d'équi-sectionnement enchaîné, effectué pour le calcul de la convolution et ainsi l'obtention des \mathbf{s}_i . Il est issu de la manière de segmentation chevauchée des blocs lors de l'estimation des matrices spectrales. L'index i référence l'échantillon central des blocs (initiaux).

Les résultats partiels \mathbf{s}_i , calculés sur le support \mathfrak{L} (sans transitoire), seront placés en cascade dans le même ordre des blocs afin de reconstituer les signaux entiers des sources séparées.

3.5 Conclusion

Nous avons présenté dans ce chapitre les algorithmes que nous avons introduits pour la séparation des mélanges convolutifs de sources audio. Ces algorithmes traitent deux situations acoustiques, la première est celle qui suppose que la complexité de l'environnement acoustique (*i.e.* de la séparation) provient seulement des pics d'échos se trouvant dans la réponse impulsionnelle du filtre de mélange (en plus à sa longueur). La deuxième, quant à elle, prend en compte le bruit de l'environnement (en plus à la complexité provenant du filtre).

Dans la première situation, l'algorithme est basé sur la non-stationnarité des signaux et opère dans le domaine de fréquence par diagonalisation conjointe des matrices spectrales des mélanges, indépendamment fréquence par fréquence. Le critère performant de diagonalisation adopté est basé sur l'information mutuelle, implémenté dans un algo-

rithme rapide et simple, ce qui donne un important avantage du point de vue coût de calcul à notre algorithme.

L'algorithme est adapté aux situations d'environnements acoustiques réels qui présentent des réponses impulsionnelles très longues et un caractère réverbérant très difficile. Cela nous a amené à nous focaliser sur l'étude du problème d'ambiguïté de la séparation, et à proposer deux solutions fondées sur la continuité de la transformée de Fourier de la matrice de mélange et sur la caractérisation des sources par leurs profils de distribution d'énergie. L'idée de continuité est intégrée directement dans l'algorithme de diagonalisation conjointe, construisant ainsi un pré-traitement des permutations qui est renforcé par la deuxième étape d'élimination des permutations qui utilise les profils d'énergie. Par ailleurs, nous avons introduit l'estimation par SFD des profils pour améliorer l'efficacité de notre algorithme.

Dans la deuxième situation où le bruit est modélisé par un signal additif (aux signaux capteurs) gaussien, l'algorithme estime les sources à partir de leurs mélanges bruités par une méthode du maximum de vraisemblance. En disposant de la modélisation mathématique de cette situation, la vraisemblance des paramètres du modèle est formulée et un algorithme itératif est utilisé pour son estimation. Cet algorithme dont les résultats montrent qu'il est efficace, est fondé sur la technique Expectation-Maximization qui opère en deux étapes, estimation des paramètres puis maximisation de la log-vraisemblance. À la convergence, les signaux sources pourront être calculés en fonction des paramètres estimés.

Nous présentons dans le chapitre suivant les résultats expérimentaux de nos algorithmes en utilisant des données réelles d'environnements acoustiques.

Chapitre 4

Simulations Numériques et Résultats

Ce chapitre est consacré à la présentation des résultats de la simulation numérique de nos algorithmes de séparation de sources (cas bruité et non bruité), que nous venons de décrire dans le chapitre précédent. À noter que nous avons appliqué nos algorithmes à des données réelles dans toutes les étapes de leur développement. Afin de montrer les améliorations des différentes étapes dans le cas non bruité, nous allons montrer les résultats expérimentaux à chaque étape en évaluant la séparation par un indice de performance. Quant au cas bruité, nous allons exposer les résultats encourageants, obtenus dans différentes configurations réelles en présentant également une évaluation à ces résultats.

4.1 Jeux de données

Pour l'expérimentation de nos algorithmes, nous avons choisi deux jeux de données de salles acoustiques réelles. Le premier jeu, est la base de réponses impulsionnelles réelles que Westner a enregistré dans le cadre de sa thèse [109] et qu'il l'a aussi doté d'une macro Matlab (roomix.m) paramétrable, elle est d'ailleurs disponible sur le net (<http://sound.media.mit.edu/ica-bench/>). Quand au deuxième jeu, c'est un ensemble de réponses impulsionnelles mesurées dans les laboratoires de l'Université McMaster dans le cadre du projet BLISS [18] pour leur utilisation dans le développement d'une aide aux malentendants, par la séparation de sources.

4.1.1 Jeu de données de A. G. Westner

Afin d'expérimenter ses algorithmes dans différentes configurations acoustiques, Westner a mesuré des réponses impulsionnelles dans une salle de conférence réverbérante à 8 positions géométriques [109, pages 44-49].

La salle mesure $3,5\text{m} \times 7\text{m} \times 3\text{m}$, où deux et demi de ses murs sont couverts de tableaux blancs et un demi mur est couvert par un écran de projection. Une table de conférence et d'autres équipements en désordre se trouvent aussi dans la salle. Pour une

acquisition optimale, Westner a placé dans la salle, 8 microphones espacés d'un demi mètre, autour de la table de conférence de sorte que l'espace de parole soit couvert. Il a aussi choisi pour les haut-parleurs 24 positions, qui représentent les localisations types pour la génération de parole (voir la figure 4.1).

Les différentes localisations des microphones et haut-parleurs dans la salle se résument comme suit :

- 16 positions autour de la table de conférence.
- 8 autres positions, pour les haut-parleurs, avec une configuration en double hauteur, dans les quatre coins de la salle.
- 8 positions espacées entre elles d'un demi-mètre pour les microphones d'enregistrement.

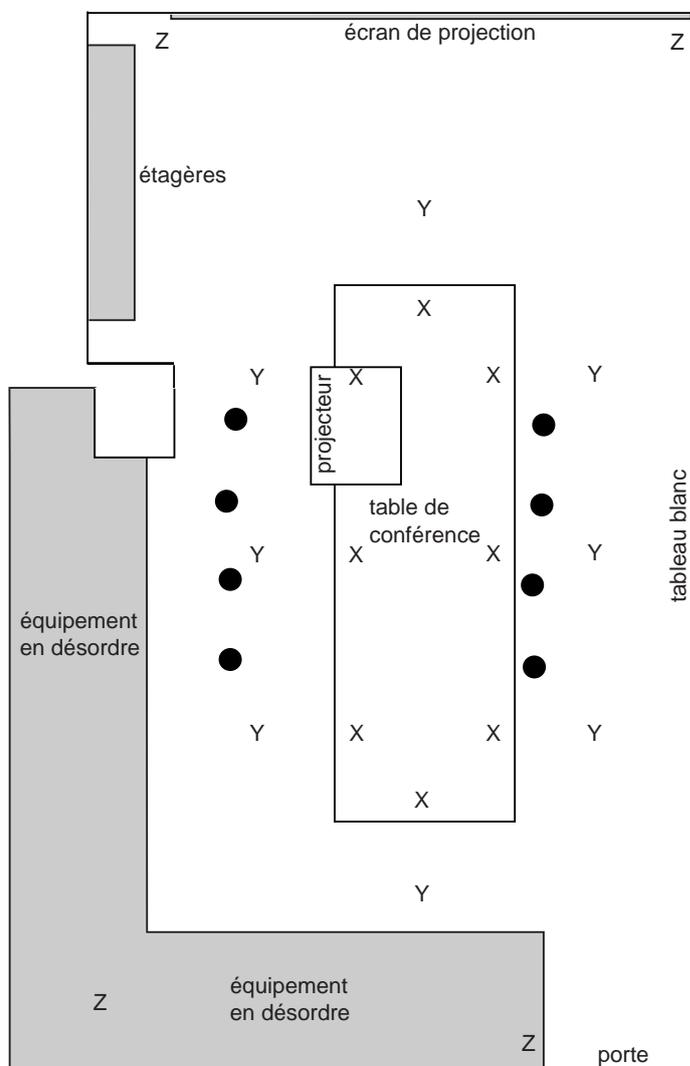


FIG. 4.1 – géométrie de la salle d'acquisition
(points : microphones, lettres : haut-parleurs)

Pour mesurer les réponses impulsionnelles, Westner envoie une excitation (séquence

à longueur maximale¹) à partir des différentes 24 positions des haut-parleurs et reçoit à chacun des 8 microphones une réponse correspondant à une propagation différente de la source. Ainsi 8 réponses impulsionnelles sont collectées, dont chacune mesure une propagation qu'un signal source peut prendre pour arriver à une localisation spécifique de la salle. De ce fait, la réponse impulsionnelle de la salle dépend des localisations simultanées de la source et le microphone de réception. Quant au calcul de la réponse impulsionnelle, il est confié à un ensemble sophistiqué à base de DSP [109, section 3.1.1].

Pour garantir la capture complète de la réponse impulsionnelle de la salle, le logiciel utilisé est configuré pour calculer la réponse impulsionnelle à une longueur de 750ms (8192 échantillons après sous-échantillonnage à 11,025kHz). Le caractère réverbérant de la salle domine sur la réponse impulsionnelle, qui malgré un prétraitement par filtrage passe-bas pour diminuer les échos et le bruit, se présente comme suit :

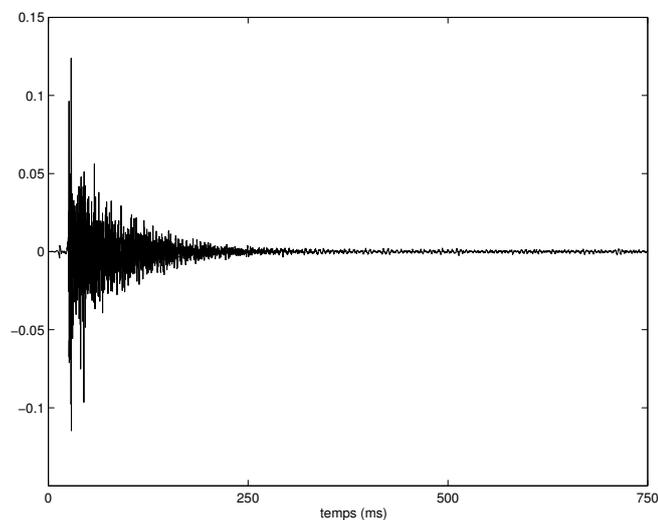


FIG. 4.2 – réponse impulsionnelle acoustique

Enfin, Westner met sur le web, en libre disposition, un code Matlab ("roommix.m") qui permet de récupérer directement en sortie, les mélanges des signaux sources par ses réponses impulsionnelles acoustiques réelles, en fonction d'un paramétrage de 8 configurations possibles.

4.1.2 Jeu de données de l'université McMaster

Dans le cadre du projet BLISS, l'université McMaster nous a fourni un jeu de données de réponses impulsionnelles réelles, mesurées en différentes localisations dans des salles acoustiques réelles de diverses dimensions et diverses caractéristiques de réverbérations [103].

La mesure est effectuée dans 3 salles, la première fait $3,4\text{m} \times 3,4\text{m} \times 2,6\text{m}$ et est couverte par une double couche de tissu à ses périphéries, la deuxième est semblable à la

¹signal aléatoire dont la distribution est similaire à celle d'un bruit rose.

première sauf qu'elle est découverte sur une dimension de $3,7\text{m} \times 3,7\text{m}$ et la troisième est une salle de cours à l'université McMaster assez réverbérante avec les dimensions $5,5\text{m} \times 10\text{m} \times 2,6\text{m}$.

Un modèle (statuette) d'un torse-tête (appelé KEMAR) est utilisé pour placer sur ses oreilles les microphones d'enregistrements. Le KEMAR est mis au centre de la salle et les réponses impulsionnelles sont prises en déplaçant un haut-parleur autour du KEMAR en 48 positions réparties : selon 8 angles ($0^\circ, 45^\circ, 90^\circ, 135^\circ, 180^\circ, 225^\circ, 270^\circ, 315^\circ$) partant d'en face le KEMAR, dans le sens des aiguilles d'une montre, avec 3 hauteurs possibles (18cm, 84cm et 165cm) du plafond jusqu'au centre du haut-parleur, et avec 2 distances possibles (0,9m et 1,8m) du KEMAR jusqu'au haut-parleur (voir la figure 4.3).

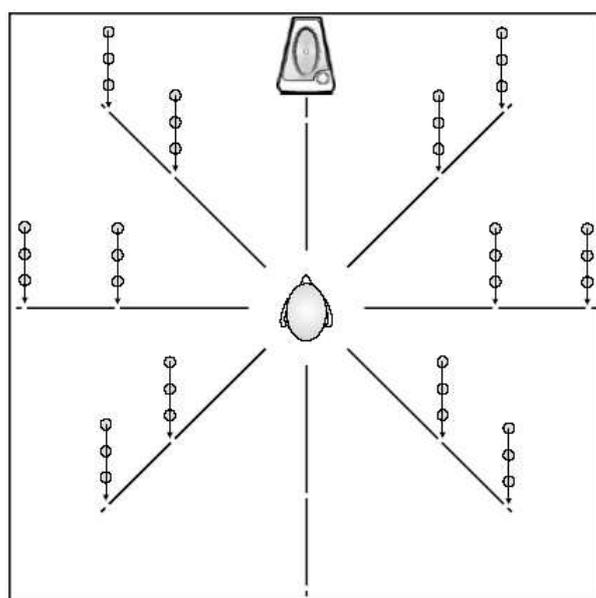


FIG. 4.3 – schéma de mesure des réponses impulsionnelles

Les réponses impulsionnelles sont enregistrées simultanément aux deux microphones en approximant le Dirac par un signal exponentiel. Pour la vérification des mesures, la propagation d'un signal de parole est enregistrée simultanément par les microphones, elle est ensuite comparée à la convolution du signal par la réponse impulsionnelle mesurée. Les réponses impulsionnelles sont fournies avec une fréquence d'échantillonnage égale à 44,100kHz et une durée d'environ 1,5s.

4.2 Indices de performance de la séparation

La mesure de performance est une tâche essentielle en séparation de sources, plusieurs définitions sont données dans la littérature [2, 117, 64, 80]. Pour évaluer la performance de séparation de nos algorithmes, nous avons adopté pour le cas non bruité deux idées naturelles permettant de construire un indicateur sur l'état des signaux séparés, et pour le cas bruité un indice approprié est aussi construit.

La première idée (cas non bruité) consiste à faire une comparaison entre les sources séparées et les sources originales par une mesure de cohérence. En effet nous savons que la cohérence entre deux signaux $x(n)$ et $y(n)$, définie par :

$$c_{xy}(f) = \frac{S_{xy}(f)}{[S_{xx}(f)S_{yy}(f)]^{1/2}}, \quad (4.1)$$

est proche de 1 quand les deux signaux sont cohérents et elle devient proche de 0 dans le cas contraire. Pour mettre en œuvre la cohérence, considérons une source séparée y_1 (cas de 2 sources, 2 capteurs) :

$$\begin{aligned} \mathbf{y} = G\mathbf{x} = GH\mathbf{s} &\implies y_1 = (GH)_{11}s_1 + (GH)_{12}s_2 \\ &= g_{11}h_{11}s_1 + g_{12}h_{21}s_1 + g_{11}h_{12}s_2 + g_{12}h_{22}s_2 \\ &= g_{11}z_{11} + g_{12}z_{12} + g_{11}z_{21} + g_{12}z_{22} \end{aligned} \quad (4.2)$$

où z_{ij} est la contribution de la source i au capteur j .

Ainsi, il est naturel que la cohérence soit calculée entre chaque source estimée et les contributions des sources originales sur un même capteur. De cette manière, étant donnée l'ambiguïté de la séparation, la cohérence va nous renseigner sur la bonne qualité (la non dégradation) des sources séparées quand elle est plus proche de 1 (0 quand il y a une permutation) et nous renseigne aussi sur les permutations restantes, car selon l'équation (4.2), la dynamique de la source bascule pour suivre l'une des dynamiques des contributions en fonction de l'occurrence des permutations. Les permutations se traduisent sur la fonction de cohérence par une discontinuité qui produit un basculement vers 1 ou 0.

La deuxième idée consiste à évaluer la qualité de la séparation à partir de la comparaison de l'ordre de grandeur de la diagonale du filtre global à celui de son anti-diagonale. La façon naturelle d'effectuer cette comparaison est de calculer le rapport entre le produit des éléments diagonaux et le produit des éléments anti-diagonaux. Nous pouvons donc définir un deuxième indicateur (pour le cas de 2 sources, 2 capteurs) comme suit :

$$r(f) = \sqrt{\frac{|(GH)_{12}(f)(GH)_{21}(f)|}{|(GH)_{11}(f)(GH)_{22}(f)|}} \quad (4.3)$$

où $(GH)_{ij}$ est l'élément ij du filtre global égale au produit matriciel GH .

Ainsi, cet indice de séparation nous indique une bonne séparation quand il est plus proche de 0 et nous indique l'occurrence des permutations quand il devient infiniment grand². Pour un renseignement plus clair de cet indicateur, nous traçons sur un même graphique l'indice et son inverse ; ainsi, les fréquences dont la qualité de séparation moins bonne sont identifiées par une valeur de l'indice, plus loin de 0 et une permutation est

²pour la bonne représentation de l'indice, toutes les valeurs infiniment grandes sont ramenées à 1 (nous utilisons $\min(r,1)$).

identifiée par un basculement opposé entre l'indice et son inverse.

Nous notons que cet indice de séparation possède l'avantage d'être inchangé par rapport à l'ambiguïté d'échelle, d'une fréquence à une autre. En effet, le facteur d'échelle est simplifié grâce au ratio définissant l'indice.

Dans le cas bruité, l'évaluation de la performance de séparation est effectuée par deux indices, le premier estime en fait le rapport signal à bruit dans le plan temps-fréquence de façon à montrer comment le contenu fréquentiel de la source séparée est affectée par le bruit. La formulation mathématique de l'indice (pour une source s_i) est naturellement donnée par :

$$SNR_i(n, f) = \frac{S_{s_i}(n, f)}{S_{[\hat{s}_i - s_i]}(n, f)} \quad (4.4)$$

Partant de l'estimation de s_i selon l'équation (3.58) et ayant le modèle $\mathbf{x}_i = A\mathbf{s}_i + \mathbf{e}_i$, il vient :

$$\begin{aligned} \hat{\mathbf{s}}_i &= D_i A^* (A D_i A^* + N)^{-1} \mathbf{x}_i \\ &= D_i A^* (A D_i A^* + N)^{-1} (A \mathbf{s}_i + \mathbf{e}_i) \\ &= B (A \mathbf{s}_i + \mathbf{e}_i) \end{aligned} \quad (4.5)$$

avec $B = D_i A^* (A D_i A^* + N)^{-1}$. Alors :

$$\begin{aligned} \hat{\mathbf{s}}_i - \mathbf{s}_i &= B (A \mathbf{s}_i + \mathbf{e}_i) - \mathbf{s}_i \\ &= B A \mathbf{s}_i + B \mathbf{e}_i - \mathbf{s}_i \\ &= (B A - I) \mathbf{s}_i + B \mathbf{e}_i \end{aligned} \quad (4.6)$$

Ainsi, la dsp $S_{[\hat{s}_i - s_i]}$ est calculée comme suit :

$$\begin{aligned} S_{[\hat{s}_i - s_i]} &= E[(\hat{\mathbf{s}}_i - \mathbf{s}_i)(\hat{\mathbf{s}}_i - \mathbf{s}_i)^*] = E[|\hat{\mathbf{s}}_i - \mathbf{s}_i|^2] \\ &= E[((B A - I) \mathbf{s}_i + B \mathbf{e}_i) ((B A - I) \mathbf{s}_i + B \mathbf{e}_i)^*] \\ &= (B A - I) S_{s_i} (B A - I)^* + B N B^* \\ &= (B A - I) D_i (B A - I)^* + B N B^* \end{aligned} \quad (4.7)$$

Enfin, le SNR est donné par :

$$\boxed{SNR_i(n, f) = 10 \log \left(\frac{D_i}{(B A - I) D_i (B A - I)^* + B N B^*} \right)} \quad (4.8)$$

L'autre évaluation est portée sur les paramètres estimés, à savoir la matrice de mélange $A(f) = \hat{H}(f)$ et le spectre des sources $D(n, f) = S_s(n, f)$. Cet indice évalue la différence entre les paramètres estimés et leurs vraies valeurs en calculant la distance quadratique entre la partie source dans le vrai spectre d'observation et son estimée ($S_x(n, f) = H(f) S_s(n, f) H^*(f) + S_b(n, f)$), ce qui constitue une bonne façon d'avoir la

distance entre les deux ensembles. La distance est normalisée à 1, elle est définie comme suit :

$$d(n, f) = (|H(f)S_s(n, f)H^*(f)| - |A(f)D(n, f)A^*(f)|)^2 \quad (4.9)$$

Cet indice normalisé à 1, indique une bonne séparation quand d est loin de 1 (proche de 0) ou le contraire sinon. Par ailleurs, avec sa formulation, cet indice a aussi l'avantage d'être inchangé par rapport à l'ambiguïté d'échelle dans le cas général (plus d'une source).

4.3 Évaluation numérique

Après avoir présenté les jeux de données et les indices de performance que nous avons utilisé pour la validation de nos algorithmes, nous décrivons dans cette section les différentes configurations de simulations effectuées dans le cas non bruité et aussi dans le cas bruité, nous présentons les différents résultats obtenus en montrant les diverses améliorations apportées (cas non bruité) et nous montrons l'évaluation des résultats par les indices de performance dans les deux cas.

4.3.1 Simulations dans le cas non bruité

Avec l'hypothèse de l'absence du bruit additif, nous traitons dans ce cas un contexte dont la difficulté principale est liée à l'environnement dans lequel se propagent les signaux. Celle-ci est modélisée par des réponses impulsionnelles longues et contenant des échos réverbérants.

4.3.1.1 Continuité de la réponse fréquentielle

Dans cette simulation nous avons utilisé le jeu de données de Westner pour construire la matrice des filtres de mélange et deux signaux audio pour construire le vecteur des sources.

Les signaux sources et les réponses impulsionnelles sont échantillonnés à 22,050kHz, les sources sont d'une durée égale à 2,7864s (61440 échantillons). la première source est un signal de parole et la deuxième est un signal de musique. Quant aux réponses impulsionnelles, nous nous sommes limités à une longueur de 512 retards. En effet, les réponses impulsionnelles fournies par roommix.m sont très longues et les coefficients des derniers retards est quasiment nulle.

La simulation avec les réponses impulsionnelles à 512 retards (sans aucune perte d'échos réverbérants) a malheureusement donné des résultats non encourageants, cela nous a donc amené à effectuer une nouvelle modification des réponses impulsionnelles en diminuant encore leur longueur à 256 retards et en atténuant les échos (voir les figures 4.4 et 4.5).

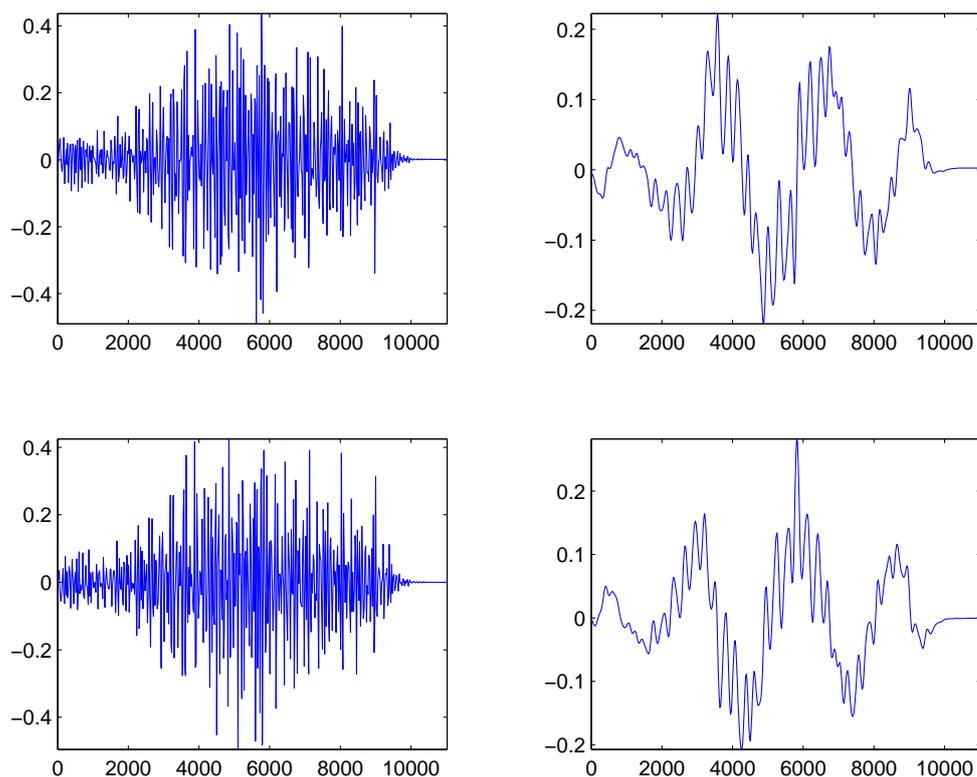


FIG. 4.4 – réponse fréquentielle du filtre de mélange h_{21} ;
à gauche : filtre à 512 retards, à droite : filtre à 256 retards ;
en haut : partie réelle, en bas : partie imaginaire.

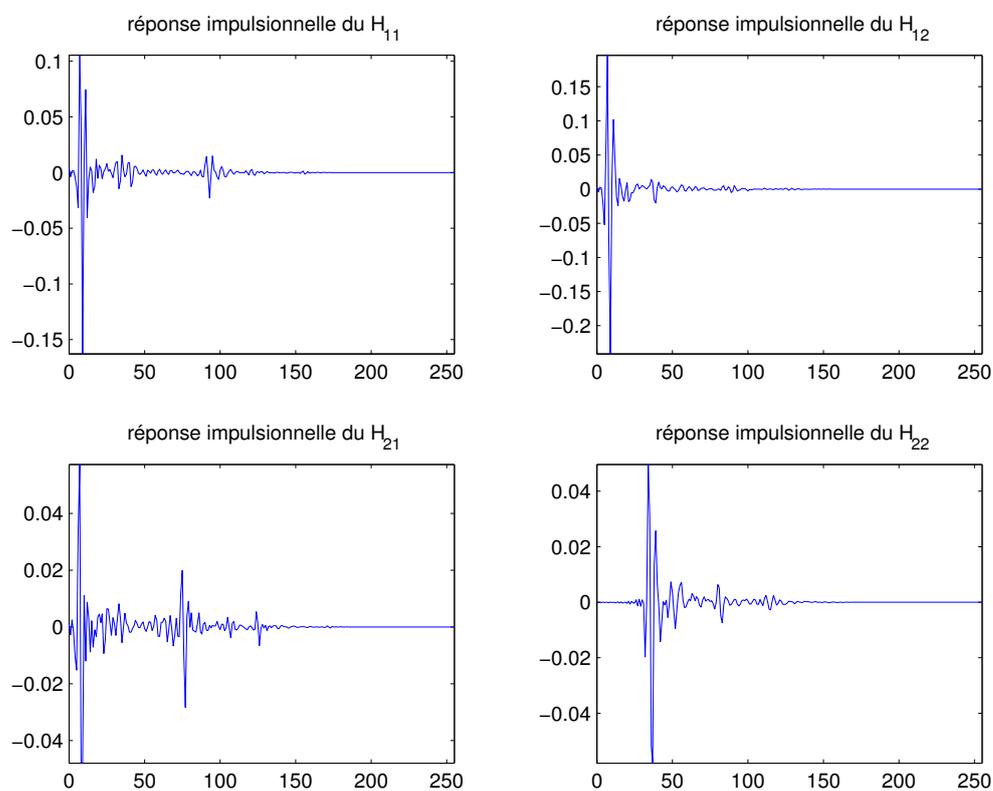


FIG. 4.5 – réponse impulsionnelle du filtre modifié

En effet, le taux des pics d'échos présents dans la version originale des réponses impulsionnelles, engendre des oscillations très rapides dans la réponse fréquentielle des filtres de mélange ; Celle-ci se représente comme sur les graphiques précédents.

L'estimation des matrices spectrales est effectuée en procédant avec des blocs de taille égale à 4096 échantillons, *i.e.* 15 blocs temporels, et des moyennes locales sur 9 fréquences. Cette configuration a donné un résultat dont l'indice de performance est présenté sur la figure 4.6. Nous rappelons que la bonne séparation est indiquée sur cet indice par des valeurs proches de 0 et les permutations occurrentes sont identifiées par une valeur égale à 1 de $\min(r(f), 1)$ (*i.e.* r est très grand) et une valeur < 1 de $\min(1/r(f), 1)$ pour la même fréquence f .

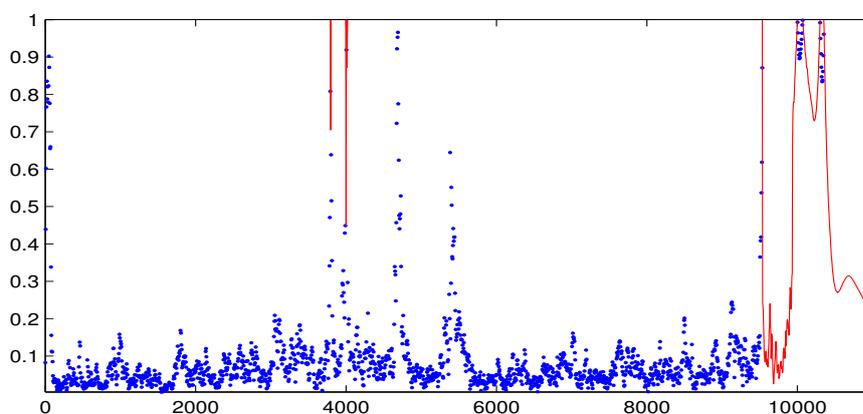


FIG. 4.6 – indice de séparation (en points) et son inverse (en trait).

À partir de cette figure nous constatons qu'il y a eu relativement peu de permutations. De plus, la faible dispersion de l'indice, ainsi que ses valeurs faibles, indiquent que le filtre de séparation était bien estimé. Les permutations occurrentes vers 4000Hz et aux HF sont probablement dues à la faible énergie du signal dans ces fréquences.

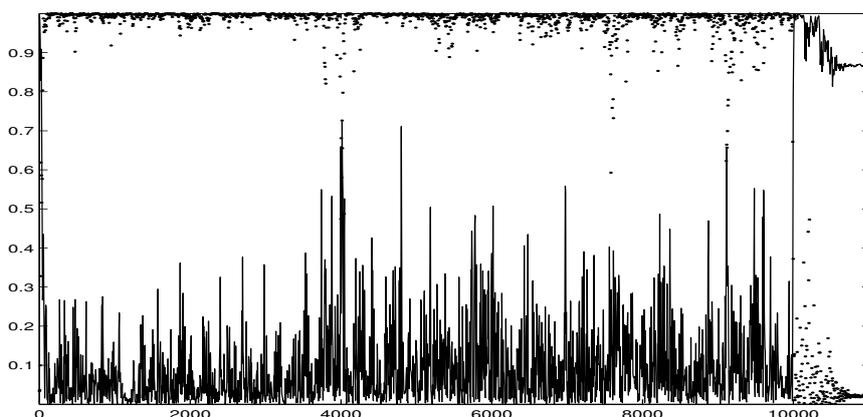


FIG. 4.7 – cohérences entre \hat{s}_1 et les contributions des sources originales.

La figure 4.7 montre la cohérence entre la première source estimée et la contribution des deux sources originales au capteur n° 1. Cet indice montre, également, une faible occurrence des permutations et un résultat de séparation qui n'est pas mauvais.

La réponse impulsionnelle du filtre global $(G * H)(n)$, présentée sur la figure³ 4.8, montre que $(G * H)_{12}(n)$ est très inférieure à $(G * H)_{11}(n)$ et $(G * H)_{21}(n)$ est très inférieure à $(G * H)_{22}(n)$, ce qui signifie que les sources sont séparées.

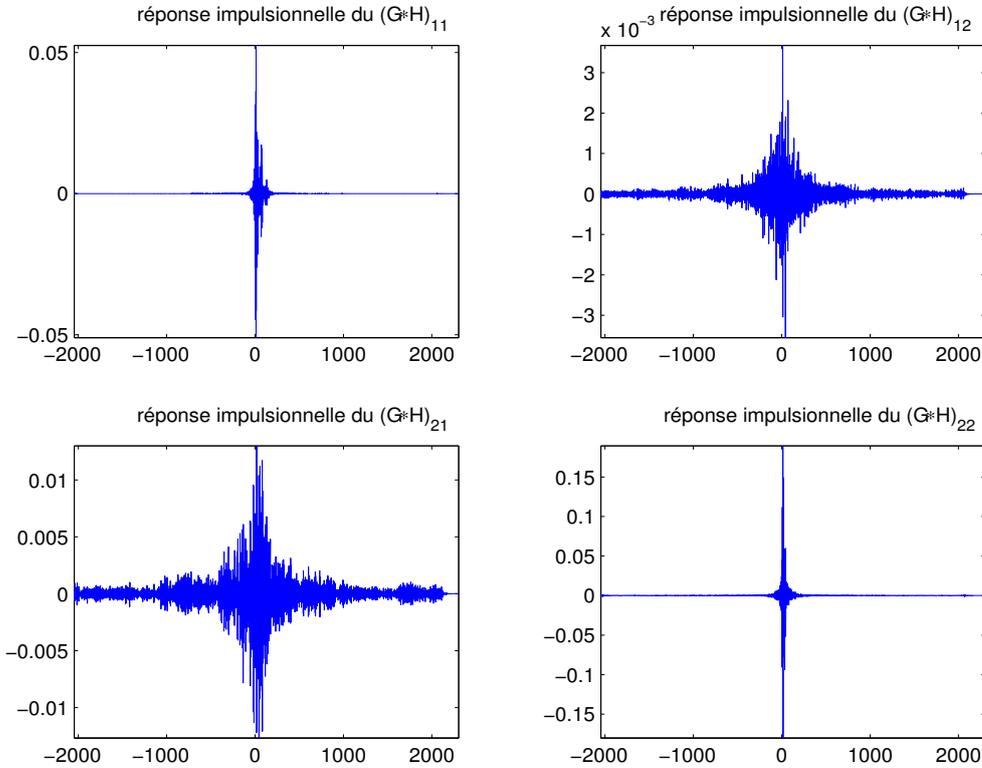


FIG. 4.8 – réponse impulsionnelle du filtre global $(G * H)(n)$.

Nous constatons que les quatre réponses impulsionnelles du filtre global ont la même forme qui a un pic au milieu de la réponse avec une décroissance de l'amplitude sur les deux côtés dans le même sens. Cela peut être expliqué par le fait que la matrice de mélange contient des filtres qui ne sont pas à phase minimale⁴. Alors, leurs filtres inverses seront non-causaux; c'est pourquoi, il est nécessaire de retarder ("translation à gauche") les échantillons des réponses impulsionnelles de la matrice de séparation pour avoir une partie causale et une partie non-causale (*i.e.* filtre non causal). De plus, la séparation tend à adapter les filtres croisés à supprimer la partie sources non désirée du signal issu des filtres directs. *i.e.* pour restituer $y_1 = (GH)_{11}s_1 + (GH)_{12}s_2$ et $y_2 = (GH)_{21}s_1 + (GH)_{22}s_2$ de sorte qu'ils ressemblent plus y_1 (estimée) à s_1 (originale) et y_2 à s_2 ; il faut que $(GH)_{12}$ soit faible devant $(GH)_{11}$ et $(GH)_{21}$ soit faible devant $(GH)_{22}$. Par ailleurs, les expériences de Westner [109, 107] montrent que l'amplitude du filtre de séparation décroît lorsque le nombre d'observations augmente. Ceci a été expliqué par le fait qu'une configuration p capteurs $\times r$ sources, additionne p copies modifiées d'un signal mélange pour reconstruire le signal source. C'est pourquoi, plus il y a des copies

³les échelles de ces graphiques tracés par Matlab ne sont pas identiques.

⁴en général, pour qu'un filtre soit à phase minimale, le premier échantillon doit être plus grand que tous les autres [107].

rajoutées ensemble plus l'amplitude de chaque copie peut être faible.

Sur le graphique suivant de cette simulation, nous présentons les signaux des sources originales, des mélanges générés et des sources résultantes de la séparation des mélanges par notre algorithme :

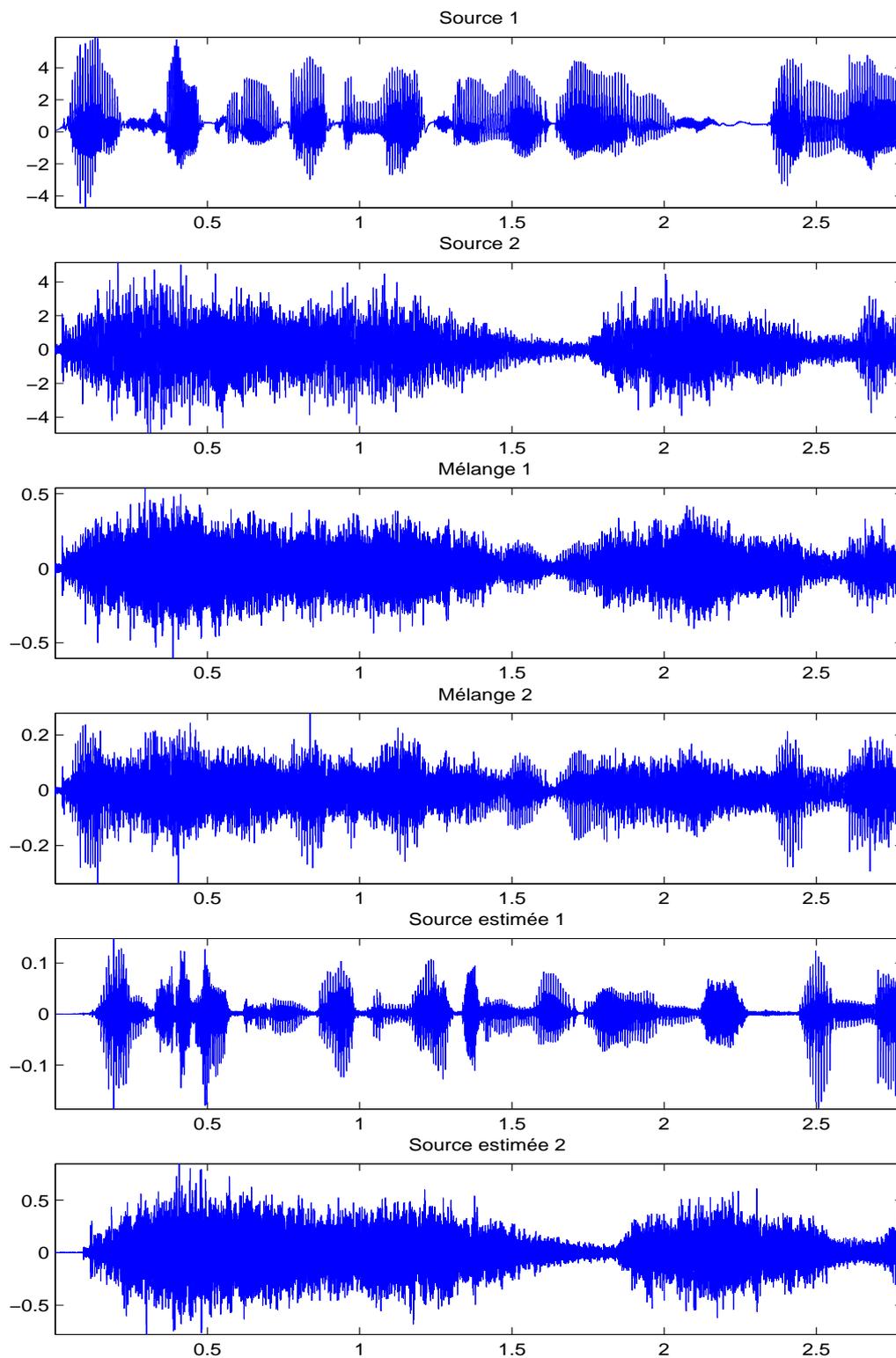


FIG. 4.9 – sources originales, mélanges et sources estimées.

4.3.1.1.1 Discussion

La présente simulation montre que notre algorithme arrive, pour certaines configurations d'environnements acoustiques, à séparer les mélanges mesurés par les capteurs. Cependant, la contrainte de continuité du filtre de mélange (de séparation aussi) n'est pas suffisante pour supprimer toutes les permutations dans le cas d'environnements très réverbérants (voir la figure 4.10). En effet, la dynamique forte de la réponse impulsionnelle de ces environnements nécessite un traitement avec une grande résolution fréquentielle afin de détecter les permutations entre $G(f_{l-1})$ et $G(f_l)$. Cela malheureusement induit un biais dans l'estimation de $G(f)$. Les sources séparées seront contaminées en conséquence, d'où la difficulté de leur affectation à la source 1 ou 2. De plus, les filtres réels de mélange sont mal conditionnés pour plusieurs fréquences (voir la figure 4.11), *i.e.* les colonnes de $H(f)$ sont presque proportionnelles à ces fréquences, ce qui rend très difficile, voire impossible la séparation à ces fréquences.

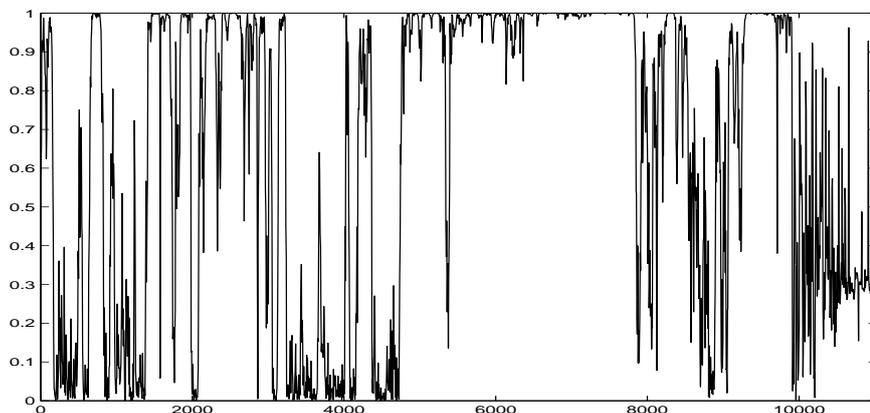


FIG. 4.10 – cohérence (filtre réel à 512 retards).

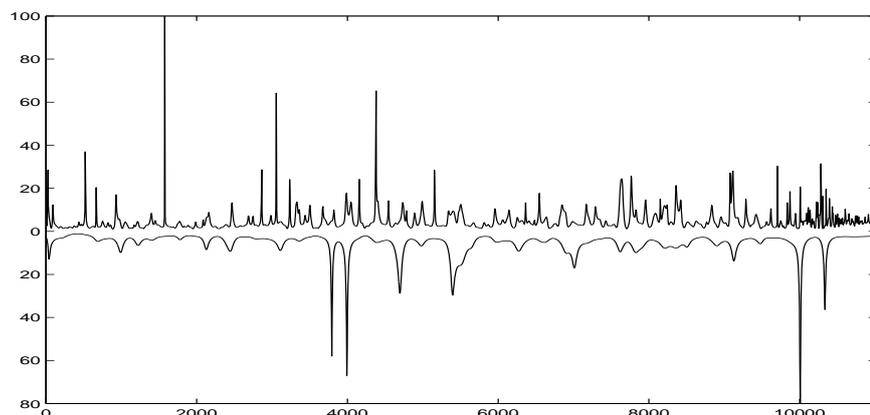


FIG. 4.11 – conditionnement de la réponse fréquentielle du filtre de mélange ;
en haut : filtre à 512 retards, en bas : filtre à 256 retards.

En conséquence, nous avons amélioré notre algorithme par une seconde étape de traitement qui renforce la séparation dans le cas d'environnements acoustiques réels. Une

simulation montrant les nouveaux résultats de notre algorithme sera présentée dans la section suivante.

4.3.1.2 Profils d'énergie par moyenne globale

Nous utilisons aussi dans cette simulation la base de données de Westner pour construire les mélanges à séparer. Quant aux sources mélangées, ce sont deux signaux de parole d'environ 2,98s, avec une fréquence d'échantillonnage égale à 11,025kHz.

Les réponses impulsionnelles utilisées sont sous-échantillonnées à 11,025kHz et coupées à 256 retards où les échos sont quasiment tous préservés. Les filtres de la matrice de mélange utilisée sont présentés sur les figures suivantes :

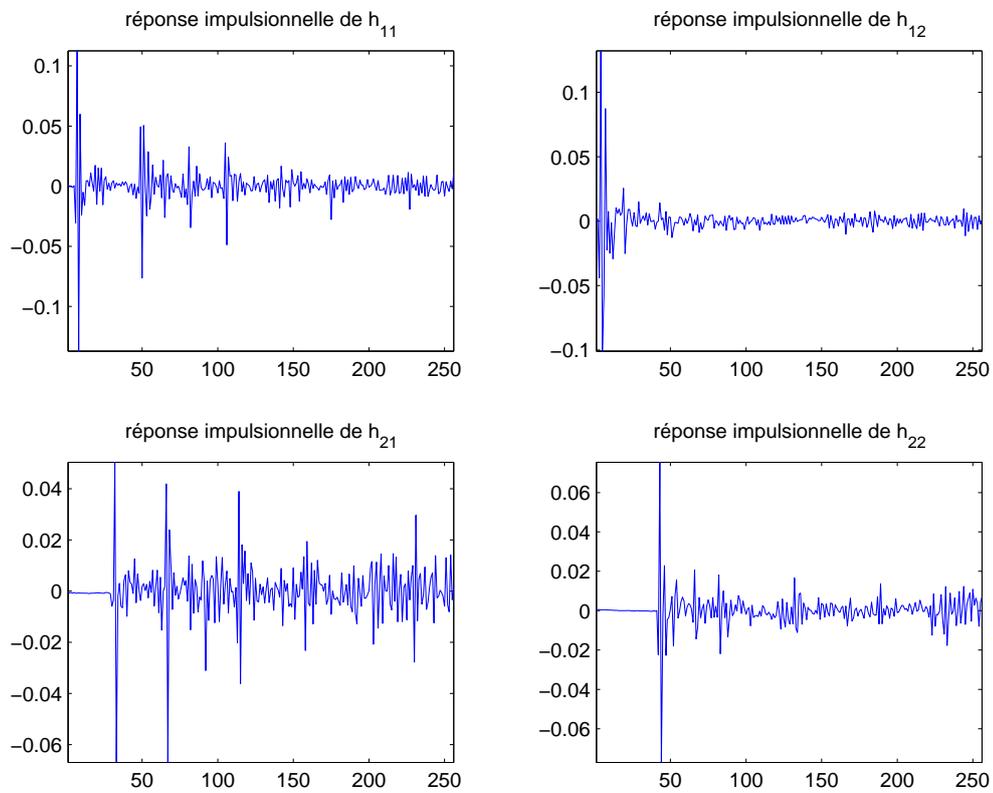


FIG. 4.12 – réponse impulsionnelle du filtre considéré.

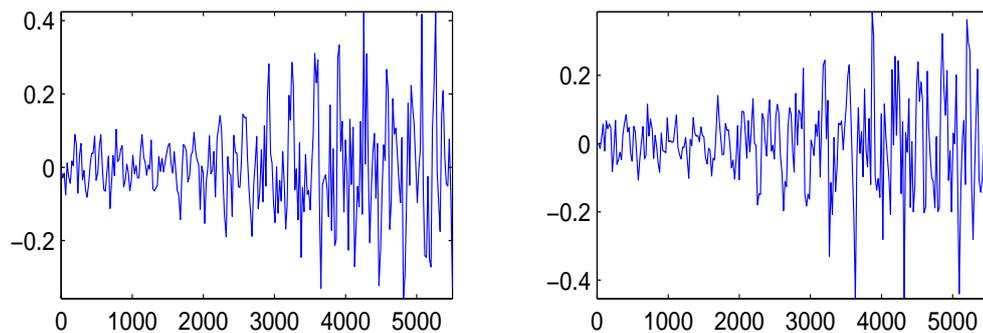


FIG. 4.13 – réponse fréquentielle de h_{21} (filtre considéré);
à gauche : partie réelle, à droite : partie imaginaire.

Dans cette simulation, nous avons procédé avec 31 blocs chevauchés d'un demi bloc. La taille des blocs est réglée à 2048 échantillons. Le moyennage du périodogramme est fixé à 5 fréquences.

Pour montrer l'amélioration apportée par la procédure d'élimination des permutations, basée sur les profils d'énergie des sources, nous avons calculé l'indice de séparation à la sortie de l'algorithme sans⁵ et avec cette procédure :

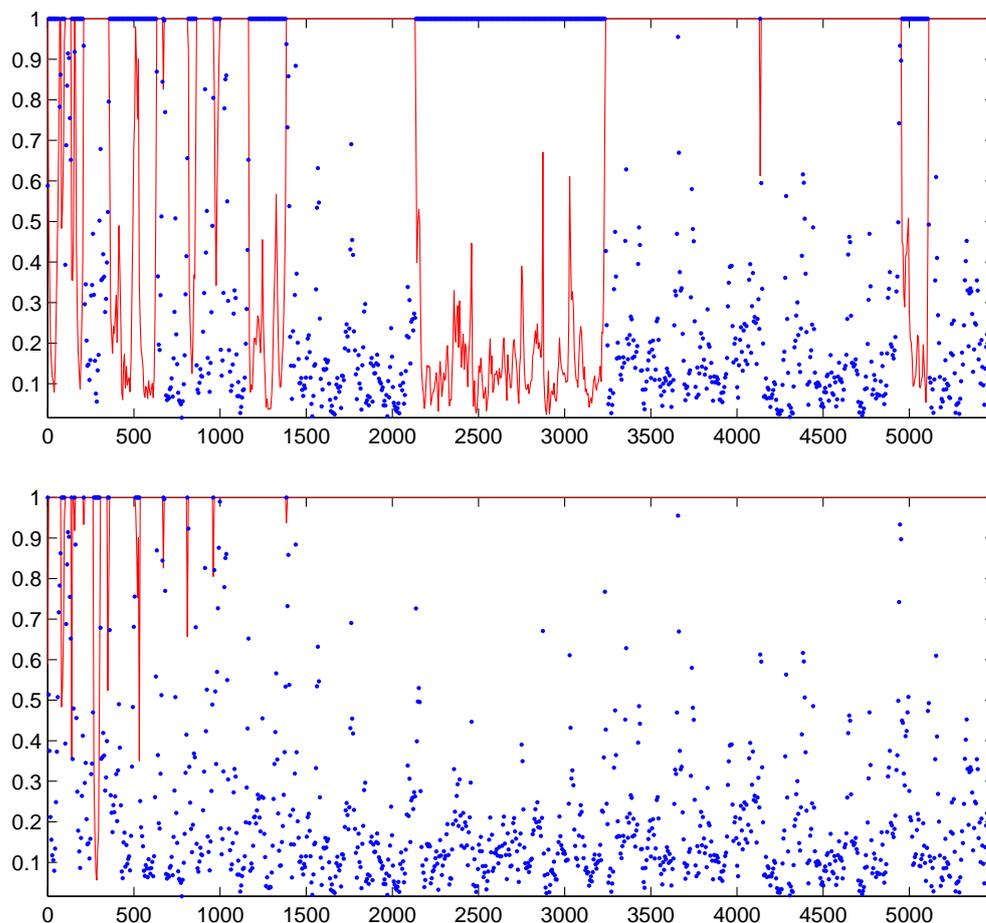


FIG. 4.14 – *indice de séparation (trait) et son inverse (points);
en haut : avant l'application de la procédure des profils d'énergie ;
en bas : après l'application de la procédure des profils d'énergie.*

En comparant les deux indices, nous remarquons que la nouvelle procédure basée sur les profils d'énergie permet d'éliminer presque toutes les permutations qui ont subsisté malgré la contrainte de continuité de la réponse fréquentielle. L'ambiguïté de permutation, se produit actuellement seulement dans certains canaux de fréquences, à l'opposé de son occurrence avant, sur des bandes entières de fréquences.

La présence de permutations après l'application de la procédure des profils est due à l'estimation actuelle du profil qui est issu des sources estimées. En effet, ces estimées peuvent être contaminées, étant donnée leur estimation effectuée seulement avec la

⁵*i.e.* seulement avec la contrainte de continuité de la réponse fréquentielle des filtres

contrainte de continuité. De plus, vue leur estimation, les profils se diffèrent des écarts-types, seulement d'une variation un peu plus lisse et les deux profils ne sont pas bien discernés à l'opposé des vrais profils (*i.e.* des sources originales). Sur la figure suivante sont tracés les profils d'énergie et leurs écarts-types.

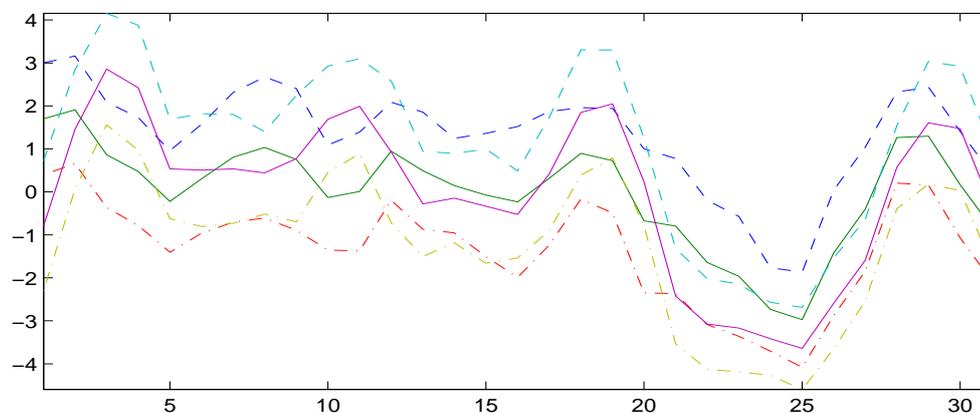


FIG. 4.15 – profils des 2 sources (en trait continu) et leur écart-type (en - - et ---).

La réponse impulsionnelle⁶ du filtre global $(G * H)(n)$, présentée sur la figure 4.16, montre que $(G * H)_{11}(n)$ est très inférieure à $(G * H)_{12}(n)$ et $(G * H)_{22}(n)$ est très inférieure à $(G * H)_{21}(n)$, ce qui signifie que les sources sont séparées avec une permutation globale.

Sur le suivant graphique de cette simulation, nous présentons les signaux des sources originales, des mélanges générés et des sources résultantes de la séparation des mélanges par notre algorithme :

4.3.1.2.1 Discussion

Avec l'exploitation des caractéristiques intrinsèques aux sources, la construction de la procédure fondée sur les profils, qui représentent la variation d'énergie des sources, a apporté une grande amélioration à notre algorithme qui réussit à séparer des sources mélangées convolutivement dans un environnement acoustique réel où les réponses impulsionnelles des filtres sont très longues et caractérisées par beaucoup de réverbérations.

Cependant, l'estimation de ces profils nécessite un affinement pour avoir des résultats meilleurs. C'est pourquoi, nous avons proposé des nouvelles méthodes d'estimation, dont les résultats de simulation seront présentés dans la prochaine section.

4.3.1.3 Profils d'énergie par moyenne locale et par SFD

Dans l'objectif de montrer les nouvelles améliorations apportées par les méthodes d'estimation des profils d'énergie, par une moyenne locale glissante et par l'utilisation

⁶leurs graphiques tracés par Matlab ne sont pas fournis avec des échelles identiques.

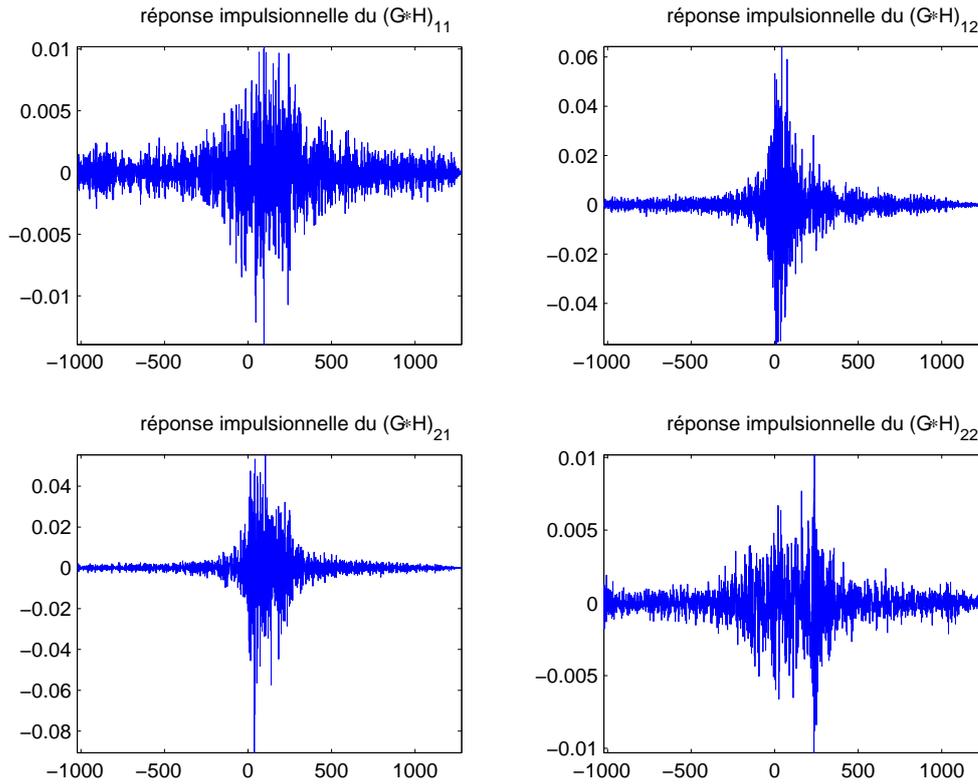


FIG. 4.16 – réponse impulsionnelle du filtre global $(G * H)(n)$.

de la Série de Fourier Discrète, nous avons choisi trois simulations pour comparer les résultats issus de l'application de ces nouvelles méthodes.

Nous avons gardé la même configuration que celle de la simulation 4.3.1.2, *i.e.* des sources d'une longueur d'environ 2,98s, échantillonnées à 11,025kHz. Les matrices spectrales sont estimées en travaillant avec des blocs apodisés de 2048 échantillons, avec des recouvrements de 75%, ce qui fait 57 blocs ; le périodogramme est moyenné sur 5 blocs.

Quant aux réponses impulsionnelles, en plus de la réponse impulsionnelle utilisée dans la simulation précédente (256 retards), nous avons aussi utilisé deux réponses impulsionnelles, de longueur 512 retards et 1024 retards, qui sont issus du jeu de données de l'université McMaster⁷ (voir les figures 4.18 et 4.19).

Afin de comparer les résultats des deux méthodes d'estimation des profils et montrer leur apport améliorant la séparation, nous avons calculé l'indice de séparation avant l'application de ces méthodes (algorithme avec la contrainte de continuité et avec l'ancienne estimation des profils) et après leur application, à chacune son indice de performance.

Avec un paramétrage optimal des deux méthodes d'estimation des profils, *i.e.* nombre de fréquences moyennées pour l'estimation par moyenne local, nombre de fréquences à filtrer (sélectionner) par l'utilisation de l'SFD, nous avons obtenu les résultats suivants :

4.3.1.3.1 Réponse impulsionnelle à 256 retards

⁷voir l'Annexe B pour la description de la méthode de leur sous échantillonnage

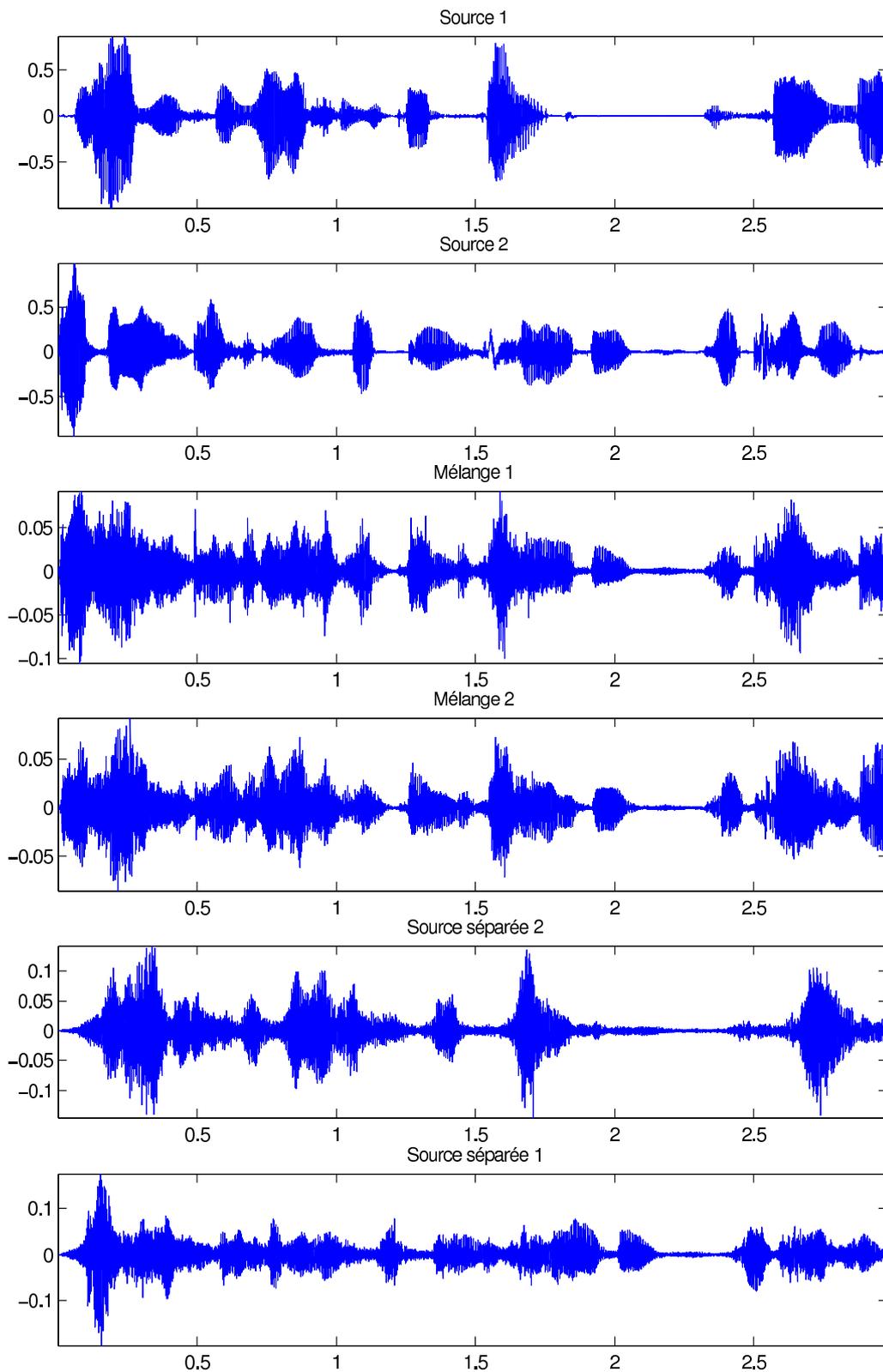


FIG. 4.17 – sources originales, mélanges et sources estimées.

Sur la figure 4.20 nous présentons les trois indices de séparation correspondant aux résultats de chaque méthode d'estimation : ancienne estimation, estimation par moyenne

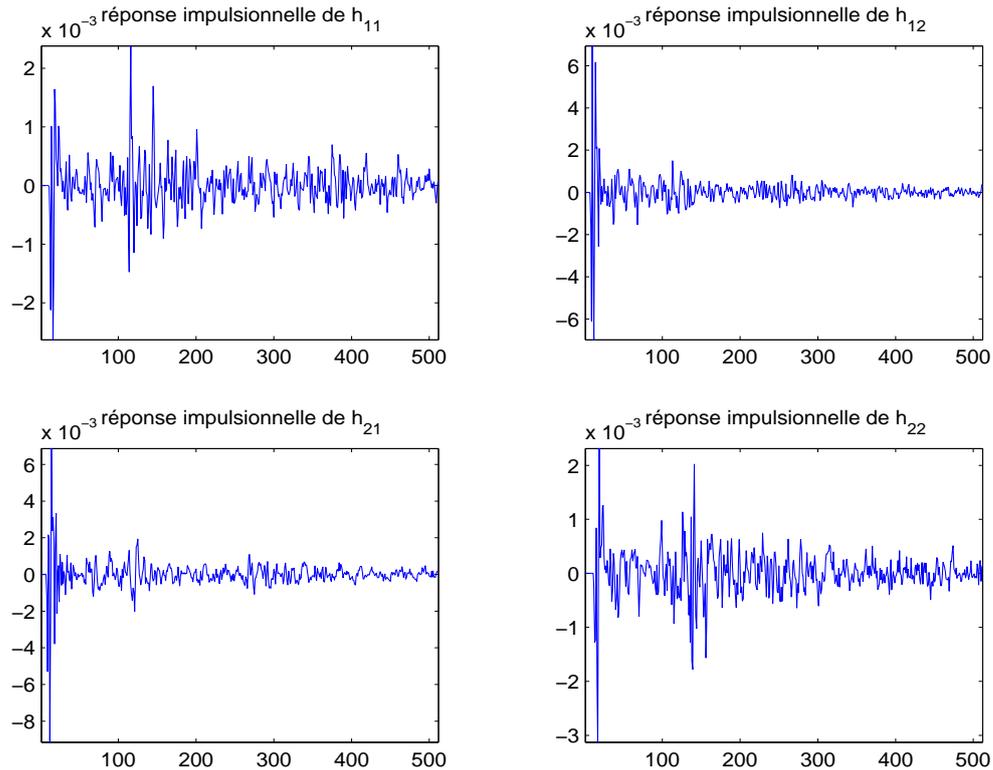


FIG. 4.18 – réponse impulsionnelle à 512 retards (jeu de données de McMaster).

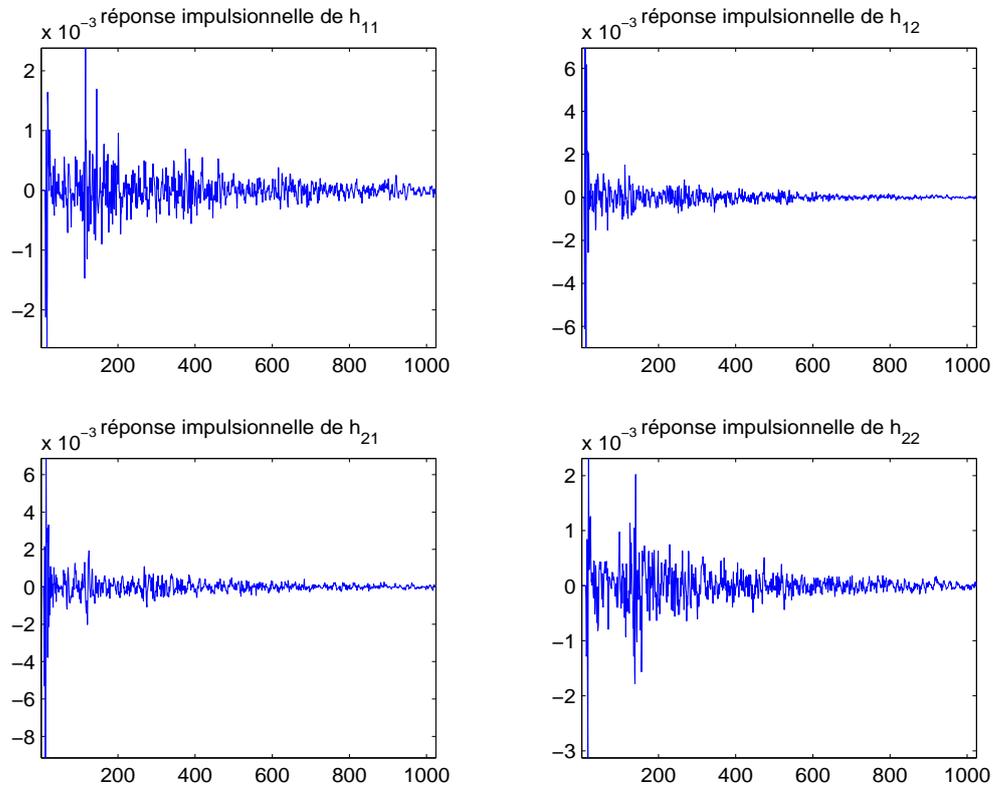


FIG. 4.19 – réponse impulsionnelle à 1024 retards (jeu de données de McMaster).

locale et estimation par l'utilisation de la SFD.

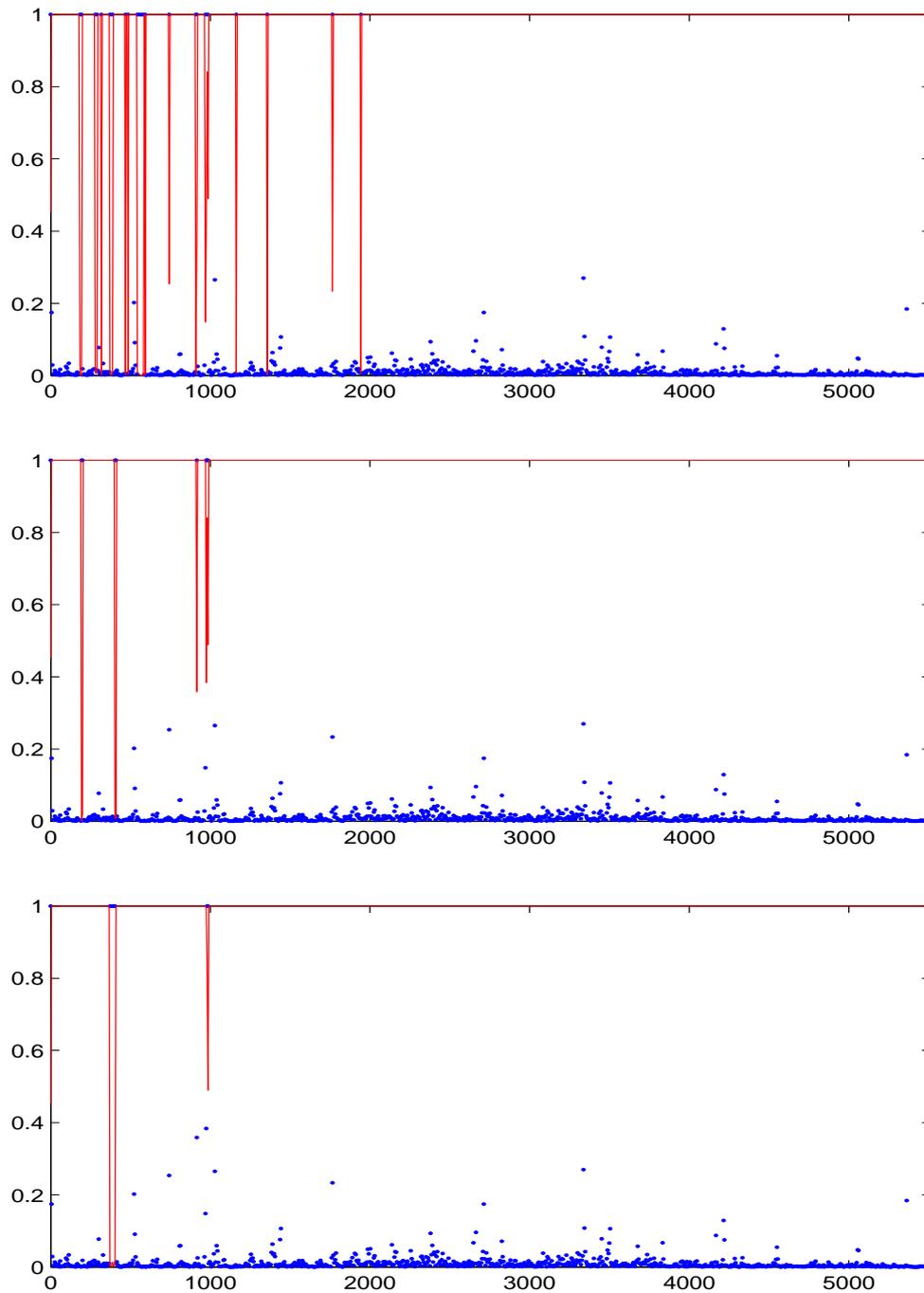


FIG. 4.20 – *indice de séparation (trait) et son inverse (points);
en haut : profils estimés par moyenne globale;
au milieu : profils estimés par moyenne locale;
en bas : profils estimés en utilisant la SFD.*

En comparant les trois indices de séparation, correspondants chacun à une méthode d'estimation de profils, nous remarquons qu'avec la nouvelle estimation une nette amélioration est apportée à la séparation. En effet, nous arrivons par leur application à supprimer les permutations qui n'ont pas été détectées par la première méthode des profils. Nous remarquons également que la méthode basée sur la SFD améliore encore le résultat par rapport à la méthode d'estimation par moyenne locale.

Néanmoins, les résultats des deux méthodes sont proches. Cependant, en pratique, le paramétrage de la méthode basée sur la SFD est plus simple par rapport à l'autre méthode. L'estimation par moyenne locale est réalisée avec une fenêtre de taille égale à 450 fréquences et l'estimation par SFD est réalisée en faisant un filtrage (sélectif) de 3 fréquences⁸.

La réponse impulsionnelle du filtre global $(G * H)(n)$, correspondant au résultat de la méthode basée sur la SFD, confirme l'accomplissement de la séparation des sources. *i.e.* $(G * H)_{12}(n)$ est très inférieure à $(G * H)_{11}(n)$ et $(G * H)_{21}(n)$ est très inférieure à $(G * H)_{22}(n)$:

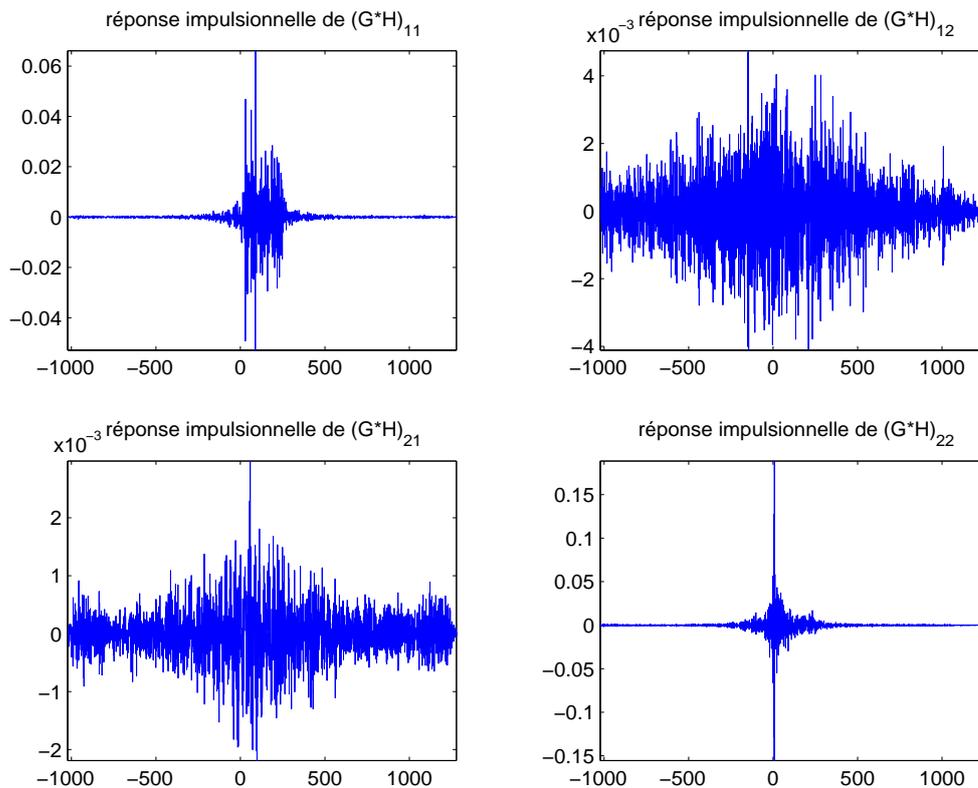


FIG. 4.21 – réponse impulsionnelle du filtre global $(G * H)(n)$ (méthode de la SFD).

Nous rappelons que l'élimination des permutations est liée à la bonne estimation des profils. Le peu de permutations qui reste, est dû soit à la mauvaise estimation des sources aux fréquences d'occurrence, soit à la ressemblance des sources à ces fréquences, ce qui peut être expliqué par le fait que l'algorithme n'arrive pas à bien affecter les sources à leur profils en ces fréquences. Nous avons présenté sur les figures 4.22 et 4.23 les profils de quelques blocs, estimés par les deux méthodes.

Les signaux des sources originales, des mélanges générés et des sources séparées en utilisant la méthode de la SFD sont tracés sur la figure 4.24.

⁸ces échantillons fréquentiels correspondent aux plus basses fréquences de la grille à 1024 points positifs.

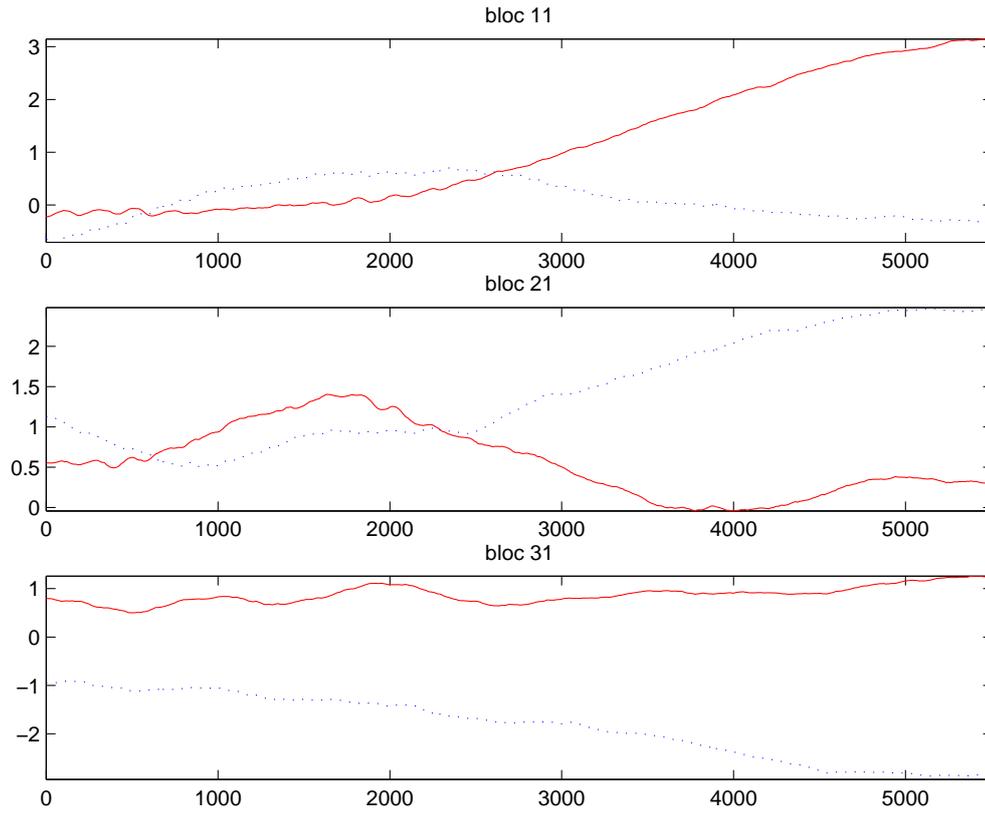


FIG. 4.22 – *profils de quelques blocs (méthode de la moyenne locale); source 1 : en trait continu, source 2 : en pointillé.*

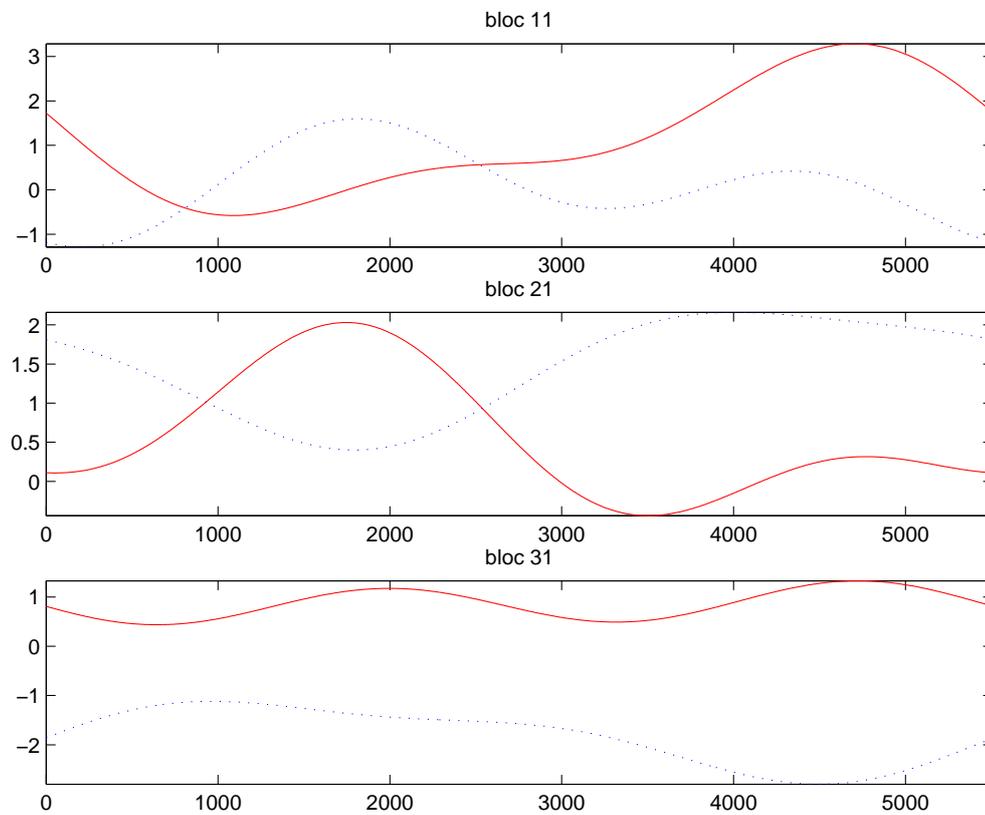


FIG. 4.23 – *profils de quelques blocs (méthode de la SFD); source 1 : en trait continu, source 2 : en pointillé.*

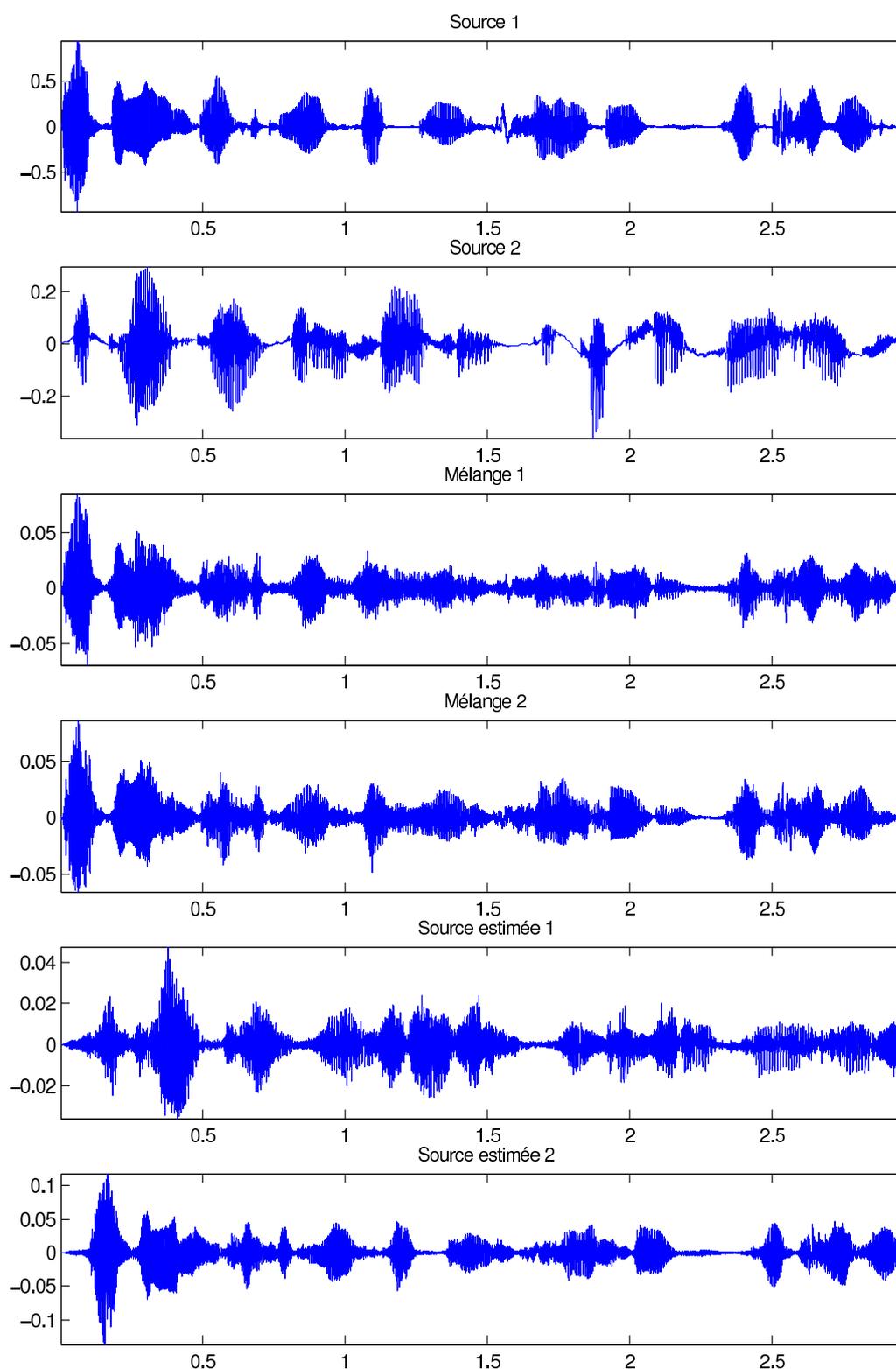


FIG. 4.24 – sources originales, mélanges et sources estimées (méthode de la SFD).

4.3.1.3.2 Réponse impulsionnelle à 512 retards

Les résultats de cette simulation sont évalués par l'indice de séparation pour les trois méthodes d'estimation des profils comme suit :

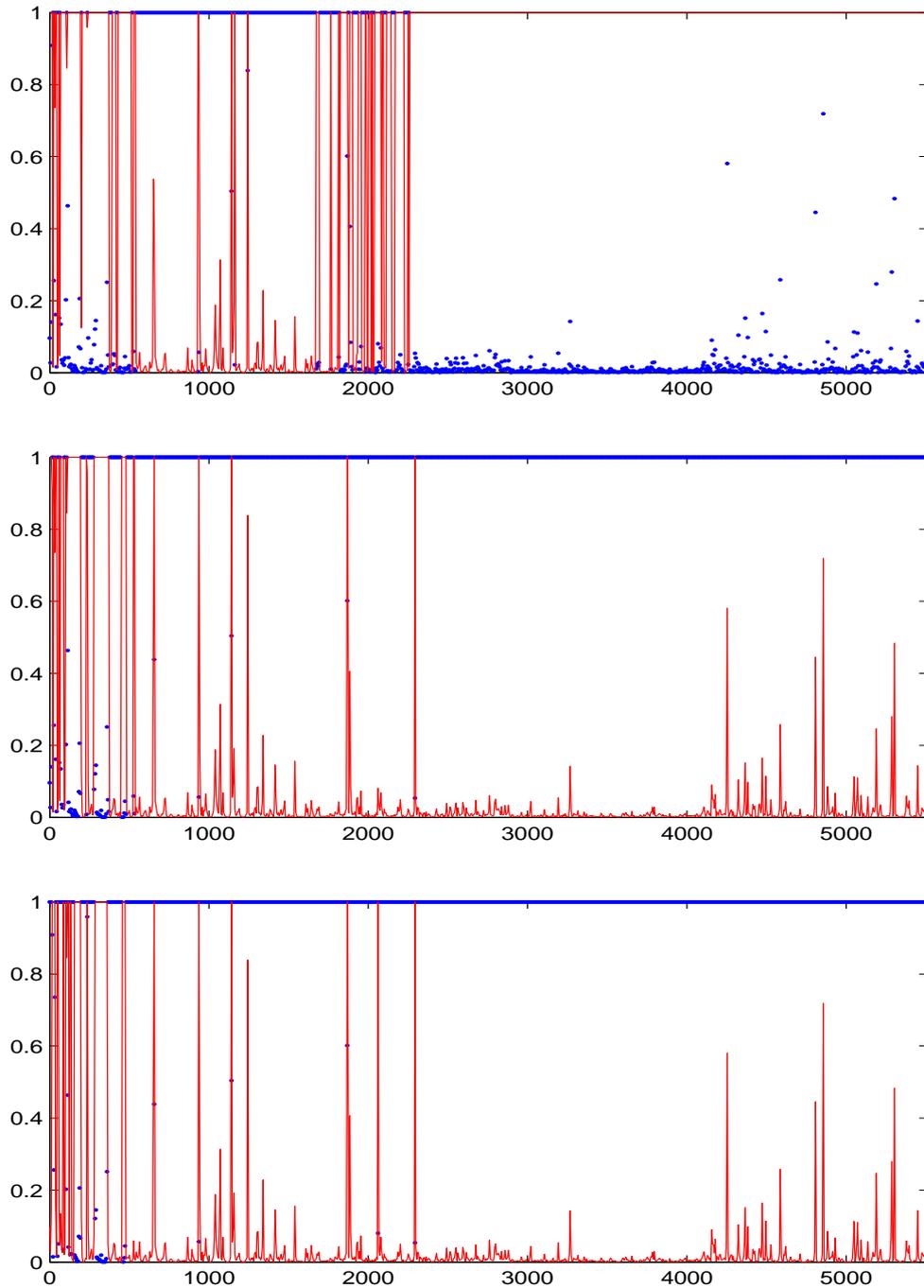


FIG. 4.25 – *indice de séparation (trait) et son inverse (points);
 en haut : profils estimés par moyenne globale;
 au milieu : profils estimés par moyenne locale;
 en bas : profils estimés en utilisant la SFD.*

Dans cette simulation la performance obtenue par les nouvelles méthodes est meilleure que celle de la méthode de la moyenne globale; Avec ces deux méthodes l'algorithme arrive à éliminer beaucoup de permutations qui, cependant, restaient avec l'ancienne méthode d'estimation des profils. Les résultats sont obtenus avec un paramétrage égale à 400 fréquences moyennées localement pour la première nouvelle méthode et un lissage à 0,39% (2 fréquences) pour la deuxième nouvelle méthode (par SFD). Par ailleurs, nous

constatons que les résultats de ces deux nouvelles méthodes sont comparables.

La réponse impulsionnelle du filtre global $(G * H)(n)$, du résultat final (méthode de la SFD) est illustrée à la figure 4.26 :

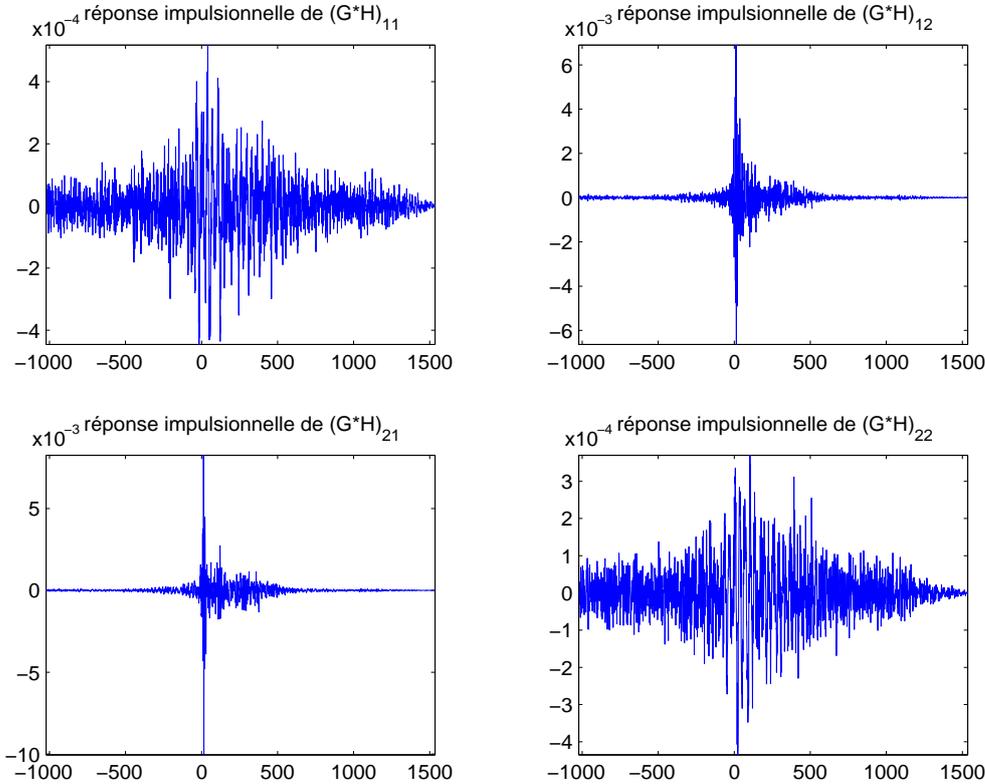


FIG. 4.26 – réponse impulsionnelle du filtre global $(G * H)(n)$ (méthode de la SFD).

Nous constatons que $(G * H)_{11}(n)$ est très inférieure à $(G * H)_{12}(n)$ et $(G * H)_{22}(n)$ est très inférieure à $(G * H)_{21}(n)$, ce qui signifie que grâce à notre algorithme les sources ont pu être séparées (avec un ordre inversé), en plus à la suppression de leurs permutations fréquentielles, qui est confirmée par le graphique 4.25. Les signaux correspondants aux sources originales, aux mélanges générés et aux sources séparées (méthode de la SFD), sont tracés sur la figure 4.29.

Les profils des blocs 11, 21 et 31, estimés par les deux méthodes sont présentés sur les deux figures 4.27 et 4.28. Nous remarquons que les profils des deux méthodes sont très similaires, ce qui explique le voisinage des performances obtenues par les deux méthodes.

Étant donnée que l'estimation de la matrice de séparation n'est pas très favorable à cause de la nature du filtre de mélange qui est caractérisé par la présence de beaucoup de pics d'échos réverbérants, les profils estimés ne peuvent pas très bien caractériser les sources. Cependant, la séparation des signaux mélanges par le nouvel algorithme reste globalement de bonne qualité au vu du peu d'erreur de permutations.

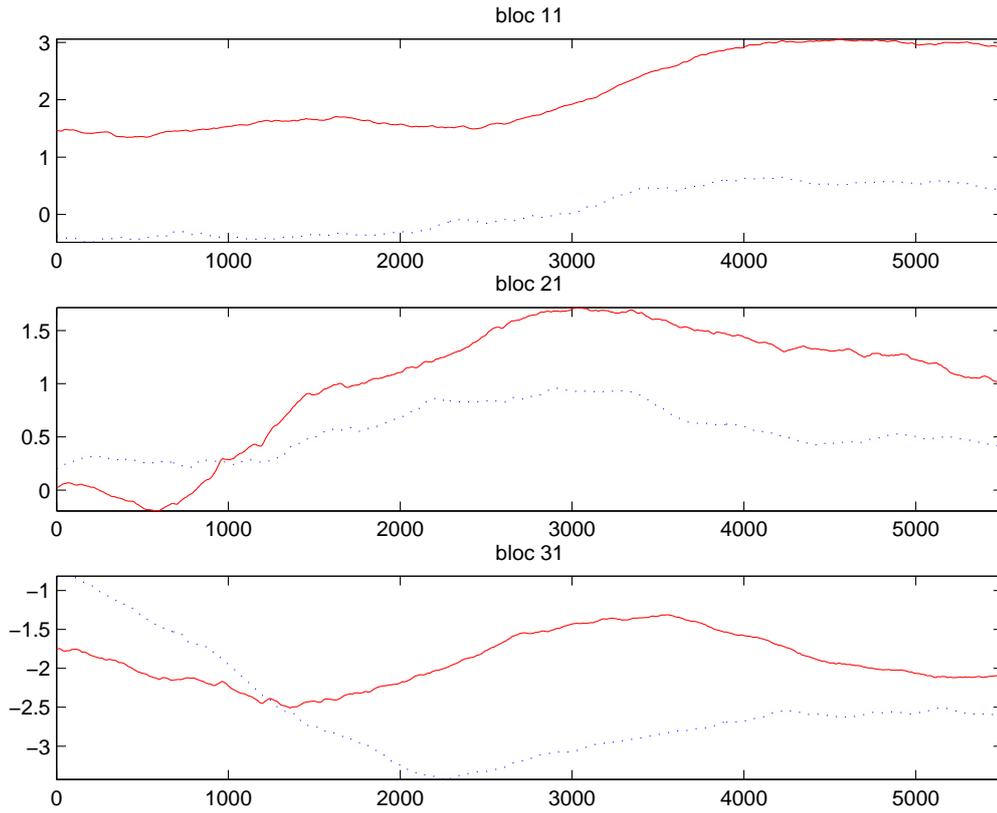


FIG. 4.27 – *profils de quelques blocs (méthode de la moyenne locale); source 1 : en trait continu, source 2 : en pointillé.*

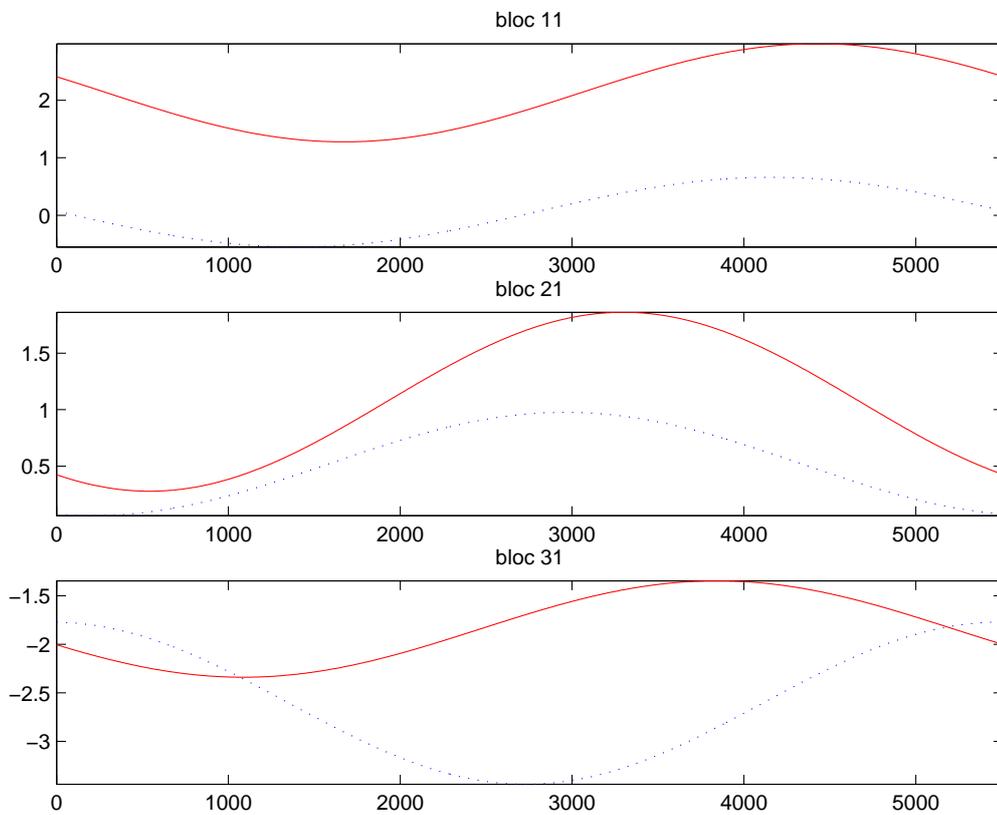


FIG. 4.28 – *profils de quelques blocs (méthode de la SFD); source 1 : en trait continu, source 2 : en pointillé.*

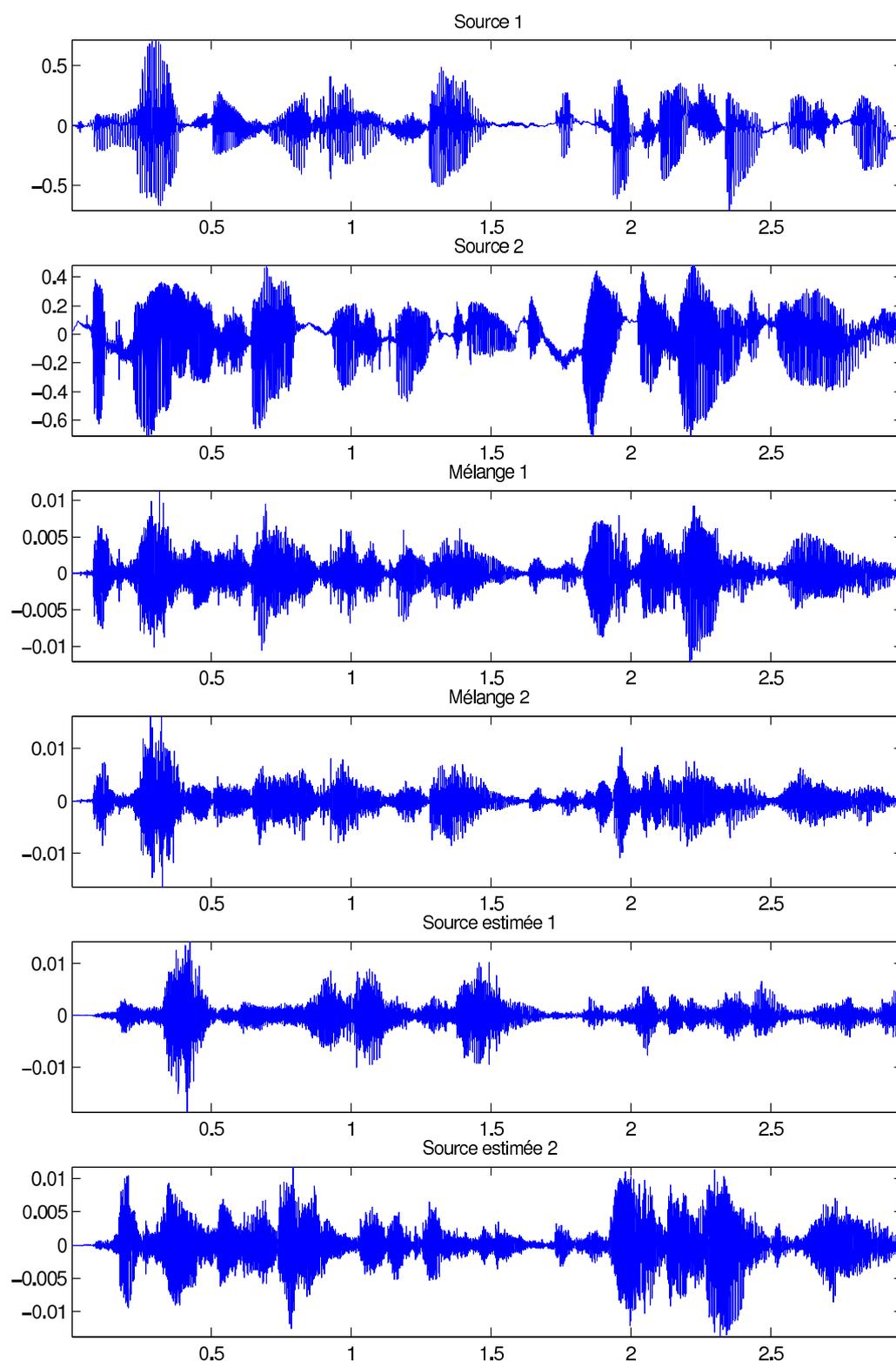


FIG. 4.29 – sources originales, mélanges et sources estimées (méthode de la SFD).

4.3.1.3.3 Réponse impulsionnelle à 1024 retards

Afin de comparer les résultats des trois méthodes dans ce contexte de séparation difficile (1024 retards), nous avons calculé les indices de séparation, qui se présentent

comme suit :

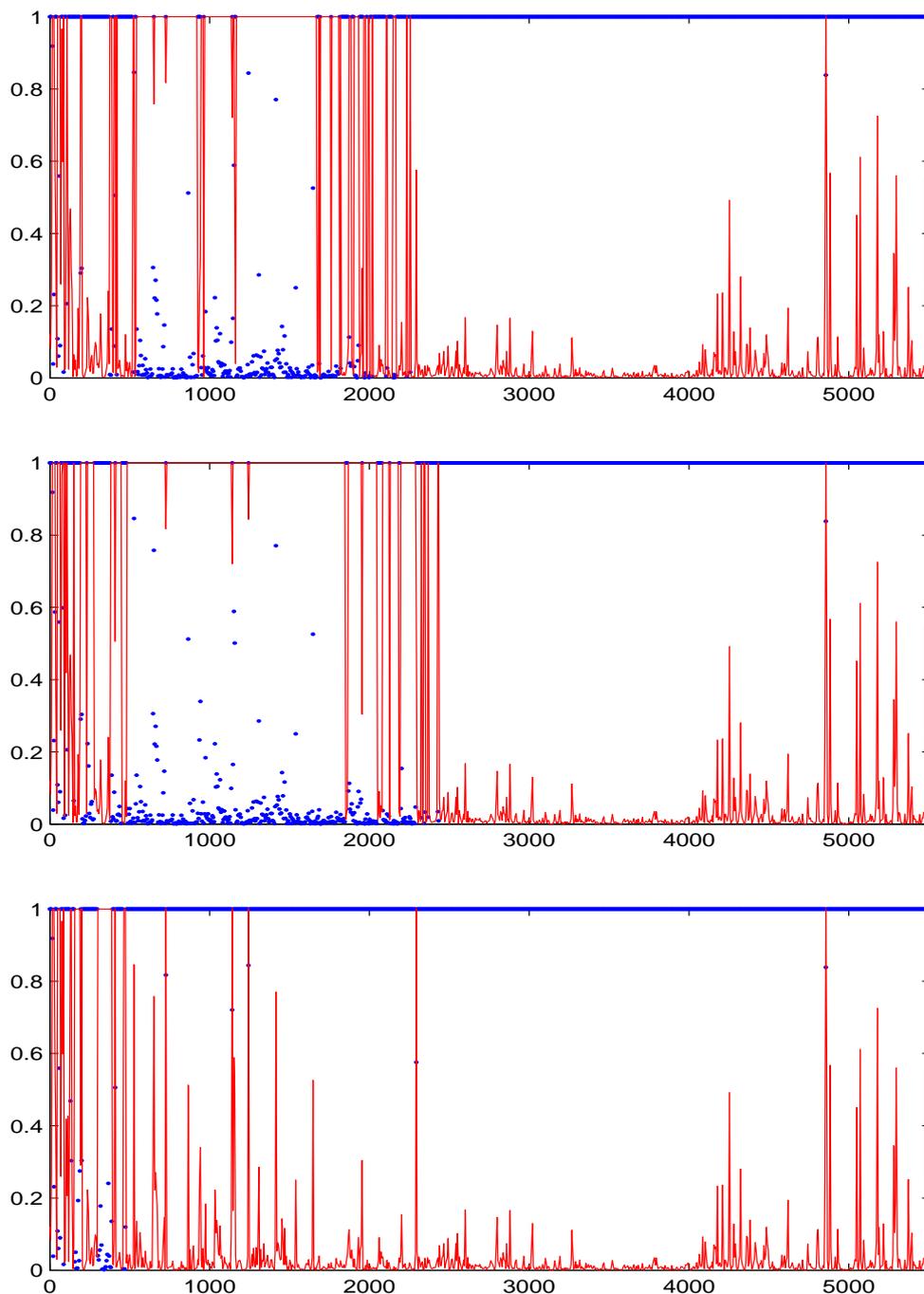


FIG. 4.30 – *indice de séparation (trait) et son inverse (points);
en haut : profils estimés par moyenne globale;
au milieu : profils estimés par moyenne locale;
en bas : profils estimés en utilisant la SFD.*

À partir de cette simulation, il s'avère que la méthode des SFD fournit un résultat meilleur que les autres méthodes. En effet, dans le cas de ce filtre long et contenant beaucoup de réverbération, les méthodes basées sur l'idée de moyennage laissent plusieurs permutations. Cependant, la méthode qui est basée sur la SFD élimine quasiment toutes les permutations, en plus de son paramétrage plus facile. À noter que la mé-

thode qui utilise la SFD a été effectuée avec un réglage correspondant à un filtrage de 3 fréquences et la méthode qui procède par moyenne locale a été effectuée avec un paramétrage correspondant à un lissage local sur 400 fréquences.

Par ailleurs, la réponse impulsionnelle du filtre global $(G * H)(n)$, confirme la séparation des deux sources mélangées, nous voyons sur la figure 4.31 que $(G * H)_{11}(n)$ est très inférieure à $(G * H)_{12}(n)$ et $(G * H)_{22}(n)$ est très inférieure à $(G * H)_{21}(n)$, ce qui signifie que les sources ont été séparées avec une permutation globale :

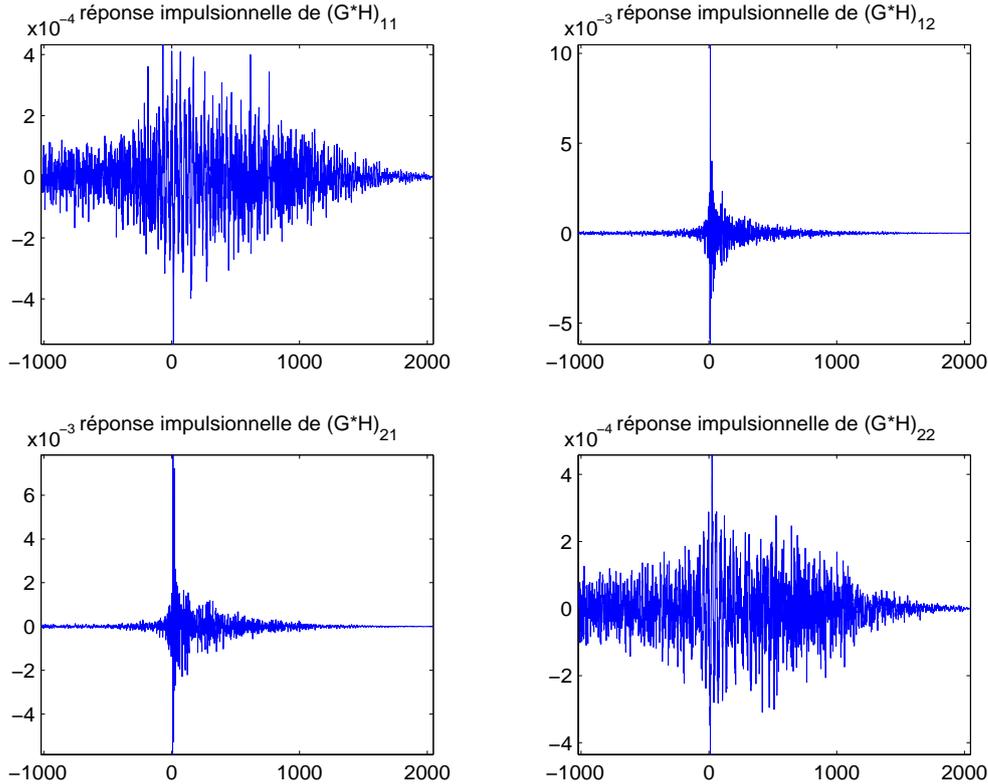


FIG. 4.31 – réponse impulsionnelle du filtre global $(G * H)(n)$ (méthode de la SFD).

Nous pouvons aussi voir le résultat de séparation de ces deux sources sur la figure 4.34 qui représente les signaux des sources originales, des mélanges générés et des sources séparées (méthode de la SFD).

Sur les figures des profils (blocs 11, 21 et 31), nous voyons également la différence entre les deux méthodes d'estimation. Nous remarquons que la méthode basée sur la SFD estime mieux les profils. En effet, sur la figure 4.33 (méthode de la SFD) les profils des deux sources sont plus différents que ceux estimés par la méthode de la moyenne (figure 4.32) et en conséquence, selon notre hypothèse sur les profils, les premiers caractérisent bien les sources et permettent à l'algorithme de détecter les permutations, à l'opposé des deuxièmes qui ne sont pas bien séparés à cause de leur mauvaise estimation.

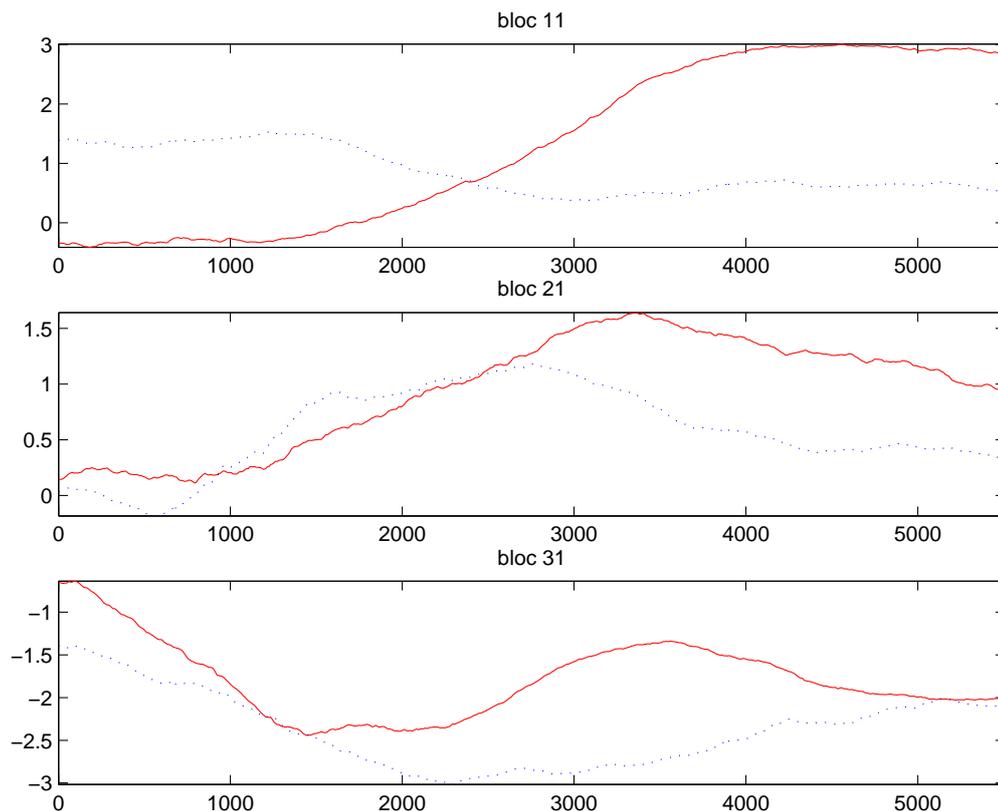


FIG. 4.32 – *profils de quelques blocs (méthode de la moyenne locale); source 1 : en trait continu, source 2 : en pointillé.*

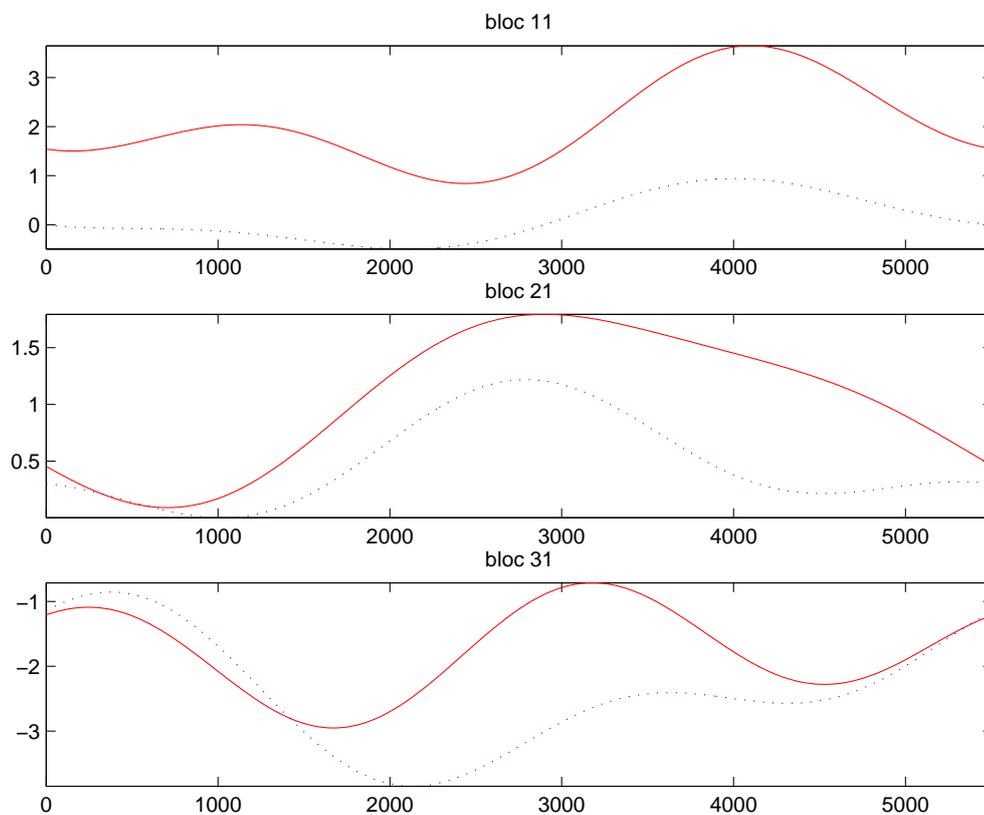


FIG. 4.33 – *profils de quelques blocs (méthode de la SFD); source 1 : en trait continu, source 2 : en pointillé.*

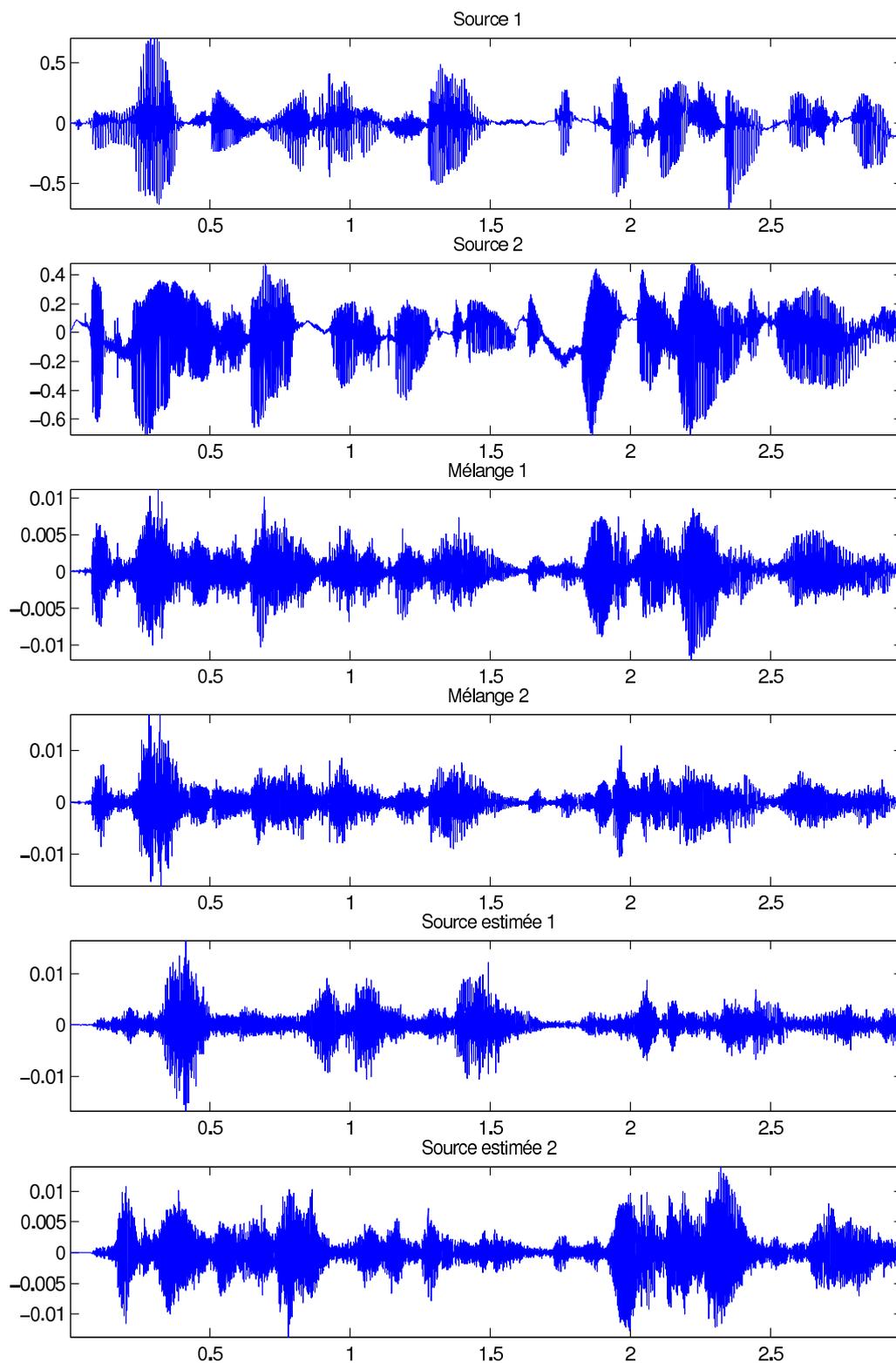


FIG. 4.34 – sources originales, mélanges et sources estimées (méthode de la SFD).

4.3.1.3.4 Discussion

Les simulations précédentes montrent l'amélioration apportée à notre algorithme par les deux nouvelles méthodes d'estimation, en particulier par la méthode qui utilise la

SFD.

Avec ces nouvelles méthodes, nous arrivons à mieux séparer les sources dans les contextes acoustiques réels qui sont caractérisés par des réponses impulsionnelles très longues et par une grande réverbération. L'algorithme arrive à supprimer presque toutes les permutations et fournit ainsi un résultat dont l'erreur d'ambiguïté de permutation est faible.

4.3.2 Comparaison avec d'autres méthodes

Nous présentons ici une comparaison entre notre algorithme (cas non bruité) et deux autres algorithmes pour la séparation de mélanges convolutifs⁹, l'algorithme de Murata, Ikeda et Ziehe [81] et l'algorithme de Parra et Spence [85, 84]. Pour cela, nous utilisons leurs codes Matlab qui sont disponibles sur le Web (voir [77, 86]).

La méthode de Murata, Ikeda et Ziehe est basée sur une approche temps-fréquence où ils appliquent la décorrélation temporelle au long des bandes fréquentielles de Fourier. Ils exploitent en réalité le contenu des signaux audio afin d'accomplir leur séparation aveugle. Quant à l'ambiguïté de permutation due à la procédure de séparation fréquence par fréquence, les auteurs fondent leur solution sur la maximisation de la corrélation entre les enveloppes des spectres, fréquence par fréquence (voir le détail de l'algorithme dans le chapitre 2).

Parra et Spence proposent de séparer les sources en diagonalisant une matrice de covariance. La procédure est effectuée dans le domaine de Fourier à chaque fréquence, où ils calculent un périodogramme par une TF glissante. Leur critère de diagonalisation est adaptatif et fondée sur le minimisation d'un critère. Le problème de permutation dans cette méthode est résolu en imposant une contrainte de continuité ("lissage") fréquentiel (voir chapitre 2).

4.3.2.1 Expériences et mesure de performance

Nous nous plaçons dans le cas de 2 sources et 2 capteurs. Nous utilisons les signaux réels de l'Université McMaster [103] pour lesquels nous simulons différents chemins de propagation. Les sources se propagent de leurs origines jusqu'aux capteurs de mesures en suivant chacune un chemin. Chaque chemin est défini par une combinaison composée d'un angle, une distance et une hauteur (voir la section 4.1.2). Nos expériences ont été réalisées avec les combinaisons dont l'angle est fixé à : 0° , 45° , 90° , 270° et 315° . Ceci correspond finalement à 20 cas de séparations possibles. Tous les signaux utilisés sont échantillonnés à 11,025kHz, avec une durée de 18048 échantillons. Les réponses impulsionnelles de la matrice de mélange sont tronquées à 1024 retards. La FFT est réalisée sur 512 points.

⁹une étude comparative plus exhaustive a été aussi réalisée pour le projet BLISS (voir le rapport technique [57])

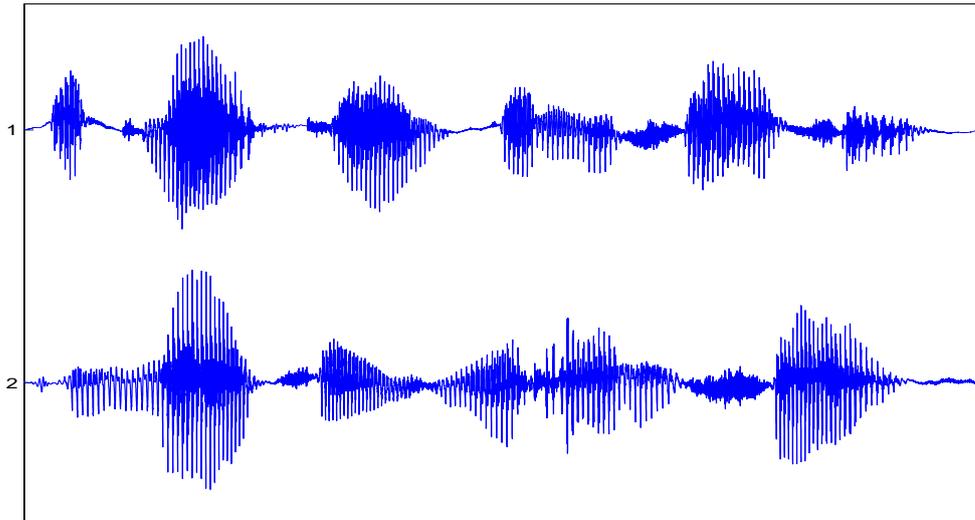


FIG. 4.35 – sources audio utilisées.

Nous faisons tourner les trois algorithmes avec les mêmes entrées et en sortie nous récupérons les données à comparer. Nous procédons ainsi à l'analyse de leur performance par le calcul de l'indice (4.10). Cette mesure de performance a été développée spécialement pour le cas de séparation aveugle de mélanges convolutifs ; l'analyse de la qualité de séparation dans cet indice est portée sur le système global [57].

4.3.2.2 Évaluation des résultats

Avant d'exposer les résultats numériques de cette comparaison, nous commençons d'abord par expliquer le principe de notre mesure de performance. Le système (filtre) global, matrice résultant du produit entre la matrice de mélange et la matrice de séparation, est à la base de cette mesure. En effet, celui-ci nous permet d'analyser la qualité de séparation obtenue par chaque méthode par rapport à la bonne séparation de mélanges convolutifs, qui est définie par une matrice diagonale pour le filtre global et sans qu'il ait d'ambiguïtés dans la séparation. Sachant que les ambiguïtés de séparation possibles sont :

- i l'ambiguïté de permutation globale (*i.e.* système global sous forme de matrice anti-diagonale ou inter-changement des lignes du filtre de séparation) ;
- ii l'ambiguïté d'échelle globale (*i.e.* lignes du système global ou du filtre de séparation multipliées par des constantes) ;
- iii l'ambiguïté de filtrage de permutation et/ou d'échelle (*i.e.* application d'un certain filtrage aux signaux séparés) ;

l'objectif reviendrait à mesurer la performance avec un indice inchangé par rapport à ces ambiguïtés.

Pour cette fin, un indice efficace est développé pour pouvoir répartir le facteur d'échelle (ambiguïté ii), circonvenir l'ambiguïté iii et mesurer enfin la similitude avec

une matrice de permutation globale (ambiguïté **i**). Cet indice est appelé "*Indice d'Amari pour les mélanges convolutifs*". Il s'applique sur le système global $C = G * H$ comme suit :

- a) cette étape traite l'ambiguïté d'échelle. Nous normalisons les lignes de C en empilant les filtres de chaque ligne dans un long vecteur que nous normalisons. Ensuite, nous remettons les éléments de ce vecteur dans leurs entrées originales. Par ceci, nous répartirons l'ambiguïté d'échelle.
- b) quant à cette étape, elle traite l'ambiguïté de filtrage. Nous calculons la norme de chaque filtre dans C et nous la stockons dans une nouvelle matrice \tilde{C} . Cela en fait résume les proportions globales des contributions entre les sources originales et les sources séparées. Ainsi, la construction de cette matrice permet de savoir que la contribution d'une source est nulle ou très faible quand l'élément correspondant dans cette matrice est faible ou nul.
- c) la dernière étape calcule le degré de séparation par l'indice d'Amari (voir l'Annexe **E**) qui mesure la similitude à une matrice de permutation :

$$\sum_i \left(\sum_j \frac{|\tilde{C}_{ij}|}{\max_k |\tilde{C}_{ik}|} - 1 \right) + \sum_j \left(\sum_i \frac{|\tilde{C}_{ij}|}{\max_k |\tilde{C}_{kj}|} - 1 \right) \quad (4.10)$$

Les résultats numériques de comparaison, obtenus par l'application des trois algorithmes sur les mélanges convolutifs des deux sources audio de la figure 4.35, sont représentés par des graphiques en niveau de gris. Chaque graphique est divisé en rectangles correspondant chacun à un chemin de propagation défini par deux angles. Le niveau de gris de chaque rectangle indique la valeur de l'indice d'Amari pour les mélanges convolutifs.

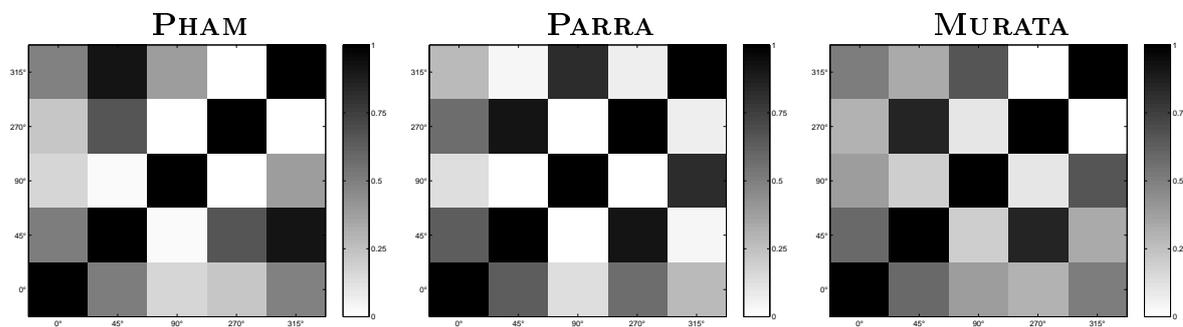


FIG. 4.36 – Indices d'Amari.

L'analyse de la performance des trois algorithmes à partir des précédents graphiques révèle que notre algorithme produit la solution la plus efficace en terme d'ambiguïté de séparation et fournit la plus bonne performance par rapport au reste des algorithmes. En effet, sur l'ensemble des expériences réalisées, la meilleure valeur de l'indice a été obtenue

par notre algorithme dans 64% des cas¹⁰, par l'algorithme de Parra dans 16% des cas et n'a jamais été obtenue par l'algorithme de Murata. Ceci est confirmé par l'intensité du niveau de gris sur les trois graphiques.

Par ailleurs, nous pouvons aussi constater que la séparation est de meilleure qualité lorsque les sources arrivent aux capteurs, de directions (angles) très différentes (de différents cotés), ce qui permet de ne pas mélanger les sources trop fortement. Plus les chemins des sources se rapprochent, plus les signaux qui arrivent aux capteurs sont mélangés et ainsi leur séparation est plus difficile.

4.3.3 Simulations dans le cas bruité

Nous nous plaçons dans le contexte difficile où les signaux capteurs sont complètement noyées dans le bruit. En effet, les simulations ont été effectuées avec des SNR qui vont de -1dB à -40dB. Nous allons présenter nos premiers résultats où l'on considère simplement la reconstruction d'une seule source à partir d'un ensemble d'observations. Ceci permet de reporter le problème d'ambiguïté à plus tard ! Par ailleurs, ce cas de figure peut être intéressant comme application de la séparation de sources car, par exemple, chez un malentendant, l'intérêt est toujours porté sur une seule source parmi tous les signaux de l'environnement, en particulier dans un contexte très bruité.

4.3.3.1 Le bruit

Au vu des signaux capteurs qui portent l'information observée, le bruit dont le caractère est aléatoire, est un signal indésirable qui réduit l'intelligibilité et la perception de l'information. En effet, la tâche de séparation de source devient significativement plus difficile en considérant la présence du bruit dans le modèle de séparation, d'ailleurs ce cas n'a pas reçu suffisamment d'attention dans la littérature [29].

Par ailleurs, selon son spectre¹¹, on distingue différents types de bruits. Dans le cadre de la simulation de notre algorithme de séparation de mélanges convolutifs de sources, nous avons procédé (par commodité) avec deux bruits faciles à générer :

4.3.3.1.1 Blanc^{12,13} Gaussien

Sa propriété remarquable est sa définition complète par ses caractéristiques statistiques du premier et deuxième ordre. Sa densité de puissance est donnée par le moment d'ordre deux ; de plus sa valeur dans une gamme de fréquences donnée ne dépend pas de

¹⁰sans compter les cas dont les deux sources se propagent à partir du même angle. Ce cas n'est pas résolvable par l'ICA.

¹¹par exemple, un bruit dont le spectre est limité du côté des hautes fréquences est appelé "*bruit rose*".

¹²cette appellation provient d'une analogie entre la composition spectrale du bruit en proportions égales à toutes les fréquences et la lumière blanche qui comprend toutes les couleurs.

¹³les bruits blancs diffèrent entre eux par leurs statistiques supérieures (à deux).

la place de cette bande de fréquence sur toute l'étendue possible du spectre. Sa fonction de corrélation est une impulsion de Dirac centrée.

4.3.3.1.2 Autorégressif Gaussien

Appelé aussi "bruit rouge"¹⁴, basé sur un modèle statistique de filtrage défini par :

$$B(n) = a_1B(n-1) + a_2B(n-2) + \dots + a_MB(n-M) + b(n) \quad (4.11)$$

où les a_i ($i = 1, \dots, M$) sont des paramètres de réglage et $b(n)$ est un bruit blanc Gaussien.

Par cette modélisation nous disposons d'une bonne manière de générer des bruits de différentes densités spectrales et fonctions d'autocorrélation. Nous optons dans nos simulations à un modèle AR du premier ordre, la formulation mathématique du bruit autorégressif de bruit gaussien est ainsi donnée par :

$$B(n) = aB(n-1) + b(n) \quad (4.12)$$

où a est le paramètre de coloration du bruit. Il est pris égal à 0.5 ($0 < a < 1$ pour un bruit rouge). L'avantage de ce bruit est qu'il permet de défavoriser les fréquences élevées de manière contrôlée.

4.3.3.2 Réponse impulsionnelle à 512 retards

La configuration de cette simulation se présente avec une fréquence d'échantillonnage des signaux sources et des réponses impulsionnelles égales à 11,025kHz. Les réponses impulsionnelles utilisées sont celles de l'université McMaster avec une longueur de 512 retards et la source reconstruite est un signal de parole d'une durée égale à 1,64s :

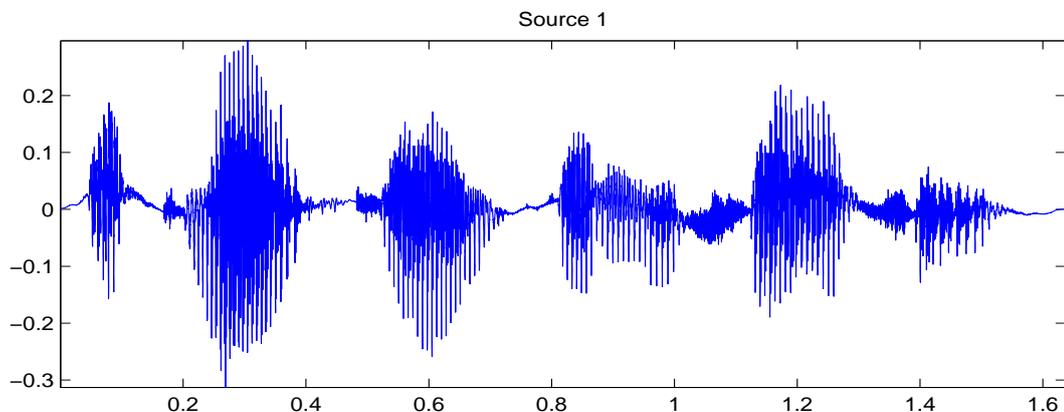


FIG. 4.37 – source originale.

¹⁴Par analogie avec la lumière, si on filtre les longueurs d'onde basses, c'est-à-dire les hautes fréquences, on enlève du bleu et du vert et la lumière prend une coloration rouge.

Avec la même méthode que celle utilisée pour le cas non bruité (méthode des profils), nous procédons à l'estimation des matrices spectrales sur des blocs de taille 2048, chevauchés à 75% et moyennés sur 5 points (blocs). Le vecteur de mélange est composé de quatre signaux capteurs fait avec des SNR qui valent respectivement -1dB, -2dB, -3dB, -4dB ; Les bruit utilisé dans cette simulation est blanc.

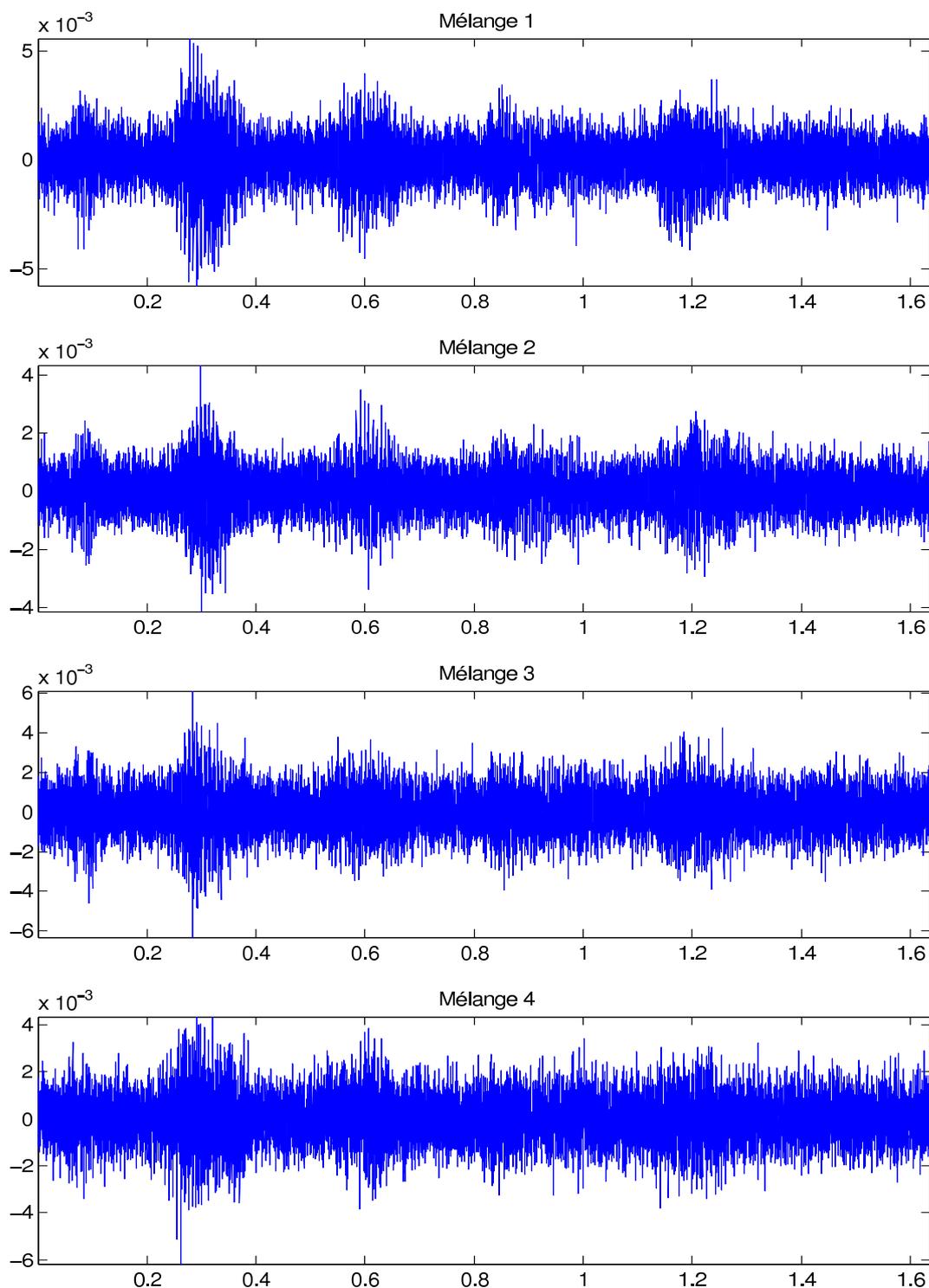


FIG. 4.38 – vecteur mélange.

Ces observations sont obtenues à la sortie des filtres de mélange, qui caractérisent

l'environnement acoustique réel dans lequel la source s'est propagée jusqu'aux capteurs. En effet, la réverbération est toujours présente par des pics d'échos sur les signaux des réponses impulsionnelles, représentées sur la figure suivante :

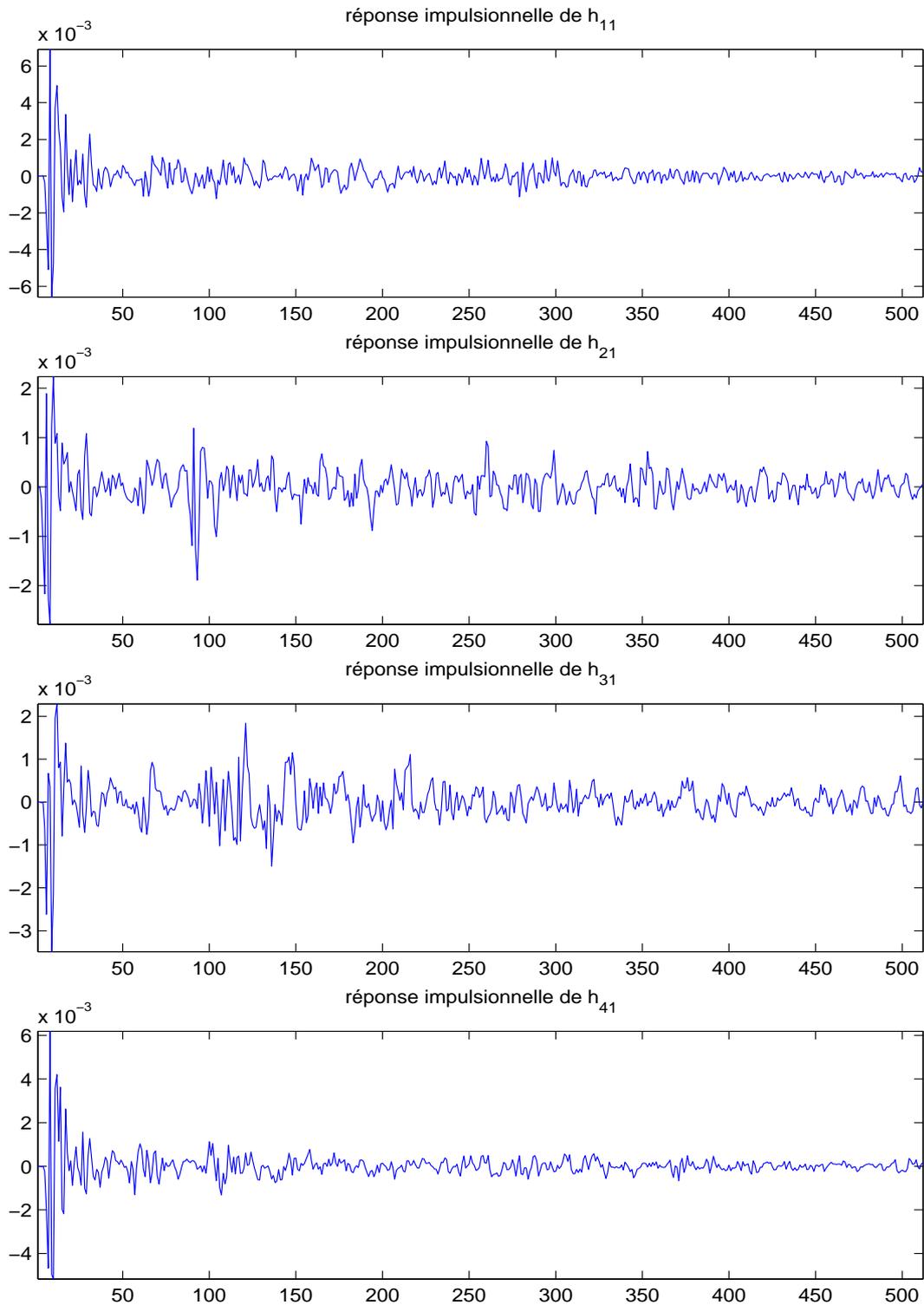


FIG. 4.39 – *filtre de mélange.*

À la figure 4.40 nous obtenons la source séparée par l'application de notre algorithme de séparation fondé sur la technique EM. Nous voyons que l'algorithme essaie de tracer

l'allure ("les bouffées") de la source originale tout en diminuant le rapport signal à bruit.

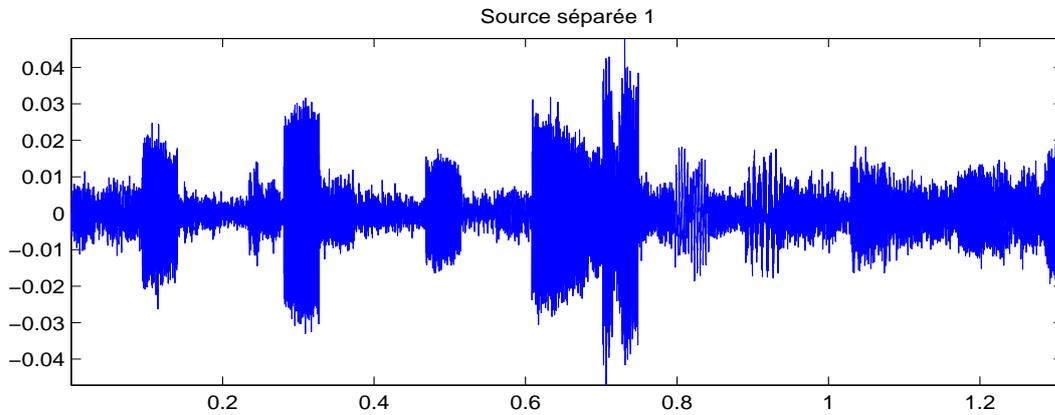


FIG. 4.40 – source séparée.

Le rapport signal à bruit évaluant la séparation est présenté sur la figure 4.41. Il révèle la faible affectation de la source séparée par le bruit dans quasiment tout le plan temps-fréquence.

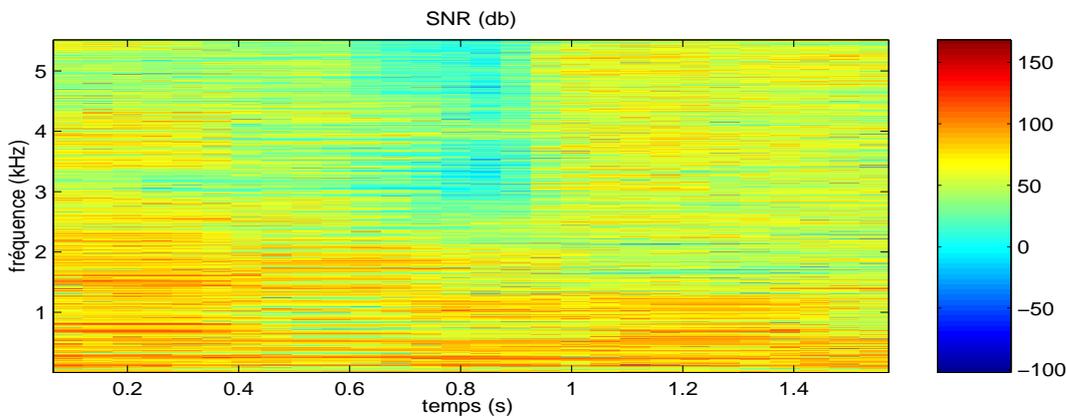


FIG. 4.41 – rapport signal à bruit.

Sur la figure 4.42 nous avons calculé l'écart quadratique entre la partie "source" $H(f)S_s(n, f)H^*(f)$ du spectre (éléments de la diagonale de la matrice spectrale) des observations¹⁵ et l'estimée (par l'algorithme EM) de cette même partie spectrale : $A(f)D(n, f)A^*(f)$. Nous constatons que la distance entre ces deux quantités est globalement faible, elle n'a que quelques valeurs significatives dans certains points du plan temps-fréquence. Ceci pourra faire un deuxième argument de validation des résultats de séparation de l'algorithme EM.

¹⁵le spectre des signaux capteurs est composé de deux parties : une qui provient du signal source et une qui provient du signal bruit (voir l'équation (3.29)).

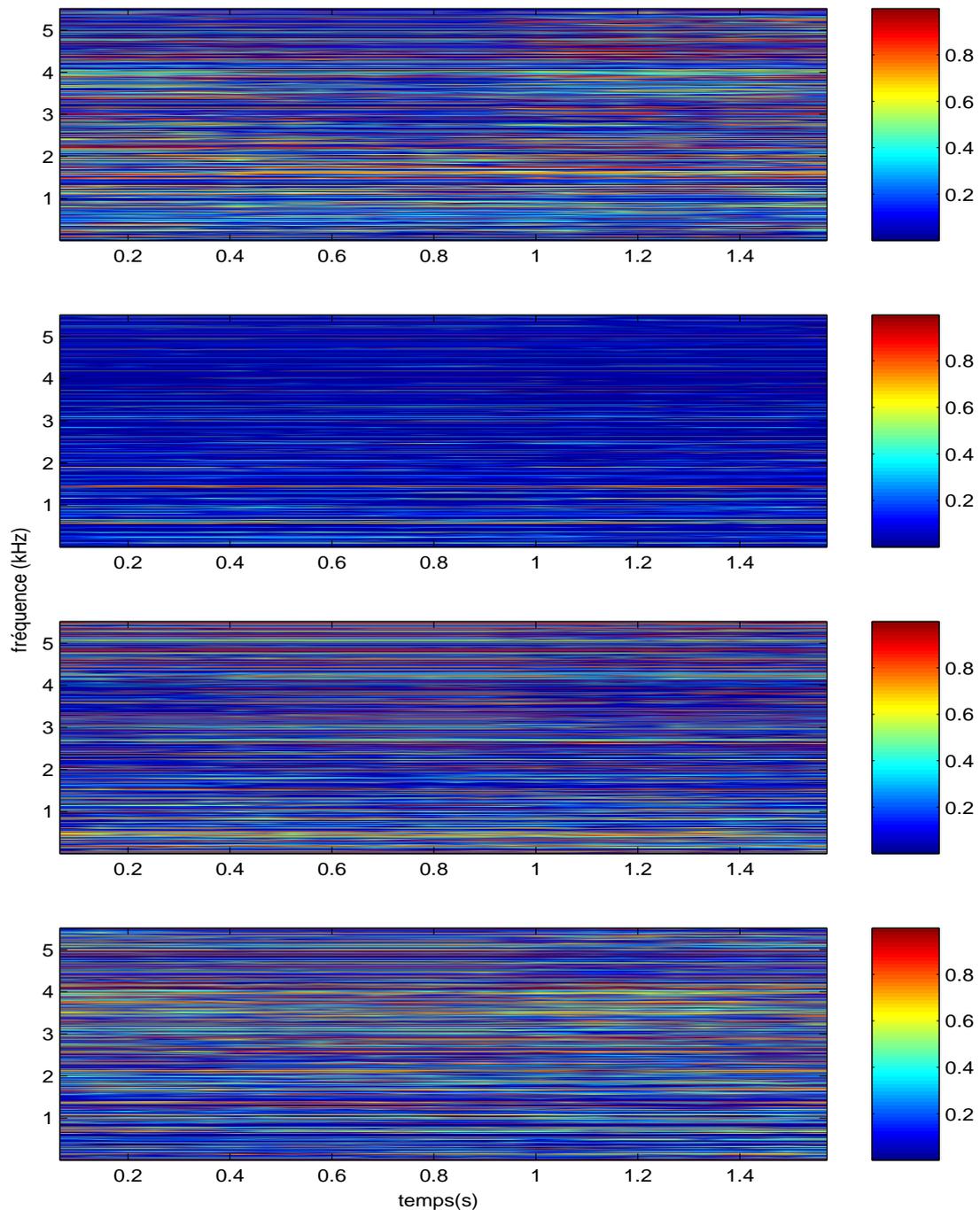


FIG. 4.42 – *distance d'évaluation (4 mélanges).*

4.3.3.3 Réponse impulsionnelle à 1024 retards

Cette simulation est effectuée avec un configuration comparable à la précédente, sauf que nous essayons à présent de voir le comportement de l'algorithme dans une situation plus complexe de point de vue de l'intensité du bruit et de la longueur de la réponse impulsionnelle de mélange. La source utilisée est de longueur égale à 1,64s (figure 4.45) ; la réponse impulsionnelle est de 1024 retards. Nous procédons de la même façon pour l'estimation du spectre-interspectre des observations (blocs de 2048 échantillons, 75% de chevauchement et 5 points pour le lissage). Tous les signaux ont une fréquence d'échan-

tillonnage égale à 11,025kHz. Les niveaux de bruit aux capteurs sont respectivement de -11dB, -15dB, -17dB, -21dB, -25dB, -27dB, -31dB, -35dB. Les observations sont tracées sur la figure 4.43.

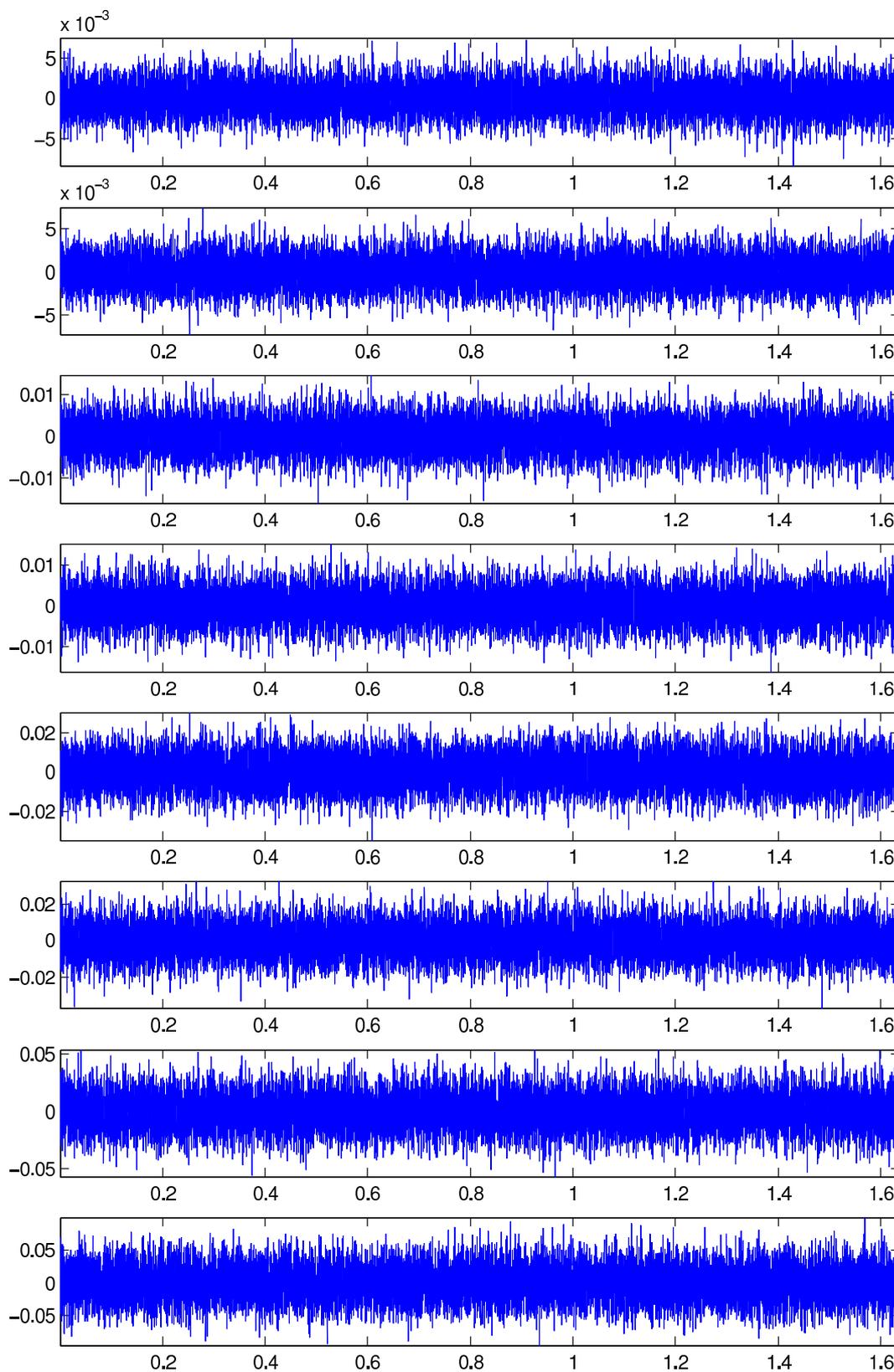


FIG. 4.43 – vecteur-mélange.

Le mélange est réalisé par les réponses impulsionnelles de McMaster, nous avons utilisé les positions qui sont les plus cohérentes (45° , 22° , 315° et 338°).

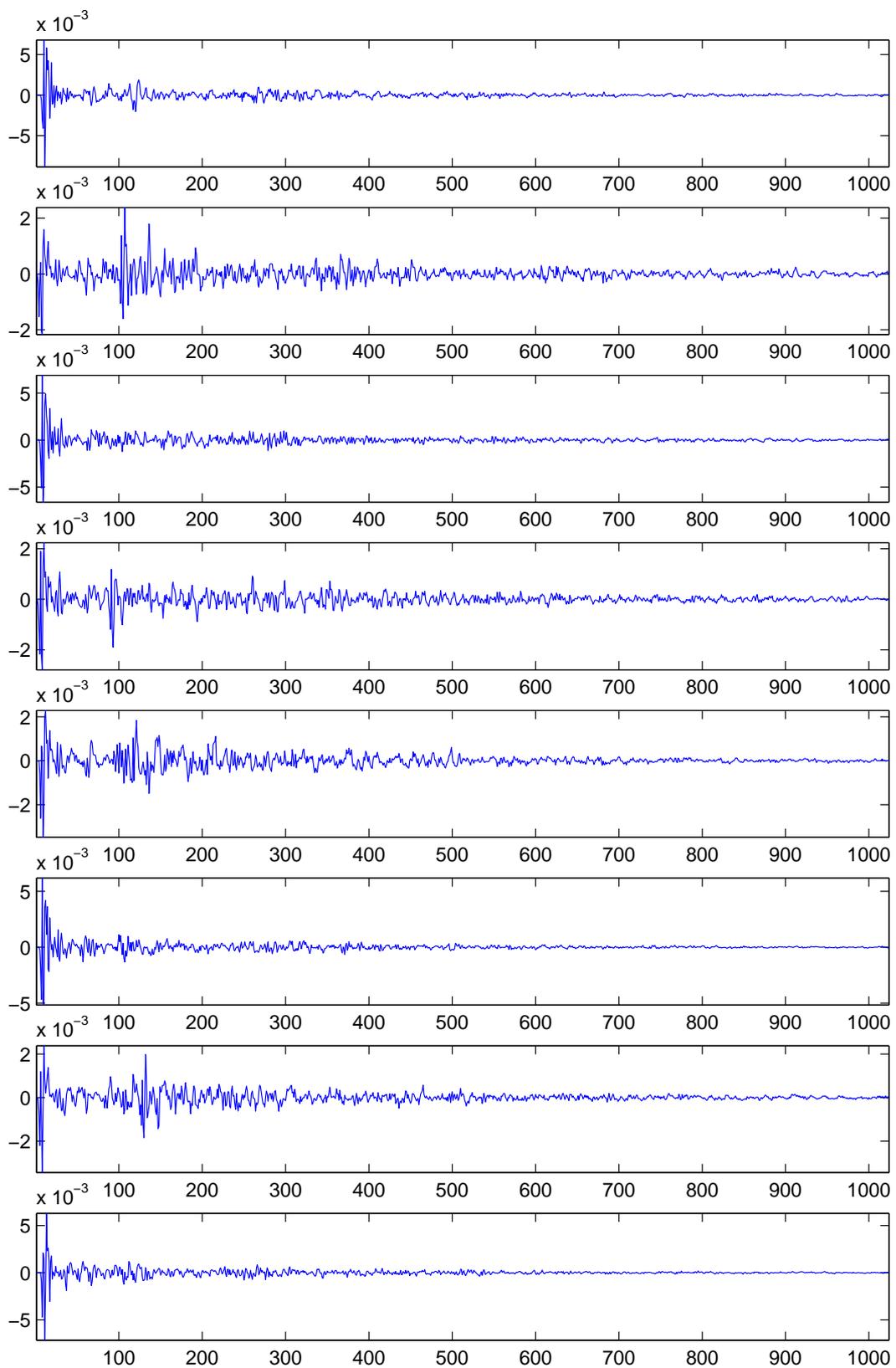


FIG. 4.44 – filtre de mélange.

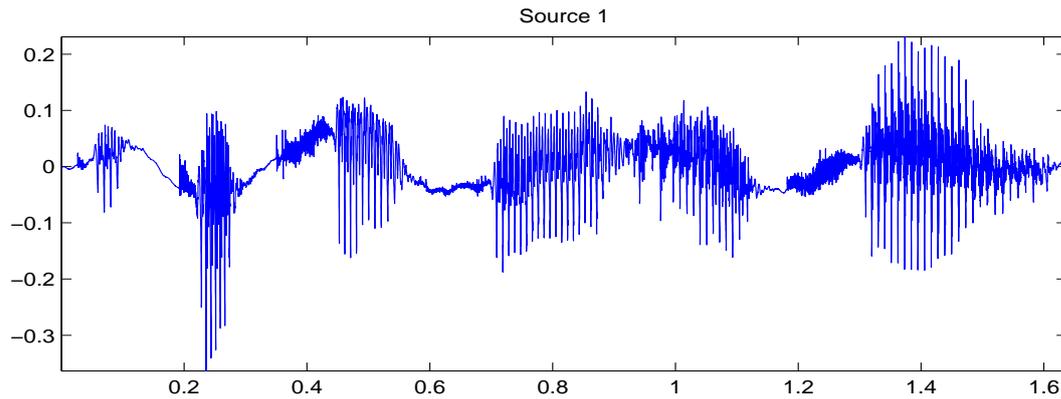


FIG. 4.45 – *source originale.*

Sur la figure suivante nous présentons le signal de la source issue de l'application de notre algorithme de séparation sur les observations précédentes :

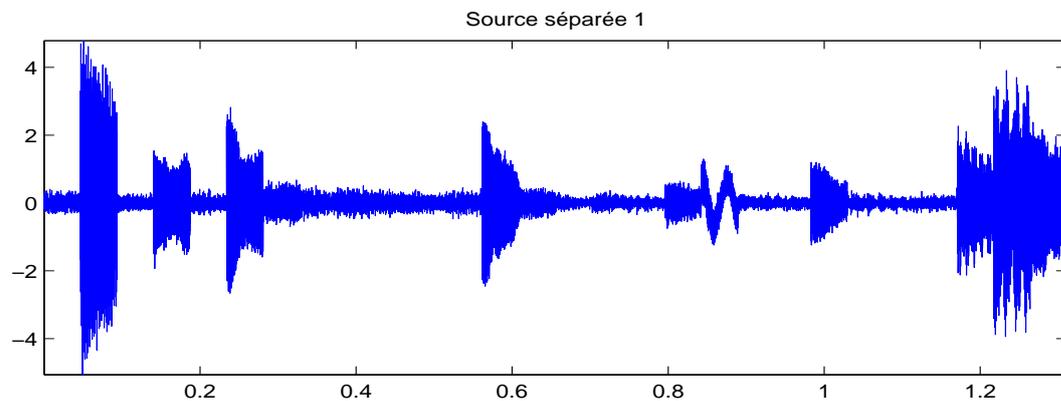


FIG. 4.46 – *source séparée.*

La performance est évaluée également pour cette configuration avec le SNR qui montre la faible présence du bruit dans la source séparée et la distance d (4.9) qui, quant à elle, montre une faible valeur de cette distance sauf à certains points du plan temps-fréquence. Elle est donnée par les graphiques ci-après :

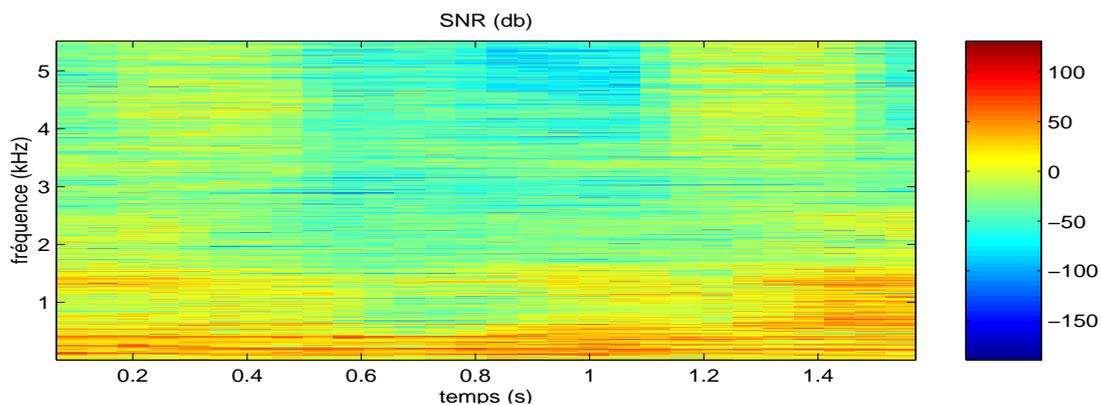


FIG. 4.47 – *rapport signal à bruit.*

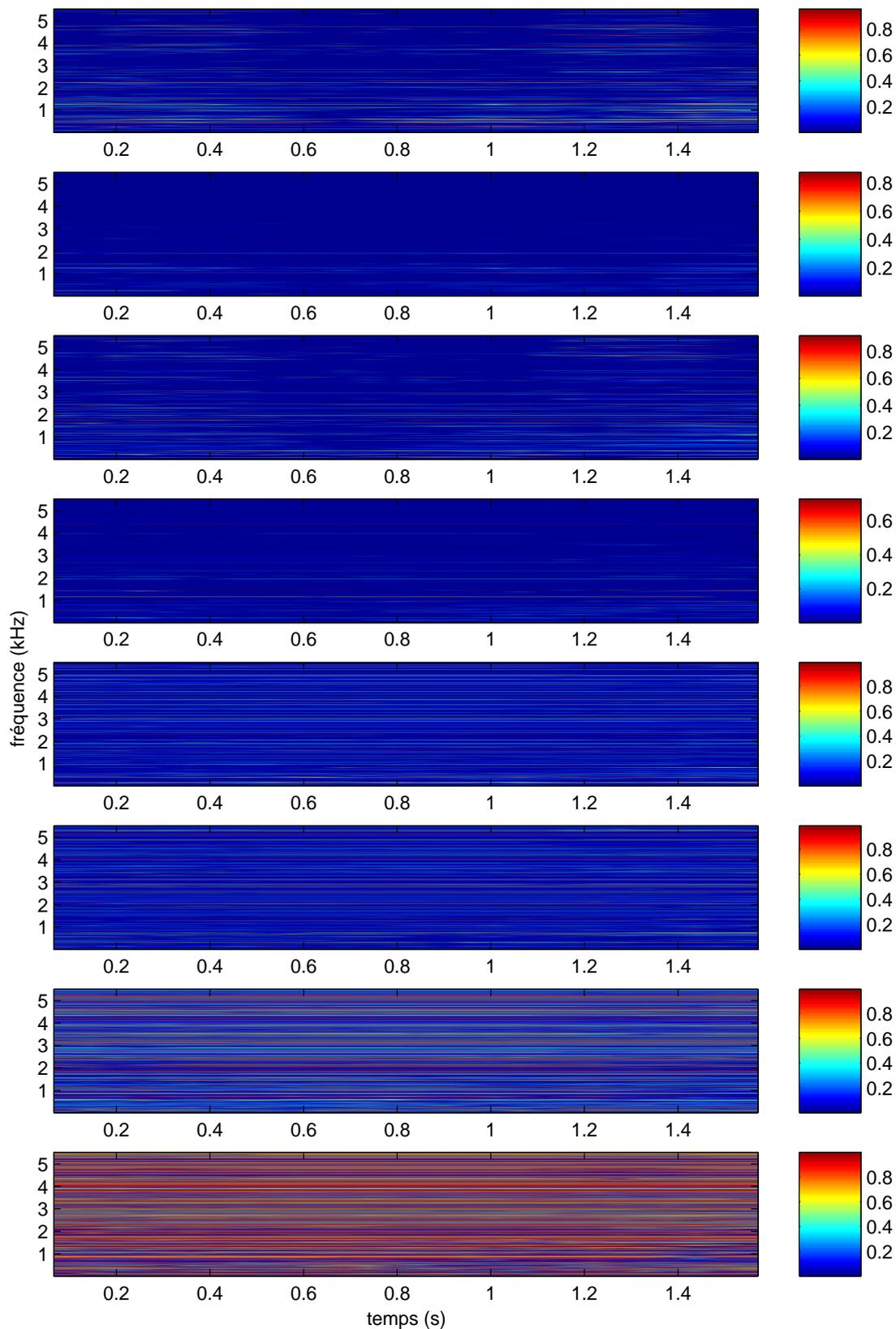


FIG. 4.48 – distance d'évaluation.

Par ailleurs, nous avons expérimenté le comportement de l'algorithme EM en présence du bruit autorégressif. Les résultats obtenus seront donnés dans la suite :

Cette simulation est réalisée avec le même paramétrage : réponse impulsionnelle de 1024 retards, blocs de 2048 échantillons, 75% de chevauchement et 5 blocs dans le moyennage. Les signaux ont une fréquence d'échantillonnage égale à 11,025kHz. Cependant, nous avons utilisé cette fois deux observations et des SNR de 3dB et 5dB. La source originale est celle présentée sur la figure suivante :

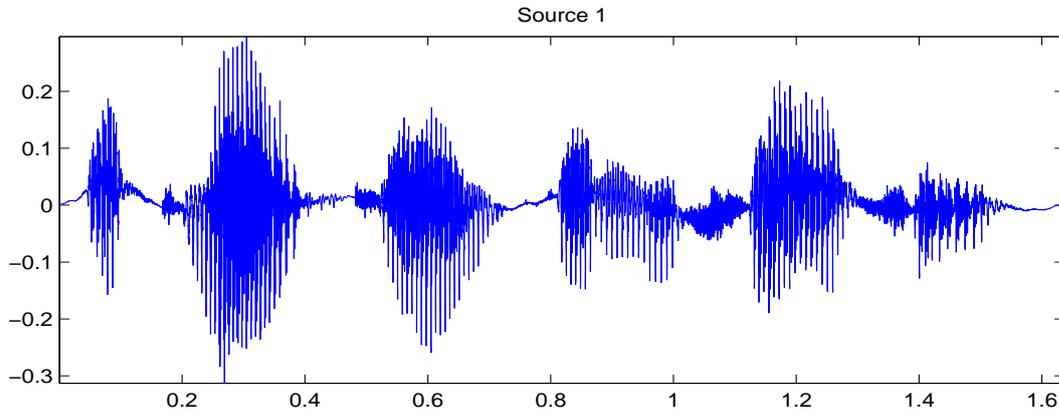


FIG. 4.49 – *source originale.*

Le filtre de mélange et les signaux observations sont tracés sur les deux figures suivantes :

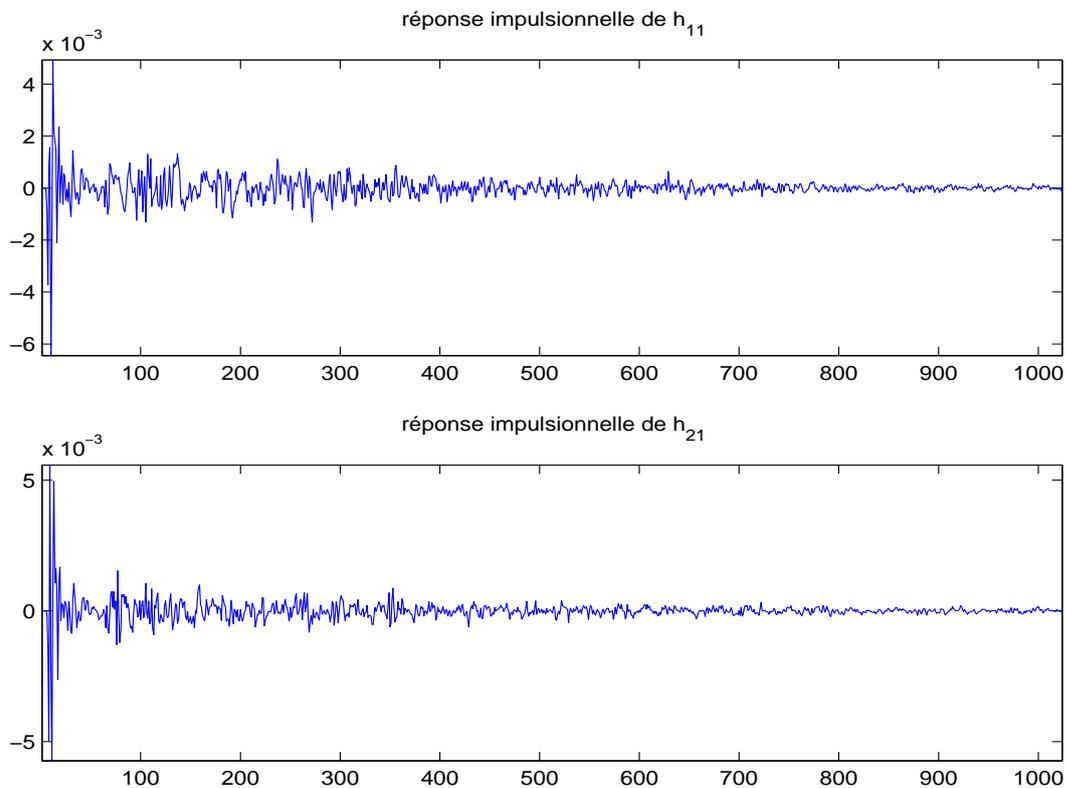


FIG. 4.50 – *filtre de mélange.*

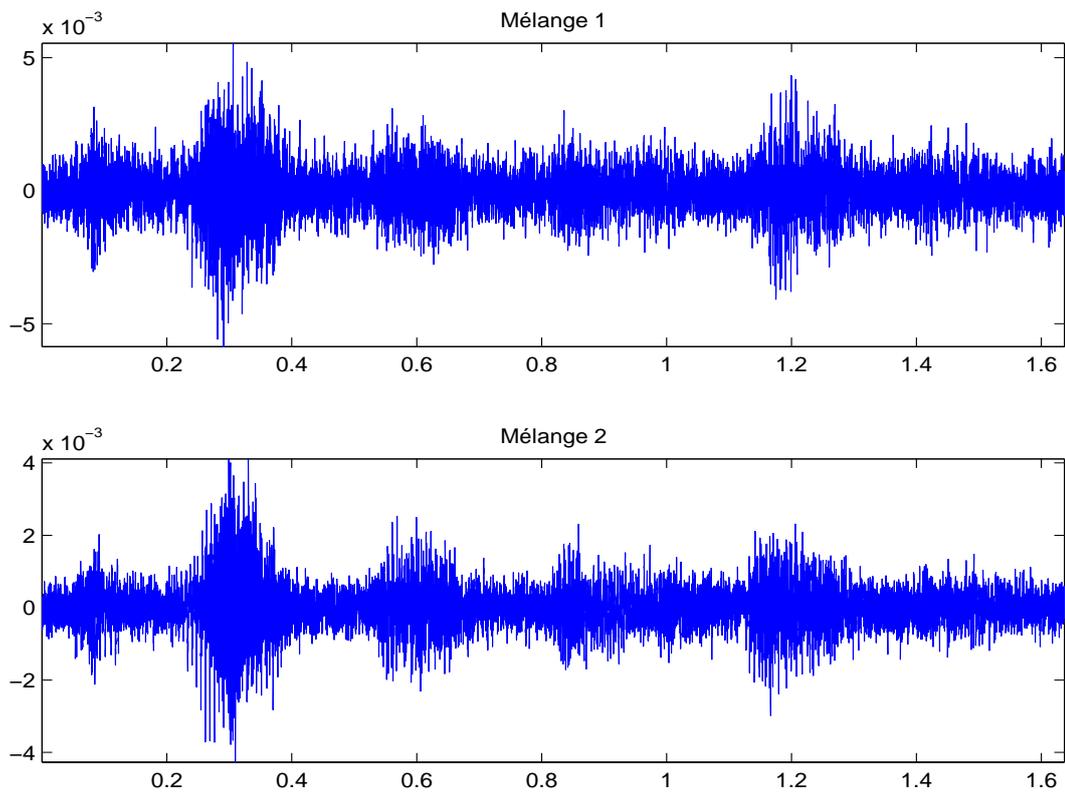


FIG. 4.51 – observations.

La performance de séparation calculée par les deux indices est illustrée par les deux graphiques suivants, qui indiquent que l’algorithme EM fournit des résultats comparables pour les deux types du bruit blanc ou autorégressif.

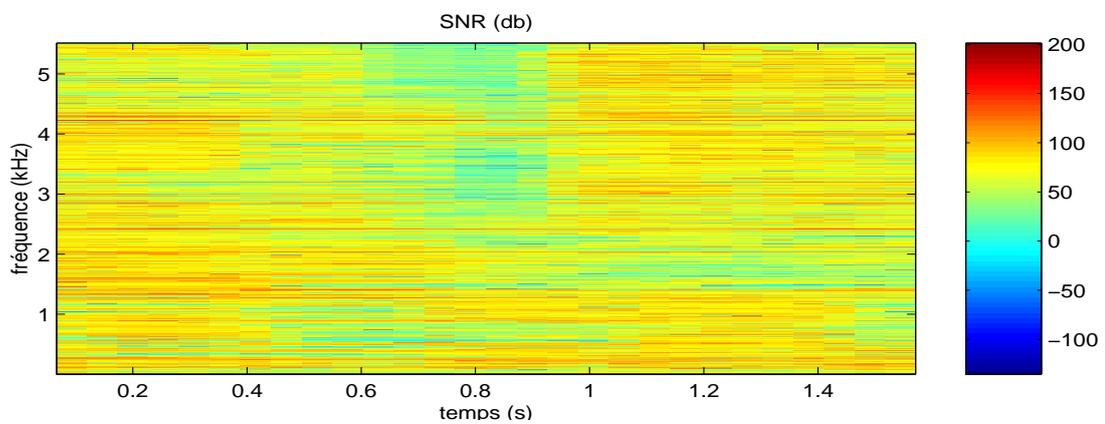


FIG. 4.52 – rapport signal à bruit.

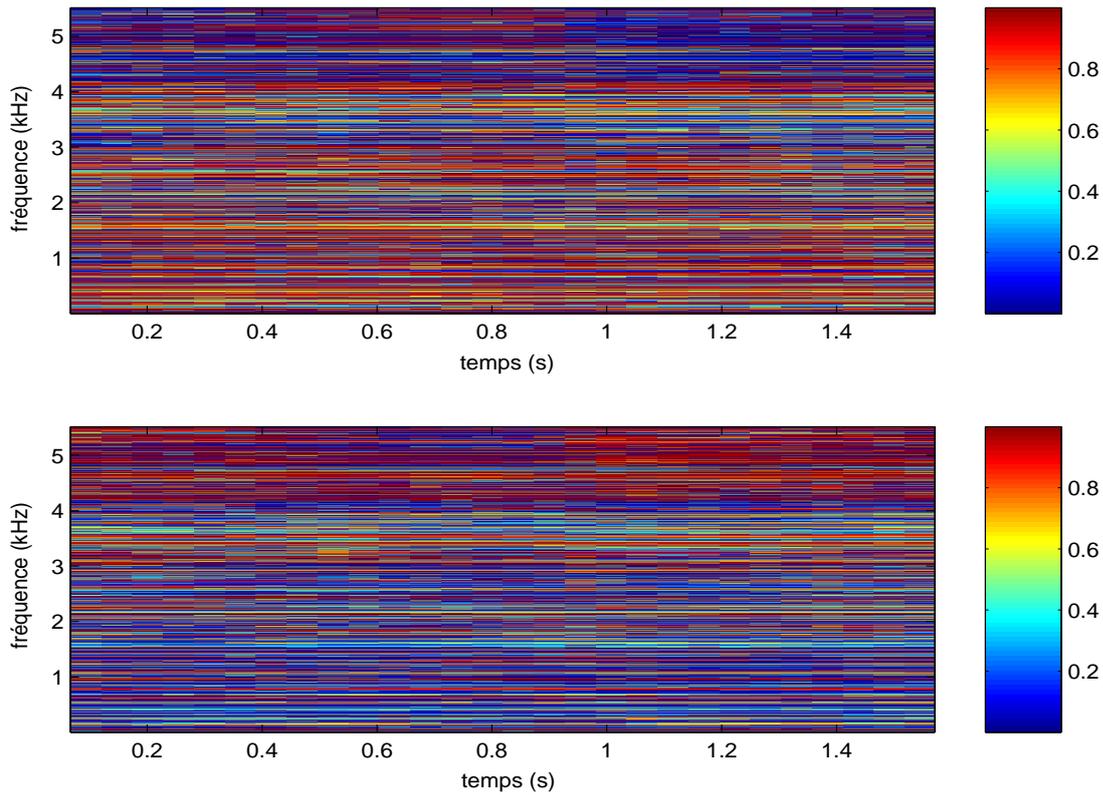


FIG. 4.53 – *distance d'évaluation.*

Enfin, la source séparée dans cette expérience est donnée dans la figure suivante :

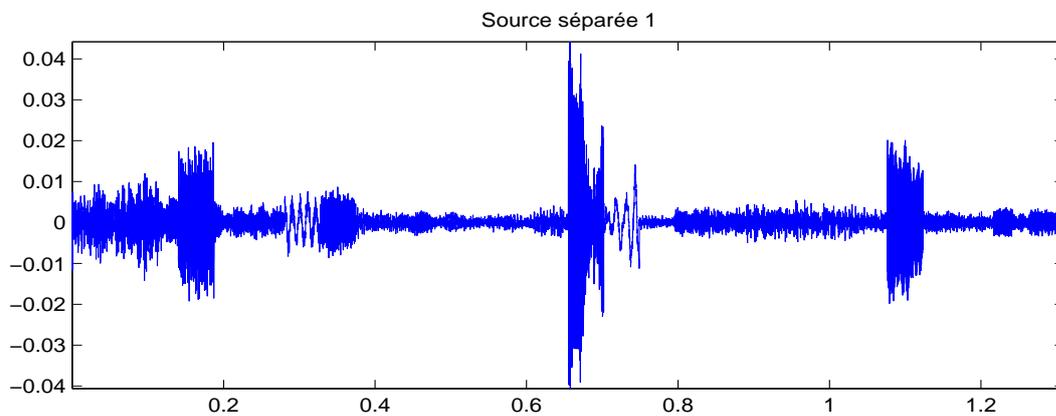


FIG. 4.54 – *source séparée.*

4.3.3.4 Discussion

À l'égard des résultats obtenus et des indices de performance, il apparaît que l'algorithme EM parvient relativement à reconstruire l'allure de la source noyée dans le bruit même pour de faible rapport signal à bruit. Cependant, l'amélioration de la qualité de la source extraite est envisageable, en jouant sur l'initialisation de l'algorithme EM. En effet, afin d'étudier le comportement et la performance de l'algorithme nous avons pro-

cedé à un test avec les vraies valeurs¹⁶, et là, l'algorithme a produit un excellent résultat. La source extraite qui ne diffère que peu de la vraie source est représentée en dessous de l'originale sur la figure suivante :

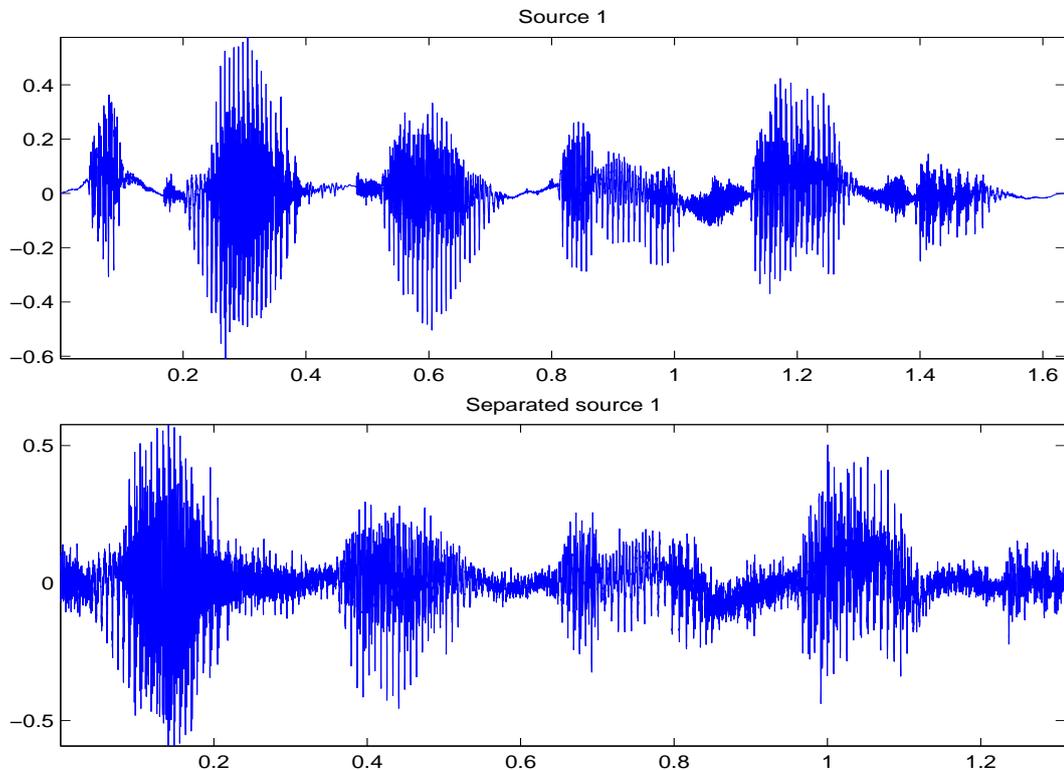


FIG. 4.55 – *séparation avec l'initialisation idéale ;
en haut : source originale, en bas : source estimée.*

4.4 Conclusion

Dans ce chapitre, nous avons présenté les résultats des simulations numériques effectuées dans le but d'étudier la performance de nos algorithmes dans des contextes acoustiques réels.

Dans le cas non bruité, la performance a été évaluée par l'ambiguïté de permutation de l'algorithme. Avec la contrainte de continuité de la réponse fréquentielle, renforcée par la technique des profils des sources, estimés par l'utilisation de la SFD, l'algorithme présente une bonne performance et une faible ambiguïté de séparation. À ce stade, les résultats obtenus sont encourageant et suggèrent d'intégrer l'algorithme dans un système d'aide aux malentendants.

En plus des avantages de l'algorithme de diagonalisation conjointe qui est à la base de notre approche de séparation des mélanges convolutifs non bruités de sources, la procé-

¹⁶en pratique, nous n'avons pas accès à ces données pour une séparation aveugle. Cette expérience a pour objectif le test et l'évaluation.

ture d'élimination des permutations en utilisant la SFD donne aussi à notre algorithme l'avantage du paramétrage facile et simple.

Dans le cas bruité, l'évaluation est calculée premièrement par le rapport signal à bruit donnant ainsi le degré d'affectation de la source séparée et deuxièmement par la distance entre le spectre des observations (partie provenant du signal source) et son estimée, ce qui donne l'écart entre l'ensemble des vrais paramètres le calculant et leurs estimées par l'algorithme EM. Malgré le contexte de séparation difficile dans ce cas, qui est dû à la présence du bruit et aux pics de réverbération présents dans la réponse impulsionnelle de mélange, en plus de la difficulté intrinsèque de la tâche de séparation car même si nous connaissons le filtre de mélange¹⁷ son inversion n'est pas évidente, l'algorithme EM retrouve l'allure de la source se trouvant dans les signaux mélanges bruités.

Nous pensons, en conséquence que des améliorations peuvent être faites sur l'initialisation de l'algorithme EM, ainsi que sur sa convergence, et sur l'arrêt de l'algorithme sur un des minima locaux.

¹⁷ceci même s'il n'y a pas de bruit.

Conclusion Générale

Cette thèse est focalisée sur l'étude de la séparation de sources dans le cas de mélanges convolutifs générés dans des environnements réels. En particulier, nous nous sommes intéressés aux cas des salles acoustiques réverbérantes.

Pour cela, nous avons montré qu'il est possible d'effectuer la séparation de signaux audio mélangés convolutivement en utilisant les statistiques d'ordre deux et en exploitant leur propriété de nonstationnarité. Sous l'hypothèse de non-stationnarité, l'algorithme développé pour le cas non bruité, rend possible la séparation des mélanges convolutifs dans le domaine de Fourier, fréquence par fréquence, en transformant le problème en un ensemble de problèmes instantanés. Il diagonalise conjointement, sous l'hypothèse d'indépendance mutuelle des sources, l'ensemble de matrices spectrales de même fréquence. Nous avons utilisé un critère basé sur l'information mutuelle gaussienne et le maximum de vraisemblance. Des travaux antérieurs ont montré par l'expérience les bonnes performances de ce critère. De plus, nous utilisons une implémentation très rapide de l'algorithme de diagonalisation conjointe.

Par ailleurs nous nous sommes focalisés sur la résolution de l'ambiguïté de permutation de la solution. Deux solutions ont été données. La première repose sur la continuité de la fonction de transfert des filtres de mélange et de séparation, qui cherche la permutation qui rend le rapport entre les matrices séparantes de deux fréquences successives le plus proche de l'identité. La deuxième exploite la propriété intrinsèque de distribution d'énergie des sources, elle construit un profil pour chaque source et elle les permute de sorte que les sources correspondent le plus possible à leurs profils. Deux variantes de cette deuxième solution ont été aussi proposées. La première estime les profils en faisant une moyenne locale par une fenêtre centrée à chaque fréquence, sa taille est un paramètre de réglage pour la qualité de la solution. Quant à la deuxième, elle estime ces profils en utilisant un lissage par Séries de Fourier Discrète, le lissage est paramétré en fonction du nombre de fréquences hautes supprimées de la décomposition effectuée. L'expérience a montré l'apport d'une amélioration significative à l'algorithme par ces deux variantes. La méthode basée sur la SFD, donne aussi à notre algorithme l'avantage d'un paramétrage facile, en plus des avantages de la procédure de diagonalisation conjointe qui est à la base de notre algorithme de séparation de mélanges convolutifs non bruités de sources.

Pour tenir compte de la présence du bruit dans l'environnement acoustique, nous avons introduit un algorithme itératif de séparation basé sur la maximisation d'un critère.

La démarche opère sur un ensemble de paramètres composé de la matrice spectrale du bruit, la matrice de mélange et la matrice spectrale des sources. Pour cela, nous avons construit un critère d'estimation basé sur le maximum de vraisemblance en partant du modèle de mélange. Pour implanter la démarche nous avons utilisé la technique EM (Expectation-Maximization) qui permet d'accomplir la tâche d'une manière simple et relativement rapide en deux étapes.

Dans ce cas, plusieurs expériences de séparation ont été simulées où nous avons utilisé différentes réponses impulsionnelles et deux types de bruit, gaussien blanc et autorégressif. Au vue de la performance et de la qualité de séparation, évaluées par l'indice du rapport signal à bruit et la distance quadratique, les premiers résultats obtenus montrent que l'algorithme développé fournit des résultats favorables malgré la complexité et la difficulté du contexte.

Nos algorithmes de séparation ont été appliqués à deux jeux de données acoustiques réels. La performance a été évaluée au terme de l'ambiguïté de permutation dans le cas non bruité et au terme du rapport signal à bruit en présence du bruit. Les simulations effectuées montrent que la réverbération des environnements acoustiques réels a un impact négatif sur la performance de séparation. Elle engendre, en effet, plusieurs copies des sources qui sont réfléchies en retard et la séparation est ainsi plus difficile, car il faut supprimer ces copies.

Plusieurs perspectives peuvent être envisagées :

- au cours de notre étude nous avons constaté que l'estimation des matrices spectrales peut influencer directement le résultat obtenu. Nous avons effectué des améliorations sur cette estimation, mais il serait très intéressant de faire une étude complète de l'estimation en fonction du biais, de la résolution fréquentielle et de la dégradation induite par l'estimation.
- exiger l'implantation et l'intégration de l'algorithme dans un système d'aide aux malentendants en temps réel et la prise en compte du mouvement des sources et des capteurs. Il est donc important d'effectuer une étude sur la mise en œuvre de l'algorithme sur DSP dans de telles situations.
- l'expérience a montré que le paramétrage de l'algorithme (cas non bruité) joue un rôle important dans l'obtention des bons résultats, nous suggérons aussi qu'une étude sur le paramétrage automatique de l'algorithme soit effectuée.
- une étude fine peut être menée pour améliorer les résultats de l'algorithme dans le cas bruité, notamment par l'initialisation de l'algorithme EM, l'étude de la convergence, la complexité et l'ambiguïté dans le cas général.

Bibliographie

- [1] K. Abed-Meraim, W. Qiu, and Y. Hua. Blind system identification. *In IEEE Processing*, 85, n° 8 :1310–1322, 1997.
- [2] S. Amari and A. Cichocki. A new learning algorithm for blind signal separation. *ANIPS. MIT Press, Cambridge MA*, pages 757–763, 1998.
- [3] S.-I. Amari. Adaptive blind signal processing-neural network approaches. *Proceeding of the IEEE*, 86 :2026–2048, 1998.
- [4] J. Anemüller and B. Kollmeier. Amplitude modulation decorrelation for convolutive blind source separation. *In Proceeding of ICA*, pages 215–220, June 2000.
- [5] M. Babaie-Zadeh. *On blind source separation in convolutive and nonlinear mixtures*. PhD thesis, Institut National Polytechnique de Grenoble, 2002.
- [6] A. Bell and T. Sejnowski. An information-maximization approach to blind separation and blind deconvolution. *Neural Computation*, 7 :1129–1159, 1995.
- [7] A. Belouchrani. *Séparation autodidacte de sources : Algorithmes, Performances et Applications à des signaux expérimentaux*. PhD thesis, Ecole Nationale Supérieure des Télécommunications, 1995.
- [8] A. Belouchrani, K. A.-Meraim, J.-F. Cardoso, and E. Moulines. A blind source separation technique using second order statistics. *IEEE Trans. on signal processing*, 45 :434–444, 1997.
- [9] A. Belouchrani and M. G. Amin. Blind source separation based on time-frequency signal representations. *IEEE Trans. on Signal Processing*, 46 :2888–2897, 1998.
- [10] A. Belouchrani and M. G. Amin. On the use of spatial time frequency distributions for signal extraction. *Multidimensional Systems and Signal Processing, Special issue of the journal*, 9, No. 4 :349–354, October 1998.
- [11] A. Belouchrani and J.-F. Cardoso. Maximum likelihood source separation for discrete sources. *In Proceeding of EUSIPCO*, 2 :768–771, 1994.
- [12] A. Belouchrani and J.-F. Cardoso. Maximum likelihood source separation for discrete sources. *In Proceeding of NOLTA*, pages 49–53, 1995.
- [13] M. Bentoumi, G. Bloch, P. Aknin, and G. Millerioux. Blind source separation for detection and classification of rail surface defects. *In Proceeding of ENDE*, 2003.

- [14] F. Berthommier and S. Choi. Evaluation of CASA and BSS models for cocktail-party speech segregation. *ICA Conference*, 2001.
- [15] N. Le Bihan, J.I. Mars, and H. Pedersen. Wavefield separation on multicomponent sensors. *62nd meeting of european association of geoscientists and engineers*, 2000.
- [16] Blind Source Separation and Applications. http://www.lis.inpg.fr/pages_perso/bliss/index.php; <http://www-lmc.imag.fr/SMS/SASI/bliss.html>.
- [17] W. Bobillet, E. Grivel, R. Guidorzi, and M. Najim. Cancelling convolutive and additive coloured noises for speech enhancement. *In Proceedings of ICASSP 2004*, Montreal, Canada, May 17-21, 2004.
- [18] J. Bondy. Technical report on McMaster University hearing-aid research. Technical report, McMaster University (BLISS), Oct. 2003. http://www.lis.inpg.fr/pages_perso/bliss/deliverables/year3.html.
- [19] H. Boumaraf. Séparation aveugle de mélanges convolutifs de sources. *In Proceeding of MajecSTIC 2005*, pages 316–322, Nov. 2005.
- [20] H. Boumaraf, D.-T. Pham, and Ch. Servièrè. Blind separation of convolutive mixture of speech sources. *Proceedings of 13th European Signal Processing Conference*, Sept. 2005.
- [21] I. Bradaric, A. P. Petropulu, and K. I. Diamantaras. On resolving the column permutation ambiguity in the estimates of mimo system response. *2000 conference on Information Sciences and Systems*, 2000.
- [22] A. Bunese-Gerstner, R. Byers, and V. Mehrmann. Numerical methods for simultaneous diagonalisation. *SIAM J. Matrix Anal. Appl.*, 14 :927–949, 1993.
- [23] X.-R. Cao and R.-W. Liu. General approach to blind source separation. *IEEE Trans. on Signal Processing*, 44 n° 3 :562–571, 1996.
- [24] V. Capdevielle. *Séparation de sources large bande à l'aide des moments d'ordre supérieur*. PhD thesis, Institut National Polytechnique de Grenoble, 1995.
- [25] V. Capdevielle, Ch. Servièrè, and J.-L. Lacoume. Blind separation of wide-band sources : application to rotating machine signals. *Proceeding of EUSIPCO*, 3 :2085–2088, 1996.
- [26] J.-F. Cardoso. Blind signal separation : statistical principle. *Proceeding of the IEEE*, 9 :2009–2025, 1998.
- [27] J.-F. Cardoso. Analyse en composantes indépendantes. *Conférence invitée. Actes des XXXIV^{èmes} Journées de Statistique, JSBL 2002, Bruxelles*, 2002.
- [28] J.-F. Cardoso and S.-I. Amari. Maximum likelihood source separation : equivariance and adaptivity. *In Proceeding of SYSID'97*, pages 1063–1068, 1997.
- [29] J.-F. Cardoso and D.-T. Pham. Optimization issues in noisy Gaussian ICA. *In Proceeding of ICA*, pages 41–48, Sept. 2004.

- [30] J.-F. Cardoso and A. Souloumiac. Blind beamforming for non gaussian signals. *IEEE processing*, 140 :362–370, 1993.
- [31] J.-F. Cardoso and A. Souloumiac. An efficient technique for blind separation of complex sources. *In Proceeding of IEEE SP Workshop on Higher-Order Stat*, pages 275–279, 1993.
- [32] J.-F. Cardoso and A. Souloumiac. Jacobi angles for simultaneous diagonalization. *SIAM J. Matrix Anal. Appl.*, 1 :161–164, 1996.
- [33] C. Cempel. *Vibroacoustic condition monitoring*. Ellis Horwood Ltd., USA, 1991.
- [34] A. Cichocki and S.-I. Amari. *Adaptive Blind Signal and Image Processing*. Wiley-Interscience, New York, 2002.
- [35] P. Comon. Independent component analysis, a new concept ? *Signal Processing*, 36, n° 3 :287–314, 1994.
- [36] P. Comon. Blind identification and source separation in 2x3 under-determined mixtures. *IEEE Trans. on Signal Processing*, 52 :1–13, 2004.
- [37] J. M. Danthèz and R. Aquilina. Separation of broadband sources concept of the labrador software. *Mechanical systems and signal processing*, 11(1) :91–106, 1997.
- [38] J. Delabrouille, J.-F. Cardoso, and G. Patanchon. Multi-detector multi-component spectral matching and applications for cmb data analysis. *Monthly Notices of the Royal Astronomical Society*, volume 346, issue 4 :1089–1102, Dec. 2003.
- [39] F. Dellaert. The expectation maximization algorithm. Technical Report GIT-GVU-02-20, College of computing, Georgia institut of technology, Feb. 2002. <http://www.cc.gatech.edu/~dellaert/>.
- [40] A. Dempster, N. Laird, and D. Rubin. Maximum likelihood from incomplete data via the em algorithm. *Journal of Royal Statistical Society*, 39 n°1 :1–38, 1977.
- [41] 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 au 4 mai 2001.
- [42] Y. Deville. Towards industrial applications of blind source separation and independent component analysis. *Proceedings of the First International Workshop on Independent Component Analysis and Signal Separation*, pages 19–24, Aussois, France, Jan. 11-15, 1999.
- [43] Y. Deville. Panorama des applications biomédicales des méthodes de séparation aveugle de sources. *Proceedings of GRETSI'2003*, 1 :31–34, Paris, France, Sept. 8-11, 2003.
- [44] K. I. Diamantaras. Oriented PCA and blind signal separation. *In ICA proceedings*, pages 609–613, apr 2003.
- [45] S. Ebihara. Blind separation for estimation of near-surface interface by gpr with time-frequency distribution. *IEICE Trans. Commun.*, E86 :3071–3081, 2003.

- [46] P. Fabry. *Techniques de séparation de sources appliquées à des signaux vibratoires de machines tournantes*. PhD thesis, Institut National Polytechnique de Grenoble, 2001.
- [47] G. Fleury. *Analyse spectrale; Méthodes non-paramétriques et paramétriques*. Ellipses, Paris, 2001.
- [48] C. Févotte. *Approche temps-fréquence pour la séparation aveugle de sources non-stationnaires*. PhD thesis, École Centrale de Nantes, 2003.
- [49] C. Févotte and C. Doncarli. Two contributions to blind source separation using time-frequency distributions. *IEEE Signal Processing Letters*, vol. 11, no. 3, 2004.
- [50] P. Gao, E.-C. Chang, and L. Wyse. Blind separation of fetal ECG from single mixture using SVD and ICA. *IEEE Proceedings*, 3 :1418–1422, 2003.
- [51] G. Gelle. *Les statistiques d'ordre supérieur appliquées à la détection et à la séparation de sources. Utilisation en analyse vibratoire et acoustique*. PhD thesis, Université de Reims Champagne-Ardennes, 1998.
- [52] G. Gelle, M. Colas, and G. Delaunay. Separation of convolutive mixtures of harmonic signals with a temporal approach. application to rotating machine monitoring. *In Proceedings of ICA'99*, 1999.
- [53] G. Gelle, M. Colas, and C. Servière. Blind source separation : A tool for rotating machine monitoring by vibration analysis. *Journal of Sound and Vibration*, 248(5) :865–885, 2001.
- [54] H. H. Goldstein and P. Horwitz. A procedure for the diagonalization of normal matrices. *Journal of the Association for Computing Machinery*, 6 :176–195, 1959.
- [55] G. H. Golub and C. F. V. Loan. *Matrix Computations*. The Johns Hopkins University Press, London, 1993 (2nd edition).
- [56] P. Granjon. *Contribution à la compensation active des vibrations des machines électriques*. PhD thesis, Institut National Polytechnique de Grenoble, 2000.
- [57] S. Harmeling, P. V. Büнау, A. Ziehe, and D.-T. Pham. Technical report on implementation of linear methods and validation on acoustic sources. Technical report, FhG, INPG and McMaster University, Sep. 2004. http://www.lis.inpg.fr/pages_perso/bliss/deliverables/year3.html.
- [58] H. Hartley. Maximum likelihood estimation from incomplete data. *Biometrics*, 14 :174–194, 1958.
- [59] J. Herault, C. Jutten, and B. Ans. Détection de grandeurs primitives dans un message composite par une architecture de calcul neuromimétique en apprentissage non supervisé. *Colloque du GRETSI*, 1 :1017–1022, 1985.
- [60] A. Hyvarinen, J. Karhunen, and E. Oja. *Independent Component Analysis*. Wiley-Interscience, New York, 2001.

- [61] C. Jutten, L. Nguyen Thi, E. Dijkstra, E. Vittoz, and J. Caelen. Blind separation of sources : an algorithm for separation of convolutive mixtures. *Int. Signal Processing Workshop on Higher Order Statistics*, pages 273–276, July 1991.
- [62] M. Kawamoto and Y. Inouye. A deflation algorithm for the blind source-factor separation of mimo-fir channels driven by colored sources. *IEEE Signal Processing Letters*, 10, n° 11 :343–346, 2003.
- [63] W. Kellermann, H. Buchner, W. Herbordt, and R. Aichner. Multichannel acoustic signal processing for human/machine interfaces - fundamental problems and recent advances. *In Proceeding of Int. Conf. on Acoustics*, 2004.
- [64] P. Kisilev, M. Zibulevsky, and Y. Y. Zeevi. A multiscale framework for blind source separation. *Journal of Machine Learning Research*, 4 :1339–1363, 2003.
- [65] M. S. Kompella, P. Davies, R. J. Bernhard, and D. A. Ufford. A technique to determine the number of incoherent sources contributing to the response of a system. *Mechanical systems and signal processing*, 8(4) :363–380, 1994.
- [66] M. Kunt. *Traitement numérique des signaux*. Dunod, Suiss, 1981.
- [67] W. A. Kuperman and G. Turek. Matched field acoustics. *Mechanical systems and signal processing*, 11(1) :141–148, 1997.
- [68] J.-L. Lacoume, P.-O. Amblard, and P. Comon. *Statistiques d'ordre supérieur pour le traitement du signal*. Dunod, Paris, 2002.
- [69] A. Mansour. *Contribution à la séparation aveugle de sources*. PhD thesis, Laboratoire de Traitement d'Images et de Reconnaissance des Formes, 1997.
- [70] A. Mansour. Blind separation for instantaneous mixture of speech signals : Algorithms and performances. *IEEE Conf.*, pages 26–32, 2000.
- [71] A. Mansour. Blind separation of sources : Methodes, assumptions and applications. *IEICE TRANS.*, E83 :1498–1512, 2000.
- [72] A. Mansour, C. Jutten, and N. Ohnishi. Kurtosis : Definition and properties. *International Conference on Multisource-Multisensor Information Fusion*, pages 40–46, 1998.
- [73] A. Mansour, C. G. Puntonet, and N. Ohnishi. A new geometrical blind separation of sources algorithm. *In SCI'2001*, pages 350–355, 2001.
- [74] J. Markus. *Electronics and Nucleonics Dictionary*. McGRAW-HILL BOOK COMPANY, USA, 1966 (3rd edition).
- [75] J. Max and J.-L. Lacoume. *Méthodes et techniques de traitement du signal et applications aux mesures physiques*. Masson, Paris, 1996.
- [76] N. Mitianoudis and M. Davies. Audio source separation of convolutive mixtures. *IEEE Trans. on Speech and Audio Processing*, xx :y, 2002.
- [77] S. Ikeda, N. Murata and A. Ziehe MATLAB code for convolutive blind source separation. <http://www.ism.ac.jp/~shiro/research/blindsep.html>.

- [78] R. Mukai, S. Araki, and S. Makino. Separation and dereverberation performance of frequency domain blind source separation. *In Proceeding of ICA*, pages 230–235, 2001.
- [79] R. Mukai, S. Araki, and S. Makino. Separation and dereverberation performance of frequency domain blind source separation for speech in a reverberant environment. *In Proceeding of Eurospeech 2001*, pages 2599–2603, 2001.
- [80] R. Mukai, S. Araki, H. Sawada, and S. Makino. Evaluation of separation and dereverberation performance in frequency domain blind source separation. *Acoustical Science and Technology*, 25, No.2 :119–126, Mar. 2004.
- [81] N. Murata and S. Ikeda. An on-line algorithm for blind source separation on speech signals. *In Proceeding of NOLTA*, 1998.
- [82] T. Nishikawa, H. Saruwatari, K. Shikano, and S. Makino. Multistage ica for blind source separation of real acoustic convolutive mixture. *In Proceeding of ICA*, pages 523–528, 2003.
- [83] D. Nuzillard and A. Bijaoui. Blind source separation and analysis of multispectral astronomical images. *Astronomy & Astrophysics Supplement Series*, Ser. 147 :129–138, 2000.
- [84] 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, n° 8 :15–24, 2000.
- [85] L. Parra and C. Spence. Convolutional blind source separation of non-stationary sources. *In Proceeding of IEEE Trans. on Speech and Audio Processing*, 8, n° 3 :320–327, May 2000.
- [86] L. Parra and C. Spence MATLAB code for convolutional blind source separation. http://ida.first.gmd.de/~harmeli/download/download_convbss.html.
- [87] D.-T. Pham. D.-T. Pham Software of Joint Approximate Diagonalization of Positive Matrices. <http://www-lmc.imag.fr/lmc-sms/Dinh-Tuan.Pham/jadiag/>.
- [88] D.-T. Pham. Blind separation of instantaneous mixture of sources via the gaussian mutual information criterion. *Signal Processing*, pages 855–870, 2001.
- [89] D.-T. Pham. Joint approximate diagonalization of positive definite matrices. *SIAM J. on Matrix Anal. and Appl.*, 22, n° 4 :1136–1152, 2001.
- [90] D.-T. Pham and J.-F. Cardoso. Blind separation of instantaneous mixtures of non stationary sources. *IEEE Trans. Signal Processing*, 49 :1837–1848, 2001.
- [91] D.-T. Pham, P. Garrat, and C. Jutten. Separation of a mixture of independent sources through a maximum likelihood approach. *In Proceeding of EUSIPCO*, pages 771–774, 1992.
- [92] D.-T. Pham, Ch. Servière, and H. Boumaraf. Blind separation of convolutional audio mixtures using nonstationarity. *In Proceeding of ICA '03*, pages 981–986, Apr. 2003.

- [93] D.-T. Pham, Ch. Servière, and H. Boumaraf. Blind separation of speech mixtures based on nonstationarity. *In Proceeding of ISSPA*, July 2003.
- [94] L. R. Rabiner and R. W. Schafer. *Digital processing of speech signals*. Prentice-Hall, New Jersey, 1978.
- [95] C. Simon, C. Vignat, Ph. Loubaton, C. Jutten, and G. d'Urso. On the convolutive mixture - source separation by the decorrelation approach. *In Proceeding of ICASSP-98, Seattle*, pages 2109–2112, May 1998.
- [96] P. Smaragdis. Blind source separation of convolved mixtures in the frequency domain. *In International Workshop on Independence & Artificial Neural Networks*, Feb. 1998.
- [97] P. Smith, E. Fotheringham, D. Anderson, and Z. Popvic. Smart antennas with optical processing for broadband blind source separation. *IEEE Topical Conference on Wireless Communication Technology and NSF Wireless Grantees Workshop*, 2003.
- [98] S. W. Smith. *The scientist and engineer's guide to Digital Signal Processing*. California Technical Publishing, San Diego, 1999.
- [99] T. Takatani, T. Nishikawa, H. Saruwatari, and K. Shikano. High-fidelity blind separation for convolutive mixture of acoustic signals using SIMO model based independent component analysis. *In Proceeding of ISSPA*, 2003.
- [100] H.-L. Nguyen Thi and C. Jutten. Blind source separation for convolutive mixtures. *Signal Processing*, 45 n°2 :209–229, Aug. 1995.
- [101] L. Tong, R. W. Liu, V. C. Soon, and Y. F. Huang. Indeterminacy and identifiability of blind identification. *IEEE Trans. on Circuit and Systems*, 38 :499–509, 1991.
- [102] L. Tong, V. Soon, Y. F. Huang, and R. Liu. Amuse : a new blind identification algorithm. *In Proceeding of IEEE ISCAS*, pages 1784–1787, 1990.
- [103] L. Trainor, R. Sonnadara, K. Wiklund, J. Bondy, S. Gupta, S. Becker, I. C. Bruce, and S. Haykin. Development of a flexible, realistic hearing in noise test environment (R-HINT-E). *Signal Processing*, 84 :299–309, 2004.
- [104] W. Wang, J. A. Chambers, and S. Sanei. A joint diagonalization method for convolutive blind separation of nonstationary sources in the frequency domain. *In Proceeding of ICA '03*, pages 939–944, 2003.
- [105] J. H. M. Wedderburn. *Lectures on Matrices*. AMS, 201 Charles Street USA, 1934.
- [106] E. Weinstein, M. Feder, and A. V. Oppenheim. Multi-channel signal separation by decorrelation. *IEEE Trans. on Speech and Audio Processing*, 1, n° 4 :405–413, 1993.
- [107] A. Westner and J. Bove. Blind separation of real world audio signals using over-determined mixtures. *In Proceeding of ICA and BSS, Aussois, France*, January 11-15 1999.

-
- [108] A. Westner and V. M. Bove. Applying blind source separation and deconvolution to real-world acoustic environments. *Proceeding 106th of Audio Engineering Society*, 1999.
- [109] A. G. Westner. *Object-Based Audio Capture : Separating Acoustically-Mixed Sounds*. PhD thesis, Massachusetts Institute of Technology, Rutgers University, 1999.
- [110] H.-C. Wu and J. C. Principe. Simultaneous diagonalization in the frequency domain (SDIF) for source separation. *In Proceeding of ICA*, pages 245–250, 1999.
- [111] J. M. F. Xavier, V. A. N. Barroso, and J. M. F. Moura. Closed-form blind channel identification and source separation in SDMA systems through correlative coding. *IEEE Journal on Selected Areas in Communications*, 16 :1506–1517, 1998.
- [112] A. Yeredor. TV-SOBI : an expansion of sobi for linearly time-varying mixtures. *In ICA'03 Proceedings*, pages 903–908, apr 2003.
- [113] A. Ypma. *Learning methods for machine vibration analysis and health monitoring*. PhD thesis, Pattern Recognition Group, Dept. of Applied Physics, Delft University of Technology, 2001.
- [114] A. Ypma, A. Leshem, and R. P. W. Duin. Blind separation of rotating machine sources : bilinear forms and convolutive mixtures. *Neurocomputing, Special Issue on ICA/BSS*, 49 :349–368, 2002.
- [115] A. Ypma and P. Pajunen. Second-order ICA in machine vibration analysis. Technical report, Lab. of Computer and Information Science, 1998. TU Helsinki.
- [116] A. Ypma and P. Pajunen. Rotating machine vibration analysis with second-order independent component analysis. *In Proceeding of ICA'99*, pages 37–42, Aussois, Jan. 1999.
- [117] A. Ziehe and K.-R. Müller. TDSEP - an efficient algorithm for blind separation using time structure. *In Proceeding of ICANN'98*, pages 675–680, 1998.

Annexes

Annexe A

Série de Fourier Discrète

A.1 Définition

Soit $x(t)$ un signal périodique de période T_0 ; $x(t)$ est décomposable en Séries de Fourier exponentielles suivant la formule suivante :

$$x(t) = \sum_{n=-\infty}^{+\infty} C_n e^{j2\pi n f_0 t} \quad (\text{A.1})$$

avec :

$$C_n = \frac{1}{T} \int_0^{T_0} x(t) e^{-j2\pi n f_0 t} dt \quad (\text{A.2})$$

Cependant, en pratique on ne dispose que de signaux discrets, qui, contrairement aux signaux analogiques, ne sont définis qu'aux instants d'échantillonnage nT_e , multiples entiers de la période d'échantillonnage T_e . Par ailleurs, les calculateurs informatiques ne peuvent réaliser de calculs qu'en un nombre fini de points discrets. La Série de Fourier Discrète est la correspondante de la décomposition en séries de Fourier dans l'ensemble des signaux discrets. La SFD décompose un signal numérique en N signaux composites à fréquences pures : $0, f_0, 2f_0, \dots, (N-1)f_0$.

A.2 Série de Fourier Discrète

A.2.1 Forme exponentielle de la SFD

Soit $x(n)$ un signal discret, de N échantillons, obtenu par échantillonnage avec une période T_e .

On peut associer à ce signal, un signal analogique $x(t)$ tel que :

$$x(t) = \sum_{k=-\infty}^{+\infty} x(k) \delta(t - kT_e), \quad (\text{A.3})$$

et avec $T_0 = NT_e$ est la durée de sa période.

D'où la décomposition de $x(n)$ en Séries de Fourier Discrètes comme suit :

$$x(t) = \sum_{n=-\infty}^{+\infty} C_n e^{j2\pi n f_0 t} \implies x(mT_e) = \sum_{n=0}^{N-1} C_n e^{j2\pi n f_0 (mT_e)} \quad (\text{A.4})$$

avec :

$$\begin{aligned} C_n &= \frac{1}{NT_e} \int_0^{T_0} \sum_{k=-\infty}^{+\infty} x(k) \delta(t - kT_e) e^{-j2\pi n f_0 t} dt \\ &= \frac{1}{NT_e} \sum_{k=0}^{N-1} x(k) \int_0^{T_0} \delta(t - kT_e) e^{-j2\pi n f_0 t} dt \\ &= \frac{1}{NT_e} \sum_{k=0}^{N-1} x(k) e^{-j2\pi n f_0 k T_e} T_e \\ &= \frac{1}{N} \sum_{k=0}^{N-1} x(k) e^{-j\frac{2\pi}{N} nk} \end{aligned} \quad (\text{A.5})$$

A.2.2 Forme trigonométrique de la SFD

Nous pouvons passer à la forme trigonométrique comme suit :

$$\begin{aligned} x(m) &= \sum_{n=0}^{N-1} C_n e^{j2\pi n f_0 m T_e} \\ &= \sum_{n=0}^{N-1} C_n e^{j\frac{2\pi}{N} nm} \\ &= C_0 + \sum_{n=1}^{\frac{N}{2}} C_n e^{j\frac{2\pi}{N} nm} + C_{\frac{N}{2}+1} e^{j\frac{2\pi}{N} (\frac{N}{2}+1)m} + \sum_{n=\frac{N}{2}+2}^{N-1} C_n e^{j\frac{2\pi}{N} nm} \end{aligned}$$

Or :

$$\begin{aligned} e^{j\frac{2\pi}{N} (\frac{N}{2}+k)} &= e^{j\frac{2\pi}{N} \frac{N}{2}} \cdot e^{j\frac{2\pi}{N} k} \\ &= -e^{j\frac{2\pi}{N} k} \\ &= -\left(e^{j\frac{2\pi}{N} (\frac{N}{2}-k)} \right)^* \end{aligned}$$

$$x(m) = C_0 + \sum_{n=1}^{\frac{N}{2}} \left\{ C_n \left(\cos 2\pi n \frac{m}{N} + j \sin 2\pi n \frac{m}{N} \right) + C_n^* \left(\cos 2\pi n \frac{m}{N} - j \sin 2\pi n \frac{m}{N} \right) \right\}$$

$$\begin{aligned}
& + C_{\frac{N}{2}+1} \left(\cos 2\pi \left(\frac{N}{2} + 1 \right) \frac{m}{N} + j \sin 2\pi \left(\frac{N}{2} + 1 \right) \frac{m}{N} \right) \\
& = C_0 + \sum_{n=1}^{\frac{N}{2}} \left\{ (C_n + C_{n^*}) \cos 2\pi n \frac{m}{N} + \{j(C_n - C_{n^*})\} \sin 2\pi n \frac{m}{N} \right\} \\
& + C_{\frac{N}{2}+1} \left\{ \cos 2\pi \left(\frac{N}{2} + 1 \right) \frac{m}{N} + j \sin 2\pi \left(\frac{N}{2} + 1 \right) \frac{m}{N} \right\} \\
& = C_0 + \sum_{n=1}^{\frac{N}{2}} \left\{ a_n \cos 2\pi n \frac{m}{N} + b_n \sin 2\pi n \frac{m}{N} \right\} + a_{\frac{N}{2}+1} \cos 2\pi \left(\frac{N}{2} + 1 \right) \frac{m}{N} \\
& + b_{\frac{N}{2}+1} \sin 2\pi \left(\frac{N}{2} + 1 \right) \frac{m}{N} \tag{A.6}
\end{aligned}$$

avec :

$$|C_n| = \frac{1}{2} \sqrt{a_n^2 + b_n^2} \quad : \quad n \in [1, \frac{N}{2}]; \tag{A.7}$$

$$C_0 = a_0; \tag{A.8}$$

$$|C_{\frac{N}{2}+1}| = \frac{1}{2} \sqrt{a_{\frac{N}{2}+1}^2 + b_{\frac{N}{2}+1}^2} \tag{A.9}$$

d'où :

$$\begin{aligned}
x(m) & = \sum_{n=0}^{\frac{N}{2}+1} \left\{ a_n \cos 2\pi n \frac{m}{N} + b_n \sin 2\pi n \frac{m}{N} \right\} \\
& = \sum_{n=0}^{\frac{N}{2}+1} \left\{ \frac{a_n}{\sqrt{a_n^2 + b_n^2}} \cos 2\pi n \frac{m}{N} + \frac{b_n}{\sqrt{a_n^2 + b_n^2}} \sin 2\pi n \frac{m}{N} \right\} \sqrt{a_n^2 + b_n^2} \tag{A.10} \\
& = \sum_{n=0}^{\frac{N}{2}+1} A_n \cos \left(2\pi n \frac{m}{N} - \varphi_n \right)
\end{aligned}$$

avec :

$$A_n = \sqrt{a_n^2 + b_n^2}, \quad A_n = |C_n| \quad : \quad n = 0, \quad A_n = 2|C_n| \quad : \quad n \in [1, \frac{N}{2}]; \tag{A.11}$$

$$\varphi_n = \arctan \frac{b_n}{a_n} \tag{A.12}$$

A.2.3 Remarques

Vu les propriétés de symétrie par rapport à $\frac{N}{2}$:

$$\begin{aligned}
\cos \left(\frac{2\pi}{N} \left(\frac{N}{2} + k \right) \right) & = \cos \left(\frac{2\pi N}{2N} + \frac{2\pi}{N} k \right) \\
& = \cos \pi \cos \frac{2\pi}{N} k - \sin \pi \sin \frac{2\pi}{N} k
\end{aligned}$$

$$\begin{aligned}
&= -\cos\left(\frac{2\pi}{N}k\right) \\
&= -\cos\left(\frac{2\pi}{N}(-k)\right) \\
&= -\cos\left(\frac{2\pi}{N}\left(\frac{N}{2}-k\right)\right),
\end{aligned}$$

$$\begin{aligned}
\sin\left(\frac{2\pi}{N}\left(\frac{N}{2}+k\right)\right) &= \sin\left(\pi + \frac{2\pi}{N}k\right) \\
&= \sin\pi \cos\frac{2\pi}{N}k + \cos\pi \sin\frac{2\pi}{N}k \\
&= -\sin\left(\frac{2\pi}{N}k\right) \\
&= \sin\left(\frac{2\pi}{N}(-k)\right) \\
&= \sin\left(\frac{2\pi}{N}\left(\frac{N}{2}-k\right)\right),
\end{aligned}$$

$$\sin\left(\frac{2\pi}{N}\frac{N}{2}\right) = \sin\pi = 0,$$

$$\sin\left(\frac{2\pi}{N}0\right) = 0$$

Ainsi :

Si $x(n)$ est paire par rapport à $\frac{N}{2}$: $x\left(m + \frac{N}{2}\right) = x\left(-m + \frac{N}{2}\right)$, avec :

$$A_n = |a_n|; \tag{A.13}$$

$$\varphi_n = \begin{cases} 0 & \text{si } a_n > 0 \\ \pi & \text{si } a_n < 0 \end{cases}; \tag{A.14}$$

$$x(m) = \sum_{n=0}^{\frac{N}{2}} |a_n| \cos\left(\frac{2\pi}{N}nm - \varphi_n\right) \tag{A.15}$$

Si $x(n)$ est impaire par rapport à $\frac{N}{2}$: $x\left(m + \frac{N}{2}\right) = -x\left(-m + \frac{N}{2}\right)$, avec :

$$a_n = \sum_{k=0}^{N-1} x(k) \cos\left(\frac{2\pi}{N}nk\right) = \sum_{k=0, k=\frac{N}{2}+1} x(k) \cos\left(\frac{2\pi}{N}nk\right) \tag{A.16}$$

\implies tous les termes s'annulent, sauf le terme 0 et le terme $\left(\frac{N}{2} + 1\right)$

Annexe B

Sous-échantillonnage par interpolation

Soit $y(t)$, le signal observé en sortie d'un système de traitement de signal dont la réponse impulsionnelle $h(t)$ est le signal qu'on souhaite sous échantillonner :

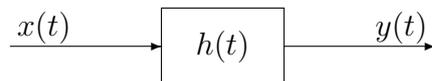


FIG. B.1 – *système de traitement de signal*

$x(t)$ est l'excitation qui a provoqué le signal discret $y(t)$ en sortie du système. Sous l'hypothèse que le système est linéaire et homogène, $y(t) = \sum_k x(k)h(n - k)$.

Notre objectif est de sous échantillonner convenablement tous le système ($x(t)$, $y(t)$ et $h(t)$) de sorte qu'il ait une adéquation et une cohérence entre l'entrée, la sortie et la réponse impulsionnelle du système avec le minimum de perte (erreur).

Si ' f ' est la fréquence d'échantillonnage de $y(t)$ et si on suppose que l'on souhaite " abaisser " cette fréquence à $((1/m) * 100)\%$ de f , la méthode consistera à avoir un échantillon sur m de l'ensemble de la séquence définissant le signal (à sous échantillonné). Ainsi, le système résultant sera défini par :

$$y_s(t) = (x_s * h_s)(t) = y(mt) = \sum_{k=-\infty}^{k=+\infty} x(k).h(mt - k) \quad (\text{B.1})$$

Le principe que nous voulons appliquer consiste à ne pas nous contenter de prendre un seul échantillon (tel qu'il est) des ' m ' échantillons. En effet, cela engendre la perte de l'information contenue dans le reste des échantillons. En revanche, il faut remplacer ces échantillons par l'information engendrée par l'ensemble, de sorte qu'il n'ait pas beaucoup de pertes.

La façon dont on réalise cela est l'interpolation ; si $\{1, 2, \dots, m-1, m\}$ sont les échantillons concernés, alors $\{2, \dots, m-1\}$ seront l'interpolation linéaire de $\{1\}$ et $\{m\}$; Cela veut dire qu'on estimera les points $\{2, \dots, m-1\}$ depuis la connaissance du $\{1^{\text{ier}}\}$ et du $\{m^{\text{ième}}\}$. Grâce à cette technique nous allons pouvoir écrire convenablement l'équa-

tion mathématique définissant le reste des $(m - 2)$ points en fonction du 1^{er} et $m^{\text{ième}}$ échantillon.

En remplaçant chaque point par son équation d'interpolation et après factorisation des termes, nous identifions la réponse impulsionnelle sous échantillonnée comme suit :

B.1 Sous-échantillonnage de 50% ($m = 2$)

Si à partir du système initial :

$$y(t) = (x * h)(t) = \sum_{k=-\infty}^{k=+\infty} x(k).h(t - k) \quad (\text{B.2})$$

Nous voulons sous échantillonné d'un rapport $\frac{1}{2}$ ($m = 2$), alors notre nouveau système sera défini par l'équation :

$$y(2t) = \sum_{k=-\infty}^{k=+\infty} x(k).h(2t - k) \quad (\text{B.3})$$

Nous allons donc prendre un échantillon sur deux, d'où le schéma d'interpolation suivant :

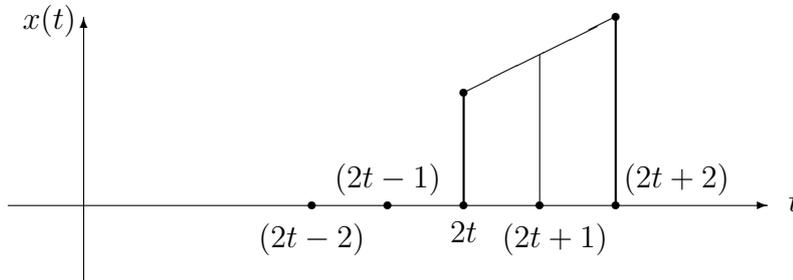


FIG. B.2 – schéma d'interpolation ($m = 2$)

Nous écrivons ainsi l'équation de la droite reliant les points $x(2t)$ et $x(2t + 2)$ et on en déduit leurs équations.

Nous divisons ensuite le domaine de variation de k en deux parties correspondant chacune à l'un des indices ($2t$ ou $(2t + 1)$) :

$$\{-\infty < k < +\infty\} = \{-\infty < 2k < +\infty\} \cup \{-\infty < (2k + 1) < +\infty\}$$

en remplaçant tous dans $y(2t)$ on obtient :

$$y(2t) = \sum_{k=-\infty}^{k=+\infty} x(2(t - k)) \left[\frac{h(2k - 1)}{2} + h(2k) + \frac{h(2k + 1)}{2} \right] \quad (\text{B.4})$$

Par comparaison avec l'équation (B.1) nous trouvons le résultat suivant :

$$h_{equiv}(k) = \frac{h(2k-1)}{2} + h(2k) + \frac{h(2k+1)}{2} \quad (\text{B.5})$$

B.2 Sous-échantillonnage de 25% ($m = 4$)

Dans cet exemple, nous prenons un point sur quatre. le schéma d'interpolation est comme suit :

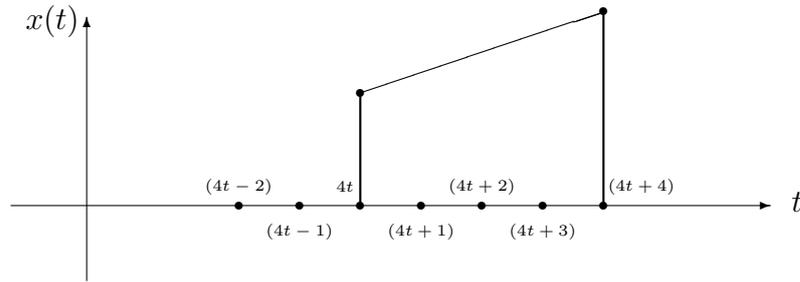


FIG. B.3 – schéma d'interpolation ($m = 4$)

Nous écrivons les équations définissant les différents points et nous procédons pour le reste du calcul de la même façon que l'exemple précédent. Ainsi :

$$h_{equiv}(k) = \frac{1}{4}h(4k-3) + \frac{2}{4}h(4k-2) + \frac{3}{4}h(4k-1) + h(4k) + \frac{3}{4}h(4k+1) + \frac{2}{4}h(4k+2) + \frac{1}{4}h(4k+3) \quad (\text{B.6})$$

B.3 Loi générale

Nous pouvons généraliser le résultat obtenu pour m quelconque où le schéma d'interpolation sera :

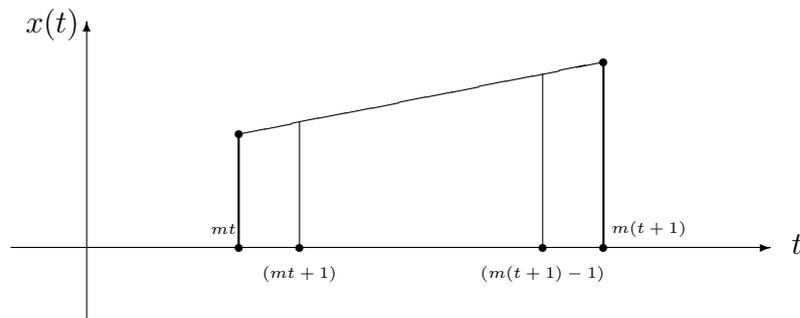


FIG. B.4 – schéma général d'interpolation

Et l'équation générale de l'un des points est :

$$x(mt + l) = \left(\frac{m-l}{m}\right) x(mt) + \left(\frac{l}{m}\right) x(m(t+1)) \quad : 1 < l < (m-1) \quad (\text{B.7})$$

Ainsi :

$$h_{equiv}(k) = \frac{1}{m} \sum_{l=-(m-1)}^{l=+(m-1)} (m \begin{bmatrix} \pm \\ \text{si} : l < 0 \\ \text{si} : l \geq 0 \end{bmatrix} l) k(mk + l) \quad (\text{B.8})$$

B.4 Calcul détaillé ($m = 4$)

Soit le schéma correspondant :

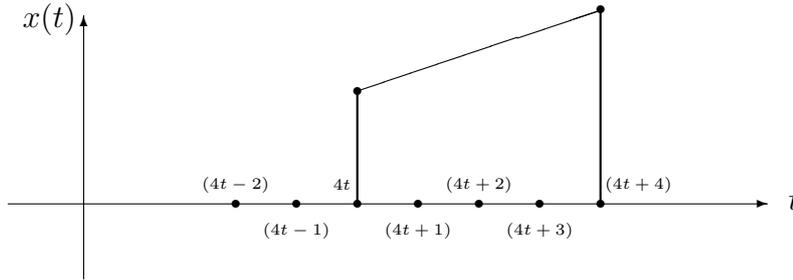


FIG. B.5 – schéma d'interpolation ($m = 4$)

L'équation de la droite d'interpolation :

$$\begin{cases} x(4t) & = a(4t) + b \\ x(4(t+1)) & = a4(t+1) + b \end{cases} \Rightarrow \begin{cases} x(4t+1) & = \frac{1}{4} [3x(4t) + x(4t+4)] \\ x(4t+2) & = \frac{1}{4} [2x(4t) + 2x(4t+4)] \\ x(4t+3) & = \frac{1}{4} [x(4t) + 3x(4t+4)] \end{cases}$$

On a :

$$\begin{aligned} y(4t) &= \sum_{k=-\infty}^{+\infty} x(k) \cdot h(4t - k) \\ &= \sum_{k=-\infty}^{+\infty} h(k) \cdot x(4t - k) \end{aligned}$$

Nous divisons ensuite l'intervalle de variation de k :

$$\begin{aligned}
y(4t) &= \sum_k h(4k).x(4t - 4k) + \sum_k h(4k + 1).x(4t - 4k - 1) \\
&\quad + \sum_k h(4k + 2).x(4t - 4k - 2) + \sum_k h(4k + 3).x(4t - 4k - 3) \\
&= \sum_k h(4k).x(4(t - k)) + \sum_k h(4k + 1).x(4(t - k) - 1) \\
&\quad + \sum_k h(4k + 2).x(4(t - k) - 2) + \sum_k h(4k + 3).x(4(t - k) - 3)
\end{aligned}$$

Or :

$$\begin{cases}
x(4t + 1) = \frac{1}{4}[3x(4t) + x(4t + 4)] \\
x(4t + 2) = \frac{1}{4}[2x(4t) + 2x(4t + 4)] \\
x(4t + 3) = \frac{1}{4}[x(4t) + 3x(4t + 4)]
\end{cases}$$

En remplaçant dans l'équation de y :

$$\begin{aligned}
\Rightarrow y(4t) &= \sum_k h(4k).x(4t - 4k) + \sum_k h(4k + 1).[x(4(t - k) - 4) + 3x(4(t - k))] \frac{1}{4} \\
&\quad + \sum_k h(4k + 2).[x(4(t - k) - 4) + x(4(t - k))] \frac{2}{4} \\
&\quad + \sum_k h(4k + 3).[3x(4(t - k) - 4) - x(4(t - k))] \frac{1}{4} \\
&= \sum_k \{x(4(t - k)).[h(4k) + \frac{3}{4}h(4k + 1) + \frac{2}{4}h(4k + 2) + \frac{1}{4}h(4k + 3)]\} \\
&\quad + \sum_k \{x(4(t - k) - 4).[\frac{1}{4}h(4k + 1) + \frac{2}{4}h(4k + 2) + \frac{3}{4}h(4k + 3)]\}
\end{aligned}$$

Et par changement de variable $\{(k \leftarrow \tau), (\tau \leftarrow (k - 1))\}$:

$$\begin{aligned}
\Rightarrow y(4t) &= \sum_k \{x(4(t - k)).[h(4k) + \frac{3}{4}h(4k + 1) + \frac{2}{4}h(4k + 2) + \frac{1}{4}h(4k + 3)]\} \\
&\quad + \sum_k \{x(4(t - k)).[\frac{1}{4}h(4k - 3) + \frac{2}{4}h(4k - 2) + \frac{3}{4}h(4k - 1)]\} \\
&= \sum_k x(4(t - k)).[h(4k) + \frac{3}{4}h(4k + 1) + \frac{2}{4}h(4k + 2) + \frac{1}{4}h(4k + 3) \\
&\quad + \frac{1}{4}h(4k - 3) + \frac{2}{4}h(4k - 2) + \frac{3}{4}h(4k - 1)]
\end{aligned}$$

$$\Rightarrow \boxed{
\begin{aligned}
h_{equiv}(k) &= \frac{1}{4}h(4k - 3) + \frac{2}{4}h(4k - 2) + \frac{3}{4}h(4k - 1) + h(4k) \\
&\quad + \frac{3}{4}h(4k + 1) + \frac{2}{4}h(4k + 2) + \frac{1}{4}h(4k + 3)
\end{aligned}
} \tag{B.9}$$

Annexe C

Maximum de vraisemblance

Maximum de vraisemblance, ou aussi appelée méthode du maximum de vraisemblance, est la procédure qui cherche la valeur d'un paramètre ou plus pour un ensemble donné de statistiques, de sorte que la distribution de vraisemblance soit maximum.

Soit X une variable aléatoire de densité de probabilité $f(x, \theta)$ connue analytiquement mais dont l'un des paramètres θ est inconnu (numériquement). Le problème consiste donc à construire une expression analytique en fonction des réalisations de cette variable dans un échantillon de taille n , permettant de trouver la valeur numérique la plus vraisemblable pour le paramètre θ .

Si $\{x_1, \dots, x_n\}$ sont des réalisations indépendantes de la variable aléatoire, on peut dire que $x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ est une réalisation d'un vecteur aléatoire $X = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix}$ dont les composantes X_i sont indépendantes deux à deux.

L'approche retenue consiste à chercher la valeur de θ qui rend le plus probable les réalisations que l'on vient d'obtenir. La probabilité d'apparition a priori de l'échantillon en question peut alors être caractérisée par le produit des probabilités d'apparition de chacune des réalisations (puisque celles-ci sont supposées indépendantes deux à deux).

$$P(X = x) = \prod_{i=1}^n f(x_i, \theta) = L(x_1, \dots, x_n, \theta) \quad (\text{C.1})$$

La méthode du maximum de vraisemblance consiste à rechercher la valeur de θ qui rend cette probabilité maximale. Le produit des valeurs $f(x_i, \theta)$ est aussi noté $L(x_1, \dots, x_n)$ et appelé *fonction de vraisemblance*. La valeur de $\hat{\theta}$ qui rend maximum la fonction de vraisemblance L est donc la solution de :

$$\frac{\partial \log L}{\partial \theta} = 0 \implies \hat{\theta} : \frac{\partial^2 \log L}{\partial \theta^2} < 0 \quad (\text{C.2})$$

L'emploi du logarithme sur la fonction L permet de passer de la maximisation d'un

produit à celle d'une somme, le résultat restant le même car la fonction logarithme est monotone, strictement croissante.

C.1 Propriétés de la fonction de vraisemblance

- $\int_{\mathbb{R}^n} L(x, \theta) dx = 1$;
- $\int_{\mathbb{R}^n} \frac{\partial L(x, \theta)}{\partial \theta} dx = 0$;
- $E \left[\frac{\partial \log L(x, \theta)}{\partial \theta} \right] = 0$;
- $E \left[\left(\frac{1}{L(x, \theta)} \frac{\partial \log L(x, \theta)}{\partial \theta} \right)^2 \right] = E \left[\frac{\partial^2 \log L(x, \theta)}{\partial \theta^2} \right]$.

Théorème C.1 *S'il existe un estimateur efficace sans biais, il sera donné par la méthode du maximum de vraisemblance.*

C.2 Cas du modèle de mélange $\mathbf{x} = A\mathbf{s}$

En supposant que la distribution de probabilité de chaque composante s_i du vecteur \mathbf{s} ait une densité de probabilité $f_i(\cdot)$, alors, la distribution du vecteur aléatoire \mathbf{s} a une densité qui est de la forme $f(\mathbf{s}) = \prod_{i=1}^n f_i(s_i)$ ¹.

On peut donc exprimer la vraisemblance (densité de \mathbf{x} sachant A) comme suit :

$$P(\mathbf{x}|A) = \frac{1}{|\det(A)|} \prod_{i=1}^n f_i([A^{-1}\mathbf{x}]_i) = \frac{1}{|\det(A)|} f(A^{-1}\mathbf{x}) \quad (\text{C.3})$$

Ainsi la log-vraisemblance $\mathcal{L}(x_1, \dots, x_n, A) = \log L(x_1, \dots, x_n, A)$ s'écrit :

$$\mathcal{L}(\mathbf{x}, A) \stackrel{\text{def}}{=} \sum_{i=1}^n \log f(x_i, A) \quad (\text{C.4})$$

$$= \sum_{i=1}^n \log f(A^{-1}x_i) - \log |\det(A)| \quad (\text{C.5})$$

¹les composantes de \mathbf{s} sont supposées indépendantes.

C.3 Cas de K variables Gaussiennes multidimensionnelles

Soient K populations d'observations (v.a.) multidimensionnelles, de taille n_1, \dots, n_K ; chaque population $\mathbf{y}_k = \{y_{ik}\}$, $i = 1, \dots, n_k$, obéit à une loi Gaussienne $\mathcal{N}_k(\mu_k, \Gamma_k)$, $k = 1, \dots, K$, ($y_{ik}, i = 1, \dots, n_k$, sont i.i.d.), où les paramètres (inconnus) μ_k et Γ_k représentent respectivement la moyenne et la matrice de covariance de la loi normale. La densité de $\mathbf{y}_k = \{y_{ik}\}$ ($i = 1, \dots, n_k$) est donnée par l'équation :

$$(2\pi)^{-n_k/2} [(\det(\Gamma_k))^{-n_k/2}] \exp \left\{ -\frac{1}{2} (\mathbf{y}_k - \mu_k)^T \Gamma_k^{-1} (\mathbf{y}_k - \mu_k) \right\} \quad (\text{C.6})$$

Si les observations $y_{1k}, \dots, y_{n_k k}$ ($k = 1, \dots, K$) sont indépendantes, la fonction log-vraisemblance $\mathcal{L}((\mu_1, \dots, \mu_K), (\Gamma_1, \dots, \Gamma_K)) = \sum_k \log \prod_i f(y_{ik})$ des observations est (à une constante près) :

$$\begin{aligned} \mathcal{L}((\mu_1, \dots, \mu_K), (\Gamma_1, \dots, \Gamma_K)) &= - \sum_{k=1}^K \left[\frac{n_k}{2} \log \det \Gamma_k + \frac{1}{2} \sum_{i=1}^{n_k} ((y_{ik} - \mu_k)^T \Gamma_k^{-1} (y_{ik} - \mu_k)) \right] \\ &= - \frac{1}{2} \sum_{k=1}^K n_k [\log \det \Gamma_k + \text{tr}(\Gamma_k^{-1} \bar{\Gamma}_k) + (\mu_k - \bar{\mu}_k)^T \Gamma_k^{-1} (\mu_k - \bar{\mu}_k)] \quad (\text{C.7}) \end{aligned}$$

où $\bar{\mu}$ et $\bar{\Gamma}$ sont respectivement la moyenne et la covariance empiriques :

$$\bar{\mu}_k = \frac{1}{n_k} \sum_{i=1}^{n_k} y_{ik}, \quad \bar{\Gamma}_k = \frac{1}{n_k} \sum_{i=1}^{n_k} ((y_{ik} - \bar{\mu}_k)(y_{ik} - \bar{\mu}_k)^T)$$

Si les v.a. sont centrés ($\mu_k = 0$) et si on dispose d'une estimation plus efficace des matrices de covariance, où les matrices estimées sont notées C_1, \dots, C_k ($C_k = \hat{\Gamma}_k$), alors (à une constante près) :

$$\boxed{\mathcal{L}(\Gamma_1, \dots, \Gamma_K) = -\frac{1}{2} \sum_{k=1}^K n_k [\log \det \Gamma_k + \text{tr}(\Gamma_k^{-1} C_k)]} \quad (\text{C.8})$$

Annexe D

Convolution sectionnée

On se propose dans cette annexe de présenter le processus de la convolution dite "sectionnée" pour la convolution de deux signaux, dont les durées ne sont pas égales et en général, l'une est beaucoup plus longue que celle de l'autre. Ces bases sont tirées de mon ancien cours de traitement de signal en ingénierie d'électronique ; les mêmes propos peuvent être trouvés dans [66, pages 103-107] et [75, pages 164-167].

Soit le signal $x(n)$ de durée L_x qu'on désire filtrer avec un filtre dont la réponse impulsionnelle $h(n)$ est de durée L_h et où L_x est beaucoup plus grande que L_h . Le calcul de la filtrée par la convolution standard, définie par :

$$y(n) = (x * h)(n) = \sum_k h(k)x(n - k), \quad (\text{D.1})$$

nécessite, rien que pour le premier échantillon de $y(n)$ d'attendre la fin de tous les produits entre les échantillons ainsi que leur somme ; Quant à son calcul par TFD, il nécessite en pratique des mémoires et un temps de calcul prohibitif. C'est pourquoi on recourt aux méthodes qui divisent le signal en plusieurs sections et où les résultats partiels seront combinés pour composer le signal résultant [66].

On se propose alors de sectionner le signal $x(n)$ en sous séquences de longueur finie L comme suit :

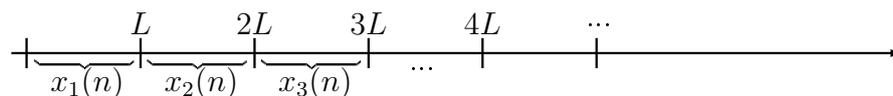


FIG. D.1 – sectionnement de $x(n)$

$$\implies x(n) = \sum_{r=1}^{L_x/L} x_r(n) \quad (\text{D.2})$$

d'où

$$\begin{aligned} y(n) &= \left(\left[\sum_{r=1}^{L_x/L} x_r \right] * h \right) (n) = \left(\sum_{r=1}^{L_x/L} x_r * h \right) (n) \\ &= \sum_{r=1}^{L_x/L} y_r(n) \end{aligned} \tag{D.3}$$

avec $y_r(n) = (h * x_r)(n)$ représente le résultat d'une convolution partielle. Cette variante des méthodes de convolution partielle est appelée "*overlap add method*".

Annexe E

Indice d'Amari

L'objectif de cette annexe est de présenter le principe de l'indice d'Amari (pour le cas instantané), utilisé dans le nouvel indice d'Amari pour les mélanges convolutifs. Nous avons employé ce nouvel indice pour l'analyse de performance dans l'étude comparative du chapitre 4.

Soit une matrice de permutation P . La multiplication (à droite) de celle-ci par une matrice diagonale D (facteur d'échelle), engendre une matrice dont les éléments sont égaux aux éléments de P multipliés en chaque colonne par l'élément correspondant de la diagonale de D :

$$\begin{aligned}
 PD &= \begin{bmatrix} 0 & 0 & \dots & 1 \\ 1 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 1 & \dots & 0 \end{bmatrix} \begin{bmatrix} \alpha & 0 & \dots & 0 \\ 0 & \beta & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \gamma \end{bmatrix} \\
 &= \begin{bmatrix} 0.\alpha & 0.\beta & \dots & 1.\gamma \\ 1.\alpha & 0.\beta & \dots & 0.\gamma \\ \vdots & \vdots & \ddots & \vdots \\ 0.\alpha & 1.\beta & \dots & 0.\gamma \end{bmatrix}
 \end{aligned} \tag{E.1}$$

Nous observons qu'il y a obligatoirement dans cette matrice un max par ligne et par colonne. Donc, si nous voudrions nous débarrasser de D (facteur d'échelle), il va falloir diviser chaque colonne par son maximum.

En conséquence, afin de mesurer la similitude de se produit, résultant de l'ambiguïté de la séparation aveugle de sources, à une matrice de permutation :

$$\begin{aligned}
 C &= GH \\
 &= PD\hat{H}^{-1}H \\
 &= PDI \\
 &= PD,
 \end{aligned} \tag{E.2}$$

Amari a proposé un indice, qui divise les éléments de chaque ligne par le max de cette ligne et en divise les éléments de chaque colonne par le max de cette colonne, pour éliminer d'abord le facteur d'échelle. De cette façon nous filtrons la forme d'une matrice de permutation. Ensuite, nous poursuivons la mesure de la similitude par le calcul de la somme des lignes et des colonnes en retranchant un par ligne et par colonne. Sachant qu'une matrice de permutation n'a qu'un élément égal à un par ligne et par colonne, l'indice d'Amari indique la meilleur similitude quand il est nul ou proche de zéro. L'indice d'Amari est ainsi formulé comme suit :

$$I_A = \underbrace{\sum_i \left(\sum_j \frac{|C_{ij}|}{\max_k |C_{ik}|} - 1 \right)}_{\text{lignes}} + \underbrace{\sum_j \left(\sum_i \frac{|C_{ij}|}{\max_k |C_{kj}|} - 1 \right)}_{\text{colonnes}} \quad (E.3)$$

Dans le pire des cas, cet indice peut être égale à $2MN - (M + N)$ pour une matrice de taille $M \times N$. En effet cela se produit dans le cas où tous les éléments de la matrice sont égaux à un. En conséquence, nous pouvons normaliser l'indice pour des valeurs comprises entre 0 et 1 en divisant I_A par $2MN - (M + N)$.

Annexe F

Estimation spectrale

L'objet de cet annexe est de présenter le fondement théorique de quelques méthodes d'estimation spectrale, ainsi que l'étude de leurs propriétés en terme de biais et de variance. Nous rappelons que l'estimation de la densité spectrale de puissance (dsp) est une étape primordiale dans nos algorithmes de séparation de sources. En effet, l'estimation spectrale permet de décrire le processus observé (mélanges) en nous informant sur son contenu fréquentiel et le degré de dépendance entre ses valeurs temporelles donné en unité de temps d'écart. Biunivoquement, la dsp et la fonction de corrélation sont liées par une transformée de Fourier (théorème de Wiener-Khintchine).

F.1 Choix du meilleur estimateur

Le choix de l'estimateur est relativement grand car il n'y a pas une méthode d'estimation *universelle*. La notion du meilleur estimateur spectral n'a pas de sens dans l'absolu. En effet, d'une part parce qu'il n'existe pas de spectre de référence dans la plupart des cas, mais surtout parce que l'optimalité d'un estimateur ne peut être énoncée que pour certaines caractéristiques spectrales, comme :

- la fidélité aux raies spectrales ;
- la réduction des lobes secondaires ;
- la fidélité au niveau moyen de l'énergie spectrale ;
- l'estimation exacte des fréquences de résonance ...

L'utilisateur devra choisir la méthode qui convient d'après les contraintes de son propre problème. Par ailleurs, le plus important paramètre d'un estimateur est sa résolution : un estimateur est dit avoir une grande résolution s'il permet de distinguer les variations rapides, par exemple d'exhiber des pics voisins au lieu de les confondre. Mais ce faisant, on augmente la covariance de l'estimateur. Il y a donc un compromis à réaliser entre les deux exigences : *grande résolution et faible variance*.

F.2 Estimateur spectral simple

L'estimateur spectral simple d'un signal $x(t)$ est la transformée de Fourier de l'estimateur $c_x(t)$ de sa fonction de corrélation :

$$S_x(f) = \sum_{t=-(K-1)}^{K-1} c_x(t) \exp(-j2\pi ft) \quad (\text{F.1})$$

Une autre terminologie consiste à appeler $S_x(f)$ *périodogramme* en écrivant :

$$S_x(f) = \frac{1}{K} |X(f)|^2 \quad (\text{F.2})$$

avec

$$X(f) = \sum_{t=0}^{K-1} x(t) \exp(-j2\pi ft)$$

F.2.1 Biais de l'estimateur simple

Le biais de l'estimateur simple est la différence entre $E[S_x(f)]$ et le spectre cherché $\Phi_x(f)$. L'espérance mathématique de l'estimateur spectral simple est donnée par :

$$\begin{aligned} E[S_x(f)] &= \sum_{t=-(K-1)}^{K-1} E[c_x(t)] \exp(-j2\pi ft) \\ &= \sum_{t=-(K-1)}^{K-1} \frac{K-|t|}{K} \phi_x(t) \exp(-j2\pi ft) \end{aligned} \quad (\text{F.3})$$

Ainsi, à cause du terme $(K-|t|)/K$, l'espérance mathématique de $S_x(f)$ n'est pas la transformée de Fourier de la fonction d'autocorrélation. Donc $S_x(f)$ est un estimateur biaisé.

F.2.2 Variance de l'estimateur simple

À cause de la complexité et la longueur du calcul de la variance de cet estimateur, nous allons omettre le détail de ce calcul. Le résultat final, valable pour toutes les fréquences est [66] :

$$\text{Var}[S_x(f)] = \Phi_x^2(f) \left[1 + \left(\frac{\sin 2\pi f K}{K \sin 2\pi f} \right)^2 \right] \quad (\text{F.4})$$

F.2.3 Commentaire

Les résultats obtenus montrent que la variance de l'estimateur spectral simple ne dépend pas de la durée d'observation K . Quelle que soit cette durée, la variance de cet estimateur reste proportionnelle au carré du spectre cherché. Par conséquent, cet estimateur n'est pas consistant.

F.3 Estimateur spectral moyenné

La méthode directe de réduire la variance d'un estimateur est de calculer une moyenne sur plusieurs estimateurs indépendants. Pour l'appliquer à l'estimateur spectral simple, il faut diviser le signal observé sur une durée K en L sections $x_l(t)$ de durée M chacune.

$$x_l(t) = x(t + (l - 1)M) \quad (\text{F.5})$$

avec $t = 0, \dots, M - 1$, $K = ML$ et $l = 1, \dots, L$

On évalue ensuite L estimateurs simples du type :

$$S_{x_l}(f) = \frac{1}{M} \left| \sum_{t=0}^{M-1} x_l(t) \exp(-j2\pi ft) \right|^2 \quad (\text{F.6})$$

avec $l = 1, \dots, L$

L'estimateur spectral moyenné est donné par :

$$\bar{S}_x(f) = \frac{1}{L} \sum_{l=1}^L S_{x_l}(f) \quad (\text{F.7})$$

F.3.1 Biases de l'estimateur moyenné

Le biais de l'estimateur moyenné est la différence entre $E[\bar{S}_x(f)]$ et le spectre cherché $\Phi_x(f)$. L'espérance mathématique de l'estimateur est donnée par :

$$E[\bar{S}_x(f)] = \frac{1}{L} \sum_{l=1}^L E[S_{x_l}(f)] = E[S_{x_l}(f)] \quad (\text{F.8})$$

F.3.2 Variance de l'estimateur moyenné

Si l'on admet que les L estimateurs simples moyennés sont indépendants, alors $\bar{S}_x(f)$ est la valeur moyenne de L observations $S_{x_l}(f)$.

$$\begin{aligned} \text{Var} &= \frac{1}{L} \text{Var}[\bar{S}_{x_l}(f)] \\ &\cong \frac{1}{L} \Phi_x^2(f) = \frac{M}{K} \Phi_x^2(f) \end{aligned} \quad (\text{F.9})$$

Cette relation montre que la variance de l'estimateur moyenné décroît proportionnellement à la durée d'observation K .

F.3.3 Commentaire

On a vu que le biais de l'estimateur moyenné peut être interprété en terme du produit de convolution du spectre cherché et d'une fonction fenêtre spectrale. Celle-ci, en général, est formée d'un pic central important et de lobes secondaires. La largeur de base du pic central, inversement proportionnelle à la durée d'observation, détermine le biais. Plus ce pic central est étroit, plus le biais est petit, car l'intégrale tend à la limite vers le spectre cherché. Donc pour diminuer le biais de $\bar{S}_x(f)$, il faut augmenter la durée M des sections $x_l(t)$. D'autre part, pour diminuer la variance de cet estimateur par rapport à celle de l'estimateur simple, il faut augmenter le nombre L de sections contribuant à la moyenne. Pour une durée d'observation globale $K = ML$ fixée, ceci conduit à un compromis.

F.4 Estimateur spectral adouci

Un autre moyen de réduire la variance de l'estimateur spectral simple est de le filtrer. On considère $S_x(f)$ comme un signal et on peut le filtrer avec un filtre dont la réponse impulsionnelle est $W(f)$. Il s'agit bien sûr d'un filtrage fréquentiel. On obtient ainsi un estimateur adouci, donné par :

$$\tilde{S}_x(f) = \int_{-1/2}^{1/2} S_x(g)W(f - g)dg \quad (\text{F.10})$$

La fonction $W(f)$ est appelée fenêtre spectrale ou *fenêtre d'apodisation*. On peut interpréter l'estimateur adouci comme la transformée de Fourier du produit de l'estimateur $c_x(t)$ et d'une fenêtre $w(t)$ de durée finie $2M - 1$.

$$\tilde{S}_x(f) = \sum_{t=-(M-1)}^{M-1} w(t)c_x(t) \exp(-j2\pi ft) \quad (\text{F.11})$$

Cette relation met en évidence les contraintes à poser sur les fonctions fenêtres. Comme la densité spectrale de puissance et toute estimation de celle-ci doivent être des fonctions réelles et positives, la fonction $w(t)$ doit être paire et sa transformée $W(f)$ doit être positive.

F.4.1 Biais de l'estimateur adouci

L'espérance mathématique de la relation précédente conduit à :

$$E[\tilde{S}_x(f)] = \int_{-1/2}^{1/2} E[S_x(g)]W(f - g)dg \quad (\text{F.12})$$

Comme $E[\tilde{S}_x(f)]$ est, à son tour, un produit de convolution, l'espérance mathématique de l'estimateur adouci est un double produit de convolution. On peut aussi l'écrire

comme la transformée de Fourier du double produit simple :

$$E[\tilde{S}_x(f)] = \sum_{t=-(M-1)}^{M-1} w_T(t)w(t)\phi_x(t) \exp(-j2\pi ft) \quad (\text{F.13})$$

avec $w_T(t) = 1 - \frac{|t|}{K}$ et $|t| < K$

Si M est relativement petit via-à-vis de K , on peut écrire approximativement en négligeant le terme $|t|/K$:

$$\begin{aligned} E[\tilde{S}_x(f)] &\cong \sum_{t=-(M-1)}^{M-1} w(t)\phi_x(t) \exp(-j2\pi ft) \\ &\cong \int_{-1/2}^{1/2} W(g)\Phi_x(f-g)dg \end{aligned} \quad (\text{F.14})$$

Le biais $B_{\tilde{S}_x}(f) = E[\tilde{S}_x(f)] - \Phi_x(f)$ est d'autant plus faible que M est grand. Inversement, une diminution de M provoque une perte de résolution fréquentielle.

F.4.2 Variance de l'estimateur adouci

La variance de l'estimateur adouci est donnée par :

$$\text{Var}[\tilde{S}_x(f)] = E[(\tilde{S}_x(f) - E[\tilde{S}_x(f)])^2] \quad (\text{F.15})$$

En substituant certains calculs et en faisant quelques approximations nous obtenons le résultat suivant :

$$\text{Var}[\tilde{S}_x(f)] = \frac{1}{2}\Phi_x^2(f) \int_{-1/2}^{1/2} W^2(g)dg \quad (\text{F.16})$$

Cette relation n'est valable que si la durée $2M - 1$ de la fenêtre $w(t)$ est étroite par rapport aux variations de $\Phi_x(f)$.

F.5 Estimateur spectral modifié

Pour diminuer la variance de l'estimateur spectral simple, deux méthodes ont été envisagées, conduisant aux estimateurs moyennés et adoucis. On peut combiner ces deux méthodes pour développer un autre estimateur spectral qui convient particulièrement bien à l'exploitation de la TFR. Cet estimateur s'appelle *estimateur spectral moyenné*. Le signal observé est de nouveau divisé en L sections de durée $M = K/L$. Toutefois, dans ce cas, chaque section est multipliée par une fonction fenêtre $w(t)$ avant le calcul de l'estimateur simple. L'équivalent de l'estimateur spectral simple est alors donné par :

$$R_{x_l}(f) = \frac{1}{MP} \left| \sum_{t=0}^{M-1} x_l(t)w(t) \exp(-j2\pi ft) \right|^2 \quad (\text{F.17})$$

avec $l = 1, \dots, L$ et $P = \frac{1}{M} \sum_{t=0}^{M-1} w^2(t)$

Le facteur de normalisation P est nécessaire pour que l'estimateur soit asymptotiquement non biaisé. L'estimateur s'exprime par :

$$\bar{R}_x(f) = \frac{1}{L} \sum_{l=1}^L R_{x_l}(f) \quad (\text{F.18})$$

F.5.1 Biais et variance de l'estimateur modifié

On peut montrer que l'espérance mathématique de l'estimateur modifié est donnée par :

$$E[\bar{R}_x(f)] = \int_{-1/2}^{1/2} \Phi_x(g)\Phi_w(f-g)dg \quad (\text{F.19})$$

où $\Phi_w(f) = \frac{1}{MP} \left| \sum_{t=0}^{M-1} w(t) \exp(-j2\pi ft) \right|^2$

Cette expression montre que, à la limite, M tendant vers l'infini, $\Phi_w(f)$ tend vers une impulsion de Dirac. Dans ce cas, l'espérance mathématique s'identifie au spectre cherché $\Phi_x(f)$. On peut également montrer que la variance de l'estimateur modifié est donnée par :

$$\text{Var}[\bar{R}_x(f)] \cong \frac{1}{L} \Phi_x^2(f) \quad (\text{F.20})$$

Par opposition à l'estimateur adouci, il n'y a, dans ce cas aucune contrainte sur la fenêtre utilisée. Comme l'adoucissement est effectué par une fonction fenêtre spectrale qui est proportionnelle au carré du module d'une transformée de Fourier, l'estimateur modifié ne conduira jamais à des valeurs négatives.

Remarque : Il existe un large nombre de fenêtres de troncature, le choix d'une forme particulière d'une fonction fenêtre dépend principalement de la largeur du pic central et de l'amplitude des lobes secondaires. Quelques noyaux (fenêtre d'apodisation) couramment employés sont :

- Noyau de troncation : $w(t) = 1$ si $|t| \leq 1$, $= 0$ sinon ;
- Noyau de Bartlett (ou Féjer) : $w(t) = 1 - |t|$ si $|t| \leq 1$, $= 0$ sinon ;
- Noyau de Tukey-Hanning : $w(t) = [1 + \cos(\pi t)]/2$ si $|t| \leq 1$, $= 0$ sinon ;
- Noyau de Parzen : $w(t) = 1 - 6t^2 + 6|t|^3$ si $|t| \leq 1/2$, $= 2(1 - |t|^3)$ si $1/2 \leq |t| \leq 1$, $= 0$ si $|t| \geq 1$.

Annexe G

Séparation de sources vibratoires

Les machines tournantes sont utilisées dans des domaines aussi variés que le transport (trains, véhicules motorisés, ...), la production électrique (alternateurs), l'industrie de production, ou encore l'électroménager. Le dysfonctionnement d'une machine dans son ensemble peut causer de gros dégâts. C'est pourquoi, le diagnostic de l'état de santé de ces machines est primordial et la disposition de méthodes qui nous indiquent la déviation des machines de leur fonctionnement normal est d'un grand intérêt.

Le but de cet annexe est de présenter la problématique des machines tournantes et l'application de la technique de séparation de sources à leurs signaux vibratoires. Nous citons également les travaux de recherche traitant ce problème.

G.1 Introduction

L'analyse de "l'état de santé" des machines tournantes permet de faire une décision sur la normalité de leur fonctionnement, ce qui nous permet d'éviter des pertes énormes en production, mais aussi d'échapper à beaucoup de risques. Pour comprendre ce contexte, citons quelques exemples. Considérons une machine tournante opérationnelle, par exemple un mixeur ménager ou un engin électrique (dans une voiture par exemple). Ces machines produisent un bruit qui semble relatif à leur vitesse de rotation, *e.g.* mettre le mixeur dans le mode rapide produit un bruit aux fréquences hautes. L'engin est encore plus complexe car plusieurs sources vibratoires dans l'engin contribuent à la vibration globale. En effet, l'engin possède une structure mécanique plus complexe. Le conducteur habituel d'une voiture est capable de reconnaître un défaut de fonctionnement provenant de l'engin, lorsqu'il entend occasionnellement un bruit bizarre dans celui habituel de la voiture. Malheureusement, malgré la sensibilité du conducteur (par rapport à toute autre personne) à reconnaître ces défauts, il est impossible d'empêcher les problèmes qu'ils engendrent ! En effet, lorsque le conducteur arrive à entendre ces défauts, ils se sont déjà développés et ont atteint un niveau avancé, car les premiers moments de naissance de ces défauts, qui sont traduits par des contributions faibles au niveau du spectre, ont été masquées par les fréquences très énergétiques provenant des

autres sources vibratoires de l'engin. Imaginons seulement ce que ce serait si le conducteur disposait d'un calculateur de bord à base d'un dsp de séparation de sources vibratoires de machines tournantes, qui serait capable de diagnostiquer ce genre d'engin et détecter tout défaut dès sa formation . . . !

Et il y a de multiples autres exemples que nous pouvons citer :

- Le constructeur d'alternateurs électriques pourrait diagnostiquer son produit grâce à un système de séparation de sources vibratoires. En effet, il pourrait détecter tout dysfonctionnement au niveau du : calibrage des éléments de ses machines, contacts entre les éléments mobiles, . . . et toute imperfection de fabrication qui peut engendrer de mauvaises conséquences dans le fonctionnement de l'ensemble.
- Un autre important contexte où la séparation de sources vibratoires joue un rôle primordial est celui du contrôle industriel comme par exemple la commande des pompes d'une station d'eau. La commande de ces machines nécessite des mesures exactes et précises, mais malheureusement les sources vibratoires peuvent bien fausser les prélèvements effectués par les capteurs. Ypma a étudié ce problème dans sa thèse [113] et avait proposé des méthodes de séparation de sources pour cela.
- En médecine, des études ont aussi été menées dans ce sens sur le monitoring biomédical.

G.2 Vibrations d'une machine tournante

Quelque soit le contexte des machines tournantes où un système de contrôle et diagnostique à base de la technique de séparation de sources pourra être intégré, l'action visée de celui-ci sera en fonction des signaux mesurés (observés) par les capteurs. La structure contenant cette machine tournante appelée "*structure vibrante*" transmettra en conséquence ses vibrations aux capteurs de mesure. Tous les éléments de la structure (moteurs ou machines tournantes, éléments de couplage ou liaison, éléments mobiles de chaque machine) vont contribuer à la création des mélanges reçus par les capteurs. La figure G.1 illustre ce phénomène.

G.2.1 Définition et mesure des vibrations

Une vibration est traduite par un mouvement oscillatoire relatif à un point de la structure vibrante. C'est la réponse à une ou plusieurs forces excitatrices qui sont exprimées par une des grandeurs physiques suivantes : un déplacement, une vitesse ou une accélération [56, page 8].

La mesure de cette vibration est réalisée par un capteur appelé *accéléromètre piézo-électrique*, dont la sortie fournit une quantité de charges électriques proportionnelles à

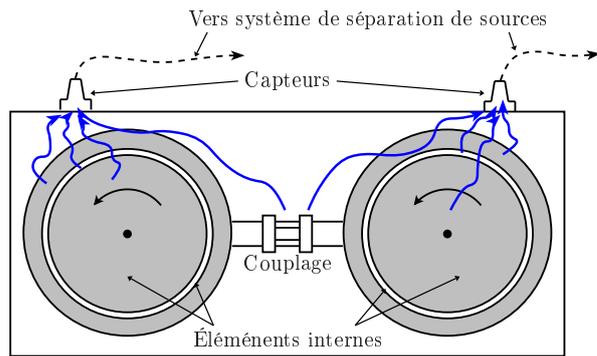


FIG. G.1 – propagation des sources vibratoires



FIG. G.2 – machine tournante

l'effet de la force d'excitation.

G.2.2 Origine des vibrations

Les vibrations d'une machine tournante résulte de son excitation par des forces qui peuvent être de différentes natures. C'est-à-dire que la structure possède une réponse mécanique qui engendre ces vibrations. La nature des forces excitatrices peut être aérodynamique, mécanique ou électromécanique.

G.3 Analyse et séparation des sources vibratoires

L'analyse et la séparation des sources vibratoires sont souvent basées sur l'estimation de leur spectre. Étant donnée la nature des forces excitatrices et leur caractère qui peut être périodique, aléatoire ou large bande, les signaux vibratoires qui leur correspondent auront le même comportement. L'allure de la densité spectrale de puissance (en fréquence) des signaux vibratoires est présentée sur la figure G.3.

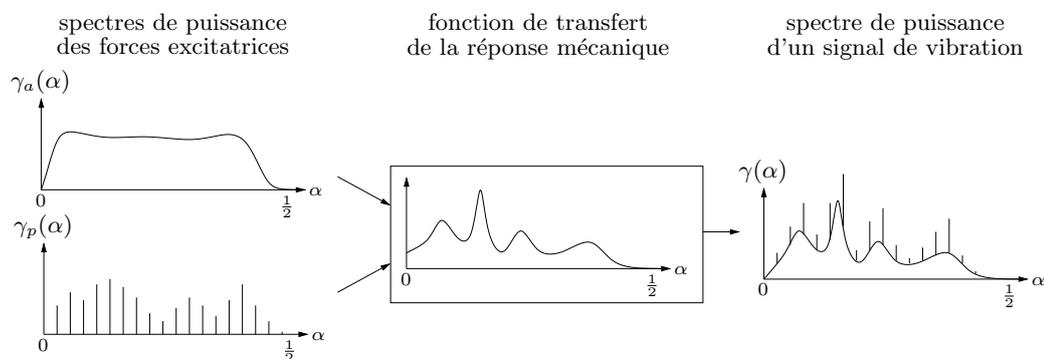


FIG. G.3 – dsp des signaux vibratoires

La densité spectrale de puissance $\gamma_p(f)$ des forces périodiques est un ensemble de raies d'amplitude variable. Celle des forces aléatoires $\gamma_a(f)$ a un caractère large bande. Par

l'intermédiaire de la réponse mécanique de la fonction de transfert, ces forces génèrent un signal vibratoire dont la dsp $\gamma(f)$ présente à la fois des raies spectrales et un caractère large bande.

G.3.1 Modèle convolutif de mélanges vibratoires

Les éléments principaux de cette modélisation sont les réponses impulsionnelles locales de la structure vibrante. Le canal de transmission entre un capteur et une source vibratoire (élément de la structure) est modélisé par un filtre linéaire [113]. Les caractéristiques modales de la structure mécanique génèrent les filtrages suivants : réverbération et dispersion des ondes vibratoires, induisant des copies modifiées et retardées des composantes fréquentielles. Dans le cas général, avec plusieurs capteurs et plusieurs sources vibratoires (interférences internes à la structure), la réponse de la structure est modélisée à la position j par :

$$\mathbf{x}_j = H_j * \mathbf{s} + \mathbf{b}_j \quad (\text{G.1})$$

où $*$ dénote le produit de convolution matriciel. Ce modèle a été introduit initialement par [33].

G.3.2 Séparation des mélanges vibratoires

Plusieurs algorithmes et travaux de recherche traitent la séparation des sources vibratoires de machines tournantes. Nous trouvons dans [113] un état de l'art complet sur ces travaux. Dans [65], le nombre de sources contribuant à la vibration multicanaux des machines tournantes est déterminé en utilisant une décomposition en valeurs singulières de la matrice spectrale des observations. De même, les travaux de [37, 67] pour la résolution du problème de séparation aveugle des sources vibratoires, sont basés sur une décomposition en valeurs singulières. Par ailleurs, Gelle *et al.* [52, 51] traitent ce problème en utilisant les statistiques d'ordre supérieur. Ypma dans sa thèse a proposé des algorithmes pour la séparation des mélanges vibratoires. Dans cette proposition [115, 116], l'auteur optimise une fonction de coût à base de statistiques d'ordre deux. Plusieurs variantes de cette fonction de coût ont été proposées. Les travaux récents [114] utilisent une forme bilinéaire pour résoudre le problème. Fabry [46] a proposé des méthodes de séparation de sources vibratoires. Il estime dans une première étape les matrices spectrales décalées des observations pour leur diagonalisation dans une deuxième étape. Il propose aussi une méthode fondée sur la maximisation du rapport signal à bruit en utilisant le filtrage de Wiener. En faisant une décomposition en valeurs singulières, il estime le nombre des sources.

Résumé

Dans cette thèse, la Séparation Aveugle de Mélanges Convolutifs de Sources est étudiée. Pour la séparation des mélanges audio, nous avons développé des méthodes nouvelles pour les cas avec bruit et sans bruit dans l'environnement de propagation. La méthode sans bruit est basée sur la diagonalisation conjointe des matrices spectrales et exploite la non stationnarité des signaux. Nous avons proposé deux techniques différentes pour résoudre le problème de permutation. La deuxième méthode, où un bruit additif est présent, est basée sur le maximum de vraisemblance. La simulation des méthodes est réalisée sur des données réelles de salles acoustiques.

Mots-clés

Séparation aveugle de sources, mélanges convolutifs, non stationnarité, bruit additif, ambiguïté de permutation, réverbération, salles acoustiques.

Abstract

In this thesis, Blind Source Separation (BSS) of Convolutional Mixtures of Sources is addressed. For separating audio mixtures, we have developed new methods for cases without noise and with noise in propagation environment. The method for case without noise is based on joint diagonalization of spectral matrices and exploit the non stationarity of signals. We have proposed two techniques in order to solve permutation problem. The second method where an additive noise is present, is based on maximum likelihood. Simulations were done with real data of acoustics room.

Keywords

Blind source separation, convolutional mixtures, non stationary, additive noise, permutation ambiguity, reverberation, acoustics room.
