



HAL
open science

Méthodes de séparation aveugle de sources fondées sur des transformées temps-fréquence. Application à des signaux de parole.

Mathieu Puigt

► **To cite this version:**

Mathieu Puigt. Méthodes de séparation aveugle de sources fondées sur des transformées temps-fréquence. Application à des signaux de parole.. Astrophysique [astro-ph]. Université Paul Sabatier - Toulouse III, 2007. Français. NNT: . tel-00270811

HAL Id: tel-00270811

<https://theses.hal.science/tel-00270811>

Submitted on 7 Apr 2008

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.

Université Toulouse III – Paul Sabatier
UFR Physique, Chimie, Automatique
Ecole doctorale Mathématiques, Informatique et Télécommunications de Toulouse

Méthodes de séparation aveugle de sources fondées sur des transformées temps-fréquence. Application à des signaux de parole

THÈSE

présentée et soutenue publiquement le 13 décembre 2007

pour l'obtention du grade de

Docteur de l'Université de Toulouse
délivré par l'Université Toulouse III – Paul Sabatier

Discipline : Traitement du Signal

par

Matthieu PUIGT

Composition du jury

- Président* : Pierre Comon, Directeur de recherche CNRS, I3S, Sophia-Antipolis
- Rapporteurs* : Ali Mansour, Enseignant-chercheur DGA, E3I2, ENSIETA, Brest
Eric Moreau, Professeur, MS, Université de Toulon-ISITV
- Examineurs* : Jean-Philippe Bernard, Chargé de recherche CNRS, CESR, Toulouse
Shahram Hosseini, Maître de conférences, LATT, Université de Toulouse
- Directeur de thèse* : Yannick Deville, Professeur, LATT, Université de Toulouse

Laboratoire d'Astrophysique de Toulouse-Tarbes — UMR 5572

Mis en page avec la classe thloria.

Remerciements

Je me souviendrai toujours de cette discussion. Yannick Deville m'avait appelé dans son bureau. Nous attendions le retour d'un organisme pour lancer officiellement ma thèse. "*J'ai une mauvaise et une bonne nouvelle.*" La mauvaise nouvelle, c'était que cet organisme ne financerait pas mes travaux... La bonne, c'est que Yannick me finançait à partir des reliquats d'anciens contrats! Mille mercis Yannick pour ce cadeau que tu me fis. Tu as su non seulement te montrer un directeur de thèse disponible et à l'écoute mais tu as aussi été un exemple pour ta rigueur scientifique, ton approche de l'enseignement et ton art d'organiser des pots de première classe!

Je tiens à remercier vivement les membres de mon jury pour leur participation. Je suis en particulier très reconnaissant envers Ali Mansour et Eric Moreau pour leur examen minutieux qui m'aura permis, je l'espère, d'améliorer la qualité de ce document. Merci à Pierre Comon d'avoir accepté d'être "Monsieur le Président". Merci à Jean-Philippe Bernard d'avoir eu le courage de lire ces quelques 200 pages d'obscures méthodes de traitement du signal et su éclairer la présentation orale de son expertise en astrophysique. Enfin, merci à Shahram Hosseini pour ses nombreux conseils et ses remarques constructives.

Je voudrais remercier aussi tous ceux qui m'ont permis de "faire un petit peu avancer la science" durant ces trois ans. Outre mes collègues et nos pauses scientifico-caféinées, je tiens à remercier en particulier Olivier Berné pour les nombreuses discussions sur les résultats que j'ai obtenus à partir de *ses* spectres astrophysiques, et Greg Hamerly qui m'a gentiment fourni son algorithme de classification G-means et qui s'est montré réactif devant mes nombreuses interrogations. Merci au bureau 54 (Karine, Paul et Guilhem) et à mes collègues de bureau (Rima et Johan) de m'avoir si souvent laissé leurs ordinateurs et de gros stocks de madeleines pendant les nuits et week-ends avant la remise du manuscrit!

Merci à mon PC qui a eu la bonne idée de ne pas me faire faux bond (et heureusement vu la fréquence des sauvegardes qu'il a subi)!

Mon parcours n'a pas été très académique puisque mes premiers travaux de recherche ont eu lieu dans un laboratoire d'acoustique, avant d'émigrer vers le Laboratoire d'Astrophysique de Toulouse-Tarbes. Alors, comme au cinéma, les remerciements des paragraphes suivants se feront par ordre d'apparition.

Merci aux "anciens du LAMI" qui n'ont pas migré avec nous et en particulier merci à Vincent, les deux Pierre, Jean-Louis et les (post-)doctorants et stagiaires de l'époque – Paola, Benoit A. (et Ingrid), Nico, Benoit M., Valentin, Jeanne, Johanna, Lionel (et Ludi). Merci pour les nombreuses pauses café (*Nespresso, what else?*), les discussions enrichissantes, les sorties à la montagne "pour poser des capteurs" (et les séances de ski/raquettes qui en découlaient), les grillades de fin d'année, les apéros du jeudi... Un merci spécial à Benoit : travailler ensemble a été un grand bonheur.

J'ai aussi une pensée pour mes compères du DEA. Arno et Vince, finalement la promo "*Complètement c... cette année*" © ne s'en sera pas trop mal sortie!

Merci aux personnels de l'OMP et du LATT, pour leur accueil et leur gentillesse. Un remerciement particulier à Josianne, Dominique, Alexandre et Eric pour leur sympathie.

Merci aux (ex-)membres de l'équipe S2I qui m'ont accueilli chaleureusement : merci à vous Eric, Sylvie, Jean, Hervé, Yannick, Shahram, Laurent, José-Philippe, Jean-François, Jean-Louis, Guilhem, Olivier, Sébastien, Karine, Rima, Ali, Christophe, Hicham, Denis, Johan, Chahinez et Truswin.

Ma vie dans ce laboratoire n'aurait pas été la même sans la présence des doctorants du LATT et de l'OMP, des (ex-)précaires et autres stagiaires. L'intégration a particulièrement été précoce et réussie puisque l'après-midi même de mon arrivée, j'ai été convié à un pot... "Special thanks" à Karine, Rima, Flo, Valérie, Denis, Guilhem, Loic, Hervé, Laurent, Paul, Johan, Alex, Ludo, Olivier, Christophe, Dass, Benjamin, Célia, Marc, Truswin, Jiri pour les nombreuses (et réussies) soirées/grillades, les pauses "à la cafet' d'en haut" (avec le Monde, le Canard et Charlie), les discussions interminables, les madeleines et autres gourmandises, les préparations des JdT et les *pirletta* maladroitement cachés ici et là...

En dehors de cette honorable bâtisse qu'est l'OMP, il m'est arrivé d'avoir une vie sociale. Je souhaite donc remercier ici certaines personnes qui me sont chères.

Merci aux amis de Perpignan (*Ze place to be*) : Pierre – suis dispo pour le grand retour des *Zippos Campent* –, Sophie et Wilfrid, Camille et Vincent, Clem, Elsa, Célia, Audrey et Mathieu, Géraldine, Delphine, Vlad, Béné, Benja, José, Carole, Frank, Corinne et Nico, Julie et Philippe, Laure et Cyril, Carine et Flo, Valé et Ped', Elie et Fred, Isa et Jordí, Julien L. et Marie-Victoria, Julien M., OJO (Olivia, Jules et Olivier), Pascal S., Cécile, son Pascal et Jade, Stef' "avec un f", Christine, Mickey, Steph et Pierre, Caroline et Pierre, Maud et Eric, Isabelle et David, les anciens de la TS1, Yvan et Lionnel.

Merci aux "toulousains" : Nass et Seb, Neus, Kirsten, Elinor, Gemma, Tom, Ian, Arno, le Vince, Hicham et ses femmes, Marc et sa petite famille, Séverine et Bruno, Patrice, les amis de Seb de l'IUFM...

Merci aux amis motards pour les superbes balades/soirées/rigolades (rayez les mentions inutiles) qu'on a vécues ensemble. En particulier, merci à la Cricri family, Feuf et Nonce, Clin d'Phare, Sir La DUF, Snake & Frog, Sp!ke, Laurent "Guignol" et les Meugeux, Philou, Djé et Daï, PCR et Photographe, le team 31 de FA2, Joel, Dams et les Motos34, les 17 pouciens, les supermotards catalans et beaucoup d'autres...

Ami lecteur, si tu as été malencontreusement oublié (et injustement lésé), passe me voir : on arrangera ça autour d'un verre...

Il est probable que je n'eus jamais fait de "thèse en karaoké" si je n'avais appris le piano avec Gilbert Lafforgue. *Monsieur Lafforgue*, vous m'avez transmis la passion de la musique, au point que j'ai voulu mêler cette passion à mon activité scientifique. Même si, de là haut, vous ne pouvez lire ces mots, merci pour tout ce que vous m'avez apporté.

Je profite de ces quelques lignes pour remercier mes parents. Vos efforts n'auront pas été vains et une bonne part de ce diplôme vous revient. Merci à mes grand-parents et à mes frères & "associées" pour tout leur soutien. Merci à toute ma famille qui a essayé de comprendre ce que je faisais.

Enfin, comment résumer en quelques mots tout ce que tu as "enduré" pour moi Sophie ? Une thèse n'y suffirait pas... Alors, simplement merci de me supporter et m'encourager si assidûment.

*A ma famille
A Sophie
Au "mini-nous" qui arrivera bientôt*

Table des matières

Table des figures	xiii
Liste des tableaux	xv
Notations	xvii
Introduction générale	1
1 Nature des mélanges	1
1.1 Présentation des différents types de mélange	1
1.2 Concept de séparation en fonction du mélange : du modèle ins- tantané au mélange convolutif	3
2 Mélanges sur- ou sous-déterminés	6
3 Plan du manuscrit	7
Chapitre 1 Panorama des méthodes de SAS	9
1.1 Critères de l'ACI	9
1.1.1 ACI par maximisation de la non-gaussianité	10
1.1.2 ACI par estimation du maximum de vraisemblance	12
1.1.3 ACI basée sur la minimisation de l'information mutuelle	13
1.1.4 ACI basée sur les statistiques d'ordre 2	14
1.1.5 ACI basée sur une approche tensorielle	16
1.2 Méthodes basées sur la factorisation en matrices non-négatives	17
1.3 Méthodes basées sur la parcimonie des sources	17
1.3.1 Méthodes issues des approches statistiques classiques	22
1.3.2 Méthodes utilisant des transformations parcimonieuses linéaires	28
1.4 Conclusion	41

Partie I	Méthodes linéaires instantanées	43
Chapitre 2	Variantes des méthodes LI-TIFROM et LI-TIFCORR	45
2.1	Méthodes LI-TEMPROM et LI-TIFROM	45
2.1.1	Approches initiales de F. Abrard et Y. Deville [2]	45
2.1.2	Les limitations de LI-TIFROM	51
2.1.3	Amélioration des méthodes LI-TEMPROM et LI-TIFROM	54
2.2	Cas N sources et N observations : utilisation de la moyenne	57
2.3	Méthodes LI-TEPCORR et LI-TIFCORR	58
2.3.1	Approches initiales de Y. Deville [74]	58
2.3.2	Variantes non-centrées de LI-TIFCORR	63
2.4	Conclusion	64
Chapitre 3	Performances des méthodes linéaires instantanées	65
3.1	Mélanges de $N = 2$ sources et $N = 2$ observations	66
3.1.1	Sources utilisées	66
3.1.2	Première série de tests avec une matrice de mélange fixée	67
3.1.3	Deuxième série de tests : influence des coefficients de mélange	72
3.1.4	Comparaison entre nos versions de LI-TIFROM et la version originale	74
3.1.5	Performances des méthodes LI-TEPCORR-C et LI-TEPCORR-NC	75
3.1.6	Comparaison avec une autre méthode à parcimonie TF : LI-TIFCOHERE	76
3.1.7	Comparaison aux méthodes de SAS classiques	77
3.2	Mélanges de $N = 4$ sources et $N = 4$ observations	77
3.2.1	Apport du calcul de la moyenne dans l'étape de détection pour $N = 4$ sources	79
3.2.2	Performances des méthodes TF	80
3.3	Performances des méthodes étendues	82
3.3.1	Cas sous-déterminé	82
3.3.2	Sources cachées	85
3.4	Tests complémentaires : application à l'astrophysique	86
3.5	Conclusion sur les approches LI de SAS	93

**Partie II Méthodes convolutives : du mélange à atténuations
et décalages au modèle convolutif général 95**

Introduction 97

Chapitre 4 Extension de LI-TIFCORR aux sources atténuées et retardées 99

4.1 Position du problème 99

4.2 Première extension de LI-TIFCORR aux mélanges AD 101

4.2.1 Version de base des étapes de détection et d'identification pour les paramètres b_{im} 102

4.2.2 Etapes de détection et identification des paramètres μ_{im} 103

4.2.3 Couplage des paramètres b_{im} et $\mu_{im'}$ 108

4.2.4 Résumé de la méthode AD-TIFFROM-CF 109

4.3 Seconde extension de LI-TIFCORR aux mélanges AD 109

4.3.1 Etapes alternatives de détection et d'identification 109

4.3.2 Résumé de la méthode AD-TIFFROM-CT 111

4.4 Conclusion 111

Chapitre 5 Apport des techniques de classification non-supervisée aux méthodes de SAS à parcimonie 113

5.1 Les méthodes de classification non-supervisée classiques 114

5.1.1 Les algorithmes hiérarchiques 114

5.1.2 Classification basée sur le partitionnement des données 116

5.1.3 Les méthodes basées sur les mélanges de densités 118

5.2 Les algorithmes proposés pour les méthodes de SAS à parcimonie . . . 120

5.2.1 CBCE : une méthode d'élimination de clusters basée sur la cardinalité 120

5.2.2 l'algorithme DEMIX de S. Arberet et al. [12, 13] 121

5.2.3 Notre approche modifiée de K-means pour la séparation de sources 122

5.2.4 Description des approches de SAS utilisant notre variante de K-means 123

5.3 Conclusion 125

Chapitre 6 Performances des méthodes à atténuations et décalages	127
6.1 Conditions des tests et critères de performances	127
6.2 Comparaison des méthodes de classification	130
6.2.1 Classification dans les méthodes à fréquence constante	130
6.2.2 Classification dans les méthodes à temps constant	133
6.3 Performances des méthodes de SAS	136
6.3.1 Comportement des méthodes AD-TIFCORR-CF et AD-TIFFROM- CF	136
6.3.2 Performances des méthodes AD-TIFCORR-CT et AD-TIFFROM- CT et comparaison avec les méthodes à fréquence constante . .	140
6.4 Conclusion	142
Chapitre 7 Extension des méthodes parcimonieuses aux mélanges convo- lutifs généraux	145
7.1 Position du problème	145
7.2 Extension de TIFCORR-NC aux mélanges convolutifs	147
7.3 Conclusion et discussion	148
Conclusion générale	149
Annexes	153
Annexe A Influence d'un faible niveau de bruit sur l'étape de détec- tion	153
Annexe B Démonstration du critère mono-source de TEMPCORR et TIFCORR	155
B.1 Démonstration du critère dans le cadre de mélanges LI	155
B.1.1 Démonstration du critère pour LI-TEMPCORR-C	155
B.1.2 Démonstration du critère pour LI-TIFCORR-C	156
B.1.3 Démonstration du critère pour LI-TIFCORR-NC	157
B.2 Démonstration du critère pour AD-TIFCORR-CT	157
Annexe C SIR des signaux mélangés et critères de performance des méthodes de SAS	159
C.1 SIR pour les mélanges LI	159
C.2 SIR pour les mélanges AD	162

C.3	SIR pour les mélanges convolutifs	163
Annexe D	Extension de LI-TIFROM aux mélanges atténués et retardés	167
D.1	Description de l'approche AD-TIFROM-CF	167
D.1.1	Étapes de détection basique et identification pour les paramètres b_{im}	168
D.1.2	Résumé global de la méthode AD-TIFROM-CF	169
D.2	Description de l'approche AD-TIFROM-CT	170
Annexe E	Atténuations négatives dans les méthodes AD	173
Annexe F	Limiter numériquement le nombre de zones d'analyse monosources : validation expérimentale	175
Bibliographie		179
Publications scientifiques de l'auteur		199

Table des figures

1	Exemple de mélange AD pour $N = 2$ sources et $P = 2$ capteurs : la propagation des ondes sonores se fait selon un trajet unique des sources aux capteurs et le temps de propagation n'est pas négligeable.	2
2	Exemple de mélange convolutif pour $N = 2$ sources et $P = 2$ capteurs. . .	3
1.1	Signal de parole parcimonieux dans deux domaines d'analyse différents. <i>A gauche</i> : domaine temporel. <i>A droite</i> : domaine temps-fréquence.	18
1.2	Classement des méthodes de SAS à parcimonie en fonction du degré de parcimonie des sources.	18
1.3	Nuage de points pour un mélange stéréo de 4 sources parcimonieuses. Dans cet exemple, les amplitudes correspondent au module des transformées TF des observations. Les droites correspondent aux directions des sources, c.-à-d. aux colonnes de la matrice de mélange.	20
1.4	Exemple de minimisation par la norme ℓ^1 du problème P_1 . En n'utilisant que deux des directions \vec{a}_1 , \vec{a}_2 et \vec{a}_3 , le chemin le plus court entre l'origine et le point C est O-B-C ou O-B'-C, où C est un point du nuage de points, \vec{a}_i la direction de la $i^{\text{ème}}$ source et A' (resp. B', A et B) est la projection de C, selon la direction \vec{a}_1 (resp. \vec{a}_2 , \vec{a}_3 et \vec{a}_3), sur la droite de direction \vec{a}_3 (resp. \vec{a}_3 , \vec{a}_2 et \vec{a}_1). Ainsi, par minimisation de la norme ℓ^1 , C sera exprimé selon les directions \vec{a}_2 et \vec{a}_3	21
2.1	Etapes des méthodes parcimonieuses LI de Y. Deville et al.	47
3.1	Ensemble des signaux sources testés.	67
3.2	Spectre d'émission des différentes composantes de poussière du MIS d'après Désert et al. [88].	86
3.3	<i>A gauche</i> : image à $7 \mu m$ extraite du cube de données de la nébuleuse Ced 201. <i>A droite</i> : spectre extrait de ce même cube sur le pixel marqué d'une croix dans l'image de gauche.	88
3.4	<i>A gauche</i> : spectres obtenus avec FastICA pour Ced 201. <i>A droite</i> : spectres obtenus par FMN pour Ced 201. Avec les deux méthodes, on reconnaît en haut le spectre des PAH et en bas celui des VSG.	88
3.5	Spectres obtenus après centrage, normalisation et ACP. Ces spectres correspondent aux observations traitées par nos méthodes de SAS.	89
3.6	Spectres obtenus avec LI-TIFROM-S2	90

3.7	Spectres obtenus avec LI-TIFCORR-C.	90
3.8	Spectres obtenus avec LI-TIFCORR-NC.	91
3.9	<i>En haut</i> : spectres obtenus avec LI-TEMPROM-S2. <i>Au milieu</i> : spectres obtenus avec LI-TEPCORR-C. <i>En bas</i> : spectres obtenus avec LI-TEPCORR-NC.	92
4.1	Phase déroulée $\phi'_2(n, \omega)$ selon la pulsation ω pour un indice temporel n fixé. Ici le signal $x_2(n)$ est juste une version décalée de quelques échantillons de $x_1(n)$, si bien que $\phi'_2(n, \omega)$ est une fonction linéaire de ω (aux erreurs d'estimation de la phase et au décrochement d'un multiple de 2π près). . .	107
5.1	Exemple de clustering hiérarchique ascendant.	115
F.1	Histogrammes de dispersion des paramètres de mélanges estimés avec AD-TIFFROM-CF, en fonction du nombre de points utilisés. <i>En haut</i> : tous les points qui vérifient l'Eq. (5.13) sont utilisés. <i>En bas</i> : Seuls les mille points qui maximisent ce critère (5.13) le sont.	176

Liste des tableaux

3.1	Performances (SIR^{out} en dB) en fonction du jeu de sources.	68
3.2	Performances des méthodes TF de SAS en fonction de la taille des fenêtres utilisées dans le calcul des TFCT (en nombre d'échantillons).	70
3.3	Performances des méthodes TF de SAS en fonction du nombre de fenêtres TF dans les zones d'analyse.	70
3.4	Performances des méthodes TF de SAS en fonction du recouvrement temporel des fenêtres TF.	71
3.5	Performances des méthodes TF de SAS en fonction de la matrice de mélange.	73
3.6	Performances des méthodes LI-TIFROM en fonction de la matrice de mélange.	75
3.7	Performances des méthodes temporelles de SAS pour chaque ensemble de deux sources de parole et performances globales moyennes.	76
3.8	Performances globales des méthodes proposées.	77
3.9	(1) SIR^{out} (en dB) des méthodes LI-TIFROM-S2, LI-TIFCORR-C et LI-TIFCORR-NC : (i) moyenne des SIR^{out} sur toutes les valeurs des paramètres ; (ii) SIR^{out} pour les paramètres optimaux. (2) SIR^{out} (en dB) des méthodes classiques.	78
3.10	Performances globales des méthodes proposées.	79
3.11	Performances des méthodes proposées pour $N = 4$ en fonction de la taille des fenêtres utilisées dans le calcul des TFCT (en nombre d'échantillons).	80
3.12	Performances des méthodes TF de SAS en fonction du nombre de fenêtres TF dans les zones d'analyse.	81
3.13	Performances des méthodes TF de SAS par rapport au recouvrement temporel des fenêtres TF.	81
3.14	Performances des méthodes TF de SAS en fonction de la taille des fenêtres TF utilisées dans le calcul des TFCT.	83
3.15	Performances de la méthode LI-TIFCORR-C pour chaque jeu de sources et performances globales pour les six jeux.	84
3.16	Performances de la méthode LI-TIFCORR-C pour chaque jeu de sources et performances globales pour les six jeux.	84
6.1	Norme de Frobenius des différences entre les matrices théoriques de paramètres b_{im} et leur estimées obtenues par AD-TIFROM-CF, en fonction de la méthode de classification utilisée, du jeu de sources et des paramètres λ et η de la matrice de mélange définie dans l'Eq. (6.2).	131

6.2	Moyennes des normes de Frobenius, obtenues par chaque méthode de classification avec AD-TIFROM-CF, en fonction de η et valeurs globales pour $\eta = 0-20$	132
6.3	Performances de AD-TIFROM-CT avec les méthodes de classification, pour chaque jeu de sources, chaque valeur de λ , $\eta = 0, 10$ ou 20 et performances moyennes pour $\eta = 0-20$	133
6.4	Performances globales de AD-TIFROM-CT, pour chaque valeur de λ et chaque méthode de classification, en fonction du nombre d' d'échantillons dans le calcul des TFCT.	135
6.5	Norme de Frobenius des différences entre les matrices théoriques de paramètres b_{im} et leurs estimées obtenues par AD-TIFCORR-CF et AD-TIFROM-CF, en fonction des paramètres λ et η de la matrice de mélange définie dans l'Eq. (6.2).	136
6.6	Performances des méthodes AD-TIFCORR-CF et AD-TIFROM-CF pour chaque jeu de sources, chaque valeur de λ , $\eta = 0, 10$ ou 20 et performances moyennes pour $\eta = 0-20$	137
6.7	Performances globales de AD-TIFCORR-CF et AD-TIFROM-CF pour les deux jeux de sources, $\eta = 0-20$ et chaque valeur de λ , en fonction du recouvrement ρ' entre les fenêtres TF.	138
6.8	Performances globales de AD-TIFCORR-CF et AD-TIFOMR-CF pour les deux jeux de sources, $\eta = 0-20$ et chaque valeur de λ , en fonction de la largeur temporelle d' des TFCT (en échantillons).	139
6.9	Performances globales des méthodes AD-TIFCORR-CF et AD-TIFROM-CF pour les deux jeux de sources, $\eta = 0-20$ et pour chaque valeur de λ , en fonction de la largeur des zones d'analyse à temps constant (en Hz).	139
6.10	Performances de AD-TIFCORR-CT et AD-TIFROM-CT, pour chaque valeur de λ , $\eta = 0, 10$ ou 20 et performances moyennes pour $\eta = 0-20$	140
6.11	Performances globales de AD-TIFCORR-CT et AD-TIFROM-CT pour les deux jeux de sources, $\eta = 0-20$ et chaque valeur de λ , en fonction du recouvrement ρ' entre les fenêtres TF.	141
6.12	Performances globales de AD-TIFCORR-CT et AD-TIFROM-CT, pour les deux jeux de sources, chaque valeur de λ , $\eta = 0-20$, en fonction de la taille d' des TFCT (en échantillons).	142
6.13	Performances globales des méthodes AD-TIFCORR-CT et AD-TIFROM-CT pour les deux jeux de sources, $\eta = 0-20$ et pour chaque valeur de λ , en fonction de la largeur des zones d'analyse à temps constant (en Hz).	143
F.1	Norme de Frobenius des différences entre les matrices théoriques de paramètres b_{im} et leur estimées obtenues par AD-TIFROM-CF, en fonction de la méthode de classification utilisée, du jeu de sources et des paramètres λ et η de la matrice de mélange définie dans l'Eq. (6.2). Le nombre de zones mono-sources n'a pas été limité.	176

Notations

Sigles

- ACI (ou *ICA*) : Analyse en Composantes Indépendantes (ou *independent component analysis*).
- ACP (ou *PCA*) : Analyse en Composantes Principales (ou *principal component analysis*).
- AD (mélanges) : mélanges à Atténuations et Décalages temporels (ou *Attenuated and Delayed mixtures*).
- AMUSE : *Algorithm for Multiple Unknown Signals Extraction* (méthode de séparation aveugle de sources).
- CBCE : *Cardinality-Based Cluster Elimination* (méthode de classification).
- -CF : (Méthodes temps-fréquence de séparation de sources à) fréquence constante (ou *Constant-Frequency time-frequency source separation methods*).
- -CT : (Méthodes temps-fréquence de séparation de sources à) temps constant (ou *Constant-Time time-frequency source separation methods*).
- DEMIX : *Direction Estimation of Mixing matrIX* (méthode de séparation aveugle de sources).
- DESPRIT : *Duet-ESPRIT* (méthode de séparation aveugle de sources).
- DUET : *Degenerated Unimixing and Estimation Technique* (méthode de séparation aveugle de sources).
- DOA : direction d'arrivée (ou *Direction Of Arrival*).
- EMD : décomposition modale empirique ou *Empirical Mode Decomposition* en anglais.
- ESPRIT : *Estimation of Signal Parameters via Rotational Invariance Techniques* (méthode d'estimation des directions d'arrivée).
- FMN (ou *NMF*) : Factorisation en Matrices Non-négatives (ou *Non-negative Matrix Factorization*).
- FOCUSS : *FOcal Underdetermined System Solver*.
- GCC-PHAT : *Generalized Cross Correlation with PHase Transform*.
- GS : *Gap Statistics* (méthode de classification).
- IR : Infra-Rouge.
- JADE : *Joint Approximative Diagonalization of Eigen-matrices* (méthode de séparation aveugle de sources).
- LI (mélanges) : mélanges Linéaires Instantanés (ou *Linear Instantaneous mixtures*).
- MAP : probabilité maximum a posteriori (ou *Maximum A posteriori Probability*).

- MENUET : *Multiple-sENsor dUET* (méthode de séparation aveugle de sources).
- MIS : Milieu Inter-Stellaire.
- PAH : hydrocarbures polycycliques aromatiques pour *Polycyclic Aromatic Hydrocarbons*.
- PDR : régions de photodissociation pour *Photo-Dissociation Regions*.
- RIF (ou *FIR*) : Réponse Impulsionnelle Finie (ou *Finite Impulse Response*).
- RII (ou *IIR*) : Réponse Impulsionnelle Infinie (ou *Infinite Impulse Response*).
- RSB (ou *SNR*) : Rapport Signal sur Bruit (ou *Signal-to-Noise Ratio*).
- SAS (ou *BSS*) : Séparation Aveugle de Sources (ou *Blind Source Separation*).
- SCA : analyse en composantes parcimonieuses (ou *Sparse Component Analysis*).
- SIR : rapport signal sur interférence pour *Signal-to-Interference Ratio*. On note $SIRI(i)$ (pour *SIR improvement*) l'amélioration entre le SIR mesuré sur l'entrée i du système de SAS, noté $SIR_{in}(i)$, et le SIR mesuré sur la sortie i , noté $SIR_{out}(i)$.
- SOBI : *Second Order Blind Identification* (méthode de séparation aveugle de sources).
- SONS : *Second Order Non-stationary source Separation* (méthode de séparation aveugle de sources).
- SSCL : *Self-Splitting Competitive Learning* (méthode de classification).
- TE (ou *TS*) : Temps-Echelle (ou *Time-Scale*).
- TF : Temps-Fréquence (ou *Time-Frequency*).
- TFD (ou *DFT*) : Transformée de Fourier Discrète (ou *Discrete Fourier Transform*).
- TFCT (ou *STFT*) : Transformée de Fourier à Court Terme (ou *Short Time Fourier Transform*).
- TEMPCORR : *TEMPoral CORRelation-based bss method* (méthode de séparation aveugle de sources). Cette méthode peut utiliser des paramètres centrés (resp. non-centrés), d'où le nom de cette variante : TEMPCORR-C (resp. TEMPCORR-NC).
- TEMPROM : *TEMPoral Ratio Of Mixtures* (méthode de séparation aveugle de sources).
- TIFCOHERE : *TIme-Frequency COHEREnce-based bss method* (méthode de séparation aveugle de sources).
- TIFCORR : *TIme-Frequency CORRelation-based bss method* (méthode de séparation aveugle de sources). Cette méthode peut utiliser des paramètres centrés (resp. non-centrés), d'où le nom de cette variante : TIFCORR-C (resp. TIFCORR-NC).
- TISCORR : *TIme-Scale CORRelation-based bss method* (méthode de séparation aveugle de sources).
- TIFROM : *TIme-Frequency Ratio Of Mixtures* (méthode de séparation aveugle de sources). Les variantes de cette méthode, proposées dans ce document, font appel à des rapports symétrisés d'où leurs noms respectifs TIFROM-S1 et TIFROM-S2.
- TT : Temps-Temps (ou *Time-Time*).
- UV : Ultra-Violet.
- VSG : très petits grains pour *Very Small Grains*.
- WDO : orthogonalité W-disjointe ou orthogonal W-disjoint (pour *W-Disjoint-Orthogonal(ity)*).

Symboles mathématiques

- T : symbole de transposition. La transposée d'une matrice M sera notée M^T tandis qu'un vecteur \underline{v} aura pour transposée \underline{v}^T .
- N : nombre de sources présentes dans les observations.
- P : nombre d'observations disponibles.
- $\underline{s}(t)$ (resp. $\underline{s}(n)$) désigne le vecteur colonne dont les éléments sont les sources $s_i(t)$ (resp. $s_i(n)$).
- $\underline{x}(t)$ (resp. $\underline{x}(n)$) désigne le vecteur colonne dont les éléments sont les observations $x_i(t)$ (resp. $x_i(n)$).
- $\underline{y}(t)$ (resp. $\underline{y}(n)$) désigne le vecteur colonne dont les éléments sont les signaux de sortie $y_i(t)$ (resp. $y_i(n)$).
- A désigne l'opérateur de mélange. Dans le cas des mélanges LI, c'est une matrice. Dans le cadre de mélanges AD ou convolutifs, cette matrice varie avec les fréquences et s'écrit $A(\omega)$.
- B désigne l'opérateur de mélange permuté et normalisé par rapport à la première observation (choisie comme référence). Dans le cas des mélanges LI, c'est une matrice. Dans le cadre de mélanges AD ou convolutifs, cette matrice varie avec les fréquences et s'écrit $B(\omega)$.
- $\|\cdot\|$: norme mathématique.
- $\mathbb{E}\{\cdot\}$: espérance mathématique.
- $\mathbb{P}(\cdot)$: probabilité.
- $f_u(\cdot)$: densité de probabilité de la variable aléatoire u . En particulier, les densités de probabilité conjointe des sources $s_i(t)$ sont notées $f_{\underline{s}}(\cdot)$ et les densités marginales $f_{s_i}(\cdot)$.
- $\mathbb{1}$: fonction indicatrice.

Introduction générale

La séparation aveugle de source (SAS) consiste à estimer un jeu de N signaux sources inconnus à partir d'un jeu de P observations. Ces observations sont des mélanges de ces sources et proviennent généralement de différents capteurs (antennes radiofréquences, microphones, caméras par exemple). Le mélange entre ces sources, qui s'effectue pendant leur propagation jusqu'aux capteurs, est inconnu. L'histoire de la SAS a été racontée dans [136] par C. Jutten et A. Taleb.

La SAS est une "discipline générique" qui permet de nombreuses applications [72, 75, 133] telles l'acoustique et le traitement audio [96, 176, 234], les communications numériques [79, 130, 216], le génie biomédical [73, 137, 204], la classification d'images [45], l'astrophysique [34, 190], l'acoustique sous-marine [161] ou encore les données financières [22, 59, 183]. Elle comporte plusieurs types de difficultés qui dépendent des paramètres que voici :

1. la nature des mélanges,
2. le nombre P d'observations par rapport au nombre N de sources.

1 Nature des mélanges

1.1 Présentation des différents types de mélange

Nous comptons dans la littérature plusieurs types de mélanges. Une première catégorie, largement étudiée, concerne les mélanges *linéaires*, c.-à-d. le cas où les observations sont des mélanges linéaires des sources. Cette catégorie est découpée en trois sous-catégories que voici.

Définition 1 *On parle de mélanges "linéaires instantanés" (LI) lorsque les valeurs des observations à un instant s'écrivent comme des combinaisons linéaires des valeurs des sources au même instant.*

En particulier, dans le cas où les sources sont des signaux mono-dimensionnels notés $s_j(t)$, les observations $x_i(t)$ sont de la forme

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

où a_{ij} est un réel constant qui correspond à l'amplitude de la contribution de la source j sur le capteur i . Ce modèle est bien adapté à divers signaux de télécommunications ou signaux

biomédicaux par exemple. Il n'est par contre pas représentatif des réalités acoustiques dans le cadre de mélanges réels. Cependant, sous réserve qu'aucun effet de post-traitement ne soit appliqué, les mélanges LI modélisent généralement bien les mélanges de sources de parole ou de musique, réalisés sur table de mixage, en studio.

Définition 2 On parle de mélanges à "atténuations et décalages temporels" (AD) lorsque les observations s'écrivent comme des combinaisons linéaires de sources décalées temporellement.

En reprenant les notations précédentes, le mélange s'écrit

$$x_i(t) = \sum_{j=1}^N a_{ij} s_j(t - t_{ij}) \quad i = 1, \dots, P, \quad (2)$$

où a_{ij} et t_{ij} sont des réels constants qui correspondent resp. à l'amplitude de la contribution et au temps de propagation de la source j jusqu'au capteur i .

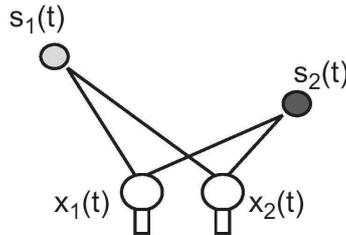


FIG. 1 – Exemple de mélange AD pour $N = 2$ sources et $P = 2$ capteurs : la propagation des ondes sonores se fait selon un trajet unique des sources aux capteurs et le temps de propagation n'est pas négligeable.

Ce type de mélange inclut en particulier le cas où des ondes sonores se propagent jusqu'aux capteurs selon un *trajet unique* (voir figure 1). En effet, nous supposons qu'il n'y a aucune réflexion (on peut comparer ces conditions à celles d'une chambre anéchoïque dans le cas de signaux acoustiques) et donc les seuls signaux reçus par les capteurs sont les ondes provenant directement des sources avec leur atténuation et leur retard respectifs.

Définition 3 On parle de mélanges convolutifs lorsque les observations s'écrivent comme la somme de sources convoluées par des filtres de propagation.

Ce type de mélange est le plus général des trois présentés jusqu'à présent puisque chaque capteur reçoit plusieurs ondes issues d'une même source qui sont toutes identiques à une atténuation et un retard près (voir figure 2) :

$$x_i(t) = \sum_{j=1}^N \sum_{k=-\infty}^{+\infty} a_{ij,k} s_j(t - t_{ij,k}) \quad i = 1, \dots, P. \quad (3)$$

On peut alors modéliser la contribution globale d'une source sur un capteur par la réponse impulsionnelle d'un filtre :

$$x_i(t) = \sum_{j=1}^N a_{ij}(t) * s_j(t) \quad i = 1, \dots, P, \quad (4)$$

où $a_{ij}(t)$ est la réponse impulsionnelle d'un filtre qui correspond à la contribution globale de la source j sur le capteur i .

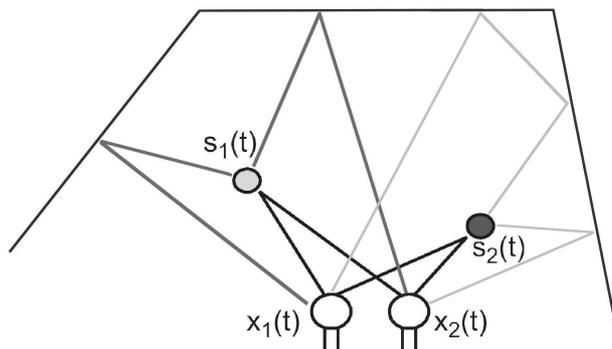


FIG. 2 – Exemple de mélange convolutif pour $N = 2$ sources et $P = 2$ capteurs.

Le second type de mélange concerne les mélanges non-linéaires, qui sont plus généraux. L'extrême complexité de ces mélanges limite son étude à des configurations bien particulières, comme par exemple le cas *post-non-linéaire* [226]. En effet, dans ce cas, le mélange est d'abord convolutif avant d'arriver sur les capteurs qui eux introduisent une non-linéarité (saturation en réception par exemple). Les méthodes que nous avons étudiées au cours de nos travaux ne concernent pas ce type de mélange. Nous n'en parlerons donc pas plus dans ce rapport.

1.2 Concept de séparation en fonction du mélange : du modèle instantané au mélange convolutif

La plupart des approches de SAS ont été développées dans le cadre de mélanges LI et permettent de séparer N sources à partir de N observations. En écrivant l'Eq. (1) sous la forme matricielle avec $P = N$,

$$\underline{x}(t) = A\underline{s}(t), \quad (5)$$

où $\underline{x}(t) = [x_1(t), \dots, x_N(t)]^T$ et $\underline{s}(t) = [s_1(t), \dots, s_N(t)]^T$, nous pouvons reconstruire les sources, dans un vecteur $\underline{y}(t) = [y_1(t), \dots, y_N(t)]^T$, aux indéterminations de permutation et d'échelle près [71]. Si nous modélisons resp. ces indéterminations par les matrices Λ (matrice de permutation) et Δ (matrice diagonale) et si on estime la matrice de mélange A (l'estimée, aux indéterminations près, est notée \hat{A}), la reconstruction des sources se fait

selon la relation suivante :

$$\underline{y}(t) = \hat{A}^{-1} \underline{x}(t) \quad (6)$$

$$= \hat{A}^{-1} A \underline{s}(t) \quad (7)$$

$$\simeq \Lambda \Delta \underline{s}(t) \quad (8)$$

Une autre façon de reconstruire les sources consiste à les extraire une à une. C'est ce que l'on appelle la *déflation*, proposée dans le cadre de la SAS par N. Delfosse et P. Loubaton [67].

Dans le cas de mélanges convolutifs, il est possible d'utiliser certaines de ces approches LI si on considère l'expression (4) dans le domaine fréquentiel [156, 192]. En supposant que les filtres de mélange sont à réponse impulsionnelle finie (filtres RIF), causaux, d'ordre maximal K et en notant $A(t)$ la matrice formée de ces réponses impulsionnelles, nous pouvons réécrire l'Eq. (4) sous forme matricielle :

$$\underline{x}(t) = A(t) * \underline{s}(t) \quad (9)$$

$$= \sum_{k=0}^K A_k \underline{s}(t - k). \quad (10)$$

L'Eq. (9) s'écrit alors dans le domaine fréquentiel :

$$\underline{X}(\omega) = A(\omega) \underline{S}(\omega). \quad (11)$$

Pour chaque pulsation ω_l , l'Eq. (11) décrit un mélange LI puisque les observations $X_i(\omega)$ s'écrivent comme des combinaisons linéaires des sources $S_j(\omega)$, ce qui correspond complètement à la définition du problème LI de SAS. A partir d'une méthode LI de SAS, il suffirait de trouver une estimée $\hat{A}(\omega_l)$ de la matrice $A(\omega_l)$ pour recombinaer les signaux sources estimés à cette pulsation. Ces signaux seraient définis par le vecteur $\underline{Y}(\omega_l)$, à l'aide de la formule suivante :

$$\underline{Y}(\omega_l) = \hat{A}^{-1}(\omega_l) \underline{X}(\omega_l) \quad (12)$$

$$= \hat{A}^{-1}(\omega_l) A(\omega_l) \underline{S}(\omega_l) \quad (13)$$

$$\simeq \Lambda(\omega_l) \Delta(\omega_l) \underline{S}(\omega_l), \quad (14)$$

où $\Lambda(\omega_l)$ est une matrice de permutation et $\Delta(\omega_l)$ est une matrice diagonale. Néanmoins, nous ne disposons alors pas d'assez d'information (un échantillon par fréquence) pour estimer la matrice de mélange. Pour pallier ce problème, une transformée temps-fréquence (TF) est utilisée à la place de la transformée de Fourier. En effet, la relation (11) reste généralement¹ vraie si elle est exprimée dans le domaine TF :

$$\underline{X}(t, \omega_l) = A(\omega_l) \underline{S}(t, \omega_l). \quad (15)$$

¹Cette relation est vérifiée avec une bonne précision si les ordres des filtres de mélange sont nettement plus petits que la largeur temporelle de la fenêtre utilisée dans la transformée TF.

Les expressions (12), (13) et (14) s'écrivent alors resp. :

$$\underline{Y}(t, \omega_l) = \hat{A}^{-1}(\omega_l) \underline{X}(t, \omega_l) \quad (16)$$

$$= \hat{A}^{-1}(\omega_l) A(\omega_l) \underline{S}(t, \omega_l) \quad (17)$$

$$\simeq \Lambda(\omega_l) \Delta(\omega_l) \underline{S}(t, \omega_l). \quad (18)$$

Il est donc possible d'utiliser des méthodes LI de SAS pour résoudre le problème convolutif. Cette famille de méthodes s'appelle les méthodes *fréquentielles* de SAS pour les mélanges convolutifs.

Malheureusement, elles apportent de nouveaux problèmes. En effet, dans l'Eq. (18), si la permutation $\Lambda(\omega)$ appliquée aux signaux n'est pas la même aux diverses pulsations, alors la version temporelle de chaque signal de sortie combinera des contributions de différentes sources, et annulera donc la séparation obtenue dans le domaine fréquentiel. Réaliser l'appariement des sources séparées à chaque pulsation est un problème difficile. En effet, pour N sources et observations, il y a $N!$ permutations possibles à chaque pulsation ω . Plusieurs approches proposent des solutions ad hoc qui résolvent l'ambiguïté de permutation partiellement. Elles peuvent majoritairement être regroupées dans deux catégories :

1. l'uniformité des coefficients de filtre,
2. l'uniformité des spectres des sources séparées.

La première exploite les connaissances a priori sur les réponses fréquentielles des filtres (uniformité des filtres de la matrice de séparation ou uniformité des filtres de la matrice de mélange), la seconde des connaissances a priori sur les sources (uniformité des spectres des sources séparées). Dans chaque groupe ci-dessus, les méthodes diffèrent dans la façon de considérer l'uniformité à travers les fréquences, en faisant varier parfois la métrique utilisée pour mesurer la distance entre les solutions à différentes fréquences. Un panorama des méthodes convolutives fréquentielles de SAS est fourni dans [156].

Dans l'Eq. (18), l'indétermination d'échelle à chaque pulsation, donnée arbitrairement par $\Delta(\omega)$, donnera un filtrage global de chaque source. Et donc, même s'il n'y avait aucune erreur dans les permutations par bande fréquentielle, les signaux temporels de sortie pourraient avoir des spectres différents des sources originales, ce qui peut-être très gênant suivant l'application.

Pour éviter l'indétermination d'ambiguïté de permutations bande à bande ci-dessus, des méthodes convolutives temporelles ont été proposées. La majorité de ces méthodes temporelles cherche à combiner les observations définies dans l'Eq. (9) afin d'extraire les sources

$$\underline{y}(t) = W(t) * \underline{x}(t) \quad (19)$$

$$= \sum_{l=0}^{L-1} W_l \underline{x}(t-l). \quad (20)$$

Au lieu de ne considérer qu'un échantillon t , l'idée de certaines approches temporelles, par exemple [19, 47, 104, 158, 159, 227, 229], est de considérer un bloc correspondant à T

échantillons. Les équations de tels blocs peuvent être réécrites ainsi :

$$\begin{aligned}
 \underline{x}(t) &= A_0 \underline{s}(t) + \cdots + A_K \underline{s}(t - K) \\
 \underline{x}(t - 1) &= A_0 \underline{s}(t - 1) + \cdots + A_K \underline{s}(t - K - 1) \\
 \underline{x}(t - 2) &= A_0 \underline{s}(t - 2) + \cdots + A_K \underline{s}(t - K - 2) \\
 &\vdots \\
 \underline{x}(t - T + 1) &= A_0 \underline{s}(t - T + 1) + \cdots + A_K \underline{s}(t - K - T + 1)
 \end{aligned} \tag{21}$$

Nous pouvons alors écrire ces séquences de vecteurs de dimension $P \times 1$ comme un vecteur de dimension $PT \times 1$:

$$\underline{\mathbf{x}}(t) = [\underline{x}^T(t), \underline{x}^T(t - 1), \dots, \underline{x}^T(t - T + 1)]^T. \tag{22}$$

De la même façon, les séquences d'entrée de dimension $N \times 1$ peuvent être réécrites comme un vecteur de dimension $N(T + K) \times 1$:

$$\underline{\mathbf{s}}(t) = [\underline{s}^T(t), \underline{s}^T(t - 1), \dots, \underline{s}^T(t - K - T + 1)]^T. \tag{23}$$

A partir des Eq. (21), (22) et (23), le mélange convolutif peut être formulé ainsi :

$$\underline{\mathbf{x}}(t) = \mathbf{A} \underline{\mathbf{s}}(t), \tag{24}$$

où \mathbf{A} a la forme suivante

$$\mathbf{A} = \begin{bmatrix} A_0 & \cdots & A_K & 0 & 0 & \cdots & 0 \\ 0 & A_0 & \cdots & A_K & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & A_0 & A_1 & \cdots & A_K \end{bmatrix} \tag{25}$$

La matrice Toeplitz par blocs \mathbf{A} est de dimension $PT \times N(T + K - 1)$. A première vue, l'Eq. (24) a la même structure qu'un problème LI de SAS, la dimension du problème ayant crû d'un facteur voisin de T . Cependant, les modèles diffèrent considérablement puisque les éléments de \mathbf{A} et $\underline{\mathbf{s}}$ sont maintenant couplés de façon spécifique.

A noter que d'autres méthodes convolutives temporelles, par exemple [235], n'utilisent pas le modèle de mélange par bloc (24).

2 Mélanges sur- ou sous-déterminés

Nous séparons dans la littérature le cas (*sur-*)*déterminé* où le nombre N de sources est inférieur ou égal au nombre P d'observations et le cas où N est plus grand que P . Ce dernier cas, dit *sous-déterminé*, ne peut souvent être traité qu'au prix de connaissances importantes sur les sources ou de limitations sur la reconstruction de celles-ci. Il intéresse beaucoup la communauté scientifique et nous trouvons dans la littérature un nombre croissant d'articles sur ces mélanges. En effet, dans le domaine du traitement audio par exemple, nous n'avons souvent que $P = 2$ observations (les signaux issus des enceintes droite et gauche d'une chaîne stéréo peuvent être vus comme des enregistrements de plusieurs instruments de musique par deux micros distincts) et $N > 2$ sources et il est donc nécessaire de trouver des approches pour séparer les sources dans ce genre de situation.

On trouve principalement deux familles d'approches :

1. Une première vision concerne la séparation partielle [2, 76, 78, 230]. Il s'agit de considérer que notre mélange contient P sources utiles et $N - P$ sources de "bruit" non-séparables. En sortie du système de SAS, nous obtenons alors les P sources d'intérêt séparées, chacune étant mélangée avec les $N - P$ sources de bruit restantes.
2. Une deuxième vision consiste à vouloir à tout prix une séparation complète des sources, au prix d'hypothèses sur celles-ci [15–17, 19, 43, 44, 85, 86, 149–151, 202, 211, 225, 228, 238, 241, 243, 248]. Nous verrons dans la section 1.3 plus de détails sur ce type de reconstruction.

3 Plan du manuscrit

Durant ces travaux de thèse, je me suis principalement intéressé à la séparation de signaux de parole et de musique, dans le cadre de mélanges LI, AD ou convolutifs, en faisant appel à des hypothèses de parcimonie. J'ai étudié les cas sur- et sous-déterminés.

Le chapitre 1 détaille les principales familles de séparation de sources (analyse en composantes indépendantes, factorisation en matrices non-négatives, analyse en composantes parcimonieuses). Pour chacune de ces familles, les principaux critères et des références de méthodes les utilisant sont donnés. Mes travaux de thèse concernant des méthodes à parcimonie, une attention particulière a été accordée aux approches basées sur cette propriété. De plus, ces méthodes étant naturellement adaptées aux mélanges sous-déterminés, un panorama des approches de reconstruction complète des sources, esquissé dans la section 2 ci-dessus, sera alors donné.

La première partie de la thèse concerne les méthodes linéaires instantanés. Après avoir décrit les méthodes qui ont servi de point de départ à mes travaux, je propose des extensions de ces approches dans le chapitre 2 ainsi qu'une validation expérimentale dans le chapitre 3.

La deuxième partie de la thèse concerne les méthodes qui ont fait l'objet de plus d'études. J'y présenterai donc des extensions des méthodes précédentes aux mélanges convolutifs. Dans le chapitre 4, je détaillerai des versions proposées pour des mélanges convolutifs simples correspondant à la propagation de sources dans un milieu anéchoïque. J'expliquerai ensuite dans le chapitre 5 comment l'utilisation de méthodes de classification non-supervisée permet d'améliorer les performances de celles-ci. Une validation expérimentale de ces approches sera proposée dans le chapitre 6. J'introduirai ensuite une méthode pour les mélanges convolutifs généraux dans le chapitre 7.

Enfin un dernier chapitre portera sur les perspectives de ces travaux.

Chapitre 1

Panorama des méthodes de SAS

Nous présentons ici une vue d'ensemble des divers critères et algorithmes de SAS. Un panorama général peut être trouvé dans [55, 56, 65, 134, 160]. La première méthode de SAS a été proposée au milieu des années 80 par C. Jutten et J. Herault dans le cadre d'études neuro-biologiques [122, 135] mais il faut attendre 1994 pour que P. Comon généralise le concept d'*Analyse en Composantes Indépendantes* (ACI ou *ICA* pour *Independent Component Analysis* en anglais) permettant d'obtenir des signaux de sortie statistiquement indépendants, égaux aux signaux sources aux indéterminations d'échelle et de permutation près dans le cas de mélanges LI. Pour ce type de mélange, ce résultat est très utile et est une des voies majeures de la SAS. De nombreux panoramas d'ACI ont été proposés [126, 127, 129, 145].

Une deuxième famille de méthodes fait l'hypothèse que les sources et les coefficients de la matrice de mélange sont positifs. Ces méthodes sont appelées des méthodes de *Factorisation en Matrices Non-négatives* (FMN ou *NMF* pour *Non-negative Matrix Factorization* en anglais). Elles connaissent une certaine attention de la part de la communauté scientifique [35] et sont particulièrement adaptées à la séparation de signaux tels des signaux de spectroscopie.

Enfin, l'*Analyse en Composantes Parcimonieuses* (SCA pour *Sparse Component Analysis* en anglais) correspond à une troisième voie de la SAS et fait l'objet d'un grand intérêt depuis une dizaine d'années [105, 180]. Les méthodes développées dans le cadre de cette thèse appartiennent à cette classe de méthodes.

Nous allons maintenant parcourir les principaux critères de ces trois familles de méthodes de SAS. La section 1.1 détaillera les critères utilisés en ACI et la section 1.2 sera consacrée aux méthodes faisant l'hypothèse de FMN. Enfin, une attention particulière sera donnée aux méthodes faisant appel à la SCA dans la section 1.3. Dans cet état de l'art, nous nous concentrerons sur les diverses approches développées pour les mélanges LI, AD et convolutifs resp. explicités dans les définitions 1, 2 et 3 de l'introduction générale.

1.1 Critères de l'ACI

Comme nous l'avons vu précédemment, l'ACI est une des voies majeures de la SAS. Un nombre très important d'articles et de critères lui ont été consacrés. Dans leur version

de base, ces méthodes font l'hypothèse que les sources sont

- statistiquement indépendantes, c.-à-d. que la densité de probabilité conjointe des sources $s_i(t)$ soit égale au produit des densités marginales :

$$f_{\underline{s}}(s_1(t), \dots, s_N(t)) = \prod_{i=1}^N f_{s_i}(s_i(t)), \quad (1.1)$$

- stationnaires,
- centrées, c.-à-d. de moyenne nulle,
- et qu'au plus une de ces sources soit gaussienne².

En notant les vecteurs de sources et d'observations $\underline{s}(t) = [s_1(t), \dots, s_N(t)]^T$ et $\underline{x}(t) = [x_1(t), \dots, x_P(t)]^T$, on peut définir de façon générale l'ACI de la manière suivante.

Définition 4 *L'ACI d'un vecteur aléatoire $\underline{x}(t)$ consiste à trouver une transformée linéaire $\underline{y}(t) = W\underline{x}(t)$ de telle façon que les signaux $y_i(t)$ ($i = 1, \dots, N$) soient aussi indépendants que possible, dans le sens de la maximisation d'une fonction $F(y_1(t), \dots, y_N(t))$ qui mesure l'indépendance.*

Nous présentons ici les principaux critères d'ACI ainsi que quelques méthodes faisant appel à ceux-ci. Certains de ces critères rentrent dans le cadre des *fonctions de contraste*, formalisées par P. Comon pour les mélanges LI dans [61] et pour les mélanges convolutifs dans [62]. Les fonctions de contrastes ont été notamment étudiées par E. Moreau et al. [174, 175].

1.1.1 ACI par maximisation de la non-gaussianité

La maximisation de la non-gaussianité est un des critères les plus utilisés dans le cadre de l'ACI. Il est fondamental puisqu'il est impossible de séparer des sources stationnaires indépendantes identiquement distribuées (iid) si la non-gaussianité de celles-ci n'est pas respectée.

Les motivations initiales sont basées sur le théorème de la limite centrale : la somme de variables aléatoires tend vers une distribution gaussienne. On considère alors qu'une combinaison linéaire de N sources aléatoires non-gaussiennes, de même distribution, est plus gaussienne que les sources elles-mêmes. En pratique, la séparation d'un mélange LI de sources s'effectue en maximisant la non-gaussianité des signaux de sortie (qui correspondent à des combinaisons linéaires des observations blanchies).

Le tout premier critère utilisé pour évaluer la non-gaussianité fut le *kurtosis* (non-normalisé³) ou auto-cumulant d'ordre 4. Il a d'abord été introduit et utilisé dans le cadre de la déconvolution [84, 240]. Il est défini par l'expression suivante

$$\text{kurt}(y(t)) = \text{cum}_4(y(t), y(t), y(t), y(t)) = \mathbb{E} \{y(t)^4\} - 3 \mathbb{E} \{y(t)^2\}^2, \quad (1.2)$$

²Je tiens à rappeler que ces hypothèses sont faites dans des versions basiques de méthodes d'ACI. Par exemple, certaines méthodes comme SOBI ne nécessitent pas l'hypothèse de gaussianité d'au plus une source.

³Il existe dans la littérature plusieurs définitions du kurtosis : certains auteurs, par exemple [127], considèrent la définition donnée ci-dessus. Au contraire, d'autres auteurs dont [163] définissent le kurtosis comme l'autocumulant d'ordre 4 normalisé, tel qu'on peut le voir dans l'Eq. (1.8).

où $y(t)$ est une variable aléatoire de moyenne nulle. Le principal défaut du kurtosis est qu'il est très sensible aux valeurs aberrantes (*outliers* en anglais) à cause du moment d'ordre 4. Notons qu'une étude du kurtosis en fonction de la distribution des sources a été proposée par A. Mansour et C. Jutten dans [163].

Une autre mesure de non-gaussianité considérée dans la littérature est l'*entropie différentielle* définie par

$$H(y(t)) = - \int f_y(\eta) \log(f_y(\eta)) \, d\eta, \quad (1.3)$$

où $f_y(\cdot)$ est la densité de la variable aléatoire $y(t)$. Plus la variable $y(t)$ sera "aléatoire" et plus grande sera sa valeur de l'entropie. En particulier, un signal gaussien maximisera l'entropie à variance donnée.

Un critère de mesure de la gaussianité peut donc être obtenu en comparant l'entropie d'un signal aléatoire normalisé et celle d'un signal suivant une distribution gaussienne de puissance unité. Ce critère, appelé *néguentropie*, est défini par

$$J(y(t)) = H(y_{\text{gauss}}(t)) - H(y(t)). \quad (1.4)$$

$J(y(t))$ est toujours positif ou nul. En particulier, il est nul si et seulement si $y(t)$ a une distribution gaussienne. C'est donc une bonne mesure de non-gaussianité, bien justifiée par les théories statistiques. L'inconvénient de la néguentropie est qu'elle nécessite l'estimation des densités de probabilité des sources qui sont inconnues! La néguentropie peut être approximée en utilisant les cumulants d'ordre supérieur :

$$J(y(t)) \simeq \frac{1}{12} \mathbb{E} \{y(t)^3\}^2 + \frac{1}{48} \text{kurt}(y(t))^2. \quad (1.5)$$

Une approche utile consiste à généraliser cette approximation fondée sur des cumulants d'ordre supérieur par l'utilisation de fonctions non quadratiques [127]. Un cas particulier correspond à remplacer resp. $y(t)^3$ et $y(t)^4$ par des fonctions G_1 et G_2 resp. impaire et paire. Le critère de néguentropie s'exprime alors de la façon suivante :

$$J(y(t)) \simeq k_1 \mathbb{E} \{G_1(y(t))\}^2 + k_2 \left(\mathbb{E} \{G_2(y(t))\} - G_2(y_{\text{gauss}}(t)) \right)^2, \quad (1.6)$$

où k_1 et k_2 sont des constantes positives et $y_{\text{gauss}}(t)$ est une variable gaussienne centrée réduite.

Une autre approche consiste à n'utiliser qu'une seule fonction non-quadratique G . Dans ce cas, l'approximation devient

$$J(y(t)) \propto \left(\mathbb{E} \{G(y(t))\} - \mathbb{E} \{G(y_{\text{gauss}}(t))\} \right)^2. \quad (1.7)$$

Toutes les méthodes basées sur la maximisation de la non-gaussianité nécessitent une normalisation, soit du critère, soit des sorties. On peut par exemple imposer aux signaux extraits d'être de norme unité. Dans ce cas, le procédé consiste à décorrélérer et normaliser les observations, afin que le signal de sortie $y(t)$ puisse être de norme unité en normalisant le vecteur d'extraction appliqué aux observations blanchies.

L'une des méthodes les plus connues basées sur la maximisation de non-gaussianité est FastICA qui est une méthode de type point-fixe. Cette méthode utilise au choix le kurtosis [128] ou la néguentropie [125]. H. Hyvärinen a aussi proposé une méthode pour approcher l'entropie [124].

La non-gaussianité a été aussi très utilisée pour la séparation de mélanges convolutifs, en utilisant le critère du kurtosis [181, 229] ou de néguentropie [87, 141, 229].

Comme nous l'avons souligné, de façon alternative, il est aussi possible de normaliser le critère. En effet, le kurtosis est dépendant du facteur d'échelle appliqué devant le signal. Afin de supprimer cette sensibilité, on peut utiliser un kurtosis normalisé défini par

$$\text{kurt}_{\text{norm}}(y(t)) = \frac{\text{kurt}(y(t))}{\mathbb{E}\{y(t)^2\}^2}. \quad (1.8)$$

Ce kurtosis normalisé a par exemple été utilisé dans le cadre de mélanges convolutifs dans [235].

1.1.2 ACI par estimation du maximum de vraisemblance

Une approche très populaire pour le modèle de l'ACI est l'estimation du maximum de vraisemblance. Le but est de trouver les paramètres du mélange qui maximisent la probabilité d'occurrence des observations. Ainsi, à partir de l'expression matricielle du mélange

$$\underline{x}(t) = A\underline{s}(t), \quad (1.9)$$

on peut exprimer la densité de probabilité du vecteur $\underline{x}(t)$ en fonction des densités de probabilité des sources supposées indépendantes et du déterminant de l'inverse $W = A^{-1}$ de la matrice de mélange

$$f_{\underline{x}}(\underline{x}(t)) = |\det W| f_{\underline{s}}(\underline{s}(t)) \quad (1.10)$$

$$= |\det W| \prod_i f_{s_i}(s_i(t)). \quad (1.11)$$

La relation (1.11) peut être exprimée par rapport aux colonnes \underline{w}_i de la matrice W et aux observations :

$$f_{\underline{x}}(\underline{x}(t)) = |\det W| \prod_i f_{s_i}(\underline{w}_i^T \underline{x}(t)). \quad (1.12)$$

Si on considère K échantillons de \underline{x} , alors la vraisemblance $L(W)$ peut être obtenue comme le produit de cette densité évalué à ces K points et peut s'exprimer comme une fonction de W :

$$L(W) = \prod_{t=1}^K |\det W| \prod_{i=1}^N f_{s_i}(\underline{w}_i^T \underline{x}(t)). \quad (1.13)$$

Il est souvent plus pratique de considérer le logarithme de la vraisemblance qui s'écrit

$$\log L(W) = \left[\sum_{t=1}^K \sum_{i=1}^N \log (f_{s_i}(\underline{w}_i^T \underline{x}(t))) \right] + K \log |\det W|, \quad (1.14)$$

ce qui donne, en divisant par K et en considérant la moyenne empirique, ici aussi notée $\mathbb{E}\{\cdot\}$,

$$\frac{1}{K} \log L(W) = \mathbb{E} \left\{ \sum_{i=1}^N \log (f_{s_i}(\underline{w}_i^T \underline{x}(t))) \right\} + \log |\det W|. \quad (1.15)$$

Comme pour la néguentropie, le principal inconvénient de la vraisemblance est que l'on ne connaît pas les densités de probabilité des sources. On peut alors supposer soit qu'elles sont connues a priori, soit qu'elles appartiennent à une famille de distributions donnée. En utilisant les fonctions score $g_i(\cdot) = (-\log(f_{s_i}(\cdot)))'$ et en supposant que le déterminant de W est constant (suite à une décorrélation et une normalisation des sources), alors maximiser la vraisemblance correspond à minimiser la néguentropie ou maximiser l'entropie, comme cela a été montré par J.F. Cardoso dans [51]. La maximisation de l'entropie à l'aide de ces fonctions score $g_i(\cdot)$, appelée *Infomax*, a été proposée par A. J. Bell et T. J. Sejnowski [29] et utilisée par exemple par T. W. Lee et al. [146] ou K. Torkkola [233]. Des approches basées sur la maximisation de la vraisemblance ont d'abord été proposées par M. Gaeta et J. L. Lacoume [101], puis entre autres par D. T. Pham et al. [199, 200], par L. C. Parra et al. [187, 191], ou par A. Belouchrani et J. F. Cardoso [33]. R. Guidara et al. ont proposé plusieurs approches qui peuvent être vues comme des extensions des méthodes [199, 200] de D. T. Pham car elles prennent en compte l'autocorrélation des sources. Leurs approches permettent de traiter des signaux stationnaires [106, 120, 121] ou non-stationnaires [107, 108] dans le cadre de mélanges LI. J. Anemüller et B. Kollmeier ont proposé une méthode fréquentielle pour les mélanges AD [9], N. Mitianoudis et M. E. Davies ont eux-aussi proposé une méthode fréquentielle mais pour les mélanges convolutifs [173] alors que E. Moulines et al. [177] et S. Deligne et R. Gopinath [68] ont proposé une approche temporelle pour les mélanges convolutifs.

1.1.3 ACI basée sur la minimisation de l'information mutuelle

L'information mutuelle I entre N variables aléatoires $y_i(t)$ est définie comme la différence entre la somme des entropies différentielles des densités marginales des variables aléatoires et l'entropie différentielle de la densité conjointe $H(\underline{y}(t))$:

$$I(y_1(t), y_2(t), \dots, y_N(t)) = \left[\sum_{i=1}^N H(y_i(t)) \right] - H(\underline{y}(t)). \quad (1.16)$$

L'information mutuelle est une mesure naturelle de la dépendance entre des variables aléatoires. Elle est toujours positive ou nulle. En particulier, elle est nulle si et seulement si les variables sont statistiquement indépendantes. Ainsi, on peut utiliser l'information mutuelle comme un critère pour l'ACI. En effet, en définissant l'ACI d'un vecteur aléatoire $\underline{x}(t)$ comme une transformée linéaire inversible

$$\underline{y}(t) = W\underline{x}(t) \quad (1.17)$$

où la matrice W est déterminée de telle façon que l'information mutuelle des composantes transformées $y_i(t)$ est minimisée, l'Eq. (1.16) peut s'écrire [127]

$$I(y_1(t), y_2(t), \dots, y_N(t)) = \sum_{i=1}^N H(y_i(t)) - H(\underline{x}(t)) - \log |\det W|. \quad (1.18)$$

Si on contraint les $y_i(t)$ à être décorrélés et de variance unité, alors [127]

$$I(y_1(t), y_2(t), \dots, y_N(t)) = \text{const.} - \sum_{i=1}^N J(y_i(t)), \quad (1.19)$$

où J est la néguentropie définie dans l'Eq. (1.4) et où le terme constant ne dépend pas de W . Cette relation (1.19) montre donc le rapport entre l'information mutuelle et la néguentropie.

Cependant, ces critères ne sont pas équivalents :

1. l'information mutuelle ne permet pas d'approche à déflation qui permet d'extraire les sources une à une, contrairement à la néguentropie.
2. Par contre, contrairement aux critères de mesure de non-gaussianité qui nécessitent des signaux non-corrélés, l'information mutuelle permet d'éviter l'étape de blanchiment.

Notons aussi que l'information mutuelle et la vraisemblance sont intimement liées [127].

Le critère d'information mutuelle a été notamment étudié par D. T. Pham [194–197], K. E. Hild et al. [94, 113–115], Z. Q. Luo et J. Lu [154], C. Jutten et al. [20, 213] ou encore M. El Rhabi et al. [205].

1.1.4 ACI basée sur les statistiques d'ordre 2

Dans le cadre général de l'ACI, la décorrélation de signaux de sortie ne suffit pas à séparer les sources. En effet, si on multiplie le vecteur source $\underline{s}(t)$ par une matrice orthogonale A quelconque, alors les observations du vecteur $\underline{x}(t)$ résultant sont décorrélées elles aussi.

Cependant, on peut quand même effectuer la séparation à l'aide des statistiques d'ordre 2 des signaux en supposant l'une des hypothèses 1 et 2 suivantes vérifiée (en alternative à l'hypothèse de non-gaussianité des sources) :

- **Hypothèse 1** *Les sources sont auto-colorées (de manière différente).*

Si l'hypothèse 1 est vérifiée, alors il est possible de séparer des sources en imposant l'intercorrélations de signaux de sortie décalés nulle :

$$\exists \tau / \mathbb{E}\{y_i(t)y_j(t - \tau)\} = 0, \quad \forall i \neq j. \quad (1.20)$$

En effet, les matrices d'autocovariance des sources centrées $\underline{s}(t)$ définies par

$$R_{\underline{s}}(\tau) = \mathbb{E}\{\underline{s}(t)\underline{s}(t - \tau)^T\}, \quad (1.21)$$

sont réelles et diagonales. Les matrices d'autocovariance des observations $\underline{x}(t)$ s'écrivent alors

$$R_{\underline{x}}(\tau) = \mathbb{E}\{\underline{x}(t)\underline{x}(t-\tau)^T\} = AR_{\underline{z}}(\tau)A^T \quad (1.22)$$

et sont simultanément diagonalisables pour divers τ . Parmi les méthodes statistiques d'ordre 2 célèbres qui supposent les sources colorées, citons AMUSE [231, 232] ou SOBI [31]. AMUSE (pour *Algorithm for Multiple Unknown Signals Extraction*) suppose que les sources sont mutuellement non-corrélées et ayant des autocorrélations normalisées différentes, c.-à-d.

$$\exists t_1, t_2 / \mathbb{E}\{s_i(t_1)s_j(t_2)\} = 0 \quad \forall i \neq j \quad (1.23)$$

et

$$\exists \tau > 0 / \frac{\mathbb{E}\{s_i(t)s_i(t-\tau)\}}{\mathbb{E}\{s_i(t)^2\}} \neq \frac{\mathbb{E}\{s_j(t)s_j(t-\tau)\}}{\mathbb{E}\{s_j(t)^2\}}. \quad (1.24)$$

Après avoir blanchi les observations (le vecteur des observations blanchies est noté $\underline{z}(t)$), la méthode utilise la fonction d'autocovariance $R_{\underline{z}}(\tau)$ pour *une seule* valeur de τ non-nulle. La matrice de mélange est estimée en reprenant le formalisme de l'Eq. (1.22) appliquée aux observations blanchies. Récemment, une version convolutive temporelle d'AMUSE a été proposée [227].

SOBI (pour *Second Order Blind Identification*) est une extension d'AMUSE. Elle consiste à diagonaliser plusieurs matrices $R_{\underline{z}}(\tau)$ pour *plusieurs valeurs différentes et non-nulles* de τ . Il suffit que la condition (1.24) soit satisfaite pour l'un des décalages temporels τ choisis pour que l'on puisse séparer les sources, ce qui représente un avantage par rapport à AMUSE qui n'utilise qu'un seul décalage. Une version convolutive temporelle de SOBI a été proposée dans [47].

– **Hypothèse 2** *Les sources sont non-stationnaires.*

Si l'hypothèse 2 est vérifiée, alors les statistiques qui varient dans le temps fournissent de l'information supplémentaire. Dans ce cas, la décorrélation des sorties entre elles à plusieurs instants suffit pour effectuer la séparation. En effet, en raison de la non-stationnarité à l'ordre 2, la covariance de $\underline{y}(t)$ dépend de l'instant t , ce qui engendre de nouvelles contraintes par rapport au simple blanchiment. Dans [223], A. Souloumiac propose de couper le signal temporel en deux et de calculer la matrice de covariance $R_{\underline{x}}(0)$ sur chaque zone temporelle, puis de diagonaliser conjointement ces deux matrices. D. T. Pham et J. F. Cardoso [198] ont étendu la méthode d'A. Souloumiac en prenant plus de deux zones pour calculer la covariance.

S. Choi et al. ont proposé une méthode, SONS [54], qui peut être vue comme une extension des méthodes de Pham et Cardoso et de SOBI : comme dans [198], le signal temporel est découpé en plusieurs zones. Mais comme pour SOBI, les observations sont d'abord blanchies puis des matrices de covariance $R_{\underline{z}}(\tau)$ ($\tau \neq 0$) sont calculées sur chacune de ces zones avant d'être conjointement diagonalisées.

Dans [118], S. Hosseini et Y. Deville proposent deux méthodes basées sur les propriétés statistiques dans le domaine fréquentiel. Ces méthodes sont appelées méthodes à décorrélation spectrale. La première est basée sur la diagonalisation d'une matrice calculée à partir des statistiques d'ordre 2 des signaux exprimés dans le domaine de Fourier. Elle est utilisée en particulier dans le cadre de signaux cyclo-stationnaires.

La deuxième méthode, étudiée aussi dans [119], fait l'hypothèse de signaux sources blancs et non-stationnaires et travaille dans le domaine fréquentiel pour obtenir des sources autocorrélées et stationnaires. Elle fait ensuite appel à des méthodes temporelles existantes, comme AMUSE ou SOBI, transposées dans le domaine fréquentiel. Tous les signaux sources (comme la parole par exemple) n'étant pas blancs, H. Saylani et al. ont remédié à ce problème en multipliant les observations par une réalisation de bruit blanc [217]. Les nouvelles sources obtenues deviennent alors non-corrélées dans le temps et permettent d'améliorer les performances (environ 20 dB de mieux) de [119].

D'autres approches à l'ordre 2 LI ou convolutives ont été proposées par K. Matsuoka et al. [139, 140, 167, 168, 182], H. Buchner et al. [49], L. Parra et C. Spence [186], D. T. Pham et al. [201] ou E. M. Fadaili et al. [95].

Un état de l'art beaucoup plus complet sur les méthodes statistiques d'ordre 2 peut par exemple être trouvé dans la thèse d'H. Boumaraf [46].

1.1.5 ACI basée sur une approche tensorielle

Comme nous venons de le voir dans la section précédente, la restriction à l'ordre 2 de l'indépendance n'est généralement pas suffisante pour séparer les sources : théoriquement, l'indépendance entre des signaux correspond à la nullité des cumulants croisés entre toutes les versions décalées de ces signaux et à tous les ordres. Cette condition ne peut toutefois pas être testée avec un coût de calcul fini. En pratique, les méthodes de SAS annulent les cumulants croisés jusqu'à un certain ordre fixé. En général, cet ordre est fixé à 4. Par exemple, le cumulant croisé $\text{cum}_4(y_i(t), y_j(t), y_j(t), y_i(t))$ est un critère d'indépendance puisqu'il est nul si deux signaux $y_i(t)$ et $y_j(t)$ sont indépendants [60]. On peut généraliser les méthodes basées sur les cumulants croisés par une approche tensorielle.

Définition 5 *Un tenseur de cumulants \mathbf{F} est un opérateur linéaire défini par les cumulants croisés d'ordre 4 des données.*

La transformation linéaire a lieu dans l'espace de dimension $N \times N$ tel que le $i, j^{\text{ème}}$ élément de la transformation, noté $\mathbf{F}_{ij}(M)$, est défini ainsi :

$$\mathbf{F}_{ij}(M) = \sum_{k,l} m_{kl} \text{cum}_4(y_i(t), y_j(t), y_k(t), y_l(t)), \quad (1.25)$$

où m_{kl} est un élément de la matrice M qui est transformée. Comme tout opérateur linéaire symétrique, le tenseur \mathbf{F} a une décomposition en valeurs propres. Une matrice propre du tenseur est, par définition, une matrice telle que

$$\mathbf{F}(M) = \lambda M. \quad (1.26)$$

Les matrices propres du tenseur \mathbf{F} seront notées par la suite M_i .

Cette décomposition en valeurs propres correspond à une diagonalisation conjointe des matrices $F(M_i)$. C'est sur un tel critère qu'est fondée la méthode JADE [50]. Ce critère peut être réécrit de la forme suivante :

$$J_{\text{JADE}} = \sum_{(ijkl) \neq (iikl)} \text{cum}_4(y_i(t), y_j(t), y_k(t), y_l(t))^2. \quad (1.27)$$

A. Yeredor a proposé une diagonalisation conjointe non-orthogonale au sens des moindres carrés et l'a appliquée à JADE [246]. Sa méthode, nommée AC-DC estime à chaque itération les paramètres de mélange (étape AC) et les statistiques des sources (étape DC). A. Yeredor a ensuite étendu AC-DC aux mélanges AD dans [245, 247] alors que A. Mertins et I. Russel en ont proposé une extension aux mélanges convolutifs généraux [172].

Une approche tensorielle LI sous-déterminée a été proposée par P. Comon dans [63].

1.2 Méthodes basées sur la factorisation en matrices non-négatives

Dans de nombreuses applications comme l'imagerie bio-médicale ou la spectroscopie, des contraintes de positivité sont imposées aux coefficients de mélanges a_{ij} de la matrice A et aux sources s_j . La factorisation en matrices non-négatives (FMN) est une famille de méthodes où une matrice X est décomposée en un produit de matrices $X = AS$. Dans le cas de la SAS, X est la matrice des observations, S correspond aux sources et A est la matrice de mélange. Historiquement, les premières méthodes ont été développées en Finlande au milieu des années 90 [10, 184] sous le nom de factorisation en matrices positives. Dans [143, 144], D. D. Lee et H. S. Seung ont publié un algorithme simple et performant de factorisation. L'algorithme calcule de manière itérative une estimation des matrices A et S afin que soit minimisée la divergence :

$$\text{div}(X|AS) = \sum_{i,j} \left\{ X_{ij} \log \left[\frac{X_{ij}}{(AS)_{ij}} \right] - X_{ij} + (AS)_{ij} \right\}. \quad (1.28)$$

Cet algorithme a par exemple été utilisé dans des applications astrophysiques [34]. De nombreuses méthodes ont été développées récemment, certaines faisant des hypothèses supplémentaires comme la parcimonie des sources par exemple [90]. Des panoramas détaillés de ces méthodes sont fournis dans [35, 55].

Citons en particulier les méthodes de B. Wang et M. D. Plumbley [238, 239] qui ont été développées pour des mélanges LI de signaux de musique. La FMN est alors appliquée aux modules des amplitudes des transformées TF des observations, dans le cas où on a plusieurs observations [238] ou dans le cas où l'on n'a qu'une seule observation [239]. Les sources sont ensuite reconstruites par masquage TF. Notez aussi qu'une FMN convolutive, éventuellement avec des contraintes de parcimonie, a été proposée et étudiée entre autre par P. Smaragdis [219] et par P. D. O'Grady et B. A. Pearlmutter [179].

1.3 Méthodes basées sur la parcimonie des sources

Comme nous l'avons vu précédemment, la plupart des méthodes de SAS sont basées sur l'ACI, qui constitue une des voies les plus étudiées pour la SAS. De façon alternative, depuis une dizaine d'années, des méthodes basées sur la *parcimonie* des sources sont apparues. Il existe de nombreuses définitions de la parcimonie dont la suivante.

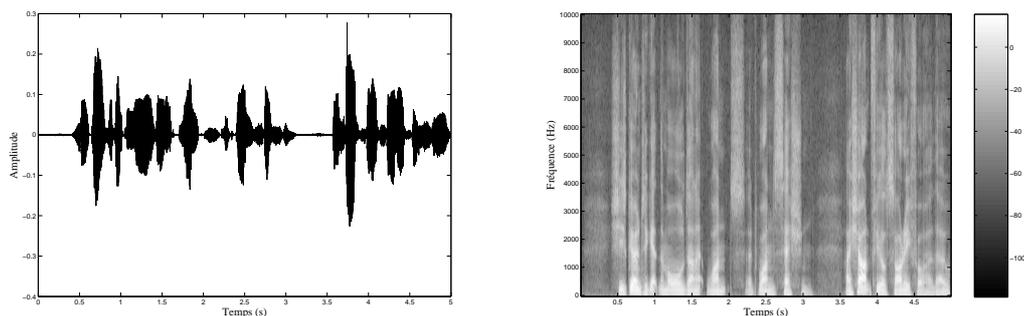


FIG. 1.1 – Signal de parole parcimonieux dans deux domaines d'analyse différents. *A gauche* : domaine temporel. *A droite* : domaine temps-fréquence.

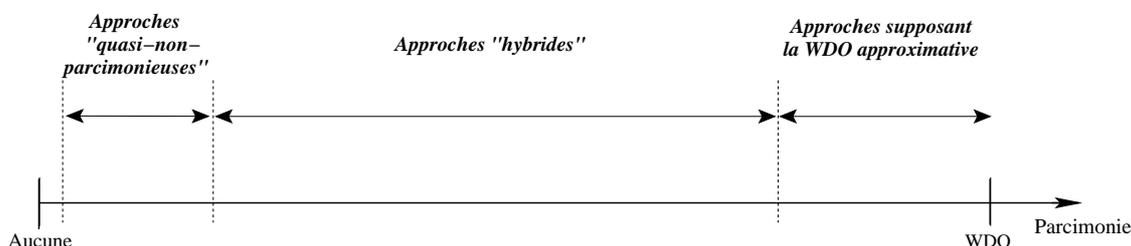


FIG. 1.2 – Classement des méthodes de SAS à parcimonie en fonction du degré de parcimonie des sources.

Définition 6 *Un signal est dit parcimonieux s'il est plus souvent nul (ou proche de zéro) que la valeur de sa variance "ne le laisserait penser" [180].*

Dit autrement, un signal vérifiant la définition 6 ne contient que "quelques" valeurs significatives. La majorité des approches de cette famille est bien adaptée aux signaux de parole et de musique, puisque ces signaux sont parcimonieux (dans les domaines temporel, fréquentiel et/ou TF par exemple – voir figure 1.1). De nombreuses transformées parcimonieuses sont proposées ou présentées dans [58, 69, 98, 116, 157]. Plusieurs panoramas ont été proposés récemment [105, 180].

Nous pouvons découper l'ensemble des méthodes parcimonieuses en trois catégories (voir figure 1.2) :

1. Une première famille utilise tous les points du domaine d'analyse car elle fait l'hypothèse que dans chacun de ces points, une seule source est présente, à de rares exceptions près. Alors que certaines méthodes supposent *vraiment* ces fortes conditions de parcimonie appelées orthogonalité W -disjointe (*WDO* pour *W-disjoint-orthogonality* en anglais), d'autres supposent que cette hypothèse ne peut être totalement vérifiée mais que les points aberrants seront peu nombreux et peu gênants pour estimer le mélange.
2. Au contraire, une deuxième famille de méthodes, dites "quasi-non-parcimonieuses", n'utilise que certaines zones "bien choisies" du domaine d'analyse, c.-à-d. des zones

mono-sources, pour estimer les mélanges. En dehors de ces zones, les sources peuvent être recouvrantes dans le domaine d'analyse. L'hypothèse de parcimonie est donc faible. En particulier, les méthodes développées dans ce document appartiennent à cette catégorie.

3. Enfin, une dernière catégorie de méthodes joue sur un degré "hybride" de parcimonie entre les méthodes des deux familles ci-dessus. En effet, elles ajoutent une mesure de confiance sur la "qualité mono-source" des points du domaine d'analyse et proposent de coupler cette mesure à des approches semblables aux méthodes WDO (approximatives).

Notons aussi l'existence de plusieurs méthodes, dont l'approche [21] proposée par M. Babaie-Zadeha et al., qui proposent d'appliquer l'ACI à des sources parcimonieuses. Ce type de méthodes ne sera pas décrit dans la suite de ce manuscrit.

Nous allons voir aussi par la suite que de nombreuses approches à parcimonie permettent de traiter les mélanges AD. Rappelons que les premières méthodes pour estimer plusieurs décalages temporels ont été entre autres proposées par B. Emile et al. [64, 92, 93].

De nombreuses approches de SCA sont bien adaptées au cas sous-déterminé. Comme nous l'avons vu dans la section 2 page 6 de ce manuscrit, certaines de ces approches proposent une reconstruction partielle des sources. C'est le cas par exemple des méthodes que nous avons développées, comme nous le verrons dans la partie I page 45 de ce manuscrit. Au contraire, d'autres approches permettent de séparer totalement les sources, en faisant appel à des critères différents, comme nous allons le voir.

1. Une famille de méthodes cherche chaque atome du plan transformé où seule une source est présente. Il suffit alors de *masquer* pour chaque source toutes les zones du domaine d'analyse où celle-ci n'est pas la seule active [238, 248]. Les versions temporelles des sources sont obtenues en utilisant la transformée parcimonieuse inverse sur celles-ci. Ce type d'approche pose un problème quand le nombre d'atomes où plusieurs sources sont présentes est important : la reconstruction des sources est alors grossière.
2. Un nombre important d'articles, par exemple [42–44, 85, 86, 149–151, 202, 211, 225, 228, 241], concerne la reconstruction des sources par minimisation sous contraintes (on suppose que la matrice de mélange a déjà été estimée⁴). En définissant la norme ℓ^q ($q \in [0, 1]$) par

$$\|\underline{u}\|_0 = \sum_{|u_i| \neq 0} |u_i|^0, \quad q = 0, \quad (1.29)$$

$$\|\underline{u}\|_q = \left(\sum_i |u_i|^q \right)^{1/q}, \quad q \in]0, 1], \quad (1.30)$$

le problème P_q de minimisation sous contraintes s'écrit

$$P_q : \quad \min_{\underline{s}} \|\underline{s}\|_q \text{ sous la contrainte } A\underline{s} = \underline{x}. \quad (1.31)$$

⁴La plupart des méthodes qui utilisent ce genre de critère pour reconstruire sont des méthodes en deux étapes. La première étape consiste à estimer la matrice de mélange. La deuxième consiste ensuite à reconstruire les sources.

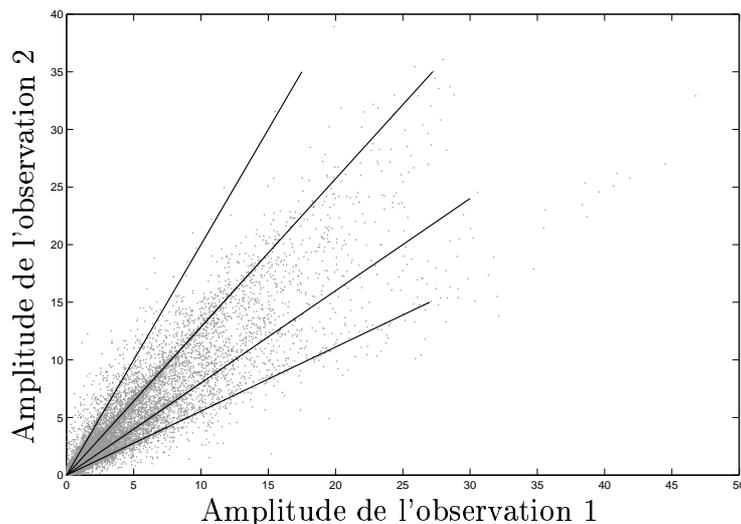


FIG. 1.3 – Nuage de points pour un mélange stéréo de 4 sources parcimonieuses. Dans cet exemple, les amplitudes correspondent au module des transformées TF des observations. Les droites correspondent aux directions des sources, c.-à-d. aux colonnes de la matrice de mélange.

Les sources estimées $\hat{\underline{s}}$ qui vérifient l'Eq. (1.31) sont alors les sources les plus parcimonieuses au sens de la norme ℓ^q (notons que dans le cas particulier où $q = 0$, la norme $\|\underline{x}\|_0$, définie dans l'Eq. (1.29), donne simplement le nombre de sources non-nulles). De nombreuses discussions ont été proposées sur la "meilleure" valeur de q . Le cas $q = 0$ donne le problème P_0 qui montre une sensibilité au bruit et plus généralement aux éléments proches de zéro mais non nuls [150].

Le cas $q = 1$ a été très étudié [149–151, 225] et correspond, lorsque la matrice de mélange A est à coefficients réels, au cas où dans chaque atome du plan transformée, au plus P sources sont présentes. D'un point de vue géométrique, dans le cadre d'un nuage de points (*scatter plot* en anglais - voir figure 1.3) où il y a $P = 2$ observations, il peut être résumé ainsi : pour tout point (x_1, x_2) entre deux directions de sources (les *directions des sources* sont les droites du nuage de points de la figure 1.3 et elles correspondent aux colonnes de la matrice de mélanges), on considère que seules les deux sources correspondant à ces deux directions sont présentes (voir figure 1.4). Dans le cas où la distribution des sources est laplacienne, les solutions à cette minimisation ℓ^1 sont les solutions de probabilité maximum à posteriori (MAP pour *Maximum A posteriori Probability* en anglais).

De façon alternative aux normes ℓ^0 et ℓ^1 , les normes ℓ^q ($q \in]0, 1[$) définies dans l'Eq. (1.30) donnent une mesure de parcimonie : une valeur plus petite de q accroît l'importance de la parcimonie de \underline{u} comme discuté dans [85].

Lorsque les coefficients de mélange et/ou les signaux sources sont à valeurs complexes, une version plus contraignante appelée ℓ^q -*basis-pursuit* a été proposée récemment [212] et rajoute une condition sur le nombre de sources présentes dans

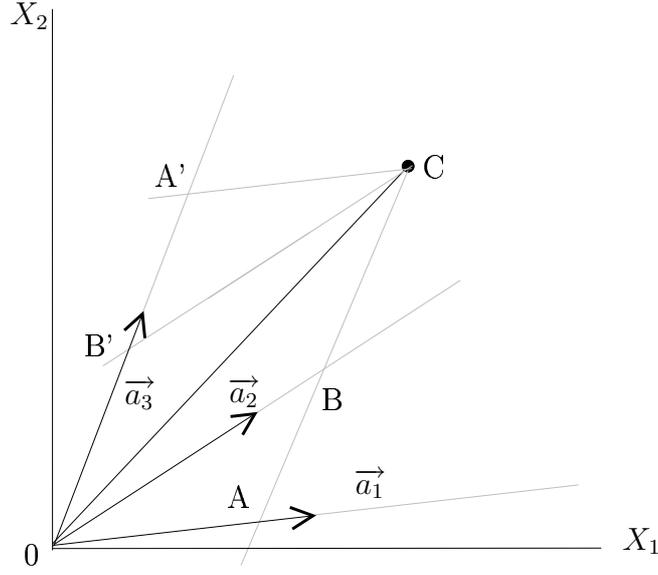


FIG. 1.4 – Exemple de minimisation par la norme ℓ^1 du problème P_1 . En n'utilisant que deux des directions \vec{a}_1 , \vec{a}_2 et \vec{a}_3 , le chemin le plus court entre l'origine et le point C est O-B-C ou O-B'-C, où C est un point du nuage de points, \vec{a}_i la direction de la $i^{\text{ème}}$ source et A' (resp. B', A et B) est la projection de C, selon la direction \vec{a}_1 (resp. \vec{a}_2 , \vec{a}_3 et \vec{a}_3), sur la droite de direction \vec{a}_3 (resp. \vec{a}_3 , \vec{a}_2 et \vec{a}_1). Ainsi, par minimisation de la norme ℓ^1 , C sera exprimé selon les directions \vec{a}_2 et \vec{a}_3 .

chaque atome du domaine d'analyse (il ne peut y avoir plus de sources que d'observations sur chaque atome⁵) :

$$\text{LqBP : } \min_{\underline{s}} \|\underline{s}\|_q \text{ sous les contraintes } A\underline{s} = \underline{x} \text{ et } \|\underline{s}\|_0 \leq P \quad (1.32)$$

- Dans [243], M. Xiao et al. montrent les limitations de la résolution du problème P_1 et proposent une approche basée sur une décomposition statistiquement parcimonieuse dans le cas de sources indépendantes. En effet, selon les auteurs, la solution la plus parcimonieuse n'est pas forcément la meilleure solution au problème de SAS sous-déterminé (par exemple, dans le cas de la figure 1.4, rien ne dit qu'en *réalité*, le point C ne dépend pas de la direction \vec{a}_1). En notant $R_{\underline{s}}(\tau)$ la matrice d'autocovariance des sources et en supposant que P sources sont présentes, on a la relation suivante :

$$R_{\underline{s}}(\tau) \simeq \begin{bmatrix} R_{\underline{s}}^{sub}{}_{P \times P} & 0_{P \times (N-P)} \\ 0_{(N-P) \times P} & 0_{(N-P) \times (N-P)} \end{bmatrix}. \quad (1.33)$$

Comme la matrice de mélange A est estimée, la sous-matrice d'autocovariance $R_{\underline{s}}^{sub}(\tau)$ peut être estimée de la façon suivante :

$$R_{\underline{s}}^{sub} \simeq A_{j_1, \dots, j_P}^{-1} R_{\underline{x}}(\tau) (A_{j_1, \dots, j_P}^{-1})^T. \quad (1.34)$$

⁵Dans le cas particulier où $q = 1$ et où la matrice de mélange est réelle, les problèmes P_1 et L1BP sont équivalents. De plus, si les observations sont réelles elles aussi, les problèmes P_q et LqBP sont eux aussi équivalents (pour $0 < q < 1$) [212].

M. Xiao et al. proposent de trouver les numéros des sources $[\hat{j}_1, \dots, \hat{j}_P]$ présentes dans le domaine d'analyse en minimisant le critère suivant

$$[\hat{j}_1, \dots, \hat{j}_P] = \arg \min \frac{\sum_{i=1}^P \sum_{j>i} |R_{\underline{s}}^{sub}(i, j)|}{\sqrt{\prod_{i=1}^P R_{\underline{s}}^{sub}(i, i)}}, \quad (1.35)$$

où $R_{\underline{s}}^{sub}(i, j)$ est le (i, j) ^{ème} élément de $R_{\underline{s}}^{sub}$. Par rapport à la résolution du problème P_1 , cette façon de reconstruire les sources est plus rapide (environ 0.4s contre environ 1200s sur un ordinateur équipé d'un processeur Intel Pentium IV 1.8G et 256MB de RAM) et donne de meilleures performances (SIR entre 1 et 10 dB supérieurs en moyenne sur les sources suivant les cas).

4. Enfin, dans [15–17, 19], A. Aïssa-El-Bey et al. proposent de projeter les sources en sous-espaces pour traiter le cas sous-déterminé. Cette méthode a été développée à la fois pour les mélanges LI et convolutifs. Pour simplifier l'explication, nous n'allons détailler ici que le cas LI, avec une transformée TF. Les auteurs supposent que dans chaque point du domaine d'analyse, au plus $P - 1$ sources sont présentes. Une fois la matrice A estimée, pour chaque point du domaine d'analyse, il cherchent les K ($K < P$) indices α_k des sources présentes en minimisant :

$$\{\alpha_1, \dots, \alpha_K\} = \arg \min_{\beta_1, \dots, \beta_K} \left\{ \|Q_{\tilde{A}} \underline{X}(t, \omega)\| \|\tilde{A}_{\beta}\| \right\} \quad (1.36)$$

où \tilde{A}_{β} est la sous-matrice de A contenant les K colonnes d'indices β_k et $Q_{\tilde{A}}$ est la projection orthogonale sur le sous-espace de A correspondant :

$$Q_{\tilde{A}} = I - \tilde{A}_{\beta} \left(\tilde{A}_{\beta}^H \tilde{A}_{\beta} \right)^{-1} \tilde{A}_{\beta}^H. \quad (1.37)$$

Nous allons maintenant détailler les diverses méthodes de SCA en fonction des hypothèses de parcimonie faites. Notons qu'il existe deux familles de décompositions parcimonieuses. En effet, historiquement, les premières décompositions parcimonieuses étaient des décompositions TF qui peuvent être choisies selon une distribution énergétique (transformation quadratique) ou une décomposition atomique (transformation linéaire). Nous allons donc découper l'étude ci-dessous entre deux sections. Dans un premier cas, nous allons étudier les méthodes faisant appel à une transformation quadratique (les méthodes de SAS correspondantes sont issues des approches statistiques classiques). Dans un second point, nous détaillerons les méthodes faisant appel à des décompositions linéaires.

1.3.1 Méthodes issues des approches statistiques classiques

A. Belouchrani et M. G. Amin [32] ont été parmi les premiers à proposer une approche orientée temps-fréquence (TF) pour répondre au problème de la SAS dans le cas des mélanges LI sur-déterminés ($N < P$). Leur méthode, qui constitue une version TF du classique algorithme SOBI, permet ainsi de réaliser la séparation des sources à partir de l'exploitation des caractéristiques TF disponibles dans les observations.

Leur principale idée est basée sur l'exploitation des différences dans l'évolution des densités spectrales de chaque source au cours du temps. Ils supposent qu'il existe des zones dans le plan TF où une seule source est présente à la fois. L'exploitation de ces zones leur permet alors d'estimer les coefficients de la matrice qu'il leur suffit d'inverser.

Les auteurs utilisent une transformée TF quadratique de la classe de Cohen [58] définie par :

$$D_{xy}(t, \omega) = \sum_{l=-\infty}^{+\infty} \sum_{m=-\infty}^{+\infty} \phi(m, l) x(t+m+l) y^*(t+m-l) e^{-j2\omega l}, \quad (1.38)$$

où ϕ représente le noyau de la transformée.

Leur approche fonctionne alors en plusieurs étapes :

- Estimation de la matrice d'autocorrélation des observations et des N plus grandes valeurs propres et vecteurs propres associés.
- Estimation de la variance de bruit, sous l'hypothèse de bruit stationnaire temporellement blanc, puis détermination d'une matrice de blanchiment des observations \hat{W} à l'aide de la matrice d'autocorrélation et de la variance du bruit.
- Calcul de K matrices $N \times N$ transformées TF D_{zz} entre les observations blanchies $z_i, i = 1, \dots, N$ pour K coordonnées (t, ω) ne contenant que des auto-termes, c.-à-d. une seule source.
- Calcul d'une matrice unitaire \hat{U} à partir de la diagonalisation conjointe des K matrices D_{zz} .
- Estimation des sources selon la relation

$$\underline{y}(t) = \hat{U}^H \hat{W} \underline{x}(t). \quad (1.39)$$

Leur approche permet ainsi de réaliser la séparation des sources à partir de l'exploitation des caractéristiques TF disponibles dans les observations mais nécessite toutefois la détermination de zones dans le plan TF où une seule source est présente. Or aucune méthode n'est proposée pour réaliser cette détermination de manière automatique, ce qui est un handicap majeur. Par ailleurs, la diagonalisation conjointe utilisée nécessite un nombre K très élevé, de l'ordre de quelques centaines, pour obtenir une séparation satisfaisante, ce qui implique un important coût calculatoire. De plus, les résultats restent modestes avec des valeurs de SIR comprises entre 14 et 26 dB.

Dans [30], A. Belouchrani et al. proposent de combiner l'approche de diagonalisation conjointe précédente à celle d'antidiagonalisation conjointe pour des mélanges LI sur-déterminés. La distribution TF d'une auto-source est définie comme

$$D_{ss}^a(t, \omega) = D_{ss}(t, \omega), \quad (1.40)$$

pour les localisations TF (t, ω) correspondant à un auto-terme. Chaque matrice $D_{ss}^a(t, \omega)$ est quasi-diagonale⁶. De même, les auteurs définissent la distribution TF d'une inter-source :

$$D_{ss}^i(t, \omega) = D_{ss}(t, \omega), \quad (1.41)$$

⁶Ces matrices ne sont pas toujours diagonales à cause d'effets de bord dus à la longueur finie des fenêtres de pondération.

si (t, ω) est la position correspondant à un inter-terme. Ainsi, puisque les éléments diagonaux de $D_{ss}^i(t, \omega)$ sont les auto-termes, cette matrice là a sa diagonale quasi-nulle. Dans la suite de cette section, par abus de langage, on parlera de "matrice auto- ou inter-source" pour définir la "matrice de distribution TF d'une auto- ou inter-source". Pour savoir quel point TF correspond à une matrice auto-source ou inter-source, les auteurs proposent de mesurer le rapport entre la trace de $D_{xx}(t, \omega)$ et sa norme. Si ce rapport est au-dessus d'un seuil défini par l'utilisateur, alors dans le point TF considéré, il n'y a qu'une seule source. Sinon, il y en a plusieurs. Les étapes de la méthode sont presque les mêmes que celle ci-dessus :

- Estimation de la matrice d'autocorrélation des observations et des N plus grandes valeurs propres et vecteurs propres associés.
- Estimation de la variance de bruit, sous l'hypothèse de bruit stationnaire temporellement blanc, puis détermination d'une matrice de blanchiment des observations \hat{W} à l'aide de la matrice d'autocorrélation et de la variance du bruit.
- Calcul de K_1 matrices D_{zz}^a ne contenant que des auto-termes et K_2 matrices D_{zz}^i ne contenant que des inter-termes.
- Calcul d'une matrice unitaire \hat{U} à partir d'une approche combinée de diagonalisation conjointe et d'antidiagonalisation conjointe des K matrices D_{zz} :

$$\hat{U} = \arg \max_V \sum_{i=1}^N \left(\sum_{k=1}^{K_1} |v_i^H D_{zz}^a v_i|^2 - \sum_{k=1}^{K_2} |v_i^H D_{zz}^i v_i|^2 \right) \quad (1.42)$$

- Estimation des sources selon l'Eq. (1.39).

Cette approche permet de relâcher la recherche d'auto-termes puisqu'on peut aussi travailler sur les inter-termes. Les performances sont légèrement meilleures (2dB de mieux sur le test comparatif) qu'avec l'algorithme précédent [32].

Dans [178], L.-T. Nguyen et al. remplacent la transformée TF de la classe de Cohen par une transformée de Wigner-Ville et proposent une extension de [32] pour les mélanges sous-déterminés. Pour cela, ils supposent que les sources sont WDO. Ils proposent de garder tous les auto-termes de puissance non-négligeable (les auto-termes sont sélectionnés selon la procédure décrite dans [30]). Pour chaque matrice $D_{xx}(t_k, \omega_k)$ ne contenant que des auto-termes, le vecteur propre $\underline{a}(t_k, \omega_k)$ et la valeur propre $\lambda(t_k, \omega_k)$ associés sont calculés. Les vecteurs propres sont ensuite groupés les uns aux autres par une mesure de distance. N clusters C_i sont ainsi obtenus, chacun correspondant aux positions des localisations TF correspondant aux auto-termes d'une source. Au final, les distributions TF de chaque source i sont estimées de la façon suivante :

$$\hat{D}_{s_i s_i}(t, \omega) = \begin{cases} \lambda(t, \omega) & \text{si } (t, \omega) \in C_i, \\ 0 & \text{sinon.} \end{cases} \quad (1.43)$$

Les auteurs proposent ensuite d'utiliser une des solutions de la littérature pour reconstruire les sources à partir de leurs distributions TF estimées respectives mais n'en testent par contre aucune dans [178].

Dans [14, 19], A. Aïssa-El-Bey et al. proposent une amélioration de [178] pour les mélanges LI sous-déterminés. L'estimation de la matrice de mélange est basée sur les mêmes paramètres que dans [178] sauf que la classification des valeurs propres est faite par l'algorithme K-means. La reconstruction des sources est basée sur la projection en sous-espaces, comme nous l'avons au début de ce panorama de méthodes parcimonieuses. Une extension convolutive a été proposée dans [14, 18].

Des approches LI et convolutives similaires mais utilisant la décomposition modale empirique (EMD pour *Empirical Mode Decomposition* en anglais) [97, 123] ont été proposées dans [14, 15, 17].

Dans [153], Y. Luo et al. proposent deux extensions de [178] qui font appel à des techniques avancées de clustering. En effet, dans le cadre de mélanges sous-déterminés, quand le nombre N de sources est inconnu, il est difficile de l'estimer correctement. Afin de résoudre ce problème, Y. Luo et al. proposent de tester deux méthodes de clustering : les GS (pour *Gap statistics* en anglais) [112] et SSCL (pour *Self-Splitting Competitive Learning*) [249]. Pour ces deux méthodes, ils considèrent un ensemble de vecteurs propres de $D_{xx}(t, \omega)$, notés $\underline{v}(t, \omega)$, où les points TF (t, ω) sont sélectionnés tels que $D_{xx}(t, \omega)$ soit de norme suffisamment grande. L'ensemble de ces vecteurs propres est noté Z . L'hypothèse de WDO est supposée vérifiée.

1. Approche utilisant les GS : Supposons que Z est partitionné en k clusters C_1, \dots, C_k . La distance moyenne deux à deux des éléments du cluster C_i est donnée par la relation suivante :

$$\zeta_i = \frac{1}{2n_i} \sum_{r, r' \in C_i} \|\underline{v}(t_r, \omega_r) - \underline{v}(t_{r'}, \omega_{r'})\|_2^2, \quad (1.44)$$

où $\|\cdot\|_2$ est la norme ℓ^2 , r et r' sont les différents indices des vecteurs du cluster C_i et n_i est le nombre de vecteurs dans le cluster C_i . En tenant compte de tous les clusters, l'erreur de dispersion des clusters mesurée dans les GS est définie par

$$E_k = \sum_{i=1}^k \zeta_i. \quad (1.45)$$

Le logarithme de l'erreur E_k est comparé à celui de l'erreur calculée à partir d'une référence Z' qui est créée selon une distribution appropriée. L'objectif est de trouver la valeur de k qui maximise la distance entre le logarithme de E_k et le logarithme de l'erreur calculée à partir de Z' , d'où le nom de *Gap* dans les GS. Les éléments de Z' sont uniformément distribués selon les composantes principales de Z . La justification de ces choix est donnée dans [112]. Pour chaque valeur de k , la méthode de clustering utilisée est K-means.

2. Méthode utilisant l'approche SSCL : L'idée de base est d'associer les clusters C_i à des prototypes \underline{p}_i qui les caractérisent. L'algorithme SSCL est initialisé avec un seul prototype choisi aléatoirement dans l'ensemble des données. Pendant la période d'apprentissage, un des prototypes (initialement le seul prototype) est coupé en deux selon une mesure de validité de division. Ce comportement d'auto-division s'arrête si aucun prototype ne peut se diviser. Après la période d'apprentissage, chaque

cluster est labélisé par un prototype situé en son centre. Appliqué à la séparation de sources, comme chaque cluster correspond à une source⁷, cela signifie que chaque prototype créé correspond à une source. Si le nombre de prototypes est inférieur au nombre N de sources, la séparation sera alors partielle, ce qui n'est pas voulu ici. Pour éviter cette situation, Y. Luo et al. proposent d'utiliser un vecteur \underline{f}_l qui guide l'apprentissage des prototypes \underline{p}_l .

De plus, dans ce même article [153] et dans [152], les auteurs proposent de relâcher l'hypothèse de WDO. Ils supposent que pour chaque point TF, au plus P sources sont présentes. Pour cela ils considèrent à nouveau l'ensemble des points TF tels que $D_{xx}(t, \omega)$ soit de norme suffisamment grande. Pour chacun de ces points TF (t_i, ω_i) , une décomposition en vecteurs propres est faite :

$$D_{xx}(t_i, \omega_i) = V\Lambda(t_i, \omega_i)V^{-1}, \quad (1.46)$$

où

$$\Lambda(t_i, \omega_i) = \text{diag}(\lambda_1, \dots, \lambda_N). \quad (1.47)$$

Si $\lambda_1 > \dots > \lambda_l \gg \lambda_{l+1} > \dots > \lambda_N$, alors le point TF (t_i, ω_i) est assigné à l'ensemble Γ_l défini par :

$$\Gamma_l = \{(t_i, \omega_i) \text{ tel que } l \text{ sources actives sont présentes}\}. \quad (1.48)$$

Chaque ensemble Γ_l ($l \geq 2$) non-vide contient l sources qui ne sont pas nécessairement les mêmes (par exemple, si $N = 3$, l'ensemble Γ_2 peut contenir des points TF où les sources des paires $\{1,2\}$, $\{1,3\}$ et $\{2,3\}$ peuvent être actives). C'est pourquoi chacun de ces ensembles Γ_l est partitionné, à l'aide d'une des deux méthodes de clustering précédentes, en sous-ensembles Γ_l^i qui contiennent chacun les mêmes l sources. L'algorithme JADE est ensuite appliqué à chacun de ces sous-ensembles Γ_l^i ($l \geq 2$) afin d'estimer les sources $\widehat{\underline{S}}(t, \omega)$ et les vecteurs de A correspondants. Les sources sont aussi séparées dans l'ensemble Γ_1 (composé de points TF où une seule source est active) en appliquant FOCUSS [103] ou en reprenant l'un des deux algorithmes faisant l'hypothèse de WDO de [153] décrits ci-dessus. Les vecteurs de A correspondants sont là encore estimés. A ce stade, nous avons un ensemble de vecteurs $\widehat{\underline{S}}(t, \omega)$ associés à des colonnes de A . Il reste alors une indétermination de permutation sur le numéro des sources extraites pour chaque atome TF (t, ω) . Les auteurs proposent de régler cette indétermination à l'aide de l'estimation de la matrice de mélange : les vecteurs de A issus de la séparation dans Γ_1 et dans les Γ_l^i sont groupés à l'aide de l'approche K-means. Les résultats de cette classification permet ainsi de reconstruire les sources dans le domaine TF, en regroupant de la même façon les valeurs estimées $\widehat{\underline{S}}(t, \omega)$ associées à ces colonnes de A .

Dans [99, 100], C. Févotte et C. Doncarli donnent une vision stochastique de l'approche déterministe [32] de A. Belouchrani et M. G. Amin en remplaçant la transformée de la classe de Cohen par un spectre spatial de Wigner-Ville. De plus, contrairement à l'approche initiale [32], ils proposent un critère pour trouver les zones du plan TF où une

⁷En réalité, chaque cluster correspond à une colonne de la matrice de mélange. Cependant, sous l'hypothèse de WDO, connaître la colonne de la matrice de mélange dans un atome TF donné correspond à connaître le numéro de la source correspondant.

seule source est présente [99, 100, 117] : en effet, si le point TF (t, ω) correspond à un auto-terme, alors $D_{ss}(t, \omega)$ est diagonale avec un seul élément non-nul dans la diagonale. Cette matrice a donc une seule valeur propre $\lambda_k(t, \omega)$ non-nulle. Ainsi, si (t, ω) correspond à un auto-terme, alors

$$C(t, \omega) \triangleq \frac{\max_k |\lambda_k(t, \omega)|}{\sum_k |\lambda_k(t, \omega)|} = 1. \quad (1.49)$$

Au contraire, si le point TF étudié correspond à un inter-terme, alors $C(t, \omega) < 1$. De plus, les valeurs propres de $D_{ss}(t, \omega)$ sont presque égales à celles de $D_{zz}(t, \omega)$. Ainsi,

$$C(t, \omega) \simeq \frac{\max_k |\lambda_k^{D_{zz}}(t, \omega)|}{\sum_k |\lambda_k^{D_{zz}}(t, \omega)|}, \quad (1.50)$$

où $\lambda_k^{D_{zz}}(t, \omega)$ sont les valeurs propres de $D_{zz}(t, \omega)$. Il suffit de chercher les points TF qui maximisent le critère (1.50) pour trouver les auto-termes et réaliser la diagonalisation conjointe.

Cette approche [99, 100] a été étendue au cas de sources corrélées dans [102] par L. Giulieri et al.

Très récemment, une approche basée sur [32] mais utilisant une transformée temps-temps (TT) a été proposée par R. M. Parry et I. Essa [188]. La représentation TT de Foote [98] partitionne le signal en petites zones temporelles, dessine l'enveloppe du spectre du signal sur ces zones et compare chaque paire de zones via une mesure de similarité. Cette transformation permet de mettre en valeur les structures répétitives des morceaux de musique (couplets, refrain) et d'y trouver l'information nécessaire à la séparation. Les auteurs de [188] proposent d'utiliser une distribution temps-temps-fréquence fonction de deux points temporels

$$D'_{xy}(t_1, t_2, \omega) = \sum_{\tau=-\infty}^{+\infty} \phi(\tau) x(t_1 + \frac{\tau}{2}) y^*(t_2 - \frac{\tau}{2}) e^{-j\omega\tau}, \quad (1.51)$$

mais ne se concentrent que sur l'aspect temporel :

$$\mathcal{D}_{x_1x_2}(t_1, t_2) = D'_{x_1x_2}(t_1, t_2, 0). \quad (1.52)$$

Ils proposent ensuite d'appliquer la méthode [32] en remplaçant la transformée définie dans l'Eq. (1.38) par celle de l'Eq. (1.52). Les auteurs ont testé leur méthode dans le cas peu favorable où les sources ne se répètent qu'une seule fois. Si les sources sont peu recouvrantes dans le domaine TF, les performances de cette approche sont comparables avec celles de l'approche TF originale. Au contraire, si les sources sont très recouvrantes dans le domaine TF, l'approche TT surpasse alors la méthode TF.

R. M. Parry et I. Essa ont aussi utilisé cette transformée TT pour détecter la présence de sources actives dans le cadre de mélanges LI sous-déterminés [189].

1.3.2 Méthodes utilisant des transformations parcimonieuses linéaires

Méthodes faisant l'hypothèse de la WDO (approximative)

DUET et ses extensions : En se basant sur des rapports TF des observations, A. Jourjine et al. [132] proposent de résoudre le problème de SAS dans le cas de mélanges sous-déterminés AD comprenant N sources et 2 observations :

$$\begin{cases} x_1(t) = \sum_{i=1}^N s_i(t) \\ x_2(t) = \sum_{i=1}^N a_{2i} s_i(t - t_i) \end{cases} \quad (1.53)$$

En supposant que les retards t_i soient *suffisamment petits* par rapport à la longueur temporelle de la fenêtre de pondération W utilisée pour calculer les transformées TF, nous obtenons la relation suivante :

$$\begin{bmatrix} X_1(t, \omega) \\ X_2(t, \omega) \end{bmatrix} = \begin{bmatrix} 1 & \dots & 1 \\ a_{21} e^{-j\omega t_1} & \dots & a_{2N} e^{-j\omega t_N} \end{bmatrix} \begin{bmatrix} S_1(t, \omega) \\ \vdots \\ S_N(t, \omega) \end{bmatrix}. \quad (1.54)$$

Leur objectif est d'estimer la matrice $A(\omega)$ définie par :

$$A(\omega) = \begin{bmatrix} 1 & \dots & 1 \\ a_{21} e^{-j\omega t_1} & \dots & a_{2N} e^{-j\omega t_N} \end{bmatrix}. \quad (1.55)$$

A. Jourjine et al. supposent que les sources sont WDO dans le plan TF associé à la transformée considérée. Ainsi, pour une fenêtre TF (t, ω) où seule la source S_i est présente, les signaux observés deviennent :

$$\begin{bmatrix} X_1(t, \omega) \\ X_2(t, \omega) \end{bmatrix} = \begin{bmatrix} 1 \\ a_{2i} e^{-j\omega t_i} \end{bmatrix} S_i(t, \omega) \quad (1.56)$$

d'où

$$\frac{X_2(t, \omega)}{X_1(t, \omega)} = a_{2i} e^{-j\omega t_i}. \quad (1.57)$$

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

$$(a_{2i}, t_i) = \left(\left| \frac{X_2(t, \omega)}{X_1(t, \omega)} \right|, -\Im \left(\log \left(\frac{X_2(t, \omega)}{X_1(t, \omega)} \right) \right) / \omega \right), \quad (1.58)$$

où \Im désigne la partie imaginaire. Il reste maintenant à déterminer les zones TF (t, ω) où chacune des N sources est présente. Pour cela, les auteurs effectuent un histogramme amplitude / phase de toutes les valeurs (a_{2i}, t_i) rencontrées. Comme les sources sont supposées WDO, cet histogramme présente des pics autour des valeurs correspondant aux paramètres de la matrice de mélange vue dans le plan TF.

Notez que les décalages temporels à estimer t_i doivent appartenir à l'intervalle $[-\pi, \pi]$ pour chaque pulsation ω . Cette limitation contraint la méthode à de faibles décalages

temporels et/ou de petites fréquences, comme cela a été souligné dans des applications de localisation des sources et/ou de séparation de sources musicales [23, 176, 236]. De plus, dans un environnement réel anéchoïque, M. Baek et U. Zölder montrent expérimentalement que les histogrammes sont étalés et donc qu'il est difficile de trouver le pic théorique [23].

Conscients de la limitation de la WDO, Jourjine et al. ont montré que cette propriété pouvait être légèrement relaxée en considérant une version approximative, en particulier pour la parole [209].

Enfin, les auteurs discutent du nombre de sources par rapport au nombre d'observations et s'en servent pour reconstruire les sources de deux façons différentes :

- dans le cas où le nombre N de sources est inférieur ou égal au nombre P d'observations, la matrice de mélange (1.55) peut directement être inversée, ce qui fournit les transformées de Fourier des sources, qui sont ensuite reconstruites dans le domaine temporel par transformées de Fourier inverse.
- Au contraire, dans le cas sous-déterminé, la matrice ne peut être inversée. Les auteurs proposent alors de reconstruire approximativement les sources à partir des zones TF associées à chacune d'entre elles.

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

1. détermination de l'histogramme amplitude / phase,
2. à partir des valeurs obtenues par l'histogramme précédent, refaire un deuxième passage afin de déterminer les zones TF associées à chaque source. Cette opération définit un *masque TF* propre à chaque source. Pour le cas sous-déterminé, l'estimation de chacune des sources est réalisée en calculant la transformée TF inverse d'une observation en ne considérant que les zones définies par chaque masque, c.-à-d. en forçant à zéro tout le domaine extérieur au masque considéré.

Afin de ne plus nécessiter ces deux étapes, cette méthode a été légèrement modifiée [27, 248] en introduisant une analyse par maximum de vraisemblance.

H. Viste et G. Evangelista ont été parmi les premiers à essayer de réduire les limitations de la WDO dans le cas particulier de sources musicales [237]. Lorsque deux instruments de musique jouent deux partitions d'un même morceau de musique, il se peut que certaines de leurs harmoniques soient recouvrantes et génèrent des battements. En acoustique, un battement est l'interférence entre deux sons de fréquences très proches, perçue comme des variations de volume. La fréquence du battement est égale à la différence des fréquences des sons. De nombreux instruments de musique génèrent des battements comme l'orgue par exemple. La version originale de DUET [132] est donc incapable de séparer de tels sons (puisque'ils sont recouvrants) et assigne les battements à l'une ou l'autre des sources. H. Viste et G. Evangelista ont proposé une technique qui permet de séparer deux harmoniques recouvrantes dans le cas de deux sources et deux capteurs. Pour cela, l'idée est de supprimer les harmoniques recouvrantes dans les sources reconstruites par DUET dans le plan TF et de chercher à les recréer en s'inspirant de la forme des enveloppes des harmoniques voisines (en fréquence) non-recouvrantes.

Dans [53], C. Choi propose une version temps-réel, inspirée de DUET, pour traiter le

cas de mélanges AD sous-déterminés dans un environnement bruité. A chaque indice temporel t , pour chaque pulsation ω , il n'y a au plus qu'un couple (a_{2i}, t_i) , défini dans l'Eq. (1.58). Afin d'éviter les points qui ne contiennent que du bruit et qui pourraient perturber l'identification des paramètres de mélange, C. Choi propose d'estimer les probabilités d'absence de sources et de s'en servir pour estimer la puissance du bruit. Dans les points qui contiennent des sources, au lieu de construire un histogramme 2D des données (a_{2i}, t_i) , il propose d'utiliser une version pondérée de la méthode de classification non-supervisée (*clustering*) K-means (appelée *soft K-means* dans le chapitre 20 de [155]) qui assigne chaque point à tous les clusters avec un certain degré. Pour reconstruire les sources dans l'environnement bruité, C. Choi propose d'estimer l'amplitude spectrale des sources par minimisation de l'erreur au sens des moindres carrés de l'estimateur $\mathbb{E} \{ |\underline{S}(t, \omega)| \mid \underline{X}(t, \omega) \}$. Enfin, comme cette méthode est appliquée dans un environnement réel peu réverbérant, et comme l'environnement réel change continuellement, les variations des puissances des signaux sources et de bruit sont estimées à l'aide d'un calcul de RSB a priori et a posteriori : à partir des probabilités d'absence de source, les variances du bruit et des sources sont remises à jour dans les zones TF correspondantes et les RSB a priori et a posteriori peuvent être calculés. En cas de changement brusque du RSB a posteriori, le RSB a priori est changé lors des échantillons suivants.

M. Baeck et U. Zölder proposent eux-aussi une version temps-réel de DUET dans [24] utilisant l'analyse par maximum de vraisemblance de [206, 248].

Pour relâcher la contrainte de WDO, S. Rickard et al. [208] ont proposé de combiner la méthode d'estimation des directions d'arrivée (DOA pour *Direction-of-Arrival* en anglais) ESPRIT [210] à DUET et appellent cette méthode DESPRIT. ESPRIT divise les P mélanges originaux ($P > 2$) en deux ensembles, éventuellement recouvrants, de m mélanges, avec $\frac{P}{2} \leq m \leq P - 1$ et permet d'estimer les directions des sources, c.-à-d. les colonnes de la matrice de mélange $A(\omega)$. Dans le cas où les micros sont placés linéairement (ce qui est supposé dans l'approche DESPRIT), $m = P - 1$. Grâce à ces deux ensembles, ESPRIT permet d'estimer les directions des sources à chaque fenêtre TF, tant qu'il n'y a pas plus de m sources actives par fenêtre, ce qui est moins contraignant que la WDO. Les auteurs proposent d'utiliser un système basé sur ESPRIT pour tracer un histogramme 2D atténuation/décalage, les centres des pics obtenus correspondant aux directions des sources. Une fois la matrice de mélange estimée, toutes les directions trouvées dans le plan TF sont assignées à la plus proche des N directions obtenues précédemment. Enfin, via une transformée TF inverse, les versions temporelles des signaux sont obtenus. De plus, contrairement à DUET, DESPRIT a besoin de $P > 2$ observations et peut donc être vu comme une extension du DUET sur le nombre d'observations. T. Melia et S. Rickard ont étendu DESPRIT aux mélanges convolutifs généraux dans [170, 171].

R. Saab et al. [211] ont eux aussi proposé une extension de DUET à $P > 2$ observations. Ils utilisent la méthode de classification K-means [169, 244]. Enfin, ils proposent de reconstruire les sources selon deux approches : une première est basée sur une minimisation ℓ^q ($0 < q < 1$) du problème P_q explicité dans l'Eq. (1.31). Une deuxième reconstruction, basée sur une minimisation ℓ^1 modifiée, est aussi proposée.

Dans [14, 16, 19], A. Aïssa-El-Bey et al. proposent d'une part une extension de DUET à $P \geq 2$ observations, mais pour les mélanges LI et d'autre part, de relâcher l'hypothèse de WDO. Dans un premier temps, ils ne considèrent que les zones qui ont une énergie non-négligeable, puis normalisent les observations dans les autres points TF. Les colonnes de la matrice de mélange sont les centroïdes des N clusters obtenus par K-means sur les observations normalisées. Pour relâcher la WDO, ils font appel à une projection en sous-espace pour reconstruire les sources, comme discutée précédemment dans ce manuscrit.

Dans [38–40], A. Blin et al. proposent une alternative intéressante à DUET, en s'inspirant du formalisme des méthodes parcimonieuses de Y. Deville et al. Leur méthode est applicable aux mélanges convolutifs :

$$x_i(t) = \sum_{j=1}^N h_{ij}(t) * s_j(t) \quad i = 1, \dots, P. \quad (1.59)$$

Ils proposent une approche fréquentielle (cf section 1.2, page 3 de ce manuscrit) et relâchent l'hypothèse de WDO. Ils proposent une reconstruction des sources proche du problème de minimisation P_1 : considérer qu'une seule source est active dans chaque fenêtre TF est trop restrictif. Supposons qu'il existe des zones TF où P sources sont présentes. Si on a au préalable estimé la matrice de mélange initiale de dimension $P \times N$ et si on a des zones où seules P sources sont actives, il est alors possible d'inverser la matrice $P \times P$ correspondante et donc de séparer ces P sources dans cette zone. Au final, les auteurs espèrent une estimation des sources beaucoup plus fine qu'avec DUET. Leur méthode est composée de trois étapes, dans le cas particulier où $P = 2$:

1. Dans une première étape, pour chaque fréquence, ils déterminent les fenêtres temporelles où une seule source est active. Pour cela, ils tracent des nuages de points des observations. Des directions des sources sont déduites de ces tracés et correspondent aux colonnes de la matrice de mélange. Considérant de petites bandes autour de ces directions, ils trouvent les fenêtres où une seule source est présente. Ils peuvent alors assigner ces points aux sources correspondantes par masquage, comme cela a été proposé pour DUET.
2. Dans un second temps, en fixant les indéterminations d'échelle en s'inspirant de l'approche proposée par Y. Deville dans [74], ils proposent d'estimer la matrice de mélange en calculant les rapports des observations dans les zones mono-sources trouvées dans les bandes autour des directions précédemment trouvées.
3. Dans une troisième étape, ils proposent de séparer $P = 2$ sources dans les fenêtres multi-sources grâce à la matrice de mélange estimée dans l'étape précédente. Pour cela, les auteurs supposent qu'entre deux directions des nuages de points, seules les deux sources correspondant aux deux directions sont présentes. Ils espèrent ainsi que le nombre de zones où $N > 2$ sources seront présentes sera faible et que les distorsions engendrées seront limitées. Dans ces zones où deux sources sont présentes, le problème de SAS est alors déterminé et il est donc possible d'inverser la matrice de mélange.

Par contre, aucune indication n'est donnée concernant l'appariement des sources, c.-à-d. la façon de résoudre l'indétermination de permutation bande à bande propres aux méthodes fréquentielles convolutives.

Dans [11], S. Araki et al. proposent une extension de DUET pour les mélanges convolutifs sous-déterminés qu'ils appellent MENUET. Ils proposent de modifier le critère (1.58) qui permet d'estimer les paramètres de mélange et utilisent la méthode de clustering K-means pour trouver des représentants de ceux-ci. A la place de l'Eq. (1.58), ils proposent de tracer des nuages de points dont les éléments $\underline{\Theta}(t, \omega)$ sont définis par :

$$\underline{\Theta}(t, \omega) = [\underline{\Theta}^L(t, \omega), \underline{\Theta}^R(t, \omega)]^T, \quad (1.60)$$

avec

$$\underline{\Theta}^L(t, \omega) = \left[\frac{|X_1(t, \omega)|}{\mathcal{A}(t, \omega)}, \dots, \frac{|X_P(t, \omega)|}{\mathcal{A}(t, \omega)} \right], \quad (1.61)$$

$$\underline{\Theta}^R(t, \omega) = \left[\frac{1}{\alpha_1 \omega} \arg \left\{ \frac{X_1(t, \omega)}{X_k(t, \omega)} \right\}, \dots, \frac{1}{\alpha_P \omega} \arg \left\{ \frac{X_P(t, \omega)}{X_J(t, \omega)} \right\} \right]. \quad (1.62)$$

Dans ces Eq. (1.61) et (1.62), le paramètre est défini par

$$\mathcal{A}(t, \omega) = \sqrt{\sum_{j=1}^P |X_j(t, \omega)|^2}, \quad (1.63)$$

k est l'indice de l'une des observations et les α_j ($j = 1, \dots, P$) sont des constantes positives. Ces constantes permettent de donner un poids plus ou moins important au taux de mélange ou à la différence de décalages temporels. Si les α_j sont grands, alors une grande importance sera donnée aux rapports de $\underline{\Theta}^L(t, \omega)$. Au contraire, si les α_j sont petits, alors l'information obtenue grâce aux différences de phases sera prépondérante. Leur méthode peut être résumée ainsi :

1. Pré-traitement : calcul des TFCT des observations.
2. Etape d'identification des paramètres $\underline{\Theta}(t, \omega)$.
3. Clustering : utilisation de la méthode K-means sur les paramètres $\underline{\Theta}(t, \omega)$. Création de N clusters C_1, \dots, C_N . Tous les points d'un même cluster correspondent à des points TF où une même source est la seule active.
4. Création des masques TF : pour chaque source s_j , un masque binaire M_j est créé :

$$M_j(t, \omega) = \begin{cases} 1 & \text{si } \underline{\Theta}^L(t, \omega) \in C_j, \\ 0 & \text{sinon.} \end{cases} \quad (1.64)$$

5. Reconstruction des sources : chaque source est reconstruite dans le domaine TF :

$$Y_j(t, \omega) = M_j(t, \omega) X_k(t, \omega), \quad (1.65)$$

puis une version temporelle est obtenue par TFCT inverse.

Récemment, afin d'éviter les approximations des masques, R. Balan et J. Rosca ont proposé de "combler les trous TF" à l'aide d'hypothèses sur les contenus fréquentiels des sources [26]. Les signaux sources sont en effet modélisés dans le domaine fréquentiel comme le produit d'une variable aléatoire de Bernoulli avec une amplitude spectrale inconnue mais déterministe. Les variables aléatoires sont modélisées par des processus de Markov de premier ordre avec des probabilités de transition obtenues à partir de bases de données.

Méthodes de Bofill et Zibulevsky : Dans [43, 44], P. Bofill et M. Zibulevsky proposent une approche basée sur les nuages de points des TFCT des observations pour estimer la matrice de mélange dans le cadre de mélanges LI sous-déterminés. Ils estiment les directions des sources à l'aide d'une fonction potentiel pondérée :

$$\Phi(\theta) = \sum_k l_k \phi(\lambda(\theta - \theta_k)), \quad (1.66)$$

où l_k et θ_k sont resp. la distance et l'angle du $k^{\text{ième}}$ point $\underline{X}(t, \omega)$ du nuage de points, $\phi(\cdot)$ est une fonction de base définie autour de ce point $\underline{X}(t, \omega)$ par :

$$\phi(\alpha) = \begin{cases} 1 - \frac{\alpha}{\pi/4} & \text{si } |\alpha| < \frac{\pi}{4}, \\ 0 & \text{sinon} \end{cases} \quad (1.67)$$

et λ est un paramètre pour ajuster la résolution. Les maxima locaux de θ correspondent aux directions des sources et permettent d'estimer la matrice de mélange. La reconstruction des sources est vue sous l'angle d'une minimisation sous la norme ℓ^1 du problème P_1 défini dans l'Eq. (1.31).

Dans [41], P. Bofill propose une extension de la méthode LI décrite ci-dessus aux mélanges AD. Contrairement à [43, 44], la méthode ne fait pas appel à une transformée TF mais à une simple Transformée de Fourier Discrète (TFD). La méthode proposée comprend trois étapes :

1. Trouver les directions des modules des sources, c.-à-d. les colonnes de a_{ij} (les coefficients a_{ij} sont supposés supérieurs à 0), grâce à un nuage de points, comme cela a été proposé dans l'approche LI.
2. Trouver les décalages temporels. Pour cela, P. Bofill propose d'estimer la différence des décalages temporels de la source j aux divers capteurs permettant d'obtenir les observations. Pour cela, il suppose que la moyenne des valeurs t_{ij} est nulle. En particulier, dans le cas où $P = 2$, $t_{1j} = -t_{2j} = \delta_j/2$ où δ_j est la différence des temps de propagation définis ci-dessus. En notant les phases des sources $\phi_j(\omega)$, nous pouvons écrire les observations de la façon suivante :

$$X_i(\omega) = \sum_{j=1}^N a_{ij} e^{j(\phi_j(\omega_l) - t_{ij}\omega)} |S_j(\omega)| \quad i = 1, \dots, P. \quad (1.68)$$

Pour une pulsation ω_l où seule la source k est présente, les observations peuvent s'écrire :

$$\begin{aligned} X_1(\omega_l) &= a_{1k} e^{j(\phi_k(\omega_l) - \frac{\delta_k}{2}\omega_l)} |S_k(\omega_l)| \\ X_2(\omega_l) &= a_{2k} e^{j(\phi_k(\omega_l) + \frac{\delta_k}{2}\omega_l)} |S_k(\omega_l)| \end{aligned} \quad (1.69)$$

Finalement, les parties réelles de l'Eq. (1.69) s'écrivent :

$$\begin{aligned}\Re\{X_1(\omega_l)\} &= a_{1k} \cos(\phi_k(\omega_l) - \frac{\delta_k}{2}\omega_l) |S_k(\omega_l)| \\ \Re\{X_2(\omega_l)\} &= a_{2k} \cos(\phi_k(\omega_l) + \frac{\delta_k}{2}\omega_l) |S_k(\omega_l)|\end{aligned}\quad (1.70)$$

Si δ_k est nul (les capteurs dont sont issues les observations sont en phase), les arguments des deux cosinus sont identiques. Supposons que $\delta_k = 0$ pour tout k , alors les directions des sources du nuage de points ($\Re\{X_1(\omega_l)\}, \Re\{X_2(\omega_l)\}$) correspondent aux directions des sources trouvées dans l'étape précédente. Par contre, si $\delta_k \neq 0$, cette correspondance disparaît. P. Boffill propose alors d'annuler les décalages temporels en décalant une observation par rapport à une autre, de façon à rendre les capteurs en phase. En conséquence, pour chaque colonne de décalages temporels à estimer, il propose de maximiser une fonction potentiel $\Phi(\hat{\theta}_l, \rho)$ (qui est le pendant de celle décrite dans l'Eq. (1.66), mais qui tient compte des décalages temporels ρ) afin d'estimer le meilleur décalage pour que les deux nuages de points ci-dessus coïncident pour une direction $\hat{\theta}_l$ donnée. $\hat{\theta}_l$ est l'estimation de la direction du module de la $k^{\text{ième}}$ source. On peut facilement montrer que la valeur estimée $\hat{\rho}_k$ qui maximise ce décalage est l'opposée à δ_k . Ainsi, en estimant toutes les valeurs $\hat{\rho}_k$, les décalages temporels t_{ij} correspondant sont eux-aussi estimés.

3. Reconstruire les sources selon le problème P₁.

Méthodes "quasi-non-parcimonieuses"

Les approches parcimonieuses nécessitant la WDO sont intéressantes car elles permettent de séparer totalement les sources mais montrent leurs limitations pratiques quand le nombre de sources est élevé. En effet, le nombre d'atomes mono-sources dans l'espace d'analyse devient réduit et la reconstruction des sources grossière. Au contraire, les approches que nous allons présenter maintenant ne cherchent qu'une (ou un petit nombre de) zone mono-source par source et accepte que les sources soient recouvrantes dans le reste du domaine d'analyse.

Y. Deville et al. [1-5, 8, 37, 74, 77, 83] ont proposé plusieurs méthodes parcimonieuses temporelles, TF ou temps-échelle (TE) de SAS, toutes basées sur les mêmes étapes mais utilisant des critères différents. Ils supposent l'existence de petites zones dans le domaine d'analyse où une source est *isolée*, c.-à-d. elle est la seule à ne pas être nulle dans ces zones dites *mono-sources*. Leurs méthodes sont basées sur les étapes suivantes (éventuellement précédées d'une pré-étape de transformation TF ou TE) :

1. Etape de détection des zones mono-sources.
2. Etape d'identification : à chaque zone mono-source trouvée correspond une colonne de la matrice de mélange.
3. Etape de reconstruction : les observations sont recombinaées afin d'obtenir les signaux de sortie.

Afin de fixer l'indétermination d'échelle, Y. Deville a proposé de réécrire les mélanges en fonction des contributions des sources dans la première observation. Une nouvelle matrice

de mélange, notée B , équivalente à la précédente matrice de mélange A est ainsi obtenue (B est composée d'éléments égaux à 1 sur la première ligne et à des rapports d'éléments de A dans les autres lignes).

Les trois méthodes développées estiment donc la matrice B :

- F. Abrard et Y. Deville ont développé deux approches parcimonieuses de SAS, basées sur les rapports d'observations, pour les mélanges LI [1–4]. Ils baptisent resp. ces méthodes LI-TEMPROM et LI-TIFROM (pour *Linear-Instantaneous BSS method based on TEMPoral (resp. TIme-Frequency) Ratios Of Mixtures* en anglais). Leurs méthodes trouvent des zones mono-sources en étudiant la variance des rapports d'observations. Ces aspects seront discutés plus en détail dans la section 2.1.1 de la partie I puisqu'ils serviront de point de départ à des améliorations proposées durant mes travaux de thèse.
- Y. Deville a aussi proposé [74, 82] deux autres approches LI de SAS (une temporelle et une TF) basées sur les mêmes étapes que LI-TEMPROM et LI-TIFROM, mais utilisant des critères basés sur la CORRÉlation, d'où le nom des approches : LI-TEMPCORR et LI-TIFCORR. Ces approches seront détaillées dans la section 2.3.1 page 58 puisque j'ai participé à leur validation expérimentale.
- B. Albouy et Y. Deville ont aussi proposé une approche TF similaire basée sur la COHERENCE, d'où son nom LI-TIFCOHERE [5, 8, 83]. Cette méthode peut-être vue comme une généralisation de LI-TIFCORR, mais le formalisme adopté est différent, comme expliqué dans [83].

Toutes ces approches peuvent être appliquées aux mélanges sous-déterminés (la reconstruction proposée est partielle) mais aussi dans le cadre de sources *cachées*, c.-à-d. le cas où il n'y a pas de zones mono-sources pour certaines des sources dans le mélange original. Elles ont montré d'excellentes performances [2, 5, 82, 83] et les versions TF surpassent les méthodes classiques de SAS [2, 82].

Toutefois, leur principal défaut commun est qu'elles nécessitent le réglage d'un nombre important de paramètres heuristiques (taille des zones d'analyse, "distance minimale" entre deux colonnes de la matrice de mélange, etc).

Les méthodes LI-TEMPROM et LI-TIFROM souffrent aussi du caractère non-symétrique de la variance des rapports d'observations qui peut parfois amener de faibles performances suivant les valeurs de la matrice de mélange (ce comportement sera décrit dans la section 2.1.2 page 51 et nous proposerons un moyen d'y remédier dans la section 2.1.3 page 54).

Durant son stage de M2R sous la direction de Y. Deville, D. Bissessur a aussi limité le nombre de paramètres heuristiques des méthodes en changeant de transformée parcimonieuse. En appliquant une transformée TE aux signaux observés, le nombre de paramètres de la transformée est beaucoup plus faible que ceux de la transformée TF utilisée jusque là. Il expérimente cette transformée sur la méthode LI-TIFCORR, d'où le nom de LI-TISCORR (pour *TIme-Scale CORRelation-based BSS method*) pour cette nouvelle méthode [37, 77]. LI-TISCORR apporte globalement de bonnes performances mais un peu en deça de celles de LI-TIFCORR. Une explication possible à cette diminution de performances pourrait venir de la parcimonie de la transformée choisie. En effet, dans [207], S. Rickard et M. Fallon montrent que la transformée en ondelettes continues est moins parcimonieuse que la TFCT.

Vous trouverez aussi dans [5] l'extension de l'approche LI-TIFCOHERE aux mélanges

AD. Cette extension fait suite à mon stage de DEA, portant sur l'extension AD de la méthode LI-TIFFROM [203], et a fait l'objet d'une étroite collaboration entre B. Albouy et moi. Comme cette extension utilise les mêmes concepts et outils que les méthodes proposées dans ce manuscrit, je n'en parlerai pas plus ici.

Dans [5–7], B. Albouy et Y. Deville proposent une méthode parcimonieuse temporelle pour les mélanges convolutifs. Comme pour toutes les méthodes parcimonieuses de Y. Deville et al., l'idée est de trouver une zone (ici temporelle) mono-source et d'estimer dans cette zone une colonne de filtres de la matrice de mélanges.

Dans [220, 221], D. Smith et al. proposent de supprimer quelques paramètres heuristiques en appliquant la méthode de clustering K-means aux approches LI-TIFFROM et LI-TIFCORR. Ils étudient aussi le comportement non-symétrique de la variance des rapports et proposent de la normaliser en la divisant par la moyenne des rapports sur ces zones d'analyse. Enfin, ils proposent une version adaptative par bloc de ces versions modifiées de LI-TIFFROM et LI-TIFCORR. Récemment [222], ils ont proposé une version de LI-TIFFROM basée sur la déflation.

Dans [242], M. Xiao et al. proposent une méthode fréquentielle de SAS LI sous-déterminée :

$$X_i(\omega) = \sum_{j=1}^N a_{ij} S_j(\omega), \quad i = 1, \dots, P. \quad (1.71)$$

Leur approche fait appel aux mêmes concepts de zones mono-sources que les approches ci-dessus, mais dans le domaine fréquentiel. Si la source S_k est la seule non-nulle à la pulsation ω_l , alors

$$X_i(\omega_l) = a_{ik} S_k(\omega_l), \quad i = 1, \dots, P, \quad (1.72)$$

et les modules des observations vérifient

$$|X_i(\omega_l)| = |a_{ik}| |S_k(\omega_l)|, \quad i = 1, \dots, P. \quad (1.73)$$

Dans le cas où $P = 2$, les auteurs projettent les vecteurs $\underline{X}(\omega_l) = [|X_1(\omega_l)|, |X_2(\omega_l)|]^T$ sur le cercle unité et obtiennent des vecteurs $\underline{u}(\omega_l) = [u_1(\omega_l), u_2(\omega_l)]^T$ où les $u_i(\omega_l)$ sont définis par

$$u_i(\omega_l) = \frac{|X_i(\omega_l)|}{\|\underline{X}(\omega_l)\|}. \quad (1.74)$$

Les auteurs proposent de mesurer l'écart entre deux valeurs de \underline{u} pour deux pulsations consécutives :

$$d_l = \underline{u}(\omega_l) - \underline{u}(\omega_{l+1}). \quad (1.75)$$

Si $d_l = 0$, alors $\underline{X}(\omega_l)$ et $\underline{X}(\omega_{l+1})$ appartiennent à une même direction d'un nuage de points de $\underline{X}(\omega)$. Les auteurs cherchent des intervalles fréquentiels où d_l prend des valeurs nulles; ces intervalles étant alors considérés comme *mono-sources*.

Après avoir trouvé tous les intervalles mono-sources, ils proposent de tracer un nuage de points des modules des observations (projetées sur le cercle unité) et de chercher

les directions des sources en appliquant une méthode de clustering. Pour reconstruire les sources, ils utilisent l'approche proposée initialement par P. Bofill et M. Zibulevsky dans [44] qui consiste en une minimisation sous la norme ℓ^1 du problème P_1 défini dans l'Eq. (1.31). En appliquant cette méthode à des mélanges de sons de flutes, ils obtiennent des SIR compris entre 20 et 60 dB.

Récemment, R. Li et M. Xiao ont étendu cette approche aux mélanges AD [148]. Cette méthode peut être vue comme une extension de [41]. Le mélange (2) s'écrit dans le domaine fréquentiel

$$X_i(\omega) = \sum_{j=1}^N a_{ij} e^{-j\omega t_{ij}} S_j(\omega) \quad i = 1, \dots, P. \quad (1.76)$$

Les atténuations a_{ij} peuvent être négatives et sont donc égales à leur module, à un terme de phase $e^{j\theta_{ij}}$ près ($\theta_{ij} \in \{0, \pi\}$). Les auteurs se placent dans le cas où $P = 2$. L'idée est d'estimer les contributions des sources dans une des deux observations.

1. Dans un premier temps, les auteurs cherchent des zones mono-sources et estiment les directions associées aux vecteurs de modules des sources, de la même façon que pour la méthode LI [242].
2. Dans une deuxième étape, ils cherchent à estimer les décalages temporels. En notant resp. les phases des observations $\varphi_i(\omega)$ et des sources $\phi_i(\omega)$, nous pouvons resp. écrire les observations et les sources en fonction de leurs modules et de leurs phases :

$$X_i(\omega) = |X_i(\omega)| e^{j\varphi_i(\omega)} \quad i = 1, \dots, P, \quad (1.77)$$

$$S_j(\omega) = |S_j(\omega)| e^{j\phi_j(\omega)} \quad j = 1, \dots, N. \quad (1.78)$$

Dans une zone mono-source définie précédemment où seule la source k est active, les observations peuvent s'écrire

$$X_i(\omega_l) \simeq |a_{ik}| e^{j(\theta_{ik} - \omega_l t_{ik})} |S_k(\omega_l)| e^{j\phi_k(\omega_l)} \quad i = 1, 2, \quad (1.79)$$

et leurs phases satisfont

$$\varphi_i(\omega_l) - \phi_k(\omega_l) = \theta_{ik} - \omega_l t_{ik} + 2k_i\pi \quad i = 1, 2, \quad (1.80)$$

où k_i est un entier arbitraire. En posant,

$$\mu_k \triangleq t_{2k} - t_{2l}, \quad (1.81)$$

$$\vartheta_k \triangleq \theta_{2k} - \theta_{1k}, \quad (1.82)$$

l'Eq. (1.80) considérée pour $i = 1$ et $i = 2$ conduit alors à

$$(\varphi_2(\omega_l) - \varphi_1(\omega_l)) - \vartheta_k + \omega_l \mu_k = 2\pi(k_2 - k_1). \quad (1.83)$$

R. Li et M. Xiao proposent d'estimer les décalages temporels en minimisant dans un premier temps la fonction coût suivante :

$$J(\mu_k, \vartheta_k) = \arg \min_{\mu_k, \vartheta_k} \left\{ \sum_{\omega_l} |1 - \exp\{-j[(\varphi_2(\omega_l) - \varphi_1(\omega_l)) - \vartheta_k + \omega_l \mu_k]\}\} \right\} \quad (1.84)$$

où $\mu_k \in \{0, \dots, K\}$ (K étant le nombre d'échantillons de la TFD), $\vartheta_k \in \{-\pi, 0, \pi\}$ et où les valeurs de ω_l sont celles qui correspondent aux zones mono-sources de la source k . Une fois μ_k et ϑ_k obtenus, les décalages temporels sont égaux à

$$\begin{cases} t_{1k} = 0, & t_{2k} = \mu_k, & \text{si } 0 \leq \mu_k < K/2 \\ t_{1k} = K - \mu_k, & t_{2k} = 0, & \text{si } \mu_k \geq K/2 \end{cases} \quad (1.85)$$

et les atténuations sont égales à

$$\begin{cases} a_{1k} = |a_{1k}|, & a_{2k} = |a_{2k}|, & \text{si } \vartheta_k = 0 \\ a_{1k} = -|a_{1k}|, & a_{2k} = |a_{2k}|, & \text{si } \vartheta_k = \pi \\ a_{1k} = |a_{1k}|, & a_{2k} = -|a_{2k}|, & \text{si } \vartheta_k = -\pi \end{cases} \quad (1.86)$$

Approches "hybrides"

Nous avons vu précédemment deux méthodes TF parcimonieuses qui présentent chacune des avantages et des inconvénients : DUET et LI-TIFFROM (et ses variantes). Le principal avantage de DUET (et ses extensions) est qu'il peut séparer complètement les sources, même dans le cas de mélanges sous-déterminés, mais la WDO (approximative) est une condition trop contraignante quand les sources sont très recouvrantes dans le plan TF. A l'opposé, l'approche LI-TIFFROM nécessite une condition beaucoup moins restrictive sur le recouvrement des sources, apporte un paramètre mesurant l'isolation d'une source, c.-à-d. la "qualité mono-source" mais impose de connaître le nombre de sources et ne permet qu'une reconstruction partielle des sources dans le cas de mélanges sous-déterminés. L'approche DEMIX [12] de S. Arberet et al. peut-être vue comme un "mélange" des méthodes DUET et LI-TIFFROM décrites précédemment.

Le modèle est un modèle LI éventuellement bruité et sous-déterminé. L'idée est de relier chaque atome du domaine d'analyse à une mesure de confiance sur la "qualité mono-source" de cet atome. Ils proposent de tracer des nuages de points particuliers : sur l'abscisse, on lit les angles des directions des sources alors que sur les ordonnées, on lit la mesure de confiance de ces angles.

Afin de reconstruire avec précision les sources, un algorithme de clustering, nommé DEMIX, est proposé : il estime à la fois le nombre de sources et les directions des colonnes de la matrice de mélange. Cet algorithme sera décrit dans la section 5.2.2 page 121.

Récemment, les auteurs ont étendu cette approche aux mélanges AD [13] en ajoutant à DEMIX une technique inspirée de GCC-PHAT [142], basée sur la corrélation, pour trouver les décalages temporels.

Dans [165], A. Master considère lui aussi le modèle de mélange AD sous-déterminé avec $P = 2$ observations, défini par l'Eq. (1.54) et que nous rappelons ici afin de simplifier la compréhension de la méthode :

$$\begin{bmatrix} X_1(t, \omega) \\ X_2(t, \omega) \end{bmatrix} = \begin{bmatrix} 1 & \dots & 1 \\ a_{21}e^{-j\omega t_1} & \dots & a_{2N}e^{-j\omega t_N} \end{bmatrix} \begin{bmatrix} S_1(t, \omega) \\ \vdots \\ S_N(t, \omega) \end{bmatrix}. \quad (1.87)$$

L'auteur propose une approche inspirée de DUET qu'il appelle DASSS (pour *Delay And Scale Substraction Scoring*) et qui donne une mesure de confiance sur l'hypothèse de WDO. La méthode peut-être résumée ainsi.

1. Dans une première étape, les paramètres de la matrice de mélange sont estimés via un histogramme amplitude/retard comme cela a été proposé pour DUET.
2. Ensuite, l'auteur propose une séparation partielle par annulation de source :

$$Y_i(t, \omega) = X_1(t, \omega) - \frac{1}{a_i} e^{+j\omega t_i} X_2(t, \omega) \quad i = 1, \dots, N. \quad (1.88)$$

Chaque signal Y_i contient donc $N - 1$ sources filtrées

$$Y_i(t, \omega) = \sum_{j \neq i} \gamma_{i,j} S_j(t, \omega), \quad (1.89)$$

où

$$\gamma_{i,j} = 1 - \frac{a_i}{a_j} e^{j\omega(t_i - t_j)}; \quad (1.90)$$

ce qui donne dans le domaine matriciel

$$\begin{bmatrix} Y_1(t, \omega) \\ Y_2(t, \omega) \\ \vdots \\ Y_N(t, \omega) \end{bmatrix} = \begin{bmatrix} 0 & \gamma_{1,2} & \cdots & \gamma_{1,N} \\ \gamma_{2,1} & 0 & \cdots & \gamma_{2,N} \\ \vdots & & \ddots & \vdots \\ \gamma_{N,1} & \gamma_{N,2} & \cdots & 0 \end{bmatrix} \begin{bmatrix} S_1(t, \omega) \\ S_2(t, \omega) \\ \vdots \\ S_N(t, \omega) \end{bmatrix}. \quad (1.91)$$

Le but de cette seconde étape est d'estimer les contributions de chaque source dans Y_i , pour chaque atome TF (t, ω) . Remarquons que grâce à la première étape de la méthode, les coefficients $\gamma_{j,i}$ sont connus. Si la source S_i est la seule présente dans la fenêtre TF (t_p, ω_l) alors

$$Y_i(t_p, \omega_l) = 0 \quad (1.92)$$

$$Y_{j \neq i}(t_p, \omega_l) = \gamma_{j,i} S_i(t_p, \omega_l) \quad (1.93)$$

$$= \gamma_{j,i} X_1(t_p, \omega_l). \quad (1.94)$$

Ce que révèle l'Eq. (1.94) est que si une source est la seule à être active dans une fenêtre TF (t_p, ω_l) , alors il est possible de prédire les N valeurs des contributions de la source active dans $Y_i(t_p, \omega_l)$, en utilisant uniquement les valeurs connues $\gamma_{j,i}$ et l'observation $X_1(t_p, \omega_l)$. Ces prédictions sont notées $\hat{Y}_i^g(t_p, \omega_l)$ et correspondent aux contributions de la source g dans Y_i ($\hat{Y}_i^g(t_p, \omega_l) = 0$). Il est alors possible de percevoir l'importance de ces contributions dans $Y_i(t_p, \omega_l)$ à l'aide d'une fonction score définie par

$$f(g) = \frac{\sum_{i=1}^N |\hat{Y}_i^g(t_p, \omega_l) - Y_i(t_p, \omega_l)|}{\sum_{i=1}^N |Y_i(t_p, \omega_l)|} \quad (1.95)$$

Pour chaque fenêtre TF (t_p, ω_l) , si l'erreur obtenue par la fonction score est "suffisamment petite", alors il est possible d'assigner la fenêtre TF à une source. Si cette erreur est trop grande, l'auteur propose un algorithme de séparation, comme nous allons le voir maintenant.

3. Dans le cas où plusieurs sources sont présentes (c.-à-d. les erreurs de modélisation de $f(g)$ sont trop importantes, quelle que soit g), A. Master propose deux solutions :

- (a) La première idée est de supposer que dans chacune de ces fenêtres TF où au moins deux sources sont présentes, seules les deux sources S_u et S_v présentant les erreurs de $f(g)$ les plus faibles sont présentes. En conséquence, pour ces fenêtres, les paramètres de mélange étant connus, il est possible d'inverser la matrice de mélange correspondante :

$$\begin{bmatrix} S_u(t, \omega) \\ S_v(t, \omega) \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ a_u e^{-j\omega t_u} & a_v e^{-j\omega t_v} \end{bmatrix}^{-1} \begin{bmatrix} X_1(t, \omega) \\ X_2(t, \omega) \end{bmatrix}. \quad (1.96)$$

- (b) Une autre idée, développée "sur le papier" dans [165] mais non testée, est de faire des hypothèses sur les sources. L'auteur suppose que les sources sont :
- soit des harmoniques, c.-à-d. que les composantes spectrales considérées sont des multiples entiers d'une fréquence fondamentale unique,
 - soit du bruit à bande limitée.

Les signaux obtenus grâce à l'équation (1.96) ne sont calculés que si chacun vérifie une des hypothèses ci-dessus. Cela implique donc de connaître le fondamental d'une source ou la bande de bruit correspondante. Pour cela, A. Master propose d'utiliser n'importe laquelle des méthodes de détection de pitch sur les signaux sources estimés dans les fenêtres mono-sources de l'étape 2 ci-dessus. Grâce aux connaissances sur les sources ainsi obtenues, l'auteur propose d'utiliser une fonction score afin de savoir à quelle source correspond chaque fenêtre TF.

4. Enfin, une fois les sources reconstruites dans le domaine TF, une version temporelle est obtenue par transformée TF inverse.

Dans [166], il place dans un contexte bayésien la situation où deux sources sont actives dans une même fenêtre TF. Tout comme les signaux $Y_i(t, \omega)$ permettent d'obtenir les valeurs des sources quand une seule est active d'après l'Eq. (1.93), il est possible de prédire les valeurs de $Y_i(t, \omega)$ quand les sources $S_u(t, \omega)$ et $S_v(t, \omega)$ sont actives :

$$Y_{i=u}(t, \omega) = \gamma_{u,v} S_v(t, \omega) \quad (1.97)$$

$$Y_{i=v}(t, \omega) = \gamma_{v,u} S_u(t, \omega) \quad (1.98)$$

$$Y_{i \neq (u|v)}(t, \omega) = \gamma_{i,u} S_u(t, \omega) + \gamma_{i,v} S_v(t, \omega) \quad (1.99)$$

Si on sait comment $S_u(t, \omega)$ et $S_v(t, \omega)$ sont distribuées, alors on sait comment $Y_{i=u}(t, \omega)$, $Y_{i=v}(t, \omega)$ et $Y_{i \neq (u|v)}(t, \omega)$ sont distribuées. A. Master suppose que les distributions des enveloppes des sources sont gaussiennes. Il obtient alors un critère probabiliste sur la présence de deux sources⁸ dans les données partiellement séparées $Y_i(t, \omega)$ qui lui permet de séparer "au mieux" les sources dans chaque fenêtre TF. Il valide cette approche dans le cas particulier de deux mélanges de trois sources de musique (violon, clarinette et violoncelle) et surpasse DUET sur cet exemple (le *SIR* de sortie moyen est de 12.0 dB contre 4.3 dB pour DUET).

⁸Il peut aussi tester la probabilité qu'une seule source soit présente et ainsi traiter les signaux "comme DUET".

1.4 Conclusion

Nous venons de présenter trois familles de méthodes de SAS : l'ACI (voir section 1.1), la FMN (voir section 1.2) et la SCA (voir section 1.3). Pour chacune de ces familles, nous avons décrit les principaux critères utilisés, aussi bien pour des mélanges LI que convolutifs (mélanges AD et mélanges convolutifs généraux).

En particulier, comme les méthodes proposées dans ce manuscrit sont basées sur la parcimonie des sources, les approches faisant appel à ce critère ont été très détaillées. Nous avons vu que nous pouvions les classer en trois sous-familles, suivant le degré de parcimonie nécessaire pour estimer la matrice de mélange.

1. Une première sous-famille fait l'hypothèse de la WDO (approximative), c.-à-d. que dans chaque atome du domaine d'analyse, au plus une source est présente (à de très rares exceptions près).
2. Au contraire, des méthodes "quasi-non-parcimonieuses" nécessitent seulement que pour chaque source, il existe de petites zones du domaine d'analyse où une seule source est présente. En dehors de ces zones, les sources peuvent être recouvrantes dans le domaine d'analyse. Les méthodes décrites dans ce document appartiennent à cette sous-famille.
3. Enfin, la troisième sous-famille concerne des méthodes "hybrides" entre ces deux extrêmes.

Dans le chapitre suivant, nous allons détailler les méthodes LI-TIFROM et LI-TIFCORR, qui ont servi de point de départ à ces travaux de thèse, ainsi que des extensions que nous avons proposées pour ces méthodes. Une évaluation de leurs performances sera décrite dans le chapitre 3.

Première partie
Méthodes linéaires instantanées

Chapitre 2

Variantes des méthodes LI-TIFROM et LI-TIFCORR

Nous avons vu dans le chapitre 1 plusieurs méthodes de SAS basées sur la parcimonie des sources. En particulier, certaines méthodes, développées par l'équipe avant mon arrivée, c.-à-d. LI-TEMPROM, LI-TIFROM, LI-TEPCORR, LI-TIFCORR, vont être décrites dans ce chapitre. En effet, elles servent de point de départ à mes travaux (leurs extensions aux mélanges plus complexes sont notamment proposées dans la partie II) et j'ai proposé pour ces approches des améliorations ou variantes.

Ce chapitre est organisé de la façon suivante : les méthodes LI-TEMPROM et LI-TIFROM seront décrites dans la section 2.1. En particulier, nous mettrons en évidence les limitations de ces méthodes dans la section 2.1.2 et proposerons une amélioration dans la section 2.1.3. La section 2.2 présentera une extension "plus robuste" au cas général N sources et N observations que celle proposée dans la méthode originale. Enfin, dans la section 2.3, nous présenterons les méthodes LI-TEPCORR et LI-TIFCORR et proposerons une variante dans la section 2.3.2.

2.1 Méthodes LI-TEMPROM et LI-TIFROM

2.1.1 Approches initiales de F. Abrard et Y. Deville [2]

Nous avons vu dans le panorama page 9 les méthodes LI-TEMPROM et LI-TIFROM, proposées par F. Abrard et Y. Deville [1–4]. Dans un but de concision, nous ne détaillons ici que l'approche TF puisque cette version a servi de comparaison avec la méthode LI-TIFCORR. Néanmoins, les concepts utilisés pour la méthode temporelle sont les mêmes que ceux de la méthode TF et l'amélioration décrite dans la section 2.1.3 ci-dessous peut aussi être appliquée à LI-TEMPROM. L'idée de base est de proposer une méthode, basée sur des rapports TF, peu gourmande en temps de calcul et qui évite les inconvénients de la WDO de DUET. F. Abrard et Y. Deville considèrent un modèle de mélange LI. Pour des raisons de simplicité, nous n'allons détailler pour l'instant que le cas deux sources / deux capteurs :

$$\begin{cases} x_1(t) &= a_{11}s_1(t) + a_{12}s_2(t) \\ x_2(t) &= a_{21}s_1(t) + a_{22}s_2(t) \end{cases} \quad (2.1)$$

où les coefficients a_{ij} de la matrice A sont constants et non nuls⁹.

L'outil TF utilisé est la transformée de Fourier à court terme (TFCT) [116] définie par :

$$X_i(t, \omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} x_i(t') h(t' - t) e^{-j\omega t'} dt' \quad (2.2)$$

où $h(t' - t)$ est une fenêtre de pondération, centrée en t . Comme cette transformée est linéaire, d'après l'Eq. (2.1), les transformées TF des observations s'écrivent alors :

$$\begin{cases} X_1(t, \omega) &= a_{11}S_1(t, \omega) + a_{12}S_2(t, \omega) \\ X_2(t, \omega) &= a_{21}S_1(t, \omega) + a_{22}S_2(t, \omega) \end{cases} \quad (2.3)$$

Le fait que tous les coefficients de mélange a_{ij} soient non-nuls implique que si la contribution d'une source est non nulle dans une observation, pour une fenêtre TF (t, ω) , alors elle est non nulle dans toutes les autres observations pour cette fenêtre TF.

Deux variantes sont proposées pour reconstruire les sources. La première consiste à extraire les sources par des combinaisons linéaires¹⁰ des observations de la forme :

$$y(t) = x_2(t) - c x_1(t). \quad (2.4)$$

En combinant les Eq. (2.1) et (2.4), ils obtiennent les valeurs c_1 et c_2 du coefficient d'annulation qui permettent resp. d'annuler les sources $s_1(t)$ et $s_2(t)$ et donc d'extraire resp. les sources $s_2(t)$ et $s_1(t)$:

$$c_1 = \frac{a_{21}}{a_{11}} \quad c_2 = \frac{a_{22}}{a_{12}}. \quad (2.5)$$

Une autre façon de reconstruire les sources consiste à estimer la matrice inverse de A , à une matrice diagonale et une matrice de permutation près. Pour cela, Y. Deville a proposé dans [74] de réécrire les mélanges selon les contributions des sources dans la première observation, à une permutation des sources près. Son approche permet ainsi de fixer les indéterminations d'échelle propres à la SAS. Nous pouvons la résumer ainsi. Soit σ une permutation donnée. Le problème de SAS posé par l'Eq. (2.1) est équivalent à :

$$\begin{cases} x_1(t) &= a_{1,\sigma(1)}s_{\sigma(1)}(t) + a_{1,\sigma(2)}s_{\sigma(2)}(t) \\ x_2(t) &= a_{2,\sigma(1)}s_{\sigma(1)}(t) + a_{2,\sigma(2)}s_{\sigma(2)}(t) \end{cases} \quad (2.6)$$

En posant :

$$s'_j(t) = a_{1,\sigma(j)}s_{\sigma(j)}(t), \quad j = 1, 2, \quad (2.7)$$

$$b_{ij} = \frac{a_{i,\sigma(j)}}{a_{1,\sigma(j)}}, \quad i, j = 1, 2, \quad (2.8)$$

⁹En réalité, les méthodes décrites à cette étape nécessitent que seuls les coefficients a_{1j} soient non nuls, parce que les versions temporelle et TF calculent des rapports dont les dénominateurs sont des éléments de la première ligne de A . Cependant, lors des améliorations décrites dans les sections 2.1.3 et 2.2, nous avons besoin que tous les éléments de A soient non nuls. Dans un but de simplicité, nous préférons généraliser la non-nullité des coefficients de A dès maintenant.

¹⁰Dans la version originale, les auteurs proposent de "soustraire" x_2 à x_1 . Ici, pour des raisons d'unité avec la suite du manuscrit, nous faisons l'inverse.

- 1) Etape de détection :** rechercher les zones mono-sources.
- Mesurer la qualité "mono-source" des zones d'analyse.
 - Classer les zones d'analyse selon la qualité mono-source décroissante.
- 2) Etape d'identification :** identifier la matrice de mélange.
- A chaque zone d'analyse mono-source correspond une colonne de B .
 - Ne garder une nouvelle colonne que si elle est "suffisamment éloignée" de celles trouvées précédemment.
- 3) Etape de reconstruction :** séparer des sources.
- Une fois la matrice B entièrement estimée, on reconstruit les sources selon l'Eq. (2.12).

FIG. 2.1 – Etapes des méthodes parcimonieuses LI de Y. Deville et al.

l'expression (2.6) s'écrit

$$\begin{cases} x_1(t) &= s'_1(t) + s'_2(t) \\ x_2(t) &= b_{21}s'_1(t) + b_{22}s'_2(t) \end{cases} \quad (2.9)$$

Notons que cette expression est le cas particulier où les décalages temporels sont nuls dans l'Eq. (1.53). La forme matricielle du système (2.9) s'écrit

$$\underline{x}(t) = B \underline{s}'(t) \quad (2.10)$$

où $\underline{x}(t)$ et $\underline{s}'(t)$ correspondent resp. aux vecteurs des observations et des sources permutées et mises à l'échelle et où

$$B = \begin{bmatrix} 1 & 1 \\ b_{21} & b_{22} \end{bmatrix}. \quad (2.11)$$

Remarquons que dans les deux reconstructions possibles, les auteurs estiment en fait exactement la même chose : les coefficients c_i étant égaux aux coefficients b_{2j} à une permutation près.

Les signaux de sortie sont obtenus par inversion de la matrice B

$$\underline{y}(t) = B^{-1} \underline{x}(t), \quad (2.12)$$

et ils correspondent aux contributions des sources dans la première observation (aux erreurs d'estimation près).

Nous allons maintenant voir quelles sont les définitions et hypothèses que posent F. Abrard et Y. Deville.

Définition 7 Une zone d'analyse TF est un jeu de fenêtres TF (t, ω) adjacentes.

Les zones d'analyse peuvent avoir n'importe quelle forme. Dans le cadre de mélanges LI, seules les zones à fréquence constante, c.-à-d. composée d'un jeu de M fenêtres (t, ω) temporellement adjacentes, sont considérées. Ces zones d'analyse sont notées (T, ω) .

Définition 8 Une source est dite "isolée" dans une zone d'analyse TF si seule cette source a une transformée TF qui n'est pas nulle partout dans cette zone. Cette zone est alors dite "zone mono-source".

Définition 9 Une source est dite "visible" dans le domaine TF s'il existe au moins une zone d'analyse TF où elle est isolée.

Hypothèse 3 Chaque source est visible dans le plan TF.

Cette hypothèse est beaucoup moins contraignante que la WDO¹¹ puisque, pour chaque source à séparer, la méthode n'a besoin que d'une petite zone TF où celle-ci est isolée.

Hypothèse 4 Il n'existe pas de zone TF où les transformées TF soient nulle partout.

Cette hypothèse est uniquement faite dans un but de simplicité : elle peut être supprimée en pratique, grâce au bruit contenu dans les enregistrements réels, comme expliqué dans [2] ou dans l'annexe A.

Hypothèse 5 Quand plusieurs sources sont actives dans un jeu donné de fenêtres TF adjacentes, elles doivent varier de telle façon que le rapport¹² $\frac{X_2(t, \omega)}{X_1(t, \omega)}$ des TFCT des observations ne soit pas constant dans cette zone TF. En particulier, i) au moins une source doit avoir des valeurs TF significativement différentes dans ces fenêtres de telle façon que la variance du rapport $\frac{X_2(t, \omega)}{X_1(t, \omega)}$ soit non-négligeable et ii) les sources ne doivent pas varier de façon cohérente, c.-à-d. proportionnellement, entre ces fenêtres.

La méthode LI-TIFROM est composée de trois étapes principales (voir la figure 2.1), précédées d'une pré-étape de calcul :

1. le calcul des TFCT,
2. l'étape de détection pendant laquelle on cherche les zones TF mono-sources,
3. l'étape d'identification où l'on calcule les coefficients c_i ou b_{2j} dans les zones mono-sources,
4. l'étape de reconstruction des sources.

L'étape de détection est basée sur le rapport suivant :

$$\alpha(t, \omega) = \frac{X_2(t, \omega)}{X_1(t, \omega)} \quad (2.13)$$

¹¹Mais l'hypothèse de WDO rend naturelle la séparation complète de mélanges sous-déterminés. La version initiale de LI-TIFROM décrite ici ne traite pas ce genre de mélanges. De plus, même dans son extension aux mélanges sous-déterminés, elle ne peut réaliser qu'une séparation partielle des sources.

¹²En réalité, [2] considère le rapport inverse $\frac{X_1(t, \omega)}{X_2(t, \omega)}$ mais pour des raisons de cohérence avec la suite de ce document et sans que cela ne change quoi que ce soit dans la compréhension de la méthode, nous allons étudier le rapport $\frac{X_2(t, \omega)}{X_1(t, \omega)}$.

Compte tenu de l'expression (2.3), l'Eq. (2.13) s'écrit :

$$\alpha(t, \omega) = \frac{a_{21}S_1(t, \omega) + a_{22}S_2(t, \omega)}{a_{11}S_1(t, \omega) + a_{12}S_2(t, \omega)} \quad (2.14)$$

Si une source est isolée dans une zone TF (T, ω_l) , ce rapport sera constant dans toute cette zone d'analyse. En effet, prenons le cas particulier où $S_1(t_p, \omega_l) = 0$ sur (T, ω_l) , c.-à-d. le cas où la source $S_2(t_p, \omega_l)$ est isolée sur cette zone. Alors

$$\alpha(t_p, \omega_l) = \frac{a_{22}}{a_{12}} \quad \forall t_p \in T, \quad (2.15)$$

et sa moyenne, définie par

$$\bar{\alpha}(T, \omega_l) = \frac{1}{M} \sum_{p=1}^M \alpha(t_p, \omega_l), \quad (2.16)$$

où M est le nombre de fenêtres TF par zone d'analyse, vaut $\frac{a_{22}}{a_{12}}$. Sa variance, définie par

$$var [\alpha] (T, \omega_l) = \frac{1}{M} \sum_{p=1}^M |\alpha(t_p, \omega_l) - \bar{\alpha}(T, \omega_l)|^2, \quad (2.17)$$

est donc nulle. Au contraire, si plusieurs sources sont actives sur une zone d'analyse et si elles vérifient l'hypothèse 5, alors la variance $var [\alpha] (T, \omega_l)$ est clairement non nulle. Les "meilleures" zones TF mono-sources sont donc considérées comme étant celles où la variance de $\alpha(t_p, \omega_l)$ prend les plus petites valeurs. Ainsi, l'étape de détection consiste à classer les zones d'analyse (T, ω_l) selon les valeurs croissantes de $var [\alpha] (T, \omega_l)$, les premières zones de cette liste ordonnée étant les "meilleures" zones mono-sources.

L'étape d'identification consiste à étudier successivement chaque zone d'analyse de la liste ordonnée définie dans l'étape précédente. Les coefficients c_i ou b_{2j} sont estimés comme la moyenne $\bar{\alpha}(T, \omega_l)$, définie par l'Eq. (2.16). Le second coefficient ainsi calculé est gardé si la distance entre celui-ci et celui trouvé précédemment est supérieure à un seuil défini préalablement. Cette étape s'arrête lorsque le nombre de coefficients trouvés est égal au nombre de sources supposé connu, c.-à-d. 2 dans cette section.

Extension au cas N sources N capteurs

Supposons maintenant que nous avons N mélanges de N sources ($N \geq 2$) :

$$x_i(t) = \sum_{j=1}^N a_{ij} s_j(t) \quad i = 1, \dots, N. \quad (2.18)$$

En reprenant le formalisme décrit dans les Eq. (2.7) et (2.8), nous obtenons :

$$\begin{cases} x_1(t) = \sum_{j=1}^N s'_j(t) \\ x_i(t) = \sum_{j=1}^N b_{ij} s'_j(t) \end{cases} \quad i = 2, \dots, N \quad (2.19)$$

et la matrice à estimer devient

$$B = \begin{bmatrix} 1 & \cdots & 1 \\ b_{21} & \cdots & b_{2N} \\ \vdots & & \vdots \\ b_{N1} & \cdots & b_{NN} \end{bmatrix}. \quad (2.20)$$

Les auteurs proposent de n'utiliser que deux observations pour détecter les zones mono-sources, c.-à-d. qu'ils considèrent deux observations $X_{k_1}(t, \omega)$ et $X_{k_2}(t, \omega)$ ($k_1 \neq k_2$) et étudient le rapport¹³

$$\alpha(t, \omega) = \frac{X_{k_1}(t, \omega)}{X_{k_2}(t, \omega)}. \quad (2.21)$$

Une liste ordonnée des zones d'analyse selon les valeurs croissantes de $\text{var}[\alpha](T, \omega_l)$ est ensuite créée. La $m^{\text{ième}}$ zone de cette liste permet d'estimer les coefficients b_{ij} de la colonne j de B comme les moyennes des rapports $\frac{X_i(t_p, \omega_l)}{X_1(t_p, \omega_l)}$ sur cette zone d'analyse. La procédure s'arrête quand les N colonnes de B sont estimées.

Mélange sous-déterminé

Nous avons vu comment les auteurs proposaient de séparer les sources dans le cas où le nombre N de sources est égal au nombre P d'observations. Que se passe-t-il quand N et P ne sont pas égaux? Le cas sur-déterminé n'apporte pas de difficulté supplémentaire puisqu'on peut par exemple choisir N des P observations et traiter un problème déterminé composé de N sources et N observations. Considérons maintenant le cas sous-déterminé, c.-à-d. où le nombre N de sources est strictement plus grand que le nombre P d'observations. Grâce aux Eq. (2.7) et (2.8), nous obtenons :

$$x_i(t) = \sum_{j=1}^N b_{ij} s'_j(t) \quad i = 1, \dots, P. \quad (2.22)$$

La matrice B à estimer est la matrice non-carrée

$$B = \begin{bmatrix} 1 & \cdots & 1 \\ b_{21} & \cdots & b_{2N} \\ \vdots & & \vdots \\ b_{P1} & \cdots & b_{PN} \end{bmatrix}. \quad (2.23)$$

L'idée proposée est de faire une séparation partielle, comme discuté dans la section 2 de l'introduction générale. Le but est donc de séparer entre elles P sources d'intérêt mélangées à $N - P$ sources "de bruit". Les étapes de détection et d'identification sont identiques à celles proposées dans l'extension N sources N capteurs précédente. Une fois la matrice B estimée, F. Abrard et Y. Deville proposent de choisir une sous-matrice carré $P \times P$ de B , de l'inverser et reconstruire ainsi, selon l'Eq. (2.12), P signaux de sortie qui correspondent aux P signaux d'intérêt à séparer, mélangés avec les $N - P$ "signaux résiduels".

¹³En particulier, nous pouvons choisir $k_2 = 1$.

Comme dans le cas déterminé, une autre reconstruction a aussi été proposée : en appliquant les combinaisons linéaires de sources (2.4) au problème (2.22), ils annulent les sources dans les observations et obtiennent une application "karaoké" de leur méthode.

Sources cachées

Une deuxième extension de la méthode LI-TIFROM concerne le cas où au moins une source n'est pas visible. La version basique de ce type d'extension peut être définie de la façon suivante, avec $P = N$ dans un but de simplicité. Supposons qu'au moins une des sources, disons $s_k(t)$, est visible. La méthode LI-TIFROM permet d'estimer la colonne correspondante de la matrice de mélange atténuée et permutée B , disons la colonne j . Cette colonne estimée est composée d'éléments $\widehat{b_{ij}}$ qui sont des estimées de

$$b_{ij} = \frac{a_{ik}}{a_{1k}} \quad i = 1, \dots, N, \quad (2.24)$$

comme vu dans l'Eq. (2.8). Ainsi, si nous calculons une version modifiée des signaux mélangés $x_i(t)$, avec $i = 2, \dots, N$, selon

$$x'_i(t) = x_i(t) - \widehat{b_{ij}}x_1(t), \quad i = 2, \dots, N, \quad (2.25)$$

les Eq. (2.18), (2.24) et (2.25) montrent que nous obtenons $N-1$ signaux qui ne contiennent aucune contribution de la source $s_k(t)$ (aux erreurs d'estimation de b_{ij} près). Le point clé est ensuite que, même si certaines sources n'étaient pas visibles dans l'ensemble initial des N sources mélangées, il se peut qu'au moins une de celles-ci devienne visible dans le nouvel ensemble des $N-1$ sources présentes dans les signaux de mélange modifiés $x'_i(t)$. Cela dépend des distributions TF de toutes les sources et apparaît si certaines sources étaient initialement cachées, c.-à-d. elles n'étaient isolées dans aucune zone d'analyse TF quand on considérait le jeu initial de N sources, mais elles sont isolées dans au moins une zone quand on considère le jeu de $N-1$ sources qui restent après annulation des contributions de la source $s_k(t)$ dans tous les signaux mélangés. Si au moins une source est visible dans ce nouvel ensemble de $N-1$ signaux de mélanges, la même procédure peut être à nouveau appliquée. Cette procédure récursive finit quand (plus aucune source n'est visible, ou quand) le nombre de signaux recombinaisonnés est égal à un, et ce signal contient alors une unique source. Cette procédure réussit donc à extraire cette source, même si toutes les sources n'étaient pas initialement visibles. Cette procédure peut ensuite être appliquée à nouveau, en sélectionnant les autres sources $s_k(t)$ à chaque étape de la récursion dans le but d'extraire d'autres sources.

Des versions plus avancées de ce type de procédure peuvent aussi être définies dans le but d'annuler, à chaque étape intermédiaire de la récursion, les contributions de toutes les sources qui sont visibles à cette étape. Cela a été illustré dans [2].

2.1.2 Les limitations de LI-TIFROM

Nous avons présenté dans la section 2.1.1 la méthode LI-TIFROM telle qu'elle a été proposée par F. Abrard et Y. Deville. Néanmoins, cette méthode souffre du côté non-symétrique du paramètre de variance (2.17) sur lequel elle est fondée [82, 220]. En effet,

prenons le cas simple deux sources deux observations et une matrice symétrique de mélange :

$$A = \begin{bmatrix} 1 & \lambda \\ \lambda & 1 \end{bmatrix}, \quad (2.26)$$

où λ est un paramètre de mélange, dont l'influence sera analysée ci-dessous. Tel que nous l'avons défini dans l'Eq. (2.13), le paramètre de détection de zones TF mono-sources s'écrit :

$$\alpha(t, \omega) = \frac{\lambda S_1(t, \omega) + S_2(t, \omega)}{S_1(t, \omega) + \lambda S_2(t, \omega)}. \quad (2.27)$$

Considérons tout d'abord une zone d'analyse *idéale*, c.-à-d. une zone TF où une des sources $S_i(t, \omega)$ ($i = 1, 2$) est strictement nulle partout. L'Eq. (2.27) montre alors que $\alpha(t, \omega)$ est constant dans cette zone. Sa variance sur cette zone est donc strictement nulle et c'est précisément cette propriété qui est utilisée dans LI-TIFROM pour détecter les zones d'analyse mono-sources. Pour des zones d'analyse *idéales*, la méthode LI-TIFROM a alors le même comportement pour les deux sources, $S_1(t, \omega)$ et $S_2(t, \omega)$.

Mais considérons maintenant une zone d'analyse mono-source *réelle* associée à la source $S_i(t, \omega)$. Cette zone *réelle* comprend une source $S_i(t, \omega)$ prédominante mais aussi des résidus de l'autre source, $S_{3-i}(t, \omega)$, très légèrement présente. Nous avons donc la relation $|S_{3-i}(t, \omega)| \ll |S_i(t, \omega)|$ sur cette zone. Nous allons voir que le comportement de LI-TIFROM, à cause du paramètre $\alpha(t, \omega)$, est différent suivant que la source isolée est $S_1(t, \omega)$ ou $S_2(t, \omega)$. Plus précisément, considérons le cas où $\lambda \ll 1$ et analysons les contributions de chaque source dans $\alpha(t, \omega)$:

- d'après l'Eq. (2.27), du point de vue de $S_1(t, \omega)$, $\alpha(t, \omega)$ correspond au rapport entre une petite et une grande composante.
- Au contraire, du point de vue de $S_2(t, \omega)$, $\alpha(t, \omega)$ correspond au rapport entre une grande et une petite composante.

Le paramètre $\alpha(t, \omega)$ n'est alors pas symétrique par rapport aux deux sources. Étudions de façon formelle ce phénomène et ses conséquences que nous venons de voir par cette analyse qualitative ci-dessus.

Considérons d'abord une zone mono-source réelle associée à $S_1(t, \omega)$, c.-à-d. une zone où $|S_2(t, \omega)| \ll |S_1(t, \omega)|$. Du point de vue de ces signaux sources, la "qualité mono-source" de cette zone d'analyse peut être mesurée à chacun de ces points TF (t, ω) par le paramètre

$$\epsilon_1 = \frac{S_2(t, \omega)}{S_1(t, \omega)}, \quad (2.28)$$

avec $\epsilon_1 \ll 1$. A partir de l'Eq. (2.27), nous pouvons obtenir le développement limité au premier ordre de $\alpha(t, \omega)$, par rapport à ϵ_1 pour n'importe quelle valeur de λ :

$$\alpha(t, \omega) \simeq \lambda \left(1 + \frac{1 - \lambda^2}{\lambda} \epsilon_1 \right). \quad (2.29)$$

Le terme du premier ordre dans l'Eq. (2.29) est responsable des variations de $\alpha(t, \omega)$ dans la zone d'analyse considérée. A cause de l'Eq. (2.29), la variance de $\alpha(t, \omega)$ dans la zone

d'analyse (T, ω) s'écrit alors, à l'approximation du premier ordre près :

$$\text{var} [\alpha] (T, \omega)_{S_1} = \lambda^2 \left(\frac{1 - \lambda^2}{\lambda} \right)^2 \text{var} [\epsilon_1] (T, \omega)_{S_1}, \quad (2.30)$$

où l'indice S_1 signifie que ces expressions s'appliquent dans une zone d'analyse mono-source associée à $S_1(t, \omega)$ et où la variance $\text{var} [\epsilon_1] (T, \omega)_{S_1}$ est définie selon l'Eq. (2.17). Cette variance mesure alors la "qualité mono-source" intrinsèque moyenne de cette zone d'analyse.

Une recherche symétrique peut être effectuée pour une zone d'analyse mono-source réelle associée à $S_2(t, \omega)$, c.-à-d. une zone où $|S_1(t, \omega)| \ll |S_2(t, \omega)|$. Du point de vue de ces signaux sources, la qualité de cette zone d'analyse peut d'abord être mesurée à chaque point TF (t, ω) par le paramètre défini ci-dessous :

$$\epsilon_2 = \frac{S_1(t, \omega)}{S_2(t, \omega)}, \quad (2.31)$$

avec $\epsilon_2 \ll 1$. Les mêmes opérations que celles faites précédemment nous permettent d'obtenir à l'approximation du premier ordre près,

$$\text{var} [\alpha] (T, \omega)_{S_2} = \frac{1}{\lambda^2} \left(\frac{1 - \lambda^2}{\lambda} \right)^2 \text{var} [\epsilon_2] (T, \omega)_{S_2}. \quad (2.32)$$

Ces résultats peuvent être interprétés ainsi. Supposons que nous avons deux zones d'analyse mono-source réelles, chacune associée respectivement à une source, dont la "qualité mono-source" est exactement la même, c.-à-d.

$$\text{var} [\epsilon_1] (T, \omega)_{S_1} = \text{var} [\epsilon_2] (T, \omega)_{S_2}. \quad (2.33)$$

Alors, malgré le côté symétrique de cette configuration de mélange, les Eq. (2.30) et (2.32) montrent que les paramètres utilisés dans LI-TIFROM apportent des comportements différents pour cette méthode dans ces deux zones d'analyse. Plus précisément, pour $\lambda \neq 1$, le paramètre de détection $\alpha(t, \omega)$ n'a pas la même variance dans ces deux zones d'analyse. De plus,

$$\frac{\text{var} [\alpha] (T, \omega)_{S_2}}{\text{var} [\alpha] (T, \omega)_{S_1}} = \frac{1}{\lambda^4}, \quad (2.34)$$

et donc cette différence croit nettement quand λ devient par exemple beaucoup plus petit que 1. Dans ce cas, $\alpha(t, \omega)$ a une plus grande variance dans les zones d'analyse associées à $S_2(t, \omega)$ que dans celles associées à $S_1(t, \omega)$. Ceci a une influence majeure sur le contenu de la liste ordonnée de zones d'analyse créée dans l'étape de détection de LI-TIFROM et son utilisation faite dans l'étape d'identification : dans cette dernière étape, LI-TIFROM prend d'abord en compte les zones d'analyse qui sont situées au début de cette zone d'analyse et qui correspondent aux faibles variances de $\alpha(t, \omega)$.

Il est donc très aisé d'estimer précisément plusieurs valeurs de la colonne correspondant à la source $S_1(t, \omega)$. En mesurant successivement les distances entre ces colonnes expérimentales, nous ne gardons qu'une seule estimation de cette colonne associée à $S_1(t, \omega)$.

LI-TIFROM parcourt ensuite progressivement la liste des zones d'analyse classées selon l'ordre croissant de la variance $var [\alpha] (T, \omega)$. Cette liste contient des zones de bonne qualité associées à $S_2(t, \omega)$, c.-à-d. des zones où $S_2(t, \omega)$ est prépondérante et où la pollution de $S_1(t, \omega)$ est faible. Ainsi, LI-TIFROM serait capable d'identifier de façon précise la colonne associée à $S_2(t, \omega)$. Toutefois, comme nous l'avons vu ci-dessus, si $\lambda \ll 1$, $\alpha(t, \omega)$ a une très grande variance dans ces zones, qui sont situées assez loin dans la liste ordonnée. Ainsi, dès que LI-TIFROM monte dans la liste, il y a un risque qu'elle donne une mauvaise détection et identification parce qu'elle aura d'abord rencontré une zone (i) qui a une plus faible variance que les zones de bonne qualité associées à $S_2(t, \omega)$ et (ii) qui donne une colonne de la matrice de mélange suffisamment éloignée de celle obtenue précédemment pour $S_1(t, \omega)$, de telle façon que cette nouvelle colonne est gardée. Quand cette situation a lieu, LI-TIFROM apporte une faible qualité de séparation, comme nous le verrons dans l'étude expérimentale proposée dans le chapitre 3. Cette étude ci-dessus est aussi valable pour la méthode LI-TEMPROM, puisque celle-ci est basée sur les mêmes principes que LI-TIFROM.

2.1.3 Amélioration des méthodes LI-TEMPROM et LI-TIFROM

Comme nous venons de le voir, les approches LI-TEMPROM et LI-TIFROM souffrent du côté non-symétrique de la variance qui peut donner de mauvais résultats en fonction de la matrice de mélange. Nous proposons de remédier à ce souci de façon très simple. Comme précédemment, nous ne détaillons que l'approche TF mais cette amélioration peut être apportée à LI-TEMPROM aussi.

En plus d'utiliser le rapport $\alpha(t, \omega)$ défini par l'Eq. (2.13), nous définissons le rapport inverse :

$$\beta(t, \omega) = \frac{X_1(t, \omega)}{X_2(t, \omega)}. \quad (2.35)$$

Nous définissons alors la moyenne et la variance de $\beta(t, \omega)$ sur une zone d'analyse (T, ω_l) :

$$\bar{\beta}(T, \omega_l) = \frac{1}{M} \sum_{p=1}^M \beta(t_p, \omega_l), \quad (2.36)$$

$$var [\beta] (T, \omega_l) = \frac{1}{M} \sum_{p=1}^M |\beta(t_p, \omega_l) - \bar{\beta}(T, \omega_l)|^2. \quad (2.37)$$

Nous proposons de rechercher les zones mono-sources à l'aide du paramètre :

$$min \{var [\alpha] (T, \omega), var [\beta] (T, \omega)\}. \quad (2.38)$$

Contrairement au seul paramètre $var [\alpha] (T, \omega)$ utilisé dans la méthode initiale, ce paramètre (2.38) ne présente plus de caractère asymétrique. En effet, considérons à nouveau la matrice A définie dans l'Eq. (2.26). Soit une zone mono-source "réelle" associée à $S_1(t, \omega)$, c.-à-d. une zone où $|S_2(t, \omega)| \ll |S_1(t, \omega)|$. Considérons le rapport ϵ_1 défini dans l'Eq. (2.28), avec $\epsilon_1 \ll 1$. A partir des définitions (2.13) et (2.35), nous pouvons obtenir

le développement limité au premier ordre de $\alpha(t, \omega)$ et $\beta(t, \omega)$, par rapport à ϵ_1 pour n'importe quelle valeur de λ :

$$\alpha(t, \omega) \simeq \lambda \left(1 + \frac{1 - \lambda^2}{\lambda} \epsilon_1 \right), \quad (2.39)$$

$$\beta(t, \omega) \simeq \frac{1}{\lambda} \left(1 + \frac{\lambda^2 - 1}{\lambda} \epsilon_1 \right). \quad (2.40)$$

Les variances respectives s'écrivent alors, à l'approximation du premier ordre près :

$$\text{var} [\alpha] (T, \omega)_{S_1} = \lambda^2 \left(\frac{1 - \lambda^2}{\lambda} \right)^2 \text{var} [\epsilon_1] (T, \omega)_{S_1}, \quad (2.41)$$

$$\text{var} [\beta] (T, \omega)_{S_1} = \frac{1}{\lambda^2} \left(\frac{\lambda^2 - 1}{\lambda} \right)^2 \text{var} [\epsilon_1] (T, \omega)_{S_1}. \quad (2.42)$$

Le critère amélioré (2.38) s'écrit alors :

$$\min \{ \text{var} [\alpha] (T, \omega)_{S_1}, \text{var} [\beta] (T, \omega)_{S_1} \} = \min \left\{ \lambda^2, \frac{1}{\lambda^2} \right\} \left(\frac{1 - \lambda^2}{\lambda} \right)^2 \text{var} [\epsilon_1] (T, \omega)_{S_1}. \quad (2.43)$$

De même, considérons une zone mono-source réelle associée à la source $S_2(t, \omega)$, c.-à-d. une zone où $S_1(t, \omega) \ll S_2(t, \omega)$, ainsi que le rapport ϵ_2 défini dans l'Eq. (2.31), avec $\epsilon_2 \ll 1$. Les développements limités au premier ordre de $\alpha(t, \omega)$ et $\beta(t, \omega)$, par rapport à ϵ_2 s'écrivent :

$$\alpha(t, \omega) \simeq \frac{1}{\lambda} \left(1 + \frac{\lambda^2 - 1}{\lambda} \epsilon_2 \right), \quad (2.44)$$

$$\beta(t, \omega) \simeq \lambda \left(1 + \frac{1 - \lambda^2}{\lambda} \epsilon_2 \right). \quad (2.45)$$

Le critère amélioré (2.38) s'écrit alors :

$$\min \{ \text{var} [\alpha] (T, \omega)_{S_2}, \text{var} [\beta] (T, \omega)_{S_2} \} = \min \left\{ \lambda^2, \frac{1}{\lambda^2} \right\} \left(\frac{1 - \lambda^2}{\lambda} \right)^2 \text{var} [\epsilon_2] (T, \omega)_{S_2}. \quad (2.46)$$

Si, comme précédemment, on suppose que les deux zones mono-source réelles ont la même qualité mono-source, ce que nous avons traduit dans l'équation (2.33), alors nous voyons que, contrairement au critère original, notre critère mixte donne la même importance à deux zones mono-sources relatives resp. à $S_1(t, \omega)$ et $S_2(t, \omega)$ et de même qualité. Plus précisément, même si $\lambda \neq 1$,

$$\frac{\min \{ \text{var} [\alpha] (T, \omega)_{S_1}, \text{var} [\beta] (T, \omega)_{S_1} \}}{\min \{ \text{var} [\alpha] (T, \omega)_{S_2}, \text{var} [\beta] (T, \omega)_{S_2} \}} = 1. \quad (2.47)$$

En raison du caractère symétrique du critère (2.38), nous appelons LI-TIFROM-S1 cette première version améliorée de TIFROM.

Une deuxième modification concerne l'étape d'identification, d'où le nom LI-TIFROM-S2 de cette variante. En effet, dans le cas de zones mono-sources réelles, les coefficients estimés b_{2j} sont plus ou moins dispersés autour des valeurs théoriques, et les valeurs des coefficients de mélange ont une influence sur cette dispersion. Considérons à nouveau le mélange (2.26) où $\lambda \ll 1$. Si la source $S_2(t, \omega)$ est isolée dans une zone mono-source "réelle", alors d'après les Eq. (2.44) et (2.45), le critère (2.38) est minimisé par le rapport $\beta(t, \omega)$. Dans cette zone mono-source (T, ω) , on a alors la relation suivante :

$$\bar{\beta}(T, \omega) = \lambda + \xi, \quad (2.48)$$

où ξ est l'erreur d'estimation du coefficient de mélange dans la zone mono-source considérée. Considérons notamment deux zones mono-sources réelles, resp. notées (T_1, ω_1) et (T_2, ω_2) , et toutes deux associées à $S_2(t, \omega)$. Supposons que ces deux zones permettent d'estimer λ avec les erreurs suivantes :

$$\bar{\beta}(T_1, \omega_1) = \lambda + \xi, \quad \xi > 0, \quad (2.49)$$

$$\bar{\beta}(T_2, \omega_2) = \lambda - \xi. \quad (2.50)$$

Ces deux zones conduisent à la même erreur relative d'estimation et la distance entre les deux paramètres d'identification associés à ces zones vaut

$$|\bar{\beta}(T_1, \omega_1) - \bar{\beta}(T_2, \omega_2)| = 2\xi. \quad (2.51)$$

Cependant, lorsqu'on va estimer les coefficients d'échelle b_{2j} correspondants dans ces zones, comme nous savons que le rapport $\beta(t, \omega)$ va minimiser le critère (2.38), nous savons alors que les coefficients $\widehat{b}_{2j(T_i, \omega_i)}$ resp. estimés les zones d'analyse (T_i, ω_i) ($i = 1, 2$) vont être égaux à l'inverse de $\bar{\beta}(T, \omega)$ dans ces deux zones :

$$\widehat{b}_{2j(T_i, \omega_i)} = \bar{\beta}(T_i, \omega_i)^{-1}, \quad i = 1, 2. \quad (2.52)$$

L'erreur relative entre les deux éléments définis dans l'Eq. (2.52) est alors égale à :

$$|\bar{\beta}(T_1, \omega_1)^{-1} - \bar{\beta}(T_2, \omega_2)^{-1}| = \left| \frac{1}{\lambda + \xi} - \frac{1}{\lambda - \xi} \right| \quad (2.53)$$

$$= \left| \frac{2\xi}{\lambda^2 - \xi^2} \right|. \quad (2.54)$$

Si λ et ξ sont très petits, alors $|\lambda^2 - \xi^2| \ll 1$ et la distance définie dans l'Eq. (2.54) sera beaucoup plus grande que celle définie dans l'Eq. (2.51). Si le seuil fixé pour choisir une nouvelle colonne de B est trop petit, alors les variations de l'erreur d'estimation sur diverses zones d'analyse peuvent nous conduire à garder par erreur deux colonnes correspondant à la même source. Pour éviter cette situation, nous proposons la démarche suivante. Si le minimum du critère (2.38) est obtenu à partir de $\alpha(t, \omega)$, alors le nouveau coefficient b_{2j} est fixé à $\bar{\alpha}(T, \omega)$ et nous le comparons directement à ceux trouvés précédemment. Au contraire, si ce minimum est obtenu à partir du rapport inverse $\beta(t, \omega)$, alors le nouveau coefficient b_{2j} est fixé à $\bar{\beta}(T, \omega)^{-1}$ et nous comparons l'inverse de ce coefficient

à l'inverse de ceux trouvés précédemment. Une version automatisée pour trouver un seuil entre les colonnes de B serait préférable, comme on en discutera dans le chapitre 5. Notons toutefois que si λ est proche de 1, alors les variantes -S1 et -S2 de LI-TIFROM doivent fournir des performances équivalentes. C'est ce que nous avons vérifié expérimentalement lors de la série de tests présentée dans le chapitre suivant.

Grâce à ces améliorations, dans le cas où le rapport $\frac{\text{var}[\alpha](T,\omega)}{\text{var}[\beta](T,\omega)}$ est proche de 1 (dans l'exemple précédent, cela correspond à λ proche de 1), nous nous attendons à avoir des performances proches de celles de la méthode originale. Dans le cas contraire (dans l'exemple précédent, cela correspond par exemple à λ proche de 0), nous nous attendons à ce que, contrairement à la version originale, notre version améliorée ne voit pas de chute de ses performances.

Nous validerons expérimentalement ces données dans la section 3.1.4 page 74.

2.2 Cas N sources et N observations : utilisation de la moyenne

Nous avons présenté dans la section 2.1.1 la version N sources N observations de LI-TIFROM. Les auteurs proposent de n'utiliser qu'un seul couple d'observations d'indices i et j pour détecter les zones d'analyse mono-sources. En effet, comme les coefficients de mélange ne sont pas nuls, les zones TF où une source est isolée sont les même pour toutes les observations. Ce phénomène est appelé dans [1, 2] la *cohérence des cartes TF*. Ainsi, si les zones mono-sources sélectionnées sont des zones "parfaites", alors la variance du rapport $\alpha(t, \omega) = X_i(t, \omega)/X_j(t, \omega)$ est nulle. De plus, cette vision est attrayante parce que le coût calculatoire est réduit par rapport à une méthode qui utiliserait un plus grand nombre de couples dans l'étape de détection. Cependant, en pratique, les signaux réels ne nous apportent que des zones mono-sources "réelles", où les résidus des autres sources créent des déviations sur la variance de $\alpha(t, \omega)$ qui n'est donc pas nulle. L'amplitude de ces écarts peut dépendre des coefficients de mélange considérés et aussi du couple d'observations considérées. Utiliser un couple unique, arbitraire peut alors apporter de plus faibles performances qu'une approche basée sur l'étude de zones mono-sources via plusieurs couples d'observations.

Nous proposons donc de calculer la moyenne de ces variances. Pour chaque fenêtre TF (t_p, ω_l) , nous utilisons les rapports $\alpha_i(t_p, \omega_l)$ et les rapports inverses $\beta_i(t_p, \omega_l)$ ainsi :

$$\alpha_i(t_p, \omega_l) = \frac{X_i(t_p, \omega_l)}{X_1(t_p, \omega_l)}, \quad i = 2, \dots, N, \quad (2.55)$$

$$\beta_i(t_p, \omega_l) = \frac{X_1(t_p, \omega_l)}{X_i(t_p, \omega_l)}, \quad i = 2, \dots, N. \quad (2.56)$$

Pour chaque zone TF d'analyse (T, ω_l) , nous pouvons alors resp. définir les moyennes et les variances de ces rapports sur les zones d'analyse :

$$\bar{\alpha}_i(T, \omega_l) = \frac{1}{M} \sum_{p=1}^M \alpha_i(t_p, \omega_l), \quad (2.57)$$

$$\overline{\beta}_i(T, \omega_l) = \frac{1}{M} \sum_{p=1}^M \beta_i(t_p, \omega_l), \quad (2.58)$$

$$\text{var} [\alpha_i] (T, \omega_l) = \frac{1}{M} \sum_{p=1}^M |\alpha_i(t_p, \omega_l) - \overline{\alpha}_i(T, \omega_l)|^2, \quad (2.59)$$

$$\text{var} [\beta_i] (T, \omega_l) = \frac{1}{M} \sum_{p=1}^M |\beta_i(t_p, \omega_l) - \overline{\beta}_i(T, \omega_l)|^2, \quad (2.60)$$

Les moyennes sur i de ces variances s'écrivent alors :

$$MVAR[\alpha] (T, \omega_l) = \frac{1}{N-1} \sum_{i=2}^N \text{var} [\alpha_i] (T, \omega_l), \quad (2.61)$$

$$MVAR[\beta] (T, \omega_l) = \frac{1}{N-1} \sum_{i=2}^N \text{var} [\beta_i] (T, \omega_l). \quad (2.62)$$

Nous considérons alors que les "meilleures" zones mono-sources sont les zones d'analyse qui minimisent :

$$\min \{MVAR[\alpha] (T, \omega_l), MVAR[\beta] (T, \omega_l)\}. \quad (2.63)$$

Nous montrerons dans la section 3.2.1 la validation expérimentale de ce choix.

2.3 Méthodes LI-TEMPCORR et LI-TIFCORR

2.3.1 Approches initiales de Y. Deville [74]

Ces méthodes, présentées initialement dans [74], vont être détaillées dans ce manuscrit car j'en ai effectué la validation expérimentale comme nous le verrons dans le chapitre 3. L'objectif est de nouveau d'estimer la matrice de mélange B définie dans l'Eq. (2.20). Ces approches semblent très proches des méthodes LI-TEMPROM et LI-TIFROM car les étapes sont les mêmes, à savoir (cf figure 2.1) :

1. une étape de détection pour trouver des zones où une seule source est présente,
2. une étape d'identification pendant laquelle sont estimées les colonnes de B ,
3. une étape de recombinaison des sources.

Toutefois, les paramètres utilisés dans les étapes de détection et identification sont ici bien différents de ceux des méthodes LI-TEMPROM et LI-TIFROM, comme nous allons le voir. Considérons le mélange (2.18). Nous supposons que les coefficients a_{1j} de la matrice de mélange sont non-nuls.

Approche temporelle

Dans cette section, nous considérons des coefficients de mélange et des signaux réels¹⁴ aléatoires et introduisons une approche de SAS temporelle statistique. Nous présentons d'abord les seules hypothèses que nous faisons par rapport aux sources dans cette approche, et les définitions associées.

Définition 10 *Une source est dite "isolée" dans une zone temporelle si seule cette source (parmi toutes les sources mélangées) a une variance non-nulle dans cette zone temporelle.*

Cette définition correspond au point de vue théorique. D'un point de vue pratique, cela signifie que les variances de toutes les autres sources sont négligeables par rapport à la variance de la source qui est isolée. Chaque zone temporelle considérée peut être réduite à un instant temporel t d'un point de vue théorique, comme nous le faisons dans l'approche statistique de cette section. En pratique, les paramètres statistiques des signaux sont estimés sur des fenêtres temporelles, et chaque zone temporelle considérée consiste en une telle fenêtre.

Définition 11 *Une source est dite "visible" dans le domaine temporel s'il existe au moins une zone temporelle où elle est isolée.*

Hypothèse 6 *Chaque source est visible dans le domaine temporel.*

L'hypothèse 6 a la conséquence suivante sur les propriétés statistiques des signaux sources aléatoires qui doivent être traités par la méthode de SAS introduite ci-dessous. Considérons la variance (théorique) à l'instant t d'une source $s_j(t)$, c.-à-d.

$$\sigma_{s_j}^2(t) = \mathbb{E} \left\{ (s_j(t) - \mathbb{E}\{s_j(t)\})^2 \right\}, \quad (2.64)$$

où \mathbb{E} est l'espérance mathématique. Grâce à l'hypothèse 6, $\sigma_{s_j}^2(t)$ est égale à zéro à certains instants t , c.-à-d. des instants où *une autre* source est isolée. De plus, elle est non-nulle pour d'autres instants, sinon le signal $s_j(t)$ aurait une variance nulle à tout instant, ce qui est exclu des configurations de l'approche de SAS considérée. Ainsi, le paramètre statistique $\sigma_{s_j}^2(t)$ d'une source $s_j(t)$ est dépendant du temps, de telle façon que ces signaux sources sont non-stationnaires¹⁵.

¹⁴Dans [74], Y. Deville fait l'hypothèse de coefficients de mélanges et de signaux complexes. Pour des raisons de simplicité et sans que cela gêne la compréhension de la méthode, nous nous limitons ici aux valeurs réelles de ceux-ci.

¹⁵Les commentaires ci-dessus correspondent à la définition théorique de la stationnarité de signaux aléatoires, basée sur leurs propriétés statistiques, qui sont définies en particulier par les espérances, associées à ces signaux. Considérons maintenant les implémentations pratiques de l'approche proposée, basées sur une réalisation unique de ces processus aléatoires. La discussion ci-dessus a alors comme conséquence les conditions pratiques suivantes. Les signaux sources doivent être non-stationnaires à long terme, mais ils doivent aussi être stationnaires à court terme de telle façon qu'il soit possible d'estimer correctement les paramètres statistiques de ces signaux sur de courtes fenêtres temporelles de la réalisation donnée. Ces contraintes sur les sources doivent alors être interprétées en conséquence. En particulier, la version pratique de la définition 10 signifie qu'une source est isolée dans une courte fenêtre temporelle si les variances empiriques de toutes les autres sources, obtenues en moyennant temporellement sur cette fenêtre, sont nulles (ou négligeables).

La condition dans l'hypothèse 6 et les conditions de non-stationnarité résultantes peuvent être considérées comme assez restrictives. Il doit être clair que cela correspond seulement à la première étape de notre approche globale, où cette hypothèse permet de concevoir une méthode de SAS simple dans cette situation spécifique. Dans les extensions TF que nous présenterons dans la section suivante, nous montrerons comment remplacer les hypothèses précédentes par d'autres hypothèses beaucoup moins contraignantes, évitant ainsi les conditions de non-stationnarité.

Définition 12 *A chaque instant t , les signaux sources actifs sont ceux qui ont une variance non-nulle.*

Hypothèse 7 *A chaque instant t , les versions centrées des signaux sources actifs sont linéairement indépendantes.*

Dans les précédents articles [74, 82, 83], Y. Deville faisait l'hypothèse que les sources étaient mutuellement non corrélées. Cependant, dans [80], il a montré que l'hypothèse 7 suffisait et permettait ainsi de traiter le cas de sources corrélées.

Dans un but de simplicité, les notations $s_j(t)$ et $x_i(t)$ que nous allons introduire ci-dessous font référence aux versions centrées¹⁶ des signaux dans cette section.

Nous déduisons maintenant une méthode de SAS en tirant parti de l'hypothèse 6, c.-à-d. du fait qu'il existe des aires temporelles où chaque source est isolée. De telles aires doivent d'abord être détectées, de façon à opérer à l'intérieur de celles-ci. Comme tous les signaux observés sont proportionnels dans n'importe laquelle de ces aires, une approche attrayante pour détecter ces aires consiste à vérifier les coefficients d'intercorrélation $\rho_{x_1 x_i}(t)$ entre les signaux observés $x_1(t)$ et $x_i(t)$, qui sont définis comme¹⁷

$$\rho_{x_1 x_i}(t) = \frac{\mathbb{E}\{x_1(t)x_i(t)\}}{\sqrt{\mathbb{E}\{x_1(t)^2\}\mathbb{E}\{x_i(t)^2\}}} \quad \forall i, 2 \leq i \leq N. \quad (2.65)$$

Plus précisément, nous montrons dans [80] et dans la section B.1.1 qu'une condition nécessaire et suffisante pour qu'une source soit isolée à l'instant t est

$$|\rho_{x_1 x_i}(t)| = 1 \quad \forall i, 2 \leq i \leq N. \quad (2.66)$$

Considérons maintenant une aire où une source, disons $s_k(t)$, est isolée. Les signaux observés (2.18) sont réduits à

$$x_i(t) = a_{ik}s_k(t) \quad i = 1, \dots, N. \quad (2.67)$$

¹⁶Le centrage est effectué ici comme couramment dans les études de SAS, c.-à-d. (i) d'abord, dans l'analyse théorique qui correspond à cette section, les sources sont supposées initialement être à moyenne nulle, ou leurs moyennes sont supposées être connues et les versions centrées correspondantes des signaux sont considérés, et ensuite (ii) en pratique, le centrage est effectué indépendamment pour chaque fenêtre temporelle étudiée, où les signaux considérés sont supposés être stationnaires à court terme, en soustrayant des signaux les moyennes estimées calculées dans ces fenêtres temporelles.

¹⁷Ici encore, dans les implémentations pratiques de cette méthode de SAS, ces coefficients d'intercorrélation sont estimés en remplaçant dans l'Eq. (2.65) les espérances $\mathbb{E}\{\cdot\}$ par les moyennes temporelles sur les fenêtres considérées, pour la réalisation des signaux considérée, en se fondant sur des hypothèses d'ergodicité.

D'autres paramètres, basés sur la corrélation, associés à ces signaux observés permettent d'identifier une partie de la matrice B . Plus précisément, quand la condition (2.67) est satisfaite,

$$\frac{\mathbb{E}\{x_i(t)x_1(t)\}}{\mathbb{E}\{x_1(t)^2\}} = \frac{a_{ik}}{a_{1k}}, \quad i = 2, \dots, N. \quad (2.68)$$

En comparant cette expression à l'Eq. (2.8), nous voyons que l'ensemble de tels paramètres de corrélation associés à la même aire temporelle mono-source et à toutes les observations indexées par i identifie une des colonnes de B (la première ligne de B n'est remplie que de 1, comme expliqué précédemment). En effectuant successivement de telles identifications de colonnes pour les aires temporelles associées à toutes les sources, nous pouvons identifier la matrice globale B , ce qui est le but de l'approche proposée.

L'étape de détection consiste donc à chercher les instants t (ou les fenêtres temporelles pratiques) où $\rho_{x_1x_i}(t)$ est très proche de 1 pour tous les i , $2 \leq i \leq N$. Dans [74], la méthode est testée dans le cas $N = 2$ et les zones temporelles sont classées selon les valeurs décroissantes de $\rho_{x_1x_2}(t)$. Cependant, aucune indication n'est donnée pour classer ces zones lorsque $N > 2$. Nous proposons d'utiliser l'approche décrite dans la section 2.2. La méthode LI-TEPCORR peut alors être résumée de la façon suivante.

1. On calcule la moyenne $\overline{|\rho_{x_1x_i}(t)|}$ sur i des $|\rho_{x_1x_i}(t)|$,
2. Les instants t (ou les intervalles temporels pratiques) sont ensuite classés par ordre décroissant de $\overline{|\rho_{x_1x_i}(t)|}$.
3. En considérant successivement chacune des aires mono-sources de la liste ordonnée, les paramètres de corrélation du côté gauche de l'Eq. (2.68) fournissent une colonne de B . Cette colonne est gardée si sa distance par rapport aux colonnes trouvées précédemment est au-dessus d'un seuil défini par l'utilisateur, montrant ainsi que l'aire temporelle considérée ne contient pas la même source que les précédentes. La procédure d'identification s'arrête lorsque le nombre de colonnes de B ainsi obtenues devient égal au nombre de sources (comme toutes les sources sont supposées visibles dans les données considérées, nous sommes capables d'identifier totalement B).
4. L'étape de reconstruction des sources consiste à recombinaison les signaux mélangés selon l'Eq. (2.12).

Cette approche pour les mélanges LI est donc une méthode de SAS TEMPorelle basée sur la CORRélation, utilisant les versions Centrées des signaux, d'où son nom "LI-TEPCORR-C" (TEPCORR pour *TEMPoral CORRelation-based method* en anglais). Nous verrons dans la section 3.1.5 les performances de cette méthode.

Approche temps-fréquence

L'approche introduite dans la section précédente est attrayante à cause de sa simplicité. On pourrait cependant considérer son application pratique limitée, parce qu'elle suppose que toutes les sources sont isolées dans des aires temporelles associées, ce qui est une condition restrictive. Mais elle ouvre la voie à des méthodes plus puissantes si nous prenons en compte maintenant les distributions TF des signaux, au lieu de leurs distributions temporelles considérées jusqu'ici. Ainsi, l'extension TF de cette approche peut être

rapidement définie comme suit. Supposons que chaque source est isolée dans une petite aire TF. En se basant sur la présentation précédente, on peut donc s'attendre à ce que les colonnes de B soient identifiables dans de telles aires.

La transformée TF considérée est de nouveau la TFCT définie par l'Eq. (2.2) et la méthode utilise par exemple des zones d'analyse TF à fréquence constante.

Comme dit précédemment, bien que les étapes de LI-TIFCORR soient les mêmes que celles de LI-TIFROM (voir figure 2.1 page 47), la méthode basée sur la corrélation utilise des paramètres différents. Ils sont associés aux zones d'analyse TF. Pour tout signal $u(t)$, la moyenne des transformées TF de ce signal sur une zone d'analyse (T, ω) est

$$\bar{U}(T, \omega) = \frac{1}{M} \sum_{p=1}^M U(t_p, \omega). \quad (2.69)$$

Pour chaque couple de signaux $u_1(t)$ et $u_2(t)$, les inter-corrélations des transformées TF de ces signaux sur les zones d'analyse considérées (T, ω) sont notées

$$C_{u_1 u_2}(T, \omega) = \frac{1}{M} \sum_{p=1}^M [U_1(t_p, \omega) - \bar{U}_1(T, \omega)] [U_2(t_p, \omega) - \bar{U}_2(T, \omega)]^*, \quad (2.70)$$

où $*$ est le complexe conjugué. Le coefficient de corrélation correspondant s'écrit :

$$c_{u_1 u_2}(T, \omega) = \frac{C_{u_1 u_2}(T, \omega)}{\sqrt{C_{u_1 u_1}(T, \omega) C_{u_2 u_2}(T, \omega)}}. \quad (2.71)$$

Pour chaque zone d'analyse (T, ω) , le vecteur comprenant les valeurs TF $S_j(t_p, \omega) - \bar{S}_j(T, \omega)$, pour $t_p \in T$, des sources $s_j(t)$ est noté $V_{s_j}(T, \omega)$ ci-dessous.

Hypothèse 8 *Pour chaque zone d'analyse (T, ω) , les vecteurs non nuls $V_{s_j}(T, \omega)$ sont linéairement indépendants (s'il existe au moins deux tels vecteurs dans cette zone).*

Ici encore, dans les précédents articles [74, 82, 83], Y. Deville supposait uniquement les sources non-corrélées :

$$\forall i \neq j, \quad C_{s_i s_j}(T, \omega) = 0. \quad (2.72)$$

Néanmoins, dans [77], il a prouvé que ces méthodes pouvaient être appliquées à des sources corrélées si l'hypothèse 8 était vérifiée.

L'étape de détection repose sur la propriété suivante. Une condition nécessaire et suffisante pour qu'une source soit isolée dans une zone d'analyse (T, ω) est

$$|c_{x_i x_1}(T, \omega)| = 1 \quad \forall i, 2 \leq i \leq N. \quad (2.73)$$

Une démonstration est fournie dans [77] ou dans la section B.1.2. En pratique, les zones d'analyse mono-source sont donc celles dont les modules des coefficients de corrélation $|c_{x_i x_1}(T, \omega)|$ sont les plus proches de 1, pour tout $i, 2 \leq i \leq N$. Comme nous l'avons souligné lors de la description de LI-TEMPCORR, dans [74], aucune indication n'est donnée pour classer ces zones quand $N > 2$. A nouveau, nous proposons donc d'utiliser l'approche

de la section 2.2 dans le cadre de LI-TIFCORR. Ainsi, nous calculons la moyenne sur i des $|c_{x_i x_1}(T, \omega)|$

$$\overline{|c_{x_i x_1}(T, \omega)|} = \frac{1}{N-1} \sum_{i=2}^N |c_{x_i x_1}(T, \omega)|. \quad (2.74)$$

Nous classons ensuite les zones d'analyse (T, ω) par valeur décroissante de $\overline{|c_{x_i x_1}(T, \omega)|}$. Les premières zones de cette liste sont considérées comme les "meilleures" zones mono-sources.

L'identification des colonnes de B s'effectue en parcourant la liste ordonnée des zones d'analyse de l'étape précédente. Sur chaque zone TF de cette liste, les éléments $b_{ij} = \frac{a_{ik}}{a_{1k}}$ d'une colonne de B sont estimés ainsi :

$$\frac{C_{x_i x_1}(T, \omega)}{C_{x_1 x_1}(T, \omega)} = \frac{a_{ik}}{a_{1k}}, \quad i = 2, \dots, N. \quad (2.75)$$

Notons que comme pour LI-TIFFROM, il est possible de traiter le cas sous-déterminé ou le cas de sources cachées avec LI-TIFCORR. Nous l'illustrerons expérimentalement dans le chapitre 3.

2.3.2 Variantes non-centrées de LI-TIFCORR

Nous avons vu dans la section 2.3.1 la méthode LI-TIFCORR proposée par Y. Deville. Nous en décrivons ici une variante un peu plus simple qui consiste à calculer les coefficients de corrélation des signaux non-centrés¹⁸. En effet, comme nous l'avons vu précédemment, la version originale de LI-TIFCORR est une extension de l'approche temporelle LI-TEPCORR. Cette méthode temporelle nécessite un centrage des coefficients de corrélation sur les zones mono-sources : en effet, le centrage permet d'éliminer les effets d'une composante continue présente dans les signaux sources. Dans l'extension TF présentée dans la section 2.3.1, ce centrage n'est plus vraiment nécessaire car il y a "peu de risques" d'avoir ce problème de composante continue pour les transformées TF des signaux. Aussi, pour simplifier légèrement la méthode, nous proposons ici de supprimer le centrage.

Mis à part ces modifications ci-dessous, les méthodes sont rigoureusement les mêmes. Nous noterons alors la méthode vue dans la section 2.3.1 LI-TIFCORR-C (C pour corrélations *Centrées*) alors que cette variante s'appellera LI-TIFCORR-NC (NC pour corrélations *Non-Centrées*).

Pour chaque couple de signaux $u_1(t)$ et $u_2(t)$, nous définissons les inter-corrélations non-centrées des transformées TF de ces signaux sur les zones d'analyse considérées (T, ω) comme

$$R_{u_1 u_2}(T, \omega) = \frac{1}{M} \sum_{p=1}^M U_1(t_p, \omega) U_2(t_p, \omega)^*. \quad (2.76)$$

¹⁸Dans [74], Y. Deville propose de ne pas centrer les signaux dans le calcul des coefficients de corrélation, uniquement pour l'étape d'identification. Nous proposons ici de ne pas les centrer non plus dans l'étape de détection.

Le coefficient de corrélation non-centré correspondant s'écrit :

$$r_{u_1 u_2}(T, \omega) = \frac{R_{u_1 u_2}(T, \omega)}{\sqrt{R_{u_1 u_1}(T, \omega) R_{u_2 u_2}(T, \omega)}}. \quad (2.77)$$

Dans cette section, pour chaque zone d'analyse (T, ω) , le vecteur comprenant les valeurs TF $S_j(t_p, \omega)$, pour $t_p \in T$, des sources $s_j(t)$ est noté $\tilde{V}_{s_j}(T, \omega)$ ci-dessous. Nous modifions en conséquence l'hypothèse 8 :

Hypothèse 9 *Pour chaque zone d'analyse (T, ω) , les vecteurs non nuls $\tilde{V}_{s_j}(T, \omega)$ sont linéairement indépendants (s'il existe au moins deux tels vecteurs dans cette zone).*

Nous montrons dans la section B.1.3 qu'une condition nécessaire et suffisante pour qu'une zone TF (T, ω) soit mono-source est alors :

$$|r_{x_i x_1}(T, \omega)| = 1 \quad \forall i, 2 \leq i \leq N. \quad (2.78)$$

Une fois une zone mono-source trouvée, l'identification (2.75) de la colonne de B correspondant à cette zone est remplacée par :

$$\frac{R_{x_i x_1}(T, \omega)}{R_{x_1 x_1}(T, \omega)} = \frac{a_{ik}}{a_{1k}}, \quad i = 2, \dots, N. \quad (2.79)$$

Nous verrons dans le chapitre 3 comment se comporte expérimentalement cette méthode. Cette variante peut bien sûr être appliquée à LI-TEPCORR et sera testée dans la section 3.1.5.

2.4 Conclusion

Dans ce chapitre, nous avons décrit et étendu deux méthodes (temporelles et) TF de SAS pour les mélanges LI. Ces méthodes, applicables aux mélanges (sur-)déterminés et sous-déterminés ainsi qu'aux sources cachées, cherchent dans le domaine d'analyse des zones où seule une source est active. Dans ces zones, elles estiment les colonnes de la matrice de mélange.

La première méthode, LI-TIFROM, est basée sur la variance des rapports d'observations pour trouver les zones mono-sources. Les colonnes de la matrice de mélange sont estimées comme des moyennes de ces rapports. Nous avons mis en évidence le côté non-symétrique de la variance des rapports d'observation et proposé une solution pour remédier à ce problème.

La seconde méthode, LI-TIFCORR, est basée sur des rapports d'inter et auto-corrélations entre les observations. Nous avons étudié une variante, basée sur des paramètres non-centrés.

Enfin, nous avons proposé une extension alternative de ces méthodes au cas général N sources et N capteurs : plutôt que de nous limiter à 2 observations pour estimer les zones mono-sources, nous avons privilégié l'utilisation de toutes les observations afin d'éviter de fausses détections.

Nous allons présenter les performances de ces méthodes dans le chapitre suivant.

Chapitre 3

Performances des méthodes linéaires instantanées

Les résultats présentés dans ce chapitre visent à illustrer les performances des méthodes LI de SAS présentées dans le chapitre précédent. La méthode LI-TIFFROM servira de point de comparaison avec la méthode LI-TIFCORR dont une analyse détaillée des performances n'avait jamais été faite précédemment. Nous en profiterons aussi pour montrer l'apport des améliorations proposées dans le chapitre précédent par rapport aux méthodes originales.

Dans un premier temps, nous allons étudier le cas $N = 2$ sources et $N = 2$ observations, d'abord pour une unique matrice de mélange (voir la section 3.1.2), puis les variations des performances en fonction de la matrice de mélange (voir la section 3.1.3). Nous détaillerons les performances des approches centrée et non-centrée de LI-TIFCORR mais aussi celles de nos variantes de la méthode LI-TIFFROM, décrites dans la section 2.1.3. Nous en profiterons pour la comparer avec la méthode originale (voir section 3.1.4). Nous verrons dans la section 3.1.5 le comportement des méthodes temporelles LI-TEMPCORR-C et LI-TEMPCORR-NC. Nous comparerons aussi les approches LI-TIFCORR-C, LI-TIFCORR-NC et LI-TIFFROM-S2 avec une autre méthode développée dans l'équipe, LI-TIFCOHERE [5, 8, 83] dans la section 3.1.6, ainsi qu'avec des méthodes classiques issues de la toolbox Matlab "ICALab" [57] (voir la section 3.1.7).

Dans un deuxième temps, nous étudierons les comportements de nos méthodes LI-TIFFROM-S2, LI-TIFCORR-C et LI-TIFCORR-NC dans le cas $N = 4$ sources et $N = 4$ observations (voir section 3.2.2).

Nous montrerons ensuite le comportement de ces méthodes dans le cas sous-déterminé et le cas où des sources sont cachées dans la section 3.3.

Enfin, nous testerons dans la section 3.4 les capacités de séparation de nos méthodes sur des signaux astrophysiques.

Le critère de performances des méthodes de SAS utilisé dans ce document est le rapport signal sur interférence (SIR - *Signal to interference ratio* en anglais) défini dans l'annexe C. Nous mesurons le SIR de sortie SIR^{out} ou l'amélioration de SIR (noté $SIRI$ - *SIR improvement* en anglais) entre entrées et sorties. Ces SIR sont liés par la relation (C.10),

exprimée en dB, que nous rappelons ici :

$$(SIRI)_{dB} = (SIR^{out})_{dB} - (SIR^{in})_{dB}, \quad (3.1)$$

où SIR^{in} est le SIR d'entrée du système de SAS. Nous considérerons dans ce chapitre qu'une méthode *échoue* à séparer les sources lorsque le $SIRI$ obtenu est négatif. Notons que ce critère nécessite par définition la connaissance préalable des sources et des matrices de mélange et ne peut donc être utilisé que pour des simulations, comme c'est le cas dans ce document et comme cela a été souligné par A. Mansour et al. dans [164].

3.1 Mélanges de $N = 2$ sources et $N = 2$ observations

Dans cette section, nous présentons un grand nombre de tests effectués dans les conditions suivantes :

- nous partons de plusieurs signaux réels de parole en anglais (voir section 3.1.1 ci-dessous),
- nous procédons à plusieurs mélanges LI artificiels de ces sources,
- nous utilisons ces signaux mélangés avec diverses méthodes de SAS : suivant les cas la version standard LI-TIFROM de F. Abrard et Y. Deville, nos versions améliorées de celle-ci, les méthodes LI-TIFCORR-C, LI-TIFCORR-NC ainsi que les méthodes classiques de SAS proposées dans le logiciel ICALab [57].

3.1.1 Sources utilisées

Lors de nos tests de SAS, nous avons utilisé des signaux de la base de données libre d'accès¹⁹ Multext. Cette base de données contient des signaux enregistrés en chambre anéchoïque, qui supprime les effets de réverbération.

Nous avons surtout travaillé avec 8 signaux de parole en anglais (voir figure 3.1), échantillonnés à 20 kHz. Les signaux 1, 2, 3, 7 et 8 correspondent à des voix masculines, tandis que les signaux 4, 5 et 6 correspondent à des voix féminines. Parmi eux, certains sont issus de la même personne (1 et 2 pour les voix masculines, 4 et 5 pour les voix féminines).

Les jeux de signaux utilisés, numérotés de 1 à 6, correspondent aux couples (1,2), (1,3), (4,5), (4,6), (1,4) et (7,8), soit :

- Jeu 1 : même voix masculines.
- Jeu 2 : voix masculines différentes.
- Jeu 3 : même voix féminines.
- Jeu 4 : voix féminines différentes.
- Jeu 5 : une voix masculine et une voix féminine.
- Jeu 6 : voix masculines différentes.

Toutes les sources correspondent à 2.5 s de parole sauf pour le jeu 6 qui correspond à 5 s de parole discontinue, les signaux du jeu 2 étant extraits de ceux du jeu 6.

¹⁹Le lien est le suivant : <http://aune.lpl1.univ-aix.fr/projects/multext/>

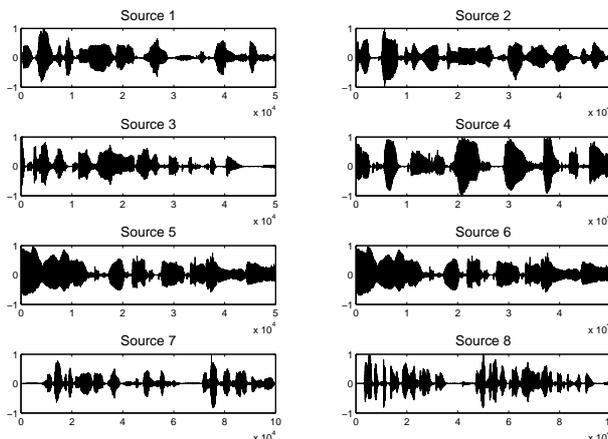


FIG. 3.1 – Ensemble des signaux sources testés.

Tous nos signaux sont centrés²⁰ et normalisés²¹ avant l'opération de mélange : cela simplifie notre problème sans que ce soit une limitation ou une contrainte de nos méthodes et les sources ont ainsi des puissances relativement similaires, ce qui est plus réaliste pour les tests de "cocktail party".

3.1.2 Première série de tests avec une matrice de mélange fixée

Dans le cas où $N = 2$ sources sont mélangées, nous avons tout d'abord étudié le cas d'un mélange "difficile" de sources. La matrice utilisée est notée

$$A_1 = \begin{bmatrix} 1 & 0.9 \\ 0.8 & 1 \end{bmatrix}. \quad (3.2)$$

D'après l'Eq. (2.8), la matrice B à estimer est

$$B = \begin{bmatrix} 1 & 1 \\ 0.8 & 1.1111 \end{bmatrix} \quad \text{ou} \quad B = \begin{bmatrix} 1 & 1 \\ 1.1111 & 0.8 \end{bmatrix}. \quad (3.3)$$

Nous faisons varier les paramètres des méthodes TF ainsi :

- Le nombre d d'échantillons dans chaque TFCT a varié géométriquement de 128 à 1024 échantillons, avec un pas de 2.
- Le nombre M de fenêtres TF par zone d'analyse a été successivement fixé à 8, 10 et 12.
- Le recouvrement temporel ρ entre deux fenêtres TF a été successivement fixé à 50%, 75% et 90%.

²⁰Le centrage considéré consiste à déduire d'un signal temporel sa valeur moyenne. Il est donc différent du centrage effectué avec certaines méthodes à parcimonie considérées dans ce document, c.-à-d. LI-TEPCORR-C et LI-TIFCORR-C, où dans chaque zone d'analyse, les observations sont centrées.

²¹La normalisation consiste à diviser chaque source par la valeur maximale de son module.

TAB. 3.1 – Performances (SIR^{out} en dB) en fonction du jeu de sources.

Méthode	Critère de perf.	Jeu de sources						
		1	2	3	4	5	6	1-6
LI-TIFROM-S2	SIR^{out} : Moy.	57.1	61.0	57.5	60.0	64.3	67.4	61.2
	Ec. type	13.0	13.5	11.6	13.8	10.0	8.5	12.3
	Min.	17.3	27.6	21.5	33.3	37.3	42.5	17.3
	Max.	72.9	85.9	71.8	80.4	83.0	80.6	85.9
LI-TIFCORR-C	SIR^{out} : Moy.	61.1	64.1	57.9	65.8	62.8	72.1	64.0
	Ec. type	11.8	7.7	7.3	7.2	4.9	6.8	8.9
	Min.	31.0	47.6	38.3	50.8	54.5	59.0	31.0
	Max.	76.7	75.9	70.8	78.4	74.3	88.6	88.6
LI-TIFCORR-NC	SIR^{out} : Moy.	61.2	65.7	61.5	66.8	63.3	71.6	65.0
	Ec. type	11.7	7.6	8.1	6.5	5.0	7.2	8.6
	Min.	31.1	47.0	40.6	52.7	56.3	61.1	31.1
	Max.	76.6	78.3	80.2	76.5	77.5	96.4	96.4

Critères de performances : valeurs moyenne, minimum et maximum et écart-type des SIR^{out} (en dB) sur tous les paramètres des méthodes de SAS.

En conséquence, la taille des zones d'analyse variait de 219 à 6656 échantillons, c.-à-d. entre 11 et 330 ms. Les autres paramètres de ces méthodes de SAS étaient constants dans ces tests, c.-à-d. :

- la fenêtre h utilisée dans les calculs des TFCT était une fenêtre de Hanning.
- Le recouvrement temporel entre les zones d'analyse adjacentes était de 50%.
- La distance entre deux colonnes identifiées de la matrice B était mesurée comme la plus grande valeur absolue entre les éléments de ces vecteurs²² qui ont le même indice, c.-à-d. en utilisant la norme l^∞ . Le seuil de distance pour accepter une nouvelle colonne de B était fixé à 0.15 (cette valeur a été choisie ainsi parce qu'elle fournissait de bons résultats dans une série de tests préliminaires; une version automatisée de cette méthode pour sélectionner ce seuil serait préférable, comme on en discutera dans le chapitre 5).

Pour chaque méthode de SAS considérée, 216 tests ont été effectués dans les conditions définies ci-dessus. Toutes les méthodes ont réussi à identifier toutes les colonnes de B , et donc à séparer les sources considérées dans tous les tests. Les résultats correspondants sont fournis dans le tableau 3.1. Ce tableau montre pour chaque jeu de sources, les SIR^{out} moyens obtenus, l'écart-type, les valeurs minimales et maximales de ces valeurs de SIR^{out} pour les méthodes LI-TIFCORR-C, LI-TIFCORR-NC et LI-TIFFROM-S2.

²²Dans le cas de LI-TIFFROM-S2, comme nous l'avons souligné dans la section 2.1.3, si le minimum du critère (2.38) est obtenu à partir de $\beta(t, \omega)$, alors la distance entre deux colonnes de la matrice B est mesurée comme la plus grande valeur absolue entre les inverses des éléments de ces vecteurs qui ont le même indice.

Nous remarquons que chaque méthode sépare les sources avec une grande précision, puisque les SIR^{out} moyens sur tous les jeux de sources sont au-dessus de 61 dB. Néanmoins, une hiérarchie semble apparaître entre ces méthodes : en effet, les méthodes basées sur la corrélation apportent de meilleures performances que LI-TIFROM-S2.

Ainsi, pour les méthodes LI-TIFCORR, les SIR^{out} moyens par jeu de sources sont compris entre 57.9 et 72.1 dB, les écarts-types sont modérés puisqu'ils varient entre 4.9 et 8.1 dB (sauf pour le premier jeu de sources, pour qui ils sont égaux à 11.8 et 11.7 dB suivant la méthode considérée, mais ce jeu est plus difficile que les applications pratiques puisqu'il correspond à deux signaux issus d'un même locuteur, ce qui implique un plus fort recouvrement fréquentiel des signaux). Le SIR^{out} minimum est égal à 38.3 dB (sauf pour le premier ensemble de sources où il est égal à 31.0 dB). Le SIR^{out} maximal sur ces tests est égal à 96.4 dB. L'approche LI-TIFCORR-NC est légèrement meilleure que LI-TIFCORR-C en termes de SIR^{out} moyens pour tous les jeux de signaux, sauf pour le jeu 6. Elle apporte aussi un SIR^{out} moyen légèrement meilleur sur l'ensemble des signaux sources, c.-à-d. 65.0 par rapport à 64.0 dB.

Concernant LI-TIFROM-S2, les SIR^{out} moyens sont compris entre 57.1 et 67.4 dB, les écarts-types sont moins faibles que pour les méthodes basées sur la corrélation puisqu'ils varient entre 8.5 et 13.8 dB. Le SIR^{out} minimum est égal à 21.5 dB (sauf pour le premier ensemble de sources où il est égal à 17.3 dB). Le SIR^{out} maximum sur ces tests est égal à 85.9 dB. Dans l'ensemble, les performances globales de LI-TIFROM-S2 sont plus faibles que celles des versions centrées et non-centrées de LI-TIFCORR.

Étudions maintenant plus en détail les performances de nos méthodes. Considérons d'abord l'influence du jeu de sources considéré sur les performances. Dans un premier temps, étudions les sources de parole continue. On pourrait s'attendre à ce que les performances des méthodes TF de SAS soient plus élevées pour les sources qui ont un faible recouvrement spectral et donc : (i) plus élevées pour les voix masculines (resp. féminines) différentes que pour le même locuteur et (ii) plus élevées quand on mélange des voix masculine et féminine plutôt que quand on mélange des voix masculines (resp. féminines) venant de locuteurs différents ou identiques. De la même façon, on s'attend à ce que le jeu 6 apporte de meilleures performances que le jeu 2, parce qu'il contient des phases de silence en plus, rendant les sources plus visibles (à n'importe quelle fréquence). Les valeurs moyennes de SIR^{out} dans le tableau 3.1 confirment toutes ces attentes, sauf une : avec les méthodes TF basées sur la corrélation, les performances obtenues quand on mélange de la parole masculine et féminine sont légèrement plus faibles que quand on mélange deux voix masculines ou deux voix féminines (mais quand même plus élevées que quand on mélange deux signaux issus du même locuteur). Cependant, quelques permutations par rapport aux mérites respectifs attendus des sources considérées pourraient être prévues parce que tous les jeux de sources considérés fournissent des SIR^{out} moyens très élevés et souvent relativement similaires, et de telles permutations ne sont pas statistiquement significatives à cause du nombre limité de signaux considérés.

Les tableaux 3.2-3.4 apportent une analyse plus détaillée de certains aspects de ces tests précédents. Ils contiennent seulement les valeurs globales sur tous les jeux de sources des critères de performances considérés, mais chacun de ces tableaux détaille les variations de ces critères par rapport à un des paramètres des méthodes de SAS (en moyennant sur les autres paramètres). Ils montrent d'abord que les SIR^{out} moyens sont assez peu

TAB. 3.2 – Performances des méthodes TF de SAS en fonction de la taille des fenêtres utilisées dans le calcul des TFCT (en nombre d'échantillons).

Méthode	Critère de perf.	Taille d des fenêtres			
		128	256	512	1024
LI-TIFROM-S2	SIR^{out} : Moy.	67.5	65.9	60.2	51.2
	Ec. type	8.1	6.5	11.0	14.8
	Min.	49.9	41.6	35.0	17.3
	Max.	85.9	83.0	80.4	79.5
LI-TIFCORR-C	SIR^{out} : Moy.	65.6	65.5	64.1	60.6
	Ec. type	8.0	6.3	9.9	10.2
	Min.	46.4	50.8	31.0	34.0
	Max.	78.4	75.6	86.5	88.6
LI-TIFCORR-NC	SIR^{out} : Moy.	67.2	66.8	63.7	62.3
	Ec. type	6.4	5.1	8.8	11.8
	Min.	50.2	54.7	31.1	34.1
	Max.	78.8	78.3	76.6	96.4

Critères de performances : les mêmes que dans le tableau 3.1.

TAB. 3.3 – Performances des méthodes TF de SAS en fonction du nombre de fenêtres TF dans les zones d'analyse.

Méthode	Critère de perf.	Nombre M de fenêtres		
		8	10	12
LI-TIFROM-S2	SIR^{out} : Moy.	62.2	61.7	59.7
	Ec. type	10.2	13.4	13.1
	Min.	35.4	17.3	21.5
	Max.	85.9	83.0	79.5
LI-TIFROM-S2 ($d \neq 1024$)	SIR^{out} : Moy.	64.6	65.8	63.2
	Ec. type	8.3	9.2	10.1
	Min.	44.1	36.0	35.0
	Max.	85.9	83.0	78.5
LI-TIFCORR-C	SIR^{out} : Moy.	63.7	63.8	64.5
	Ec. type	7.8	8.9	10.0
	Min.	38.2	42.3	31.0
	Max.	79.1	86.5	88.6
LI-TIFCORR-NC	SIR^{out} : Moy.	65.8	64.7	64.6
	Ec. type	7.6	8.9	9.3
	Min.	39.4	39.9	31.1
	Max.	81.6	96.4	84.1

Critères de performances : les mêmes que dans le tableau 3.1.

TAB. 3.4 – Performances des méthodes TF de SAS en fonction du recouvrement temporel des fenêtres TF.

Méthode	Critère de perf.	Recouvrement ρ		
		50%	75%	90%
LI-TIFFROM-S2	SIR^{out} : Moy.	54.7	63.1	65.8
	Ec. type	15.5	10.7	6.1
	Min.	17.3	34.7	50.0
	Max.	80.6	85.9	83.0
LI-TIFFROM-S2 ($d \neq 1024$)	SIR^{out} : Moy.	60.5	67.3	65.8
	Ec. type	12.2	7.1	5.9
	Min.	35.0	51.5	49.9
	Max.	80.6	85.9	83.0
LI-TIFCORR-C	SIR^{out} : Moy.	62.1	65.7	64.1
	Ec. type	10.9	8.1	7.1
	Min.	31.0	34.0	46.4
	Max.	86.5	79.1	88.6
LI-TIFCORR-NC	SIR^{out} : Moy.	62.6	66.4	66.1
	Ec. type	10.5	7.7	6.9
	Min.	31.1	34.1	50.2
	Max.	80.2	79.1	96.4

Critères de performances : les mêmes que dans le tableau 3.1.

sensibles à la taille des TFCT, sauf pour LI-TIFFROM-S2. En effet, on s'aperçoit pour cette méthode que dans le cas où $d = 1024$, les performances sont bien moins bonnes (même si elles restent parfaitement acceptables puisque le SIR^{out} moyen est égal à 51.2 dB) que pour les autres valeurs de d . Quand $d \neq 1024$, on remarque que les performances de LI-TIFFROM-S2 sont alors plus proches des valeurs des méthodes LI-TIFCORR (le SIR^{out} moyen pour d variant de 128 à 512, l'écart-type, les SIR^{out} minimal et maximal sont alors resp. égaux à 64.6, 9.2, 35.0 et 85.9 dB). Pour la suite de cette étude détaillée, comme les performances de LI-TIFFROM-S2 sont perturbées par cette valeur $d = 1024$, nous avons rajouté une série de lignes qui ne tient pas compte de ce cas aux tableaux 3.3 et 3.4. Nos paramètres optimaux pour LI-TIFFROM-S2 seront choisis en ne tenant compte que de ces données-là. Toutes nos méthodes sont peu sensibles au nombre M de fenêtres TF dans les zones d'analyse. Par contre, nous voyons dans le tableau 3.4 qu'un faible recouvrement engendre une légère baisse des performances des méthodes TF. Cela est sûrement dû au fait que les "chances" de trouver de bonnes zones d'analyse mono-sources sont plus faibles lorsque le recouvrement n'est pas important car le pavage pour trouver les zones TF mono-sources est moins dense.

Le tableau 3.2 montre que le nombre d'échantillons dans les fenêtres TF doit être fixé de préférence à 256 (ou 128) dans le but d'optimiser la moyenne, l'écart type et la valeur minimale des SIR^{out} . Un compromis entre ces paramètres peut être obtenu en utilisant

aussi 10 TFCT par zone d'analyse²³ et un recouvrement de 75% entre ces fenêtres. Ces tableaux montrent aussi que LI-TIFCORR-NC apporte des SIR^{out} moyens légèrement meilleurs que LI-TIFCORR-C dans la plupart des cas considérés, ce qui est en accord avec les performances globales respectives de ces méthodes déduites du tableau 3.1.

3.1.3 Deuxième série de tests : influence des coefficients de mélange

Dans une seconde série de tests, nous avons voulu étudier l'influence des coefficients de mélange sur les performances de nos méthodes. Pour cela, nous avons utilisé des matrices symétriques de la forme

$$A = \begin{bmatrix} 1 & \lambda \\ \lambda & 1 \end{bmatrix}, \quad (3.4)$$

où nous avons considéré successivement des valeurs réelles différentes pour le terme λ , de façon à faire varier le taux de mélange et donc le SIR^{in} associé aux signaux observés $x_i(t)$. Plus précisément, la forme matricielle définie dans l'Eq. (3.4) nous donne ceci :

$$SIR^{in} = \frac{1}{\lambda^2} \quad (3.5)$$

comme vu dans l'Eq. (C.17) dans l'annexe C.1. Les valeurs que nous avons considérées pour λ sont $\lambda = 0.01, 0.1, 0.5$ et 0.9 . Les matrices A correspondantes sont resp. notées A_2, A_3, A_4 et A_5 et les SIR^{in} associés sont fournis dans le tableau 3.5.

Les jeux de sources et les gammes de paramètres des méthodes sont les mêmes que ceux de la section 3.1.2. Chaque méthode TF de SAS a donc été testée dans 1080 configurations. Mis à part le cas particulier $\lambda \leq 0.1$, chacune de ces méthodes a réussi à identifier B dans tous les tests. Contrairement à la section 3.1.2, le SIR^{in} varie dans les tests considérés ici. L'Eq. (3.1) montre alors que la correspondance entre SIR^{out} et $SIRI$ n'est pas fixée. Les valeurs de ces deux paramètres sont donc fournies dans le tableau 3.5. Parmi ces deux paramètres, les performances des méthodes de SAS doivent être exprimées de préférence en termes de leur comportement de sortie, défini par les SIR^{out} . Le tableau 3.5 montre des différences nettes entre les méthodes suivant les valeurs de λ . Compte tenu des différences dans l'étape pour garder une colonne estimée de B entre LI-TIFFROM-S2 et les autres méthodes TF, nous allons d'abord nous concentrer sur la première variante de TIFFROM et les deux approches basées sur la corrélation.

²³Les performances peuvent être plus faibles quand on utilise un plus grand nombre de fenêtres TF par zone d'analyse pour la raison suivante. Considérons les meilleures aires TF mono-sources, c.-à-d. les aires où les sources interférentes ont les plus petites valeurs de TFCT par rapport à la source d'intérêt. Ce sont les zones où la matrice de mélange peut être identifiée avec la plus grande précision. Considérons le cas où ces zones sont "petites". Elles sont exploitées dans nos méthodes de SAS quand la taille des zones d'analyse utilisées est plus petite que la taille des aires TF mono-sources. De hautes performances sont ainsi obtenues. Dans le cas contraire, si nos méthodes sont utilisées avec des zones d'analyse plus grandes que ces aires mono-sources, elles ne peuvent pas identifier la matrice de mélange dans ces aires. La matrice de mélange est alors identifiée dans d'autres aires (qui peuvent inclure ces aires), en donnant ainsi de moins bonnes performances. Accroître le nombre de TFCT par zones d'analyse augmente aussi la taille de ces zones et peut aussi amener à une telle dégradation de performances.

TAB. 3.5 – Performances des méthodes TF de SAS en fonction de la matrice de mélange.

Méthode	Critère de perf.	Matrice de mélange				
		A_2 ($\lambda = 0.01$)	A_3 ($\lambda = 0.1$)	A_4 ($\lambda = 0.5$)	A_1	A_5 ($\lambda = 0.9$)
	SIR^{in}	40.0	20.0	6.0	1.4	0.9
LI-TIFFROM-S1	SIR^{out}	5.4	43.4	60.5	61.2	61.2
	$SIRI$	-34.6	23.4	54.5	59.8	60.3
LI-TIFFROM-S2	SIR^{out}	47.2	56.5	60.8	61.2	61.2
	$SIRI$	7.2	36.5	54.7	59.8	60.3
LI-TIFCORR-C	SIR^{out}	-0.4	46.1	63.7	64.0	64.0
	$SIRI$	-40.4	26.1	57.6	62.6	63.0
LI-TIFCORR-NC	SIR^{out}	0.8	52.1	64.4	65.0	65.0
	$SIRI$	-39.2	32.1	58.4	63.6	64.1

Critères de performances : SIR^{in} (en dB) et valeurs moyennes des SIR^{out} et $SIRI$ (en dB) sur tous les paramètres des méthodes et sur tous les jeux de sources.

Quand $\lambda \geq 0.5$, le SIR^{out} moyen des méthodes est compris entre 60.5 et 65.0 dB dans tous les cas, ce qui est très bon. L'Eq. (3.1) nous permet de voir qu'au contraire, le $SIRI$ varie significativement avec A , ce qui est confirmé dans le tableau 3.5.

Quand $\lambda \leq 0.1$, le SIR^{out} est plus faible, en particulier quand $\lambda = 0.01$. Ainsi, l'approche LI-TIFFROM-S1 et les méthodes basées sur la corrélation voient leurs performances baisser (le SIR^{out} moyen varie entre -0.4 et 52.1 dB). De plus, pour $\lambda = 0.1$, LI-TIFFROM-S1 voit son SIR^{out} moyen inférieur de 3 à 10 dB par rapport aux deux autres approches. La méthode LI-TIFCORR-NC est alors la plus performante des trois puisqu'elle n'échoue que 19 fois sur 216 contre 31 fois pour les deux autres approches. Si on regarde en détail les performances des méthodes basées sur la corrélation pour ce cas, on constate que pour chaque couple de sources, les versions centrées et non-centrées de LI-TIFCORR donnent des performances globales assez proches sauf pour les jeux 3 et 5 où LI-TIFCORR-C a de plus grandes difficultés à estimer la matrice de mélange. De même, l'approche LI-TIFFROM-S1 présente des performances assez proches de celles de LI-TIFCORR-C sauf pour le jeu 5 où elle a beaucoup de mal à estimer la matrice de mélange et le jeu 6 où elle l'estime moins précisément.

Cette dégradation des performances avec les valeurs de λ peut être due au fait que : (i) les paramètres de mélanges estimés b_{ij} sont plus sensibles aux erreurs d'estimation quand λ est petit, et (ii) ces paramètres d'identification couvrent alors une large gamme, ce qui peut réduire les performances de nos méthodes pour sélectionner quelles colonnes estimées expérimentales de la matrice de mélange sont gardées.

Le phénomène (ii) tend à être vérifié par les performances de LI-TIFFROM-S2 : en effet, contrairement aux autres méthodes, la méthode LI-TIFFROM-S2 réussit à estimer la matrice de mélange dans tous les cas sauf un quand $\lambda = 0.1$. Quand $\lambda = 0.01$, on voit très nettement qu'elle est meilleure que les trois autres approches : alors que cette

version améliorée n'échoue que 45 fois sur 216 dans l'estimation de la matrice de mélange, les approches LI-TIFFROM-S1, LI-TIFCORR-C et LI-TIFCORR-NC échouent resp. 127, 139 et 135 fois. Cela montre donc que les bonnes performances de LI-TIFFROM-S2 dans ces cas particuliers sont surtout dues au fait que la mesure de distance entre les colonnes dépend du rapport qui minimise le critère (2.38), comme expliqué dans la section 2.1.3. De plus, dans le cadre de sources plus fortement mélangées (matrices A_1 et A_5), les deux variantes de LI-TIFFROM permettent exactement la même qualité de séparation, ce qui est en accord avec les résultats escomptés.

Nous voyons donc un classement différent de celui proposé dans la section 3.1.2 : si on tient compte de toutes les matrices de mélange, la méthode LI-TIFFROM-S2 est celle qui globalement donne les meilleures performances, suivie par LI-TIFCORR-NC, LI-TIFCORR-C et LI-TIFFROM-S1. Par contre, si on ne considère que les cas où les sources sont les plus mélangées, c.-à-d. $\lambda \geq 0.5$, alors, comme dans la section 3.1.2, la méthode LI-TIFCORR-NC est celle qui donne globalement les meilleures performances, suivie de LI-TIFCORR-C puis LI-TIFFROM (-S1 ou -S2). Notons toutefois que ces différences de performances pour $\lambda \geq 0.5$ sont minimales : en effet, il y a moins de 5 dB d'écart entre les SIR^{out} moyens de LI-TIFFROM-S1 et de LI-TIFCORR-NC, ce qui est faible compte tenu des performances des méthodes (SIR^{out} supérieurs à 60 dB en moyenne).

3.1.4 Comparaison entre nos versions de LI-TIFFROM et la version originale

Comme nous l'avons vu dans la section 2.1.2, la méthode originale de F. Abrard et Y. Deville est sensible aux coefficients de mélange, à cause de l'aspect asymétrique de la variance des rapports des observations. En particulier dans le cas de matrices symétriques définies par l'Eq. (3.4), la version originale de LI-TIFFROM est très sensible aux faibles valeurs de λ . Notre approche LI-TIFFROM-S2 évite ce souci grâce à l'étude conjointe des rapports $\frac{X_i(n,\omega)}{X_1(n,\omega)}$ et des rapports inverses $\frac{X_1(n,\omega)}{X_i(n,\omega)}$ et d'une mesure de distance qui tient compte de ces rapports inverses²⁴. Ceci est expérimentalement vérifié dans le tableau 3.6 qui présente les mêmes tests que ceux effectués dans la section 3.1.3.

Alors que notre approche a un SIR^{out} élevé pour toutes les valeurs de λ considérées (le SIR^{out} moyen varie de 47.2 à 61.2 dB), la méthode originale voit son SIR^{out} chuter fortement avec les valeurs de λ (le SIR^{out} varie de -33.7 à 61.0 dB), en particulier dans le cas $\lambda \leq 0.1$ où la méthode n'arrive pas à estimer correctement la matrice B 332 fois sur 432 contre seulement 46 fois pour notre version améliorée. Pour les matrices A_1 ou A_5 , les SIR^{out} moyens des deux versions de la méthode sont par contre très proches, ce qui était prévisible compte tenu de notre analyse faite dans la section 2.1.3 et des coefficients de mélange appliqués.

²⁴Notez cependant que la version LI-TIFFROM-S1 présente elle-aussi de meilleures performances que la version originale. Notamment, quand $\lambda = 0.1$, nous voyons nettement l'apport des rapports inverses dans l'étape de détection. Comme nous l'avons souligné précédemment, les excellentes performances de LI-TIFFROM-S2 pour les très petites valeurs de λ sont dues aux modifications du critère de sélection d'une colonne de la matrice de mélange.

TAB. 3.6 – Performances des méthodes LI-TIFROM en fonction de la matrice de mélange.

Méthode	Critère de perf.	Matrice de mélange				
		A_2 ($\lambda = 0.01$)	A_3 ($\lambda = 0.1$)	A_4 ($\lambda = 0.5$)	A_1	A_5 ($\lambda = 0.9$)
	SIR^{in}	40.0	20.0	6.0	1.4	0.9
LI-TIFROM	SIR^{out}	-33.7	18.2	59.5	61.0	61.0
	$SIRI$	-73.7	-1.8	53.4	59.6	60.1
LI-TIFROM-S2	SIR^{out}	47.2	56.5	60.8	61.2	61.2
	$SIRI$	7.2	36.5	54.7	59.8	60.3

Critères de performances : les même que ceux du tableau 3.5.

3.1.5 Performances des méthodes LI-TEPCORR-C et LI-TEPCORR-NC

Nous avons aussi vérifié les performances des méthodes LI-TEPCORR-C et LI-TEPCORR-NC pour les mêmes signaux sources et la même matrice de mélange A_1 que dans la section 3.1.2. Les paramètres de ces méthodes de SAS ont été sélectionnés ainsi : le nombre d'échantillons dans chaque fenêtre temporelle variait géométriquement de 256 à 4096. Le recouvrement entre ces fenêtres était successivement fixé à 50%, 75% et 90%. Les performances globales résultantes sur les deux paramètres sont données dans le tableau 3.7 et confirment nos attentes. Tout d'abord, les performances obtenues par nos approches temporelles sont significativement plus faibles que nos deux extensions TF (resp. 51.3 et 52.0 dB pour les versions centrées et non centrées contre 64.0 et 65.0 dB). Plus important, ces approches temporelles identifient de façon moins précise les matrices de mélange dans un nombre significatif de configurations. Cela se vérifie à la fois dans le plus grand écart-type de son SIR^{out} (20.8 et 21.2 dB contre 8.9 et 8.6 dB) et dans son SIR^{out} minimum vraiment plus bas, c.-à-d. 3.5 dB pour les deux versions (contre 31.0 et 31.1 dB). Plus précisément, si nous considérons d'abord seulement les sources de parole continue, le SIR^{out} minimum pour un jeu donné de sources est compris entre 3.5 et 14.7 dB, suivant le jeu considéré.

Notons aussi que le seul jeu de sources à partir duquel nos approches temporelles n'apportent pas de performances plus faibles que leurs extensions TF est le jeu 6 : le SIR^{out} moyen est alors resp. de 72.0 et 72.3 dB (contre 72.1 et 71.6 dB pour les méthodes TF), leurs écarts-types sont resp. de 6.1 et 6.0 dB (contre 6.8 et 7.2 dB) et leur minima de 60.4 dB pour les deux versions (contre 59.0 et 61.1 dB). Ces résultats sont encore en accord avec nos considérations théoriques précédentes : le jeu 6 est le seul jeu de sources correspondant à de la parole discontinue, où les longues phases de silence existent donc et sont exploitées par LI-TEPCORR-C et LI-TEPCORR-NC pour obtenir de hautes performances.

Enfin, remarquons que, comme pour les approches TF, la version non centrée de TEMPCORR apporte des performances légèrement meilleures que la version centrée.

TAB. 3.7 – Performances des méthodes temporelles de SAS pour chaque ensemble de deux sources de parole et performances globales moyennes.

Méthode de SAS	Critère de perfs.	Jeu de sources						
		1	2	3	4	5	6	1-6
LI-TEPCORR-C	SIR^{out} : Moy.	42.4	49.6	39.3	53.0	51.8	72.0	51.3
	Ec. type	19.5	17.1	20.3	24.3	18.5	6.1	20.8
	Min.	10.2	14.7	6.1	14.4	3.5	60.4	3.5
	Max.	66.4	66.9	63.6	79.7	67.7	82.0	82.0
LI-TEPCORR-NC	SIR^{out} : Moy.	43.4	51.3	39.5	52.4	52.9	72.3	52.0
	Ec. type	20.0	18.8	20.6	23.9	19.3	6.0	21.2
	Min.	10.2	14.7	6.1	14.4	3.5	60.4	3.5
	Max.	68.4	73.8	64.5	76.3	70.8	85.3	82.3

Critères de performances : les mêmes que dans le tableau 3.1.

Ces tests confirment pour les signaux de parole, excepté dans les situations restrictives comprenant de la parole discontinue, que nos méthodes TF doivent être préférées à nos approches temporelles de façon à optimiser les performances.

3.1.6 Comparaison avec une autre méthode à parcimonie TF : LI-TIFCOHERE

Dans cette section, nous proposons de comparer nos méthodes avec la méthode LI-TIFCOHERE proposée dans [5, 8, 83]. Cette approche, créée par notre équipe, correspond initialement à une approche différente des méthodes LI-TIFFROM et LI-TIFCORR par les critères utilisés pour réaliser la détection des zones TF mono-sources et l'identification des coefficients, resp. basées sur l'utilisation de la fonction de cohérence et le rapport de densités (auto-) et inter-spectrales de puissance (DSP/DISP). Cependant, malgré des formalismes initiaux très différents, LI-TIFCOHERE est théoriquement identique [83] à la version centrée de LI-TIFCORR si le calcul des DSP et DISP est basé sur l'utilisation du périodogramme moyenné, et à condition de choisir les mêmes valeurs de paramètres dans ces deux méthodes. Comme les algorithmes de ces méthodes ont été développés séparément, les types de paramètres ne sont pas rigoureusement les mêmes. Une étude permet de comparer les performances de cette méthode avec LI-TIFCORR-C, mais aussi avec LI-TIFFROM-S2. Les paramètres et signaux sont les mêmes que dans la section 3.1.2, sauf que nous avons fixé la taille des zones d'analyse entre 512 et 2048 points. Les $SIRI$ moyens globaux sont donnés dans le tableau 3.8. Globalement, les performances obtenues sont très proches pour toutes les méthodes. C'était prévisible, compte tenu du fait que LI-TEMPCOHERE et LI-TIFCORR-C sont quasi-identiques, aux modifications mineures près. Plus en détail, nous remarquons que, même si LI-TIFCOHERE donne un $SIRI$ moyen très légèrement supérieur à nos méthodes, son écart-type est légèrement supérieur. Ces différences peuvent s'expliquer par le fait que l'utilisation pour LI-TIFCOHERE d'une fenêtre de pondération différente lors de l'estimation des DSP et DISP et par des fluctua-

TAB. 3.8 – Performances globales des méthodes proposées.

Méthode	$SIRI$	Ec. type
LI-TIFROM-S2	64.5	6.6
LI-TIFCORR-C	64.6	6.7
LI-TIFCOHERE	64.7	7.8

Critères de performances : valeurs moyennes de $SIRI$ (en dB) et écart-type obtenus sur tous les couples de signaux de parole et toutes les combinaisons de paramètres.

tions statistiques dans les performances associées aux différentes segmentations du plan TF lors de l'implémentation des méthodes.

3.1.7 Comparaison aux méthodes de SAS classiques

Nous avons aussi comparé les performances de nos méthodes TF de SAS et de plusieurs approches de SAS accessibles dans le logiciel ICAlab [57], dans les mêmes conditions que les tests de la section 3.1.2. Les résultats ainsi obtenus sont montrés dans le tableau 3.9. Alors que les $SIRI^{out}$ obtenus par toutes les méthodes classiques sont comprises entre environ 0 et 40 dB, ils sont supérieurs à 60 dB pour nos méthodes TF (et même très proches de 70 dB pour leurs paramètres optimaux définis avant, après avoir moyenné les $SIRI^{out}$ sur tous les jeux de sources). Pour les signaux de parole, nos approches TF basées sur les zones d'analyse mono-sources sont beaucoup plus performantes que toutes les méthodes disponibles dans le logiciel ICAlab (incluant plusieurs approches basées sur la corrélation, et incluant SONS, qui a été explicitement développé pour les signaux non-stationnaires, d'où son nom "*second order non-stationary source separation*" en anglais).

3.2 Mélanges de $N = 4$ sources et $N = 4$ observations

En supplément des tests avec deux mélanges de deux sources présentés dans la section 3.1, nous décrivons ici le comportement des méthodes TF dans le cas d'un plus grand nombre de sources. En effet, plus le nombre de sources est important et plus il est difficile de trouver des zones TF mono-sources. Bien que nous nous attendons à avoir une baisse des performances, nous espérons tout de même que celles-ci resteront élevées. Nous avons utilisé un seul ensemble de quatre sources : les sources 1, 3, 4 et 6 présentées dans la section 3.1.1 et correspondant à des signaux de parole continue de deux voix masculines et deux voix féminines. Nous les avons mélangés avec une matrice 4×4 définie comme une extension de la matrice introduite dans l'Eq. (3.4), c.-à-d. :

$$A = \begin{bmatrix} 1 & \lambda & \lambda^2 & \lambda^3 \\ \lambda & 1 & \lambda & \lambda^2 \\ \lambda^2 & \lambda & 1 & \lambda \\ \lambda^3 & \lambda^2 & \lambda & 1 \end{bmatrix}. \quad (3.6)$$

TAB. 3.9 – (1) SIR^{out} (en dB) des méthodes LI-TIFFROM-S2, LI-TIFCORR-C et LI-TIFCORR-NC : (i) moyenne des SIR^{out} sur toutes les valeurs des paramètres ; (ii) SIR^{out} pour les paramètres optimaux. (2) SIR^{out} (en dB) des méthodes classiques.

Méthode de SAS		SIR^{out}
LI-TIFFROM-S2	Moy. des paramètres	61.2
	Paramètres opt.	68.6
LI-TIFCORR-C	Moy. des paramètres	64.0
	Paramètres opt.	69.4
LI-TIFCORR-NC	Moy. des paramètres	65.0
	Paramètres opt.	71.7
AMUSE		30.5
ECV2		31.5
EVD24		23.6
SOBI		31.5
SOBI-RO		35.7
SOBI-BPF		28.4
SONS		36.2
JADE-op		2.4
JADETD		24.1
FPICA	Tangente hyperbolique	39.5
	Gaussienne	41.0
	Cubique	41.9
	Cum. d'ordre 5	25.1
	Cum. d'ordre 6	28.4
PEARSON opt.		42.2
SANG		40.5
NG-FICA		35.7
ThinICA		39.0
ERICA		37.3
SIMBEC		38.8
UNICA		37.3
FOBI-E		16.6
SYM-WHITE		20.1

Ce tableau contient les performances globales pour les six jeux de sources.

TAB. 3.10 – Performances globales des méthodes proposées.

Nombre d'observations pour l'étape détection	Méthode	SIR^{out}	Ec. type
Deux observations	LI-TIFROM	-31.0	32.0
	LI-TIFROM-S2	-16.0	33.0
Toutes les observations (moyenne)	LI-TIFROM	37.0	13.6
	LI-TIFROM-S2	37.9	11.0

Critères de performances : valeurs moyennes de SIR^{out} (en dB) et écart-type obtenus sur toutes les combinaisons de paramètres.

Nous nous focalisons ici sur les résultats obtenus pour la valeur de λ correspondant aux signaux les plus mélangés, c.-à-d. $\lambda = 0.9$. Les paramètres des méthodes TF de SAS considérées ont varié de la même façon que précédemment.

Dans un premier temps, nous allons étudier l'apport de l'utilisation de la moyenne dans l'étape de détection, comme discuté dans la section 2.2 page 57.

Nous montrerons ensuite les performances des méthodes LI-TIFCORR-C, LI-TIFCORR-NC et de notre variante de LI-TIFROM, en faisant appel à la moyenne des paramètres dans l'étape de détection.

3.2.1 Apport du calcul de la moyenne dans l'étape de détection pour $N = 4$ sources

Dans cette section, nous montrons expérimentalement l'amélioration obtenue par nos méthodes en utilisant la moyenne de nos critères mono-sources dans l'étape de détection lorsque $N > 2$. Comme discuté dans la section 2.2, F. Abrard et Y. Deville n'utilisaient qu'un seul couple d'observations (2.17) pour détecter les zones mono-sources. Nous avons proposé au contraire d'utiliser toutes les observations disponibles en moyennant, comme indiqué dans l'Eq. (2.61), les rapports d'observations. Nous proposons donc de tester la méthode originale de LI-TIFROM dans ces deux conditions.

De plus, nous proposons de tester notre version LI-TIFROM-S2 en n'utilisant d'une part que les observations $X_1(t, \omega)$ et $X_2(t, \omega)$ dans l'étape de détection, c.-à-d. le critère (2.38), puis en utilisant d'autre part tous les rapports d'observations.

Dans le tableau 3.10, nous voyons que la moyenne permet d'augmenter considérablement les performances des méthodes. En effet, la méthode originale et notre version améliorée échouent resp. 28 et 21 fois sur 36 en n'utilisant que les observations $X_1(t, \omega)$ et $X_2(t, \omega)$. Au contraire, elles réussissent toujours quand tous les rapports sont étudiés : les SIR^{out} sont alors beaucoup plus élevés (resp. 37.0 et 37.9 dB) et les écart-types sont réduits (resp. 13.6 et 11.0 dB).

TAB. 3.11 – Performances des méthodes proposées pour $N = 4$ en fonction de la taille des fenêtres utilisées dans le calcul des TFCT (en nombre d'échantillons).

Méthode	Critère de perf.	Taille d des fenêtres				
		128	256	512	1024	tous
LI-TIFFROM-S2	SIR^{out} : Moy.	45.9	43.7	35.3	26.6	37.9
	Ec. type	3.4	5.4	7.5	13.0	11.0
	Min.	41.4	36.8	24.2	5.7	5.7
	Max.	50.7	50.8	46.0	45.4	50.8
LI-TIFCORR-C	SIR^{out} : Moy.	47.0	42.7	40.4	36.0	41.5
	Ec. type	2.2	5.7	3.2	6.1	6.0
	Min.	44.5	35.2	35.9	26.7	26.7
	Max.	51.3	49.8	45.2	42.6	51.3
LI-TIFCORR-NC	SIR^{out} : Moy.	48.1	45.1	41.4	38.2	43.2
	Ec. type	1.2	5.4	2.9	9.1	6.5
	Min.	46.8	36.7	37.2	26.8	26.8
	Max.	50.0	52.3	45.2	51.7	52.3

Critères de performances : les mêmes que dans le tableau 3.1.

3.2.2 Performances des méthodes TF

Nous décrivons maintenant le comportement des méthodes LI-TIFFROM-S2, LI-TIFCORR-C et LI-TIFCORR-NC dans le cas où le nombre de sources est supérieur à 2.

Les trois méthodes ont réussi à identifier toutes les colonnes de la matrice B définie dans l'Eq. (3.6) dans tous ces tests. Comme nous ne considérons qu'un seul jeu de sources et une seule matrice de mélange, tous les tests sont effectués avec le même SIR^{in} , qui est égal à -3.65 dB. Comme dans la section 3.1.2, la différence entre SIR^{out} et SIR^{in} est constante et nous ne considérons que le SIR^{out} ci-dessous. Ses valeurs moyennes pour les méthodes LI-TIFFROM-S2, LI-TIFCORR-C et LI-TIFCORR-NC sont resp. égales à 37.9, 41.5 et 43.2 dB.

Les tableaux 3.11-3.13 détaillent la dépendance des SIR^{out} par rapport à chaque paramètre des méthodes de SAS, en moyennant par rapport aux autres paramètres. Ces tableaux montrent que, ici encore, les méthodes proposées doivent être utilisées avec 10 (ou 8 pour LI-TIFFROM-S2) fenêtres TF par zone d'analyse, avec 90% (ou 75% pour LI-TIFCORR-C) de recouvrement. Concernant la taille d des fenêtres TF :

- si nous considérons toutes les configurations de test visibles dans les tableaux 3.11-3.13, baisser la taille des fenêtres dans le calcul des TFCT à 128 échantillons donne de meilleures performances pour le SIR^{out} moyen et pour son écart-type, alors que les performances globales étaient plutôt meilleures pour les fenêtres de 256 échantillons dans les tests avec deux sources dans la section 3.1.2.
- Mais les performances sont encore ici meilleures avec 256 échantillons qu'avec 128 si nous nous focalisons sur le cas où les valeurs optimales définies ci-dessus sont utilisées pour les autres paramètres des méthodes de SAS, c.-à-d. 10 TFCT (8 pour

TAB. 3.12 – Performances des méthodes TF de SAS en fonction du nombre de fenêtres TF dans les zones d’analyse.

Méthode	Critère de perf.	Nombre M de fenêtres		
		8	10	12
LI-TIFFROM-S2	SIR^{out} : Moy.	40.6	37.0	36.2
	Ec. type	9.1	11.4	12.6
	Min.	17.8	10.0	5.7
	Max.	50.7	50.8	48.8
LI-TIFCORR-C	SIR^{out}	41.3	42.6	40.6
	Ec. type	7.0	5.2	5.8
	Min.	26.6	31.0	27.9
	Max.	51.3	49.8	47.9
LI-TIFCORR-NC	SIR^{out}	43.0	44.3	42.3
	Ec. type	6.6	6.6	6.7
	Min.	26.8	29.8	28.3
	Max.	50.0	52.3	49.4

Critères de performances : les mêmes que ceux du tableau 3.1.

TAB. 3.13 – Performances des méthodes TF de SAS par rapport au recouvrement temporel des fenêtres TF.

Méthode	Critère de perf.	Recouvrement ρ		
		50%	75%	90%
LI-TIFFROM-S2	SIR^{out} : Moy.	30.4	38.3	45.0
	Ec. type	13.9	6.5	5.7
	Min.	5.7	25.9	35.4
	Max.	50.2	46.7	50.8
LI-TIFCORR-C	SIR^{out} : Moy.	38.3	43.3	42.9
	Ec. type	7.5	5.0	3.7
	Min.	26.6	35.1	35.9
	Max.	51.3	49.8	48.1
LI-TIFCORR-NC	SIR^{out} : Moy.	39.0	43.7	46.9
	Ec. type	7.5	4.6	4.8
	Min.	26.8	36.1	37.2
	Max.	47.7	49.3	52.3

Critères de performances : les même que ceux du tableau 3.1.

LI-TIFFROM-S2) avec 90% (75% pour LI-TIFCORR-C) de recouvrement dans ces zones. Plus précisément, les SIR^{out} resp. donnés par LI-TIFCORR-C, LI-TIFCORR-NC et LI-TIFFROM-S2 sont 47.7, 46.9 et 50.7 dB pour des fenêtres de 128 échantillons, 49.8, 49.3 et 49.1 dB pour des fenêtres de 256 échantillons. Notez que LI-TIFCORR-C présente des performances légèrement meilleures que LI-TIFCORR-NC dans ces conditions particulières alors que suivant la valeur de la taille d des TFCT, LI-TIFFROM-S2 présente les meilleures ou moins bonnes performances des trois méthodes dans ces configurations.

Ces résultats sont alors cohérents avec ceux obtenus dans la section 3.1.2 pour les mélanges de deux sources. La possibilité d'obtenir des performances légèrement meilleures avec des fenêtres TF plus courtes (c.-à-d. 128 échantillons) quand quatre sources sont mélangées peut être expliquée ainsi. Quand les signaux observés sont des mélanges d'un grand nombre de sources arbitraires, les zones TF où une source est isolée tendent à devenir plus petites, parce que plus de sources se recouvrent dans les observations considérées. Des fenêtres TF plus courtes peuvent alors être utilisées de façon à ne pas rater ces petites zones mono-sources dans l'étape de détection de nos approches mono-sources.

3.3 Performances des méthodes étendues

3.3.1 Cas sous-déterminé

Nous illustrons ici les performances de nos méthodes étendues aux mélanges sous-déterminés. Ces extensions ont été décrites dans la section 2.1.1 pour la version initiale de LI-TIFFROM mais peuvent être appliquées à toutes nos méthodes TF de SAS, comme nous l'avons vu précédemment. Nous avons donc effectué des tests avec trois sources de parole parmi celles que nous avons vues dans la section 3.1.1, c.-à-d. les signaux courts (1,2,3), (1,3,5), (1,4,7), (1,5,7) et (5,6,7), que nous avons mélangés grâce à la matrice

$$A_6 = \begin{bmatrix} 1 & 0.9 & 0.7 \\ 0.8 & 1 & 1 \end{bmatrix}. \quad (3.7)$$

Nous avons fait varier les paramètres d , M et ρ de la même façon que dans la section 3.1.2. Le tableau 3.14 contient les résultats globaux ainsi obtenus, et leurs variations en fonction du nombre d d'échantillons dans chaque TFCT. Puisque le mélange considéré est sous-déterminé, les sources ne peuvent être extraites exactement par nos méthodes mais ces approches réussissent à estimer la matrice atténuée / permutée B , comme expliqué précédemment. Le paramètre utilisé ici pour mesurer les performances se réfère donc à la qualité de l'estimation de B et est constitué de la norme de Frobenius²⁵ de la différence entre la matrice théorique B et son estimée \hat{B} . Le tableau 3.14 montre la valeur moyenne, l'écart-type, les valeurs maximale et minimale de cette norme de Frobenius sur les divers jeux de sources explicités ci-dessus, en fonction de la taille d des TFCT. Ici encore, les performances varient suivant la méthode choisie. Nous remarquons que notre variante de LI-TIFFROM, LI-TIFCORR-C et LI-TIFCORR-NC estiment aussi bien la matrice B

²⁵La norme de Frobenius d'une matrice M est définie comme la somme des modules au carré des éléments de M .

TAB. 3.14 – Performances des méthodes TF de SAS en fonction de la taille des fenêtres TF utilisées dans le calcul des TFCT.

Méthode	Critère de perf.	taille d des fenêtres				
		128	256	512	1024	tous
LI-TIFFROM-S2	$\ B - \hat{B}\ _F$: Moy.	0.0018	0.0029	0.0072	0.0447	0.0141
	Ec. type	0.0012	0.0023	0.0126	0.0906	0.0487
	Min.	0.0002	0.0004	0.0003	0.0005	0.0002
	Max.	0.0050	0.0088	0.0672	0.4674	0.4674
LI-TIFCORR-C	$\ B - \hat{B}\ _F$: Moy.	0.0019	0.0022	0.0039	0.0062	0.0036
	Ec. type	0.0014	0.0012	0.0027	0.0037	0.0037
	Min.	0.0002	0.0003	0.0002	0.0005	0.0002
	Max.	0.0054	0.0046	0.0134	0.0273	0.0273
LI-TIFCORR-NC	$\ B - \hat{B}\ _F$: Moy.	0.0017	0.0024	0.0038	0.0063	0.0036
	Ec. type	0.0013	0.0015	0.0037	0.0093	0.0053
	Min.	0.0003	0.0005	0.0002	0.0005	0.0002
	Max.	0.0052	0.0080	0.0144	0.0527	0.0527

Critères de performances : valeur moyenne (sur les cinq jeux de sources), écart-type, minimum et maximum de la norme de Frobenius de la différence entre la matrice théorique B et son estimée \hat{B} en fonction des valeurs de d .

puisque les performances sont similaires pour $d \leq 256$: les valeurs moyennes, écart-types et valeurs maximales sont bien plus faibles que les éléments de la matrice B , ce qui signifie que l'estimation est bonne. Par contre, quand $d = 512$ et surtout $d = 1024$, suivant les cas, la méthode LI-TIFFROM-S2 estime mal la matrice de mélange alors que les performances des approches basées sur la corrélation sont généralement excellentes. Cela se traduit par une distance moyenne entre matrice théorique et matrice estimée environ 2 à 7 fois plus importante avec LI-TIFFROM-S2 qu'avec les méthodes basées sur la corrélation. De même, la variance est alors 4 à 10 fois plus importante. Notons toutefois que même si les versions centrées et non-centrées de LI-TIFCORR offrent de meilleures performances, les valeurs maximales de la norme de Frobenius sont tout de même relativement importantes (la version centrée semble d'ailleurs un peu plus performante avec un écart-type et une valeur maximale plus faibles que la version non-centrée). Ces résultats sont en accord avec ceux obtenus dans la section 3.1.2 : ils montrent que les méthodes basées sur la corrélation semblent moins sensibles à la taille des TFCT que l'approche LI-TIFFROM-S2.

TAB. 3.15 – Performances de la méthode LI-TIFCORR-C pour chaque jeu de sources et performances globales pour les six jeux.

Méthode	Critère de perf.	Jeu de sources					
		1	2	3	4	5	6
LI-TIFCORR-C	Norme : Moy.	5.21 e-5	4.61 e-5	14.15 e-5	9.98 e-5	9.66 e-5	5.20 e-5
	Ec. type	3.12 e-5	3.71 e-5	11.72 e-5	9.13 e-5	7.63 e-5	2.86 e-5
	Min.	0.65 e-5	0.42 e-5	2.01 e-5	0.76 e-5	1.12 e-5	0.99 e-5
	Max.	13.03 e-5	12.88 e-5	45.07 e-5	37.37 e-5	33.58 e-5	13.75 e-5

Critères de performances : valeur moyenne (sur toutes les valeurs de paramètres de la méthode de SAS), écart-type, minimum et maximum de la norme de la différence entre les valeurs théorique et estimée de la seule colonne identifiable dans la première étape de la récursion.

TAB. 3.16 – Performances de la méthode LI-TIFCORR-C pour chaque jeu de sources et performances globales pour les six jeux.

Méthode	Critère de perf.	Jeu de sources					
		1	2	3	4	5	6
LI-TIFCORR-C	Norme : Moy.	0.0094	0.0032	0.1069	0.0029	0.0041	0.0012
	Ec. type	0.0182	0.0036	0.5984	0.0035	0.0031	0.0008
	Min.	0.0007	0.0004	0.0006	0.0001	0.0008	0.0002
	Max.	0.0866	0.0181	3.5973	0.0178	0.0125	0.0033

Critères de performances : valeur moyenne (sur toutes les valeurs de paramètres de la méthode de SAS), écart-type, minimum et maximum de la norme de Frobenius de la différence entre les valeurs théorique et estimée de la sous-matrice de mélange 2×2 dans la deuxième étape de récursion.

3.3.2 Sources cachées

Nous présentons ici la validation expérimentale des approches récursives développées dans le cas de source(s) invisible(s) que nous avons détaillées dans la section 2.1.1. L'approche LI-TIFFROM ayant déjà été testée dans [2] et les versions centrées et non-centrées de TIFCORR ayant des paramètres très proches, nous n'avons testé que la méthode LI-TIFCORR-C avec six jeux de sources. Chacun de ces ensembles contenait trois sources qui étaient mélangées selon la matrice

$$A_7 = \begin{bmatrix} 1 & 0.9 & 0.9^2 \\ 0.9^3 & 1 & 0.9^3 \\ 0.9^2 & 0.9 & 1 \end{bmatrix}. \quad (3.8)$$

Chacun des jeux de sources considérés consistait en deux signaux de parole correspondant à l'un des six jeux de sources définis dans la section 3.1.1, et d'un bruit gaussien artificiel indépendant et identiquement distribué (iid). Nous nous sommes assurés que la première zone mono-source détectée correspond à ce dernier signal parce qu'il est présent dans tout le plan TF et rend donc les deux autres signaux de parole invisibles. Au contraire, il existe des zones dans le plan TF où les contributions des deux signaux de parole sont négligeables par rapport à celle du bruit. Ce signal est donc visible grâce à ces zones. La méthode récursive LI-TIFCORR-C contient alors deux étapes :

1. Nous identifions la seule colonne identifiable de la matrice de mélange B . Cette colonne correspond à la source de bruit. Nous en déduisons ensuite deux signaux de mélange modifiés comme expliqué dans l'Eq. (2.25). Ces signaux ne contiennent que les deux signaux de parole.
2. Nous appliquons ensuite la méthode standard LI-TIFCORR-C à ces deux mélanges de deux sources, où ces deux signaux de parole sont maintenant visibles, comme dans la section 3.1. Nous identifions alors la sous-matrice de mélange 2×2 correspondante et nous séparons ces deux signaux de parole²⁶.

Dans ces tests, nous avons fait varier de la même façon que dans la section 3.1.2 le nombre d d'échantillons dans chaque fenêtre temporelle, le nombre M de ces fenêtres par zone d'analyse et le recouvrement temporel ρ entre ces fenêtres. Les résultats ainsi obtenus pour chaque jeu de sources et les performances globales pour tous les jeux sont donnés dans les tableaux 3.15 et 3.16 qui correspondent resp. aux deux étapes ci-dessus de la méthode de SAS récursive proposée. Nous mesurons ici les performances obtenues à chaque étape de la même façon que pour les tests précédents : nous calculons la norme de la différence entre les valeurs théorique et estimée du vecteur ou de la matrice de mélange qui est identifié(e) dans l'étape considérée dans la récursion. Les tableaux 3.15 et 3.16 montrent que ces normes sont presque toujours²⁷ plus petites que les éléments

²⁶Nous pouvons alors extraire aussi le signal de bruit iid en recombinaison des deux signaux de parole extraits avec un mélange initial des trois sources, en utilisant une approche similaire à l'Eq. (2.25).

²⁷Des performances plus faibles sont obtenues avec le jeu de sources numéro 3 mais, encore une fois, cela correspond à une situation plus difficile que dans des applications réelles puisque : (i) les deux signaux sources considérés correspondent à la même personne et peuvent donc avoir un fort recouvrement dans le domaine TF et (ii) dans le but d'illustrer clairement les performances de notre méthode récursive,

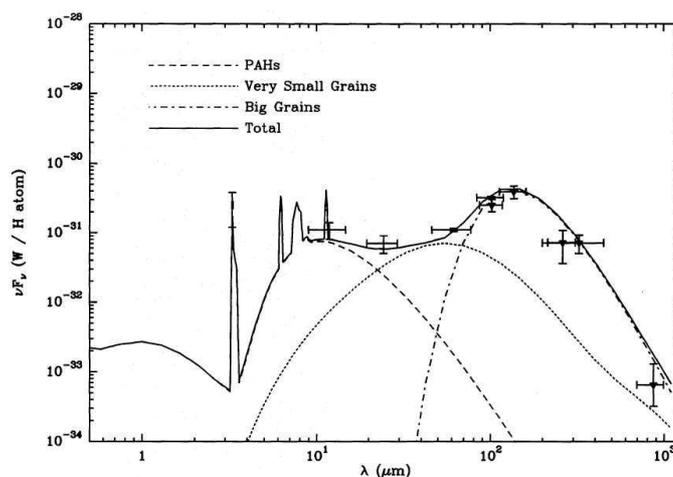


FIG. 3.2 – Spectre d’émission des différentes composantes de poussière du MIS d’après Désert et al. [88].

des parties correspondantes de la matrice de mélange. Cela démontre la capacité de la méthode récursive LI-TIFCORR-C à identifier précisément la matrice de mélange, et donc à extraire les sources dans cette configuration où deux sources étaient invisibles avant qu’elles ne soient révélées par la première étape de notre récursion.

3.4 Tests complémentaires : application à l’astrophysique

Comme de nombreuses méthodes parcimonieuses de SAS, les méthodes présentées dans ce document sont bien adaptées aux mélanges de signaux de parole et de musique, ce qui était l’une des applications majeures envisagées lors du démarrage de cette thèse au Laboratoire d’Acoustique, Métrologie et Instrumentation de Toulouse. Comme la deuxième période de ces travaux a été réalisée au Laboratoire d’Astrophysique de Toulouse–Tarbes, nous avons aussi testé nos méthodes sur les données astrophysiques que voici.

Les régions de notre galaxie situées entre les étoiles sont regroupées sous l’appellation de milieu interstellaire (MIS). Bien que ce milieu soit généralement très dilué, on y trouve de la matière qui a été enrichie par les éjections des étoiles en fin de vie (géantes rouges ou supernovæ). Cette matière n’est pas répartie uniformément dans le MIS mais concentrée dans des nuages de gaz et de petits grains de poussière. Ces nuages présentent la particularité d’absorber la lumière émise par les étoiles dans l’ultraviolet (UV) pour en

nous avons intentionnellement mélangé les deux signaux de parole à une source artificielle qui couvre *tout* le domaine TF, alors que dans la plupart des applications de parole, chaque source est uniquement présente dans une partie du domaine TF, de telle façon que la première étape de la récursion utilisée pour supprimer la source visible apporte ensuite de plus petits résidus de cette source qu’ici dans les signaux de mélanges modifiés résultants qui sont utilisés dans la deuxième étape de la récursion.

Notez au contraire que le jeu 6 apporte des performances globales plus hautes. Cela confirme la capacité de cette méthode de SAS à tirer parti des situations réalistes où quelques sources contiennent des silences.

réémettre une partie dans l'infrarouge (IR). L'observation de cette réémission dans l'IR pousse à penser [88] à la cohabitation de plusieurs composés chimiques appelés "grains" (voir figure 3.2) :

- les hydrocarbures polycycliques aromatiques (PAH pour *Polycyclic Aromatic Hydrocarbons* en anglais) : ce sont des molécules carbonées de grande taille. Leur signature IR serait située dans une bande comprise entre 3.3 et 17 μm [147]. Les PAH existent sous forme neutre, notée PAH⁰, ou ionisée, notée PAH⁺.
- Les très petits grains (VSG pour *Very Small Grains* en anglais), dont la taille varie de 1 à quelques nanomètres, ont une signature IR comprise entre 20 et 70 μm [88].
- Les gros grains dominent l'émission de poussière dans les plus grandes longueurs d'onde de l'IR [88].

De nombreux travaux, dont [34], proposent d'étudier la présence de tels composés chimiques dans les spectres IR des régions de photodissociation (PDR pour *Photo-Dissociation Regions* en anglais) qui sont des zones du MIS baignées par un flux UV suffisamment dense pour dissocier les molécules. En pratique, les données issues des observations d'une PDR sont représentées sous forme de cube : deux dimensions de ce cube correspondent aux coordonnées spatiales (pixels) de la PDR étudiée, la troisième coordonnée correspond aux longueurs d'onde du spectre IR en ce pixel (voir figure 3.3).

Pour un pixel fixé de la PDR, le spectre IR peut s'écrire sous certaines conditions comme une combinaison linéaire des spectres des composés chimiques présents. En considérant les spectres situés dans plusieurs pixels différents, on observe alors plusieurs "mélanges" des spectres des différents composés chimiques. Séparer les réponses IR des composés chimiques dans les PDR correspond donc simplement à réaliser la SAS LI des spectres IR observés pour divers pixels du cube. En notant $x_i(\lambda)$ un spectre observé issu du cube de données et $s_j(\lambda)$ le spectre d'un composé chimique présent dans la région de la PDR étudiée, nous voyons alors que les "observations" et les "sources" sont des fonctions de la longueur d'onde λ liées par la relation suivante :

$$x_i(\lambda) = \sum_{j=1}^N a_{ij} s_j(\lambda) \quad i = 1, \dots, P. \quad (3.9)$$

Nous retrouvons donc la définition (1), sauf que la variable temporelle t a été remplacée par la longueur d'onde λ .

Dans [34], O. Berné et al. appliquent les méthodes FastICA [125] et la FMN de Lee et Seung [143] sur un objet observé par le télescope Spitzer²⁸. Cet objet est la nébuleuse Ced 201 qui présente en son centre une étoile excitatrice entourée d'une "coquille" où se produit la photodissociation. Dans cet objet, le champ UV est relativement faible, et on s'attend à n'observer que des spectres de PAH et de VSG, c.-à-d. deux "sources" au sens de la SAS. Le cube observé est de taille 39 × 29 × 331 mais O. Berné et al. ne conservent que 370 (sur 1131) pixels jugés "intéressants", c.-à-d. des pixels pour lesquels les spectres ont une intensité au-dessus d'un seuil fixé par l'utilisateur pour les longueurs d'onde 9.3 et 20.6 μm . Avant d'appliquer les méthodes de SAS, chaque spectre est normalisé²⁹ (les signaux

²⁸<http://www.spitzer.caltech.edu/>

²⁹Contrairement à la normalisation proposée pour les sources de parole dans les sections précédentes, la normalisation est ici faite par rapport à l'énergie du spectre considéré.

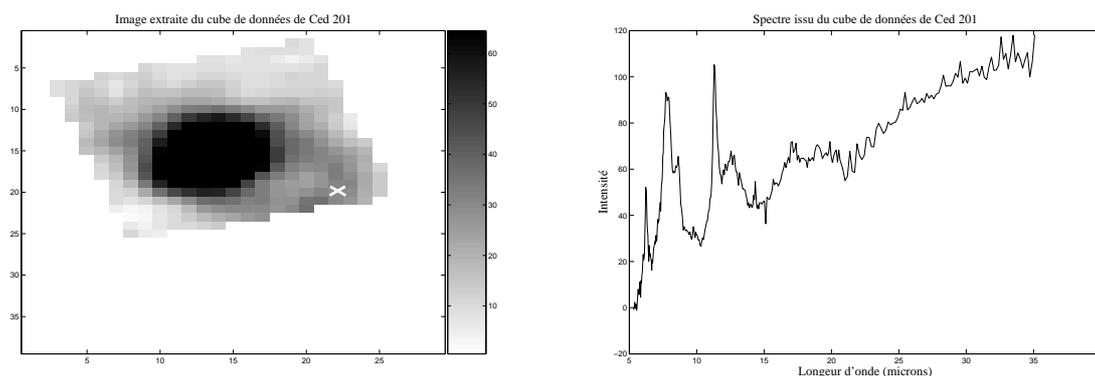


FIG. 3.3 – *A gauche* : image à $7 \mu\text{m}$ extraite du cube de données de la nébuleuse Ced 201. *A droite* : spectre extrait de ce même cube sur le pixel marqué d'une croix dans l'image de gauche.

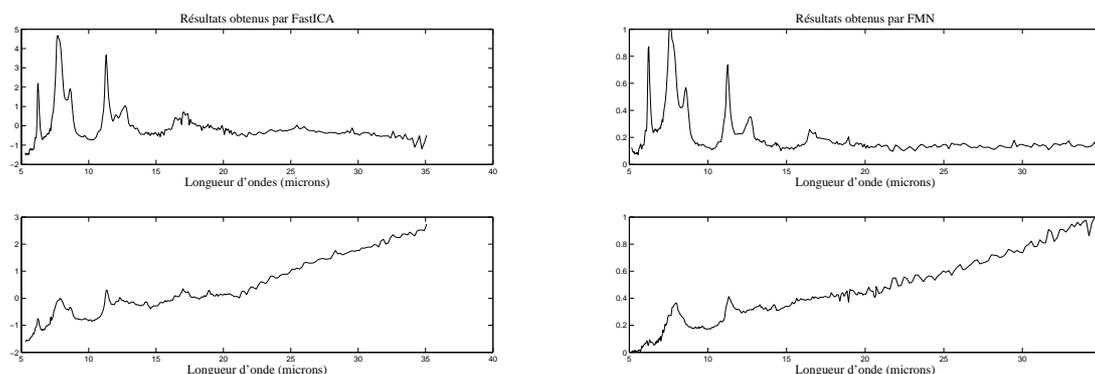


FIG. 3.4 – *A gauche* : spectres obtenus avec FastICA pour Ced 201. *A droite* : spectres obtenus par FMN pour Ced 201. Avec les deux méthodes, on reconnaît en haut le spectre des PAH et en bas celui des VSG.

traités par FastICA sont d'abord centrés) et une analyse en composantes principales (ACP) est réalisée sur ces 370 spectres pour n'en obtenir au final que deux. Les résultats obtenus avec FastICA ou par FMN appliquées à ces deux spectres par O. Berné et al. sont donnés dans la figure 3.4. Pour les deux méthodes, les courbes du haut représentent des spectres de PAH. Nous remarquons notamment les raies correspondant aux longueurs d'onde de résonance des molécules de PAH. Les courbes du bas correspondent aux spectres de VSG : elles sont caractérisées par un *continuum* dans les hautes longueurs d'onde (les grains apportent un effet de "lissage" des propriétés de résonance). Remarquons que la FMN apporte une meilleure qualité de séparation, notamment pour les VSG puisque des résidus du spectre de PAH sont encore présents sur le signal correspondant obtenu par ACI.

Nous désirons séparer les sources à partir des deux spectres obtenus après centrage,

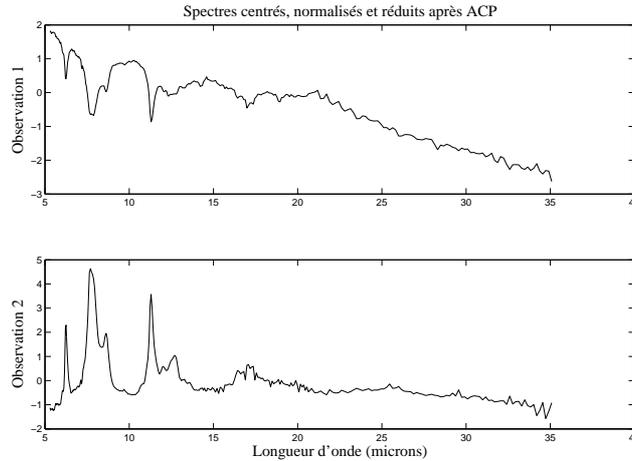


FIG. 3.5 – Spectres obtenus après centrage, normalisation et ACP. Ces spectres correspondent aux observations traitées par nos méthodes de SAS.

normalisation et ACP (voir figure 3.5). Comme nous ne sommes pas sûrs de pouvoir estimer *complètement* la matrice de mélange, nous avons annulé les contributions des sources dans la deuxième observation, par combinaison linéaire des observations selon l'Eq. (2.4). En effet, dans le cas particulier où $N = 2$ sources, nous pouvons facilement voir, à partir des Eq.(2.4) et (2.5), que les signaux extraits sont égaux aux sources, aux erreurs d'estimation et aux indéterminations classiques près.

Notons que :

1. les mélanges proviennent de données réelles et les sources sont inconnues. Nous pouvons alors uniquement juger les spectres estimés de façon qualitative, en les comparant avec ceux obtenus par O. Berné et al. mais ne pouvons pas estimer de SIR.
2. Contrairement aux sources de parole, les spectres à séparer sont peu parcimonieux. En conséquence, nous nous attendons à avoir du mal à trouver des zones mono-sources. De plus, chaque spectre n'a que 331 échantillons, ce qui risque de compliquer la recherche de zones mono-sources : utiliser de "grandes" zones d'analyse risque de rendre les sources invisibles. Au contraire, la pertinence de nos critères mono-sources utilisés sur un petit nombre de points est à discuter.
3. Les spectres sont souvent non-nuls. Cela signifie que nos méthodes temporelles non-centrées, c.-à-d. LI-TEPCORR-NC et LI-TEMPROM-S2, ne devraient pas pouvoir séparer les spectres.

En conséquence, nous ne nous attendons pas à avoir de meilleures performances que les méthodes utilisées par O. Berné et al. mais nous voulons montrer que, même si les conditions ne sont pas favorables à nos méthodes, ces dernières peuvent quand même séparer certaines sources.

Un grand nombre de tests, avec les méthodes temporelles et TF, a été réalisé. Dans la majorité des cas, nous n'avons estimé au plus qu'une seule source. En général, la

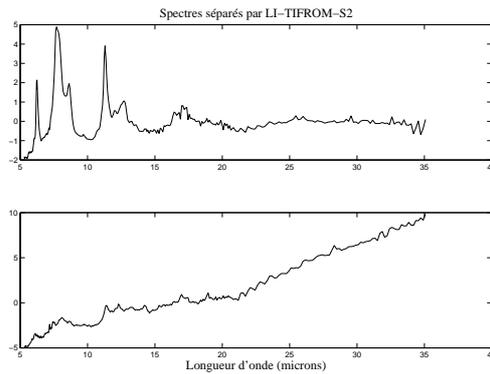


FIG. 3.6 – Spectres obtenus avec LI-TIFROM-S2

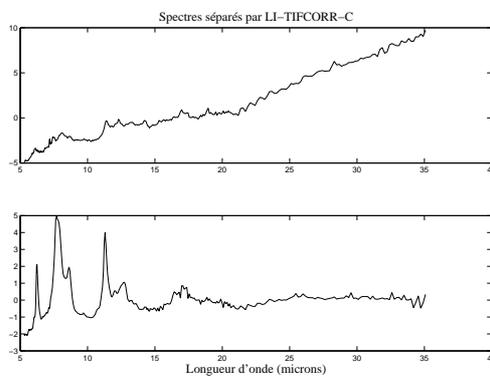


FIG. 3.7 – Spectres obtenus avec LI-TIFCORR-C.

seconde source extraite correspondait à une estimée grossière de la première. Cependant, pour certaines valeurs particulières de paramètres, nos méthodes TF ont toutes réussi à estimer la matrice de mélange et reconstruire les sources. Cela montre à nouveau que nos méthodes TF tirent parti de la parcimonie de la TFCT et sont plus performantes que leurs variantes temporelles. Pour obtenir ces signaux de sortie, nous avons fixé les paramètres suivants.

1. Pour LI-TIFROM-S2, nous avons utilisé 66 échantillons pour le calcul des TFCT. Le recouvrement ρ entre deux fenêtres TF adjacentes a été fixé à 90%. Enfin, il y avait $M = 10$ fenêtres par zones d'analyse.
2. Pour LI-TIFCORR-C, nous avons utilisé 40 échantillons pour le calcul des TFCT. Le recouvrement a été fixé à $\rho = 90\%$ et on a testé des zones d'analyse de $M = 6$ fenêtres TF.
3. Pour LI-TIFCORR-NC, nous avons utilisé 16 échantillons pour le calcul des TFCT. Le recouvrement a été fixé à $\rho = 90\%$. Enfin, il y avait $M = 10$ fenêtres par zones d'analyse.

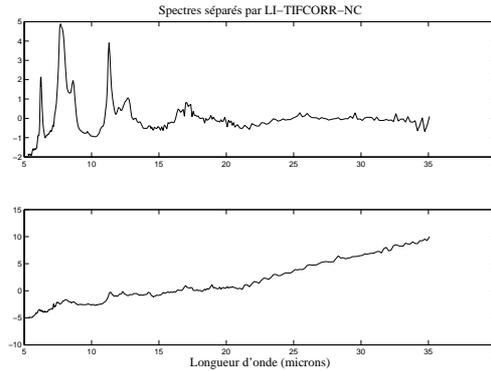


FIG. 3.8 – Spectres obtenus avec LI-TIFCORR-NC.

Les signaux estimés sont présentés dans les figures 3.6-3.8. Remarquons qu'ils ressemblent bien aux spectres théoriques de PAH et VSG. En particulier, le spectre estimé des VSG est plus "propre" que celui obtenu avec FastICA (les résidus de spectres de PAH ne sont pas présents avec nos méthodes TF).

En guise de comparaison, nous présentons les signaux extraits avec les méthodes temporelles dans la figure 3.9. Pour ces tests, les zones d'analyse contenaient 12 échantillons et étaient recouvrantes de 90%. Les spectres obtenus avec LI-TEMPCORR-C sont deux spectres de VSG, l'une des deux colonnes de mélange ayant été estimée de façon plus précise que l'autre, comme nous l'avons expliqué ci-dessus. Les autres résultats obtenus sont intéressants : le deuxième spectre extrait par les méthodes LI-TEMPCORR-NC et LI-TEMPROM-S2 est un spectre de PAH. Le premier ressemble à celui des PAH mais il est suivi d'un continuum négatif, qui n'a aucun sens physique. Il faut bien garder en tête que nos méthodes temporelles non-centrées, c.-à-d. LI-TEMPROM-S2 et LI-TEMPCORR-NC, ne sont théoriquement pas adaptées à ces signaux. En effet, supposons qu'une source soit constante et que l'autre source varie sur une zone d'analyse. Alors cette zone sera détectée comme une zone mono-source mais les critères d'identification non-centrés de ces méthodes engendreront une mauvaise estimation des paramètres de mélange. C'est la situation que l'on rencontre sur ces spectres dans la figure 3.9. Les signaux séparés par LI-TEMPCORR-C correspondent à une estimée grossière et fine du spectre des VSG.

Pour conclure, notons bien que les résultats obtenus par nos méthodes TF ont été réalisés pour des paramètres particuliers. Néanmoins, alors que les données n'étaient pas vraiment adaptées à ces méthodes, elles arrivent souvent à extraire une source. Cela laisse à penser que si, dans le futur, le nombre d'échantillons disponibles sur ces spectres augmente de manière significative, nos méthodes à parcimonie pourront concurrencer les approches de FMN (ou les méthodes statistiques classiques). Dans la suite de ce manuscrit, nous n'appliquerons nos méthodes qu'à des signaux de parole.

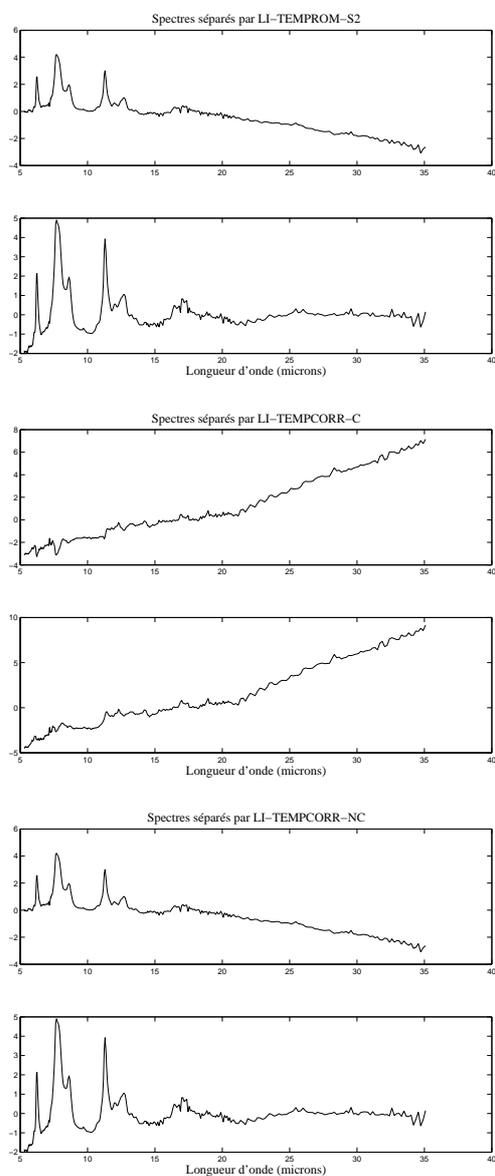


FIG. 3.9 – *En haut* : spectres obtenus avec LI-TEMPROM-S2. *Au milieu* : spectres obtenus avec LI-TEPCORR-C. *En bas* : spectres obtenus avec LI-TEPCORR-NC.

3.5 Conclusion sur les approches LI de SAS

Dans ce chapitre, nous avons présenté les performances des méthodes temporelles et TF décrites dans le chapitre précédent. Nous avons montré que les performances obtenues sont excellentes pour les sources de parole.

- Dans le cas où $N = 2$, les SIR^{out} moyens sont généralement supérieurs à 60 dB (et même proches de 70 dB quand les paramètres sélectionnés sont optimaux). Ces valeurs sont bien supérieures à celles obtenues par les méthodes classiques disponibles dans le logiciel ICALAB (entre 0 et 40 dB).
- Quand $N = 4$, les performances obtenues approchent 50 dB alors que le nombre de zones d'analyse mono-sources est réduit par rapport au cas $N = 2$.
- Une fois les "bons" paramètres réglés, ces méthodes montrent d'excellentes capacités d'estimation de la matrice de mélange dans le cas de mélanges sous-déterminés et/ou de sources cachées.

De plus, les performances atteintes lorsque les signaux sont des spectres contenant peu de points sont correctes alors que la nature de ces signaux ne favorise pas nos approches.

Cependant, ces méthodes souffrent d'un nombre important de paramètres heuristiques (même si certains de ces paramètres ont relativement peu d'influence sur les performances générales des méthodes). En particulier, le seuil qui indique si une nouvelle colonne estimée n'a pas déjà été trouvée est très handicapant et occasionne de fortes baisses de performances dans certains cas. Une version automatisée serait préférable. Cet aspect sera discuté dans la suite de ce manuscrit.

Enfin, même si les performances obtenues sont excellentes, la nature des mélanges limite naturellement le champ d'application de ces approches, notamment dans le domaine audio. Des versions étendues à des mélanges plus réalistes seront proposées dans la partie suivante.

Deuxième partie

Méthodes convolutives : du mélange à
atténuations et décalages au modèle
convolutif général

Introduction

Nous avons présenté dans la première partie de ce document des méthodes LI de SAS, originales, basées sur la parcimonie et qui présentaient d'excellentes performances pour des signaux de parole. Cependant, la nature des mélanges ne les rend pas applicables à des signaux audio réels (mis à part dans de rares situations comme les enregistrements en studio où les mélanges sont réalisés sur table de mixage).

Nous présentons donc ici plusieurs extensions de ces méthodes à des mélanges plus réalistes. Dans le chapitre 4, nous allons présenter les premières extensions de ces méthodes à des mélanges convolutifs simples : les mélanges à atténuations et décalages temporels (AD)³⁰. Ces mélanges correspondent physiquement à des enregistrements en chambre anéchoïque. Dans la littérature, on considère qu'ils modélisent de façon correcte les mélanges réels dans un environnement peu réverbérant.

Nous avons aussi vu dans la première partie de ce manuscrit les limitations de l'utilisation de paramètres heuristiques. Nous allons donc voir comment améliorer les méthodes LI précédentes et leurs extensions à l'aide d'algorithmes de classification non-supervisée dans le chapitre 5.

Les performances de ces méthodes AD seront décrites dans le chapitre 6.

Enfin, nous étendrons ces méthodes au cas convolutif général dans le chapitre 7.

³⁰Dans la suite de ce document, par abus de langage, nous parlerons de "mélanges AD" au lieu de "mélanges à AD". De façon générale, suivant le contexte, "AD" correspondra aux termes "atténuations et décalages" ou aux adjectifs "atténué(es) et décalé(es)".

Chapitre 4

Extension de LI-TIFCORR aux sources atténuées et retardées

Dans ce chapitre, nous présentons les premières extensions des méthodes LI-TIFCORR-C et LI-TIFCORR-NC à des mélanges convolutifs simples : les mélanges à atténuations et décalages temporels (AD). Pour alléger la présentation, nous nous limitons ici à l'extension de LI-TIFCORR-NC mais les modifications apportées peuvent être appliquées à la version centrée sans aucune difficulté. En conséquence, dans la suite de ce manuscrit, nous oublions la terminaison "-NC". Ainsi, notre méthode AD-TIFCORR est applicable aux mêmes mélanges que DUET sans avoir la contrainte de la WDO (mais AD-TIFCORR n'est pas, à cette étape, applicable aux mélanges sous-déterminés). Des extensions similaires de LI-TIFFROM aux mélanges AD sont présentées dans l'annexe D. J'ai en outre appliqué ce principe à une autre méthode de SAS, en collaboration avec B. Albouy, dans le cadre d'un contrat préalable à ma période de thèse. Nous avons ainsi développé une extension AD de la méthode LI-TIFCOHERE proposée par B. Albouy. La méthode AD-TIFCOHERE ainsi obtenue a été décrite dans la thèse de B. Albouy [5].

Dans la section 4.1, nous poserons le problème de la séparation de mélanges AD. Ensuite, une première extension de LI-TIFCORR à ce type de mélanges sera proposée dans la section 4.2. Comme cette extension utilise des zones d'analyse à fréquence constante (*Constant Frequency* en anglais) pour estimer certains paramètres, nous appellerons cette version AD-TIFCORR-CF. Au contraire, une autre extension n'utilisant que les zones d'analyse à temps constant (*Constant Time* en anglais) sera nommée AD-TIFCORR-CT dans la section 4.3.

4.1 Position du problème

Dans cette partie, nous considérons la configuration de N sources et $P = N$ observations. Dans le cas LI que nous avons vu précédemment, les observations $x_i(t)$ s'écrivaient en fonction des sources $s_j(t)$ selon la relation (2.18). Pour des raisons d'unité avec la suite

de ce chapitre, nous discrétisons ici cette relation :

$$x_i(n) = \sum_{j=1}^N a_{ij} s_j(n) \quad i = 1, \dots, N. \quad (4.1)$$

où les coefficients d'échelle a_{ij} sont constants et représentent l'atténuation durant la propagation de la source j à l'observation i .

Dans ce chapitre, nous considérons des mélanges plus généraux, qui tiennent aussi compte du temps de propagation entre les sources et les observations, au moyen de décalages temporels entiers n_{ij} . Le modèle étendu aux mélanges AD qui en résulte s'écrit :

$$x_i(n) = \sum_{j=1}^N a_{ij} s_j(n - n_{ij}) \quad i = 1 \dots N. \quad (4.2)$$

Dans un but de simplicité, nous supposons que les coefficients a_{ij} sont réels et strictement positifs (cela correspond à la propagation directe, c.-à-d. sans réflexion – une extension aux coefficients potentiellement négatifs est discutée dans l'annexe E) et que les sources sont à valeurs réelles. La transformée de Fourier de l'Eq. (4.2) s'écrit :

$$X_i(\omega) = \sum_{j=1}^N a_{ij} e^{-j\omega n_{ij}} S_j(\omega) \quad i = 1 \dots N. \quad (4.3)$$

Nous obtenons alors sous forme matricielle

$$\underline{X}(\omega) = A(\omega) \underline{S}(\omega) \quad (4.4)$$

avec $\underline{S}(\omega) = [S_1(\omega) \dots S_N(\omega)]^T$, $\underline{X}(\omega) = [X_1(\omega) \dots X_N(\omega)]^T$ et

$$A(\omega) = \begin{bmatrix} a_{11}e^{-j\omega n_{11}} & \dots & a_{1N}e^{-j\omega n_{1N}} \\ a_{21}e^{-j\omega n_{21}} & \dots & a_{2N}e^{-j\omega n_{2N}} \\ \vdots & & \vdots \\ a_{N1}e^{-j\omega n_{N1}} & \dots & a_{NN}e^{-j\omega n_{NN}} \end{bmatrix}. \quad (4.5)$$

Idéalement, notre méthode de SAS a pour but d'obtenir la matrice de mélange $A(\omega)$ ci-dessus, cette dernière étant supposée inversible à toutes les fréquences. Néanmoins, cela peut seulement être fait aux indéterminations de filtrage et aux permutations inhérentes au problème de SAS près. Nous étendons donc aux mélanges AD l'approche qui a été introduite pour LI-TIFCORR [74] et décrite dans la section 2.1.1 : nous considérons une permutation arbitraire $\sigma(\cdot)$, appliquée aux indices j des signaux sources. Nous obtenons ainsi les sources permutées $s_{\sigma(j)}(n)$. Nous introduisons alors des versions AD de ces signaux, égales à leurs contributions dans la première observation, c.-à-d. :

$$s'_j(n) = a_{1,\sigma(j)} s_{\sigma(j)}(n - n_{1,\sigma(j)}). \quad (4.6)$$

L'équation de mélange (4.2) peut alors être réécrite de la façon suivante :

$$x_i(n) = \sum_{j=1}^N a_{i,\sigma(j)} s_{\sigma(j)}(n - n_{i,\sigma(j)}) \quad (4.7)$$

$$= \sum_{j=1}^N b_{ij} s'_j(n - \mu_{ij}), \quad (4.8)$$

avec

$$\begin{cases} b_{ij} &= \frac{a_{i,\sigma(j)}}{a_{1,\sigma(j)}} \\ \mu_{ij} &= n_{i,\sigma(j)} - n_{1,\sigma(j)} \end{cases} . \quad (4.9)$$

Cette équation de mélange (4.8) a la même forme que l'Eq. (4.2) sauf que les sources et paramètres $s_j(n)$, a_{ij} et n_{ij} sont ici remplacés par $s'_j(n)$, b_{ij} et μ_{ij} . En utilisant la même approche que ci-dessus, nous obtenons l'expression matricielle dans le domaine de Fourier

$$\underline{X}(\omega) = B(\omega) \underline{S}'(\omega), \quad (4.10)$$

où $\underline{S}'(\omega) = [S'_1(\omega) \cdots S'_N(\omega)]^T$ et

$$B(\omega) = \begin{bmatrix} 1 & \cdots & 1 \\ b_{21}e^{-j\omega\mu_{21}} & \cdots & b_{2N}e^{-j\omega\mu_{2N}} \\ \vdots & & \vdots \\ b_{N1}e^{-j\omega\mu_{N1}} & \cdots & b_{NN}e^{-j\omega\mu_{NN}} \end{bmatrix}. \quad (4.11)$$

Nous allons maintenant introduire des méthodes pour estimer $B(\omega)$. Les signaux de sortie peuvent alors être obtenus dans le domaine fréquentiel en calculant

$$\underline{Y}(\omega) = B^{-1}(\omega) \underline{X}(\omega), \quad (4.12)$$

où $\underline{Y}(\omega) = [Y_1(\omega) \cdots Y_N(\omega)]^T$ est le vecteur de transformées de Fourier des signaux de sorties. Les versions temporelles de ces signaux sont ensuite obtenues en appliquant une transformée de Fourier inverse à $\underline{Y}(\omega)$.

4.2 Première extension de LI-TIFCORR aux mélanges AD

Nous introduisons maintenant une première approche pour étendre la méthode LI-TIFCORR aux mélanges AD définis dans la section 4.1. La méthode TF de SAS ainsi introduite dans cette section utilise des zones d'analyse à fréquence constante pour estimer une partie des paramètres de mélanges. Cette méthode est donc appelée AD-TIFCORR-CF. Comme pour LI-TIFCORR que nous avons présentée dans les sections 2.3.1 et 2.3.2, ses principales étapes consistent à détecter des zones mono-sources TF et ensuite identifier à l'intérieur de celles-ci les paramètres de $B(\omega)$ que nous avons introduits dans l'Eq. (4.11), c.-à-d. les paramètres b_{im} et μ_{im} qui définissent la $m^{\text{ième}}$ colonne de $B(\omega)$. Dans cette section, nous allons successivement décrire :

- la méthode que nous proposons pour détecter les zones mono-sources (à fréquence constante) et pour identifier les paramètres b_{im} à l'intérieur de celles-ci (voir la section 4.2.1 dessous),
- les méthodes que nous proposons pour détecter les zones mono-sources (à temps constant) et pour identifier les paramètres $\mu_{im'}$ à l'intérieur de celles-ci (voir la section 4.2.2 ci-dessous),
- la procédure pour coupler les paramètres b_{im} et $\mu_{im'}$ identifiés ci-dessus (voir la section 4.2.3 ci-dessous),
- et la méthode globale qui résulte de ces principes (voir la section 4.2.4 ci-dessous).

4.2.1 Version de base des étapes de détection et d'identification pour les paramètres b_{im}

Comme expliqué ci-dessus, la méthode de SAS que nous introduisons ici inclut d'abord une étape de détection pour trouver les zones TF mono-sources. En partant des équations de mélange AD (4.2), nous avons déduit dans la section 4.1 les équations de mélange (4.3) dans le domaine fréquentiel. Cette relation entre les sources et les observations reste presque exacte quand elle est exprimée dans le domaine TF si les décalages temporels sont *suffisamment petits* par rapport à la largeur temporelle des fenêtres $h(\cdot)$ utilisées dans le calcul des TFCT définies dans l'Eq. (2.2).

Dans ce chapitre, nous supposons que cette condition est vérifiée et donc que les TFCT des observations peuvent être exprimées par rapport aux TFCT des sources suivant :

$$X_i(n, \omega) = \sum_{j=1}^N a_{ij} e^{-j\omega n_{ij}} S_j(n, \omega) \quad i = 1 \dots N. \quad (4.13)$$

Cela peut être comparé au cas des mélanges LI, où les équations de mélange temporel (4.1) donnent (sans aucune approximation) la relation suivante dans le domaine TF :

$$X_i(n, \omega) = \sum_{j=1}^N a_{ij} S_j(n, \omega) \quad i = 1 \dots N. \quad (4.14)$$

La relation (4.13) pour les mélanges AD est ainsi la même que l'expression (4.14) pour les mélanges LI, excepté le coefficient global appliqué à chaque source. La propriété qui a été utilisée dans la section 2.3.2 dans l'étape de détection de l'approche pour les mélanges LI ne dépend pas des valeurs de ces coefficients et s'applique donc encore ici. En d'autres termes, nous considérons encore une zone d'analyse TF qui est constituée d'un jeu de M fenêtres TF éventuellement recouvrantes qui correspondent à la même pulsation ω_l et à différentes positions temporelles adjacentes n_p . Ce jeu de M points temporels n_p , avec $p = 1, \dots, M$, est de nouveau noté T et la zone TF correspondante s'écrit alors (T, ω_l) . La version discrète de l'Eq. (2.76) s'écrit alors

$$R_{u_1 u_2}(T, \omega) = \frac{1}{M} \sum_{p=1}^M U_1(n_p, \omega_l) U_2(n_p, \omega_l)^*. \quad (4.15)$$

En appliquant la preuve générale de [77] à l'Eq. (4.13), on montre directement qu'une condition nécessaire et suffisante pour qu'une source soit isolée dans une zone d'analyse (T, ω_l) est

$$|r_{x_1 x_i}(T, \omega_l)| = 1 \quad \forall i \in \{2, \dots, N\}. \quad (4.16)$$

En pratique, comme pour la méthode LI, pour chaque zone d'analyse, nous calculons la moyenne sur i , notée $\overline{|r_{x_1 x_i}(T, \omega_l)|}$, des $|r_{x_1 x_i}(T, \omega_l)|$. Nous classons ensuite ces zones par valeur décroissante de $\overline{|r_{x_1 x_i}(T, \omega_l)|}$. Les premières zones de cette liste ordonnée sont donc les "meilleures" zones TF mono-sources.

Si une source $S_k(n, \omega)$ est isolée dans une zone d'analyse TF (T, ω_l) alors les Eq. (4.13) et (4.15) nous permettent d'obtenir la relation

$$I_i(T, \omega_l) = \frac{R_{x_1 x_1}(T, \omega_l)}{R_{x_1 x_1}(T, \omega_l)} = \frac{a_{ik}}{a_{1k}} e^{-j\omega_l(n_{ik} - n_{1k})} = b_{im} e^{-j\omega_l \mu_{im}} \quad (4.17)$$

avec b_{im} et μ_{im} définis dans l'Eq. (4.9) et $k = \sigma(m)$. Puisque nous supposons que tous les coefficients a_{ik} sont réels et positifs, les coefficients d'échelle b_{im} qui en résultent sont eux aussi réels et positifs. En conséquence, le module du paramètre $I_i(T, \omega_l)$ obtenu dans l'Eq. (4.17) est égal à b_{im} .

L'étape d'identification pour les coefficients b_{im} consiste donc à successivement considérer les premières zones d'analyse de la liste ordonnée définie ci-dessus. Pour chacune de ces zones, les coefficients estimés b_{im} associés à une colonne de $B(\omega)$ prennent pour valeur $|I_i(T, \omega_l)|$. Une nouvelle colonne de b_{im} est gardée si sa distance par rapport aux autres colonnes trouvées précédemment est au-dessus d'un seuil ϵ_1 défini par l'utilisateur. L'étape d'identification est finie lorsque N colonnes de coefficients d'atténuation mis à l'échelle sont trouvées.

Dans le chapitre 5, nous présenterons une version modifiée de l'étape d'identification des paramètres b_{im} .

4.2.2 Etapes de détection et identification des paramètres μ_{im}

Une méthode d'identification naturelle mais inadaptée

Grâce à l'expression (4.17) dans les zones d'analyse mono-sources, nous avons utilisé les modules de ces paramètres pour identifier les coefficients d'échelle b_{im} . En considérant encore cette expression (4.17), une idée naturelle pour estimer les décalages temporels μ_{im} consiste alors à tirer parti de la phase de $I_i(T, \omega_l)$. Ce qui peut être obtenu en pratique est l'angle, noté $\vartheta_i(T, \omega_l)$, associé à la phase de $I_i(T, \omega_l)$ et qui est situé dans l'intervalle $[-\pi, \pi]$. Nous avons alors la relation suivante d'après l'Eq. (4.17)

$$-\omega_l \mu_{im} = \vartheta_i(T, \omega_l) + 2q_{im}(T, \omega_l)\pi \quad (4.18)$$

et donc

$$\mu_{im} = -\frac{\vartheta_i(T, \omega_l) + 2q_{im}(T, \omega_l)\pi}{\omega_l}, \quad (4.19)$$

où $q_{im}(T, \omega_l)$ est le multiple inconnu associé à 2π dans l'Eq. (4.19) qui amène une indétermination sur la valeur de μ_{im} .

De façon alternative, nous pouvons par exemple étudier les rapports $\alpha_i(T, \omega_l)$, définis dans l'Eq. (2.55), qui s'écrivent ici

$$\alpha_i(n, \omega) = \frac{X_i(n, \omega)}{X_1(n, \omega)} = \frac{\sum_{j=1}^N a_{ij} e^{-j\omega n_{ij}} S_j(n, \omega)}{\sum_{j=1}^N a_{1j} e^{-j\omega n_{1j}} S_j(n, \omega)}. \quad (4.20)$$

Une idée naturelle pour estimer les décalages temporels consiste alors à considérer la phase des rapports $\alpha_i(n, \omega)$ dans les zones d'analyse mono-sources. Considérons une zone (T, ω_l) où la source $S_k(n, \omega)$ est isolée. Nous avons la relation suivante qui est vérifiée :

$$\alpha_i(n_p, \omega_l) = \frac{a_{ik}}{a_{1k}} e^{-j\omega_l(n_{ik} - n_{1k})} = b_{im} e^{-j\omega_l \mu_{im}}, \quad (4.21)$$

avec $k = \sigma(m)$. A nouveau, les valeurs d'angles obtenues, notées $\psi_i(n, \omega)$, sont situées dans l'intervalle $[-\pi, \pi]$. A cause de l'Eq. (4.21), nous avons :

$$-\omega_l \mu_{im} = \psi_i(n_p, \omega_l) + 2r_{im}(n_p, \omega_l)\pi \quad (4.22)$$

et donc :

$$\mu_{im} = -\frac{\psi_i(n_p, \omega_l) + 2r_{im}(n_p, \omega_l)\pi}{\omega_l}, \quad (4.23)$$

où $r_{im}(n_p, \omega_l)$ est le multiple associé à 2π dans l'Eq. (4.23) qui amène une indétermination sur la valeur de μ_{im} .

Notons toutefois que les deux phases ci-dessus ne sont pas équivalentes. En effet, avec l'approche basée sur la phase de $I_i(T, \omega)$, d'après l'Eq. (4.19), nous n'avons qu'une seule valeur $\vartheta_i(T, \omega_l)$ de la phase sur toute la zone d'analyse. Au contraire, si on étudie la phase du rapport $\alpha_i(n, \omega)$ sur toute cette zone d'analyse, alors d'après l'Eq. (4.23), nous avons M valeurs $\psi_i(n_p, \omega_l)$. Nous verrons ci-dessous les conséquences de cette différence.

Ces deux approches, resp. basées sur la phase de $I_i(T, \omega)$ et sur la phase de $\alpha_i(n, \omega)$, ne rendent donc pas possible l'estimation des paramètres μ_{im} car elles conduisent toutes deux à l'indétermination définie ci-dessus. Néanmoins, elles suggèrent une approche modifiée pour éviter cette indétermination, que nous présentons maintenant.

Détection des zones mono-sources à temps constant et identification des paramètres μ_{im}

Il faut se souvenir que les angles associés aux phases des transformées de Fourier rapides (TFR) ou des TFCT souffrent du type d'indétermination vu ci-dessus quand ils sont considérés individuellement, mais des techniques bien connues existent pour introduire de la "cohérence" entre les phases des points fréquemment adjacents des TFR ou TFCT, en déroulant ces phases par rapport à la fréquence. En combinant ce déroulement de phase aux principes d'identification basés sur la phase que nous avons introduits ci-dessus, nous suggérons une approche modifiée de l'identification des paramètres μ_{im} . Cette nouvelle approche peut être basée sur la phase de $I_i(T, \omega)$ considérée dans la zone TF (T, ω) ou sur celle de $\alpha_i(n, \omega)$ considérée sur les fenêtres TF (n, ω) .

Etudions d'abord l'approche basée sur la phase de $I_i(T, \omega)$. Nous considérons indépendamment chaque position temporelle T associée à ces zones TF et pour chaque position,

nous déroulons la phase de $I_i(T, \omega)$ pour tous les points TF associés, fréquemment adjacents. Nous allons maintenant montrer que cette phase déroulée permet d'identifier les paramètres μ_{im} , si elle est considérée sur un sous-ensemble connexe des points TF ci-dessus où une source est isolée. Cela signifie que nous devons d'abord détecter des zones d'analyse mono-sources, mais il doit être clair que nous avons à considérer ici un autre type de zones d'analyse que celles que nous avons préalablement utilisées pour identifier les paramètres b_{im} : à cause du principe de déroulement de phase dont nous avons besoin ici, nous considérons des zones d'analyse qui correspondent à un jeu de zones TF ayant la même position temporelle et des positions fréquentielles adjacentes. Nous avons alors à développer non seulement une méthode d'identification pour les paramètres μ_{im} , mais aussi une méthode pour détecter les zones d'analyse mono-sources à temps constant. Ces deux méthodes peuvent en fait être déduites des principes ci-dessus, comme nous allons le voir.

Considérons une zone d'analyse à temps constant, c.-à-d. un jeu de M' zones TF correspondant à la même position temporelle T' et à des pulsations adjacentes $\omega_{l'}$, avec $l' = 1, \dots, M'$. Ce jeu de points fréquentiels est noté Ω ci-dessous et la zone TF correspondante est alors notée (T', Ω) . Comme indiqué précédemment, ceci doit être différencié des zones d'analyse (T, ω_l) que nous utilisons pour l'identification des coefficients d'échelle b_{im} et qui sont constituées de fenêtres TF temporellement adjacentes. Focalisons-nous sur une zone (T', Ω) où la source $S_{k'}(n, \omega)$ est isolée. L'expression (4.13) des observations en n'importe quel point de cette zone d'analyse est alors réduite à :

$$X_i(n_{p'}, \omega_{l'}) = a_{ik'} e^{-j\omega_{l'} n_{ik'}} S_{k'}(n_{p'}, \omega_{l'}), \quad \forall n_{p'} \in T', \forall \omega_{l'} \in \Omega, \forall i \in \{1, \dots, N\}. \quad (4.24)$$

Le paramètre d'identification (4.17) s'écrit alors :

$$I_i(T', \omega_{l'}) = \frac{a_{ik'}}{a_{1k'}} e^{-j\omega_{l'}(n_{ik'} - n_{1k'})} \quad (4.25)$$

$$= b_{im'} e^{-j\omega_{l'} \mu_{im'}} \text{ avec } k' = \sigma(m'). \quad (4.26)$$

La phase de $I_i(T', \omega_{l'})$ dans les zones d'analyse à temps constant est alors reliée aux paramètres $\mu_{im'}$. Plus précisément, dans la zone d'analyse mono-source (T', Ω) où la source $S_{k'}(n, \omega)$ est isolée, si nous considérons la phase déroulée $\theta'_i(T', \omega_{l'})$ de $I_i(T', \omega_{l'})$ dans cette zone, nous avons :

$$-\omega_{l'} \mu_{im'} = \theta'_i(T', \omega_{l'}) + 2q'_{im'}(T')\pi, \quad (4.27)$$

où $q'_{im'}(T')$ est un entier inconnu. Le multiple associé à 2π dans l'expression (4.27) apporte encore une indétermination. Néanmoins, cette indétermination est différente de celle qui était présente dans l'Eq. (4.18) : grâce au déroulement de phase que nous avons appliqué, l'entier $q'_{im'}(T')$ ne dépend pas de la fréquence. L'Eq. (4.27) montre que la courbe associée aux variations de la phase $\theta'_i(n_{p'}, \omega_{l'})$ par rapport à $\omega_{l'}$ dans une zone d'analyse mono-source (T', Ω) est une droite et que sa pente ne dépend pas de la valeur de $q'_{im'}(T')$ et est égale à $-\mu_{im'}$. Cette pente apporte donc un moyen d'identifier $\mu_{im'}$, sans aucune indétermination de phase. Cependant, la taille des zones d'analyse (T', Ω) ainsi créées est grande par rapport à celles des zones à fréquence constante (T, ω_l) étudiées jusqu'alors et les possibilités de trouver une source isolée dans de telles zones sont donc plus limitées pour (T', Ω) que pour (T, ω_l) .

De façon alternative, nous proposons alors d'appliquer l'approche ci-dessus sur la phase du rapport $\alpha_i(n, \omega)$. Nous considérons ainsi indépendamment chaque position temporelle n et nous déroulons la phase de $\alpha_i(n, \omega)$ pour tous les points TF associés, fréquemment adjacents. Considérons une zone d'analyse à temps constant qui est ici un jeu de M' fenêtres TF correspondant à la même position temporelle $n_{p'}$ et à des pulsations adjacentes $\omega_{l'}$, avec $l' = 1, \dots, M'$. En notant à nouveau ce jeu de points fréquentiels Ω , la zone TF correspondante s'écrit alors $(n_{p'}, \Omega)$. Sur une zone $(n_{p'}, \Omega)$ où la source $S_{k'}(n, \omega)$ est isolée, le paramètre d'identification (4.20) s'écrit donc :

$$\alpha_i(n_{p'}, \omega_{l'}) = \frac{a_{ik'}}{a_{1k'}} e^{-j\omega_{l'}(n_{ik'} - n_{1k'})} \quad (4.28)$$

$$= b_{im'} e^{-j\omega_{l'} \mu_{im'}} \text{ avec } k' = \sigma(m'). \quad (4.29)$$

Comme la phase de $I_i(T', \omega_{l'})$, celle de $\alpha_i(n_{p'}, \omega_{l'})$ dans les zones mono-sources à temps constant $(n_{p'}, \Omega)$ est reliée aux paramètres $\mu_{im'}$. Plus précisément, en reprenant le principe de l'Eq. (4.27), appliqué à la phase déroulée $\phi'_i(n_{p'}, \omega_{l'})$ de $\alpha_i(n_{p'}, \omega_{l'})$ dans la zone $(n_{p'}, \Omega)$, nous obtenons :

$$-\omega_{l'} \mu_{im'} = \phi'_i(n_{p'}, \omega_{l'}) + 2q'_{im'}(n_{p'})\pi. \quad (4.30)$$

Comme pour l'Eq. (4.27), le multiple associé à 2π dans l'expression (4.30) apporte une indétermination qui ne dépend pas de la fréquence. La courbe associée aux variations de la phase $\phi'_i(n_{p'}, \omega_{l'})$ par rapport à $\omega_{l'}$ dans une zone d'analyse $(n_{p'}, \Omega)$ est alors une droite (voir la figure 4.1) et sa pente, qui ne dépend pas de la valeur de $q'_{im'}(n_{p'})$, est égale à $-\mu_{im'}$. Cette pente apporte donc un moyen d'identifier $\mu_{im'}$, sans aucune indétermination de phase. De plus, les zones d'analyse $(n_{p'}, \Omega)$ utilisées ici présentent une largeur temporelle beaucoup plus petite que les zones (T', Ω) utilisées dans l'approche basée sur la phase de $I_i(T', \omega_{l'})$, ce qui doit nous permettre de trouver plus facilement des zones d'analyse à temps constant mono-sources.

En conclusion de l'analyse précédente, notre méthode pour identifier le jeu de paramètres $\mu_{im'}$ associé à une colonne de $B(\omega)$ s'effectue ainsi. Dans la zone d'analyse mono-source à temps constant sélectionnée, pour chaque signal observé d'indice i , nous considérons les M' points qui ont deux coordonnées, resp. définies comme les pulsations $\omega_{l'}$ et les valeurs correspondantes $\phi'_i(n_{p'}, \omega_{l'})$ de la phase déroulée du paramètre d'identification. Nous déterminons la droite de régression par moindres carrés associée à ces points. L'estimée du paramètre $\mu_{im'}$ est alors définie comme l'opposé du coefficient directeur de cette droite de régression. Plus précisément, $\mu_{im'}$ est arrondi à l'entier le plus proche de l'opposé de ce coefficient directeur, puisque le décalage temporel $\mu_{im'}$ est défini dans le domaine discret et est donc entier.

Cette approche peut alors être étendue pour *détecter* les zones d'analyse mono-sources à temps constant qui sont utilisées dans l'étape d'identification que nous venons de décrire, comme nous allons le voir maintenant. Nous considérons successivement toutes les zones d'analyse à temps constant $(n_{p'}, \Omega)$ qui peuvent être associées aux points TF disponibles, c.-à-d. les zones correspondant à toutes les valeurs de $n_{p'}$ et à des zones de fréquences adjacentes Ω qui contiennent chacune M' points fréquentiellement adjacents $\omega_{l'}$. Dans chaque zone $(n_{p'}, \Omega)$, nous déterminons la droite de régression définie ci-dessus et l'erreur quadratique moyenne des points $(\omega_{l'}, \phi'_i(n_{p'}, \omega_{l'}))$ par rapport à leur droite de

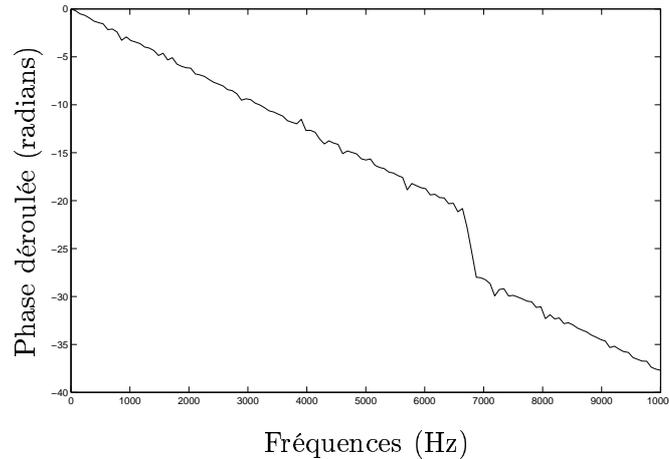


FIG. 4.1 – Phase déroulée $\phi'_2(n, \omega)$ selon la pulsation ω pour un indice temporel n fixé. Ici le signal $x_2(n)$ est juste une version décalée de quelques échantillons de $x_1(n)$, si bien que $\phi'_2(n, \omega)$ est une fonction linéaire de ω (aux erreurs d'estimation de la phase et au décrochement d'un multiple de 2π près).

régression associée. Les meilleures zones mono-sources sont celles qui apportent la plus petite erreur de régression³¹. Ces zones peuvent être ensuite utilisées de différentes façons pour identifier les paramètres $\mu_{im'}$, par exemple en classant ces zones par ordre croissant de leur erreur quadratique moyenne ou en utilisant des techniques de classification (en faisant par exemple appel à une des techniques proposées dans le chapitre 5). Elles apportent enfin un ensemble de vecteurs colonne. Chacun de ces vecteurs contient les valeurs $\mu_{im'}$, qui correspondent à toutes les observations d'indice i et à la source d'indice $k' = \sigma^{-1}(m')$ qui est isolée dans la zone d'analyse considérée. Ces vecteurs sont obtenus dans un ordre arbitraire, suivant quelle source est isolée dans la zone d'analyse considérée (c'est le problème de permutation connu en SAS). Le même type de commentaire s'applique à l'identification des paramètres b_{im} que nous avons précédemment obtenus avec les méthodes décrites dans la section 4.2.1. Il faut rappeler que nous voulons identifier les paramètres qui définissent chaque colonne de $B(\omega)$, si bien que nous devons déterminer quelle colonne de paramètres b_{im} va avec quelle colonne de paramètres $\mu_{im'}$, c.-à-d. correspond à la même source. Ce n'est pas encore défini dans cette étape de notre discussion, puisque ces deux types de paramètres ont été obtenus indépendamment et chacun d'entre eux dans un ordre arbitraire, comme mentionné plus haut. Une étape finale doit donc être

³¹Les procédures standard de déroulement de phase ont parfois des effets parasites, c.-à-d. elles gardent une discontinuité de la phase égale à un multiple de 2π à quelques (rares) fréquences (voir la figure 4.1). La description ci-dessus montre que cela n'est pas un problème majeur dans notre approche : si une telle discontinuité apparaît dans une zone d'analyse mono-source, il en résultera une forte erreur de régression dans cette zone. Un très petit ensemble de zones mono-sources peuvent ainsi être écartées dans la détection de zones mono-sources. Cela ne rend pas la méthode défectueuse pour autant, parce qu'en pratique les signaux apportent suffisamment de zones mono-sources et même si quelques-unes sont oubliées, d'autres sont utilisées à la place.

ajoutée à notre approche, dans le but de coupler chaque colonne de décalages temporels $\mu_{im'}$ à une colonne de coefficients d'atténuation b_{im} . Ceci définira comment nous utilisons les zones d'analyse mono-sources à temps constant et les paramètres correspondants $\mu_{im'}$ qui sont obtenus par la méthode que nous avons décrite dans cette section.

4.2.3 Couplage des paramètres b_{im} et $\mu_{im'}$

Une méthode d'identification alternative pour les paramètres b_{im}

Dans la section 4.2.2, nous avons introduit une méthode pour détecter les zones d'analyse mono-sources à temps constant et nous avons montré comment la phase des paramètres $\alpha_i(n, \omega)$ dans ces zones pouvait être utilisée pour identifier les décalages temporels $\mu_{im'}$. Nous notons ici que les modules de ces paramètres dans ces zones permettent aussi d'identifier les paramètres $b_{im'}$: l'Eq. (4.28) montre que, pour toute pulsation $\omega_{i'}$ d'une telle zone, le module de $\alpha_i(n_{p'}, \omega_{i'})$ est égal à $b_{im'}$. Ce dernier paramètre peut donc être identifié comme la valeur moyenne du module de $\alpha_i(n, \omega)$ sur une zone d'analyse mono-source à temps constant.

La valeur ainsi obtenue est notée $b'_{im'}$ ci-dessous, de façon à la distinguer de la valeur b_{im} obtenue à l'aide des méthodes que nous avons introduites dans la section 4.2.1. L'approche alternative que nous proposons dans cette section est attrayante car chaque zone d'analyse considérée permet d'obtenir les paramètres $b'_{im'}$ et $\mu_{im'}$ correspondant à la *même* source. Elle apporte donc une solution au couplage de ces types de paramètres. Néanmoins, nos premiers tests expérimentaux ont montré que les valeurs des paramètres $b'_{im'}$ ainsi obtenues estimaient moins précisément les paramètres de mélange réels que les valeurs b_{im} que nous avons décrites dans la section 4.2.1. Nous allons donc introduire une approche modifiée qui tire parti des deux types de paramètres.

Couplage des paramètres b_{im} et $(b'_{im'}, \mu_{im'})$

En tirant parti des principes définis précédemment, nous introduisons maintenant une méthode pour coupler finalement les paramètres b_{im} et $\mu_{im'}$. Cette méthode consiste à :

1. déterminer les coefficients d'atténuation b_{im} comme expliqué dans la section 4.2.1,
2. déterminer indépendamment les couples $(b'_{im'}, \mu_{im'})$ comme expliqué dans les sections 4.2.2 et 4.2.3,
3. et ensuite coupler les paramètres $\mu_{im'}$ aux paramètres b_{im} grâce aux $b'_{im'}$. Voici comment ce couplage est effectué. L'identification des paramètres b_{im} nous permet d'obtenir N colonnes de paramètres, chacune associée à une source différente. Dans la détection des zones d'analyse mono-sources à temps constant, nous gardons un nombre de zones significativement plus grand que N , en sélectionnant toutes les zones où l'erreur quadratique moyenne obtenue à partir de la droite de régression associée est plus petite qu'un seuil ϵ_2 défini par l'utilisateur. Pour chacune de ces zones, nous identifions les deux colonnes qui contiennent resp. les paramètres $b'_{im'}$ et $\mu_{im'}$ correspondant à cette zone. Nous considérons ensuite que les paramètres $b'_{im'}$ et b_{im} sont resp. des estimées grossière et précise des coefficients d'échelle associés à la matrice de mélange et nous couplons chaque colonne de $b'_{im'}$ à la plus proche

colonne³² de b_{im} . Vu que les paramètres $b'_{im'}$ ont été couplés avec les paramètres $\mu_{im'}$, ces derniers paramètres sont ainsi couplés aux N colonnes de paramètres b_{im} . Pour chaque élément (associé à l'observation i) dans chacune de telles colonnes de b_{im} , nous devons enfin ne garder qu'une seule valeur du paramètre $\mu_{im'}$. Ceci est effectué pour chaque élément de la façon suivante : de toutes les valeurs $\mu_{im'}$ qui ont été associées à cet élément, nous gardons la valeur qui a le plus grand nombre d'occurrences.

4.2.4 Résumé de la méthode AD-TIFROM-CF

Suite à la description ci-dessus, la méthode AD-TIFROM-CF que nous proposons ici contient les étapes suivantes :

1. L'étape de pré-traitement consiste à déduire les TFCT $X_i(n, \omega)$ des signaux de mélange, selon la version à temps discret de l'expression (2.2).
2. Nous considérons ensuite uniquement les paramètres b_{im} . Nous détectons d'abord les zones d'analyse mono-sources à fréquence constante (T, ω_l) et nous identifions ensuite les paramètres b_{im} dans ces zones, en utilisant la procédure basique décrite dans la section 4.2.1.
3. Nous considérons ensuite les paramètres $b'_{im'}$ et $\mu_{im'}$. Nous calculons d'abord les droites de régression et leurs erreurs quadratiques moyennes associées sur les zones d'analyse TF à temps constant $(n_{p'}, \Omega)$ définies dans la section 4.2.2. Nous détectons ensuite les zones mono-sources et identifions les paramètres $b'_{im'}$ et $\mu_{im'}$ dans ces zones, en utilisant la procédure décrite dans les sections 4.2.2 et 4.2.3.
4. Nous couplons ensuite les paramètres b_{im} et $\mu_{im'}$, comme expliqué dans la section 4.2.3. Tous les paramètres, associés à la matrice $B(\omega)$, qui sont requis sont disponibles à cette étape.
5. En utilisant ces paramètres, nous obtenons enfin les sources extraites des observations, selon l'Eq. (4.12).

4.3 Seconde extension de LI-TIFCORR aux mélanges AD

4.3.1 Etapes alternatives de détection et d'identification

Dans la section 4.2.1, nous avons introduit une méthode qui exploite les zones d'analyse à fréquence constante (T, ω_l) , basée sur le coefficient de corrélation $r_{x_1x_i}(T, \omega_l)$, dans le but de détecter des zones mono-sources. Nous étions alors capables de calculer de très bonnes estimations de b_{im} mais ne pouvions pas estimer de façon précise les décalages temporels μ_{im} correspondants dans les mêmes zones. C'est pourquoi nous avons présenté dans la section 4.2.2 une méthode, utilisant les zones TF à temps constant $(n_{p'}, \Omega)$, et

³²Ici encore, un seuil ϵ_3 défini par l'utilisateur est utilisé pour ignorer toute colonne de paramètres $b'_{im'}$ telle que toutes ses distances par rapport aux N colonnes de paramètres b_{im} soient au-dessus de ce seuil.

nous avons introduit une nouvelle approche pour trouver les zones TF mono-sources, basée sur les droites de régression des phases déroulées de $\alpha_i(n, \omega)$ et les erreurs de régression associées. Les zones à temps constant sont attrayantes car elles couplent naturellement les deux types de paramètres identifiés en leur sein, c.-à-d. $b'_{im'}$ et $\mu'_{im'}$. Mais nous avons indiqué dans la section 4.2.3 que les paramètres $b'_{im'}$ obtenus dans des zones sélectionnées selon les erreurs de régression estimaient de façon moins précise les mélanges que les valeurs b_{im} ci-dessus. Nous allons ici combiner les avantages des deux types d'approche ci-dessus en considérant des zones à temps constant couplées aux paramètres $b'_{im'}$ et $\mu'_{im'}$, et en détectant lesquelles de ces zones contiennent une seule source. Nous appelons cette méthode, qui utilise *seulement* des zones à temps constant, AD-TIFCORR-CT, par opposition à l'approche AD-TIFCORR-CF décrite dans la section précédente, qui utilise aussi des zones d'analyse à fréquence constante.

Nous présentons maintenant en détail les principes de cette méthode. Nous devons d'abord modifier notre précédent critère de corrélation $R_{u_1u_2}$ défini dans l'Eq. (2.76). Nous utilisons ici

$$R'_{u_1u_2}(n, \Omega) = \frac{1}{M'} \sum_{l'=1}^{M'} |U_1(n, \omega_{l'}) U_2^*(n, \omega_{l'})|. \quad (4.31)$$

Le coefficient de corrélation correspondant s'écrit alors

$$r'_{u_1u_2}(n, \Omega) = \frac{R'_{u_1u_2}(n, \Omega)}{\sqrt{R'_{u_1u_1}(n, \Omega) R'_{u_2u_2}(n, \Omega)}}. \quad (4.32)$$

Nous faisons l'hypothèse suivante :

Hypothèse 10 *Quand plusieurs sources sont actives dans la même zone d'analyse, elles doivent varier de telle façon que le module d'au moins un des rapports des TFCT des observations $\frac{X_i(n, \omega)}{X_1(n, \omega)}$ ne prenne pas la même valeur dans toutes ces fenêtres.*

Grâce à l'hypothèse 10, nous montrons dans la section B.2 de l'annexe B qu'une condition nécessaire et suffisante pour qu'une source soit isolée dans une zone d'analyse $(n_{p'}, \Omega)$ est

$$|r'_{x_1x_i}(n_{p'}, \Omega)| = 1 \quad \forall i \in \{2, \dots, N\}. \quad (4.33)$$

Les paramètres de mélange $b'_{im'}$ sont alors identifiés en utilisant la méthode basique définie dans la section 4.2.1, à part que les zones d'analyse (T, ω_l) sont ici remplacées par les zones $(n_{p'}, \Omega)$. Ainsi, le paramètre $I_i(T, \omega_l)$, défini dans l'Eq. (4.17), est remplacé par

$$I'_i(n_{p'}, \Omega) = \frac{R'_{x_ix_1}(n_{p'}, \Omega)}{R'_{x_1x_1}(n_{p'}, \Omega)}. \quad (4.34)$$

L'identification des paramètres $\mu'_{im'}$ est obtenue de la façon suivante. Dans chaque zone mono-source où une colonne est sélectionnée (et gardée), la colonne correspondante de paramètres $\mu'_{im'}$ est simultanément identifiée, en utilisant encore la méthode basée sur les droites de régression décrite dans la section 4.2.2.

4.3.2 Résumé de la méthode AD-TIFROM-CT

En conclusion de la description ci-dessus, la méthode AD-TICORR-CT que nous proposons aussi pour les mélanges AD contient les étapes suivantes :

1. L'étape de pré-traitement consiste à déduire les TFCT $X_i(n, \omega)$ des signaux de mélanges, selon la version à temps discret de l'Eq. (2.2).
2. Nous détectons ensuite les zones d'analyse mono-sources à temps constant $(n_{p'}, \Omega)$ et nous identifions les paramètres $b'_{im'}$ et $\mu_{im'}$ dans ces zones, en utilisant la procédure introduite dans la section 4.3.1. Tous les paramètres requis associés à la matrice estimée $B(\omega)$ sont disponibles à cette étape.
3. En utilisant les paramètres ci-dessus, nous déduisons enfin les sources extraites des observations, selon l'Eq. (4.12).

4.4 Conclusion

Dans ce chapitre, nous avons étendu de deux façons différentes la méthode LI-TIFCORR aux mélanges AD. Une première méthode, notée AD-TIFCORR-CF, utilise des zones d'analyse à fréquence constante (T, ω_l) pour détecter les zones mono-sources et estimer les paramètres b_{im} . Nous avons ensuite montré que l'estimation d'une large gamme de décalages temporels μ_{im} associés n'était possible que dans des zones d'analyse à temps constant. La deuxième étape de la méthode consiste donc à considérer les zones d'analyse à temps constant $(n_{p'}, \Omega)$. Quand ces zones sont mono-sources, les variations de la phase des rapports d'observations TF en fonction de la pulsation définissent une droite dont le coefficient directeur est l'opposé du décalage temporel $\mu_{im'}$ à estimer. A l'aide des droites de régression et de l'erreur associée, nous avons alors trouvé les zones mono-sources, estimé les décalages temporels $\mu_{im'}$ et les coefficients d'échelle $b_{im'}$ associés. Grâce à ces derniers, nous pouvons enfin associer chaque décalage temporel $\mu_{im'}$ estimé au paramètre correspondant b_{im} , estimé dans l'étape précédente de la méthode.

La deuxième méthode consiste à ne considérer que les zones à temps constant $(n_{p'}, \Omega)$, d'où son nom AD-TIFCORR-CT. Dans ces zones, nous avons proposé un critère déduit de celui de la première étape de AD-TIFCORR-CF mais transposé aux zones à temps constant. Ce critère permet de trouver les zones mono-sources et d'estimer les paramètres $b_{im'}$ dans ces zones. Dans ces mêmes zones, nous pouvons ensuite estimer les décalages temporels $\mu_{im'}$ correspondant à l'aide des droites de régression de la phase des rapports d'observation.

Cependant, les méthodes proposées ici font toujours appel à divers paramètres heuristiques. Nous allons donc étudier dans le chapitre suivant l'apport d'approches de classification non-supervisée et voir les améliorations qu'elles peuvent apporter aux méthodes décrites dans ce chapitre.

Chapitre 5

Apport des techniques de classification non-supervisée aux méthodes de SAS à parcimonie

Nous avons proposé dans le chapitre 4 des méthodes très simples pour estimer les coefficients d'échelle b_{im} et les décalages temporels $\mu_{im'}$. Ces approches, qui ont d'abord été introduites pour les coefficients b_{im} dans le cas de mélanges LI, apportent d'excellents résultats, comme vu dans le chapitre 3. Appliquées à des mélanges plus complexes qui tiennent compte de retards, elles sont aussi satisfaisantes pour des mélanges de $N = 2$ sources avec des "retards modérés" (inférieurs à un dixième de la longueur des TFCT dans nos tests du chapitre 6). Cependant, pour $N > 2$ ou quand les retards deviennent très grands dans le cas $N = 2$, nous avons obtenu de mauvais résultats pour un nombre significatif de tests expérimentaux. Plus précisément, les paramètres b_{im} des colonnes de $B(\omega)$ ont alors été identifiés dans des zones d'analyse qui étaient sélectionnées parce qu'elles étaient au début de la liste ordonnée créée dans l'étape de détection, mais ces colonnes identifiées ne correspondaient pas à la "véritable" matrice de mélange (aux indéterminations classiques près), si bien que les signaux de sortie ne correspondaient pas aux sources séparées. Nous proposons alors de tenir compte de plus de zones d'analyse et donc de plus de paramètres estimés b_{im} dans l'étape d'identification en utilisant des méthodes de classification non-supervisée (ou *clustering* en anglais).

De plus, nous avons souligné dans la première partie de ce manuscrit les limitations dues aux paramètres heuristiques tels le seuil minimal de distance entre deux colonnes estimées de la matrice de mélange. Idéalement, les méthodes de classification permettront de se passer de ceux-ci. Dans la section 5.1, nous allons décrire quelques méthodes classiques de classification. Puis, dans la section 5.2, nous étudierons et proposerons des méthodes spécialement dédiées aux approches parcimonieuses décrites dans ce document.

5.1 Les méthodes de classification non-supervisée classiques

Durant les 50 dernières années, un grand nombre de méthodes de classification non-supervisée ont été proposées pour diverses applications [109, 244]. Considérons Q objets ou individus multidimensionnels³³ x_i . L'ensemble de ces individus, noté $\underline{x} = \{x_1, \dots, x_Q\}$, est appelé *nuage de points*. Le but des méthodes de classification non-supervisée est de répartir de façon automatique les individus du nuage de points \underline{x} en k groupes, appelés *classes* ou *clusters* et notés C_j . Chaque cluster C_j est associé à son *centroïde* c_j , c.-à-d. un point de l'espace des individus. Le nombre k de clusters n'est pas forcément connu a priori et un nombre important de méthodes, comme par exemple [91, 110, 111, 193, 214, 244], tendent à l'estimer automatiquement.

La plupart des méthodes appartiennent à une des familles suivantes :

1. les méthodes hiérarchiques,
2. les méthodes basées sur le partitionnement dur ou flou des données,
3. les méthodes basées sur les mélanges de densités (estimation des densités de probabilité via les mélanges de densités)
4. les méthodes basées sur la théorie des graphes,
5. les méthodes basées sur les techniques de recherche combinatoire,
6. les méthodes basées sur les réseaux de neurones.

Nous allons maintenant présenter quelques algorithmes classiques, mais un panorama très complet est proposé dans [244].

5.1.1 Les algorithmes hiérarchiques

Ces algorithmes font parti des grands classiques et sont groupés en deux catégories :

1. les algorithmes hiérarchiques ascendants,
2. les algorithmes hiérarchiques descendants.

Dans le premier cas, l'idée est que chaque cluster est progressivement absorbé par le plus proche cluster [131]. A l'initialisation, chaque individu à traiter est son propre cluster. A chaque itération, les clusters les plus proches sont groupés en un cluster plus général. La procédure s'arrête lorsque le nombre voulu de clusters finaux k est obtenu (voir figure 5.1).

Au contraire, les algorithmes hiérarchiques descendants partent d'un unique cluster et le divisent jusqu'à ce que le nombre voulu de clusters finaux soit obtenu. Le problème de ces méthodes-ci est qu'elles nécessitent l'utilisation d'algorithmes ascendants pour savoir comment couper un cluster. En conséquence, elles sont souvent moins intéressantes que les méthodes ascendantes, sauf pour des applications particulières où elles sont naturelles à utiliser.

³³Remarquons que chaque individu est potentiellement un vecteur, suivant le nombre de paramètres qu'il contient. Afin de simplifier les écritures, nous ne notons pas ces individus \underline{x}_i mais x_i .

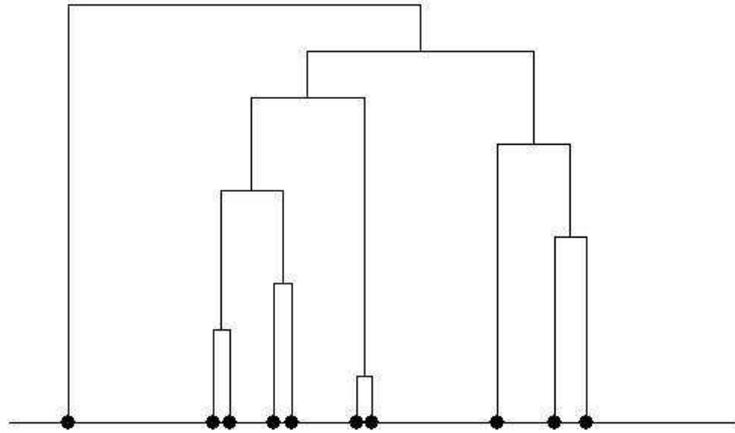


FIG. 5.1 – Exemple de clustering hiérarchique ascendant.

La difficulté des méthodes hiérarchiques consiste à trouver l'*ultramétrie*, c.-à-d. la métrique qui permet de mesurer la distance entre les clusters, la plus adaptée aux individus à classer. Les ultramétries classiques sont :

- la distance minimale entre les individus des clusters (*single linkage*),
- la distance maximale entre les individus des clusters (*complete linkage*),
- la distance moyenne entre tous les individus de tous les clusters (*group average linkage*),
- la distance entre les centroïdes (*centroid linkage*).

De façon générale, les limitations des méthodes de classification hiérarchique sont les suivantes :

1. Une fois qu'un individu est assigné à un cluster, il n'est plus possible de revenir en arrière, ce qui signifie que les méthodes hiérarchiques ne sont pas capables de corriger une assignation erronée d'un individu.
2. La complexité des méthodes est de l'ordre de Q^2 , ce qui limite leur champ d'action à des données de petite taille.
3. Suivant l'ultramétrie choisie, les problèmes suivants peuvent apparaître :
 - (a) sensibilité au bruit et aux points aberrants (notamment quand l'ultramétrie utilisée est la distance minimale entre les individus),
 - (b) difficulté à gérer des clusters de taille différente et les formes convexes (en particulier quand l'ultramétrie utilisée est la distance maximale ou la distance moyenne entre tous les individus),
 - (c) tendance à couper les gros clusters (si l'ultramétrie utilisée est par exemple la distance maximale ou la distance moyenne entre tous les individus).

5.1.2 Classification basée sur le partitionnement des données

L'algorithme K-means et quelques versions modifiées

De toutes les méthodes de clustering qui ont été proposées, K-means [169] est sûrement la plus connue et la plus utilisée, en particulier en SAS [11, 14–19, 53, 149, 150, 211, 212, 220, 221], notamment parce que c'est un algorithme simple et plutôt rapide (complexité d'ordre Q) qui permet de résoudre les problèmes de classification. La procédure permet de classer un nuage de points selon un nombre k de clusters fixé a priori. L'algorithme peut être défini de la façon suivante.

1. Définir k centroïdes.
2. Assigner chaque individu au groupe qui a le plus proche centroïde.
3. Recalculer les nouveaux k centroïdes comme les barycentres des clusters définis dans l'étape 2.
4. Répéter les étapes 2 et 3 jusqu'à ce que les k centroïdes ne bougent plus.

Réaliser la classification d'individus x_i dans des classes C_j de centroïdes c_j à l'aide de la méthode K-means revient à minimiser la fonction de coût suivante :

$$J = \sum_{j=1}^k \sum_{i=1}^Q \mathbb{1}_{i,j} \|x_i - c_j\|^2, \quad (5.1)$$

où $\mathbb{1}_{i,j}$ est la fonction indicatrice qui vaut 1 si $x_i \in C_j$ et 0 sinon.

Même si cette méthode est simple et peut donner de bons résultats, elle a trois gros désagréments :

1. K-means est très sensible aux points aberrants (*outliers*) : même si un individu x_i est très éloigné d'un centroïde, il est quand même forcé dans un cluster et "distord" donc la forme de celui-ci.
2. La convergence des centroïdes dépend de l'initialisation des k centroïdes et les optima sont locaux. Une stratégie générale consiste à lancer plusieurs fois l'algorithme avec des initialisations aléatoires et garder les pics des histogrammes des centroïdes finaux trouvés³⁴.
3. Nous devons connaître le nombre de clusters³⁵ avant de lancer la méthode.

Dans le but d'éviter ces limitations, quelques méthodes, inspirées de K-means, ont été proposées.

- Tom C. Bakker Schut et al. [25] ont proposé de résoudre les possibles erreurs dues à l'initialisation des centroïdes. Pour cela, ils utilisent K-means, initialisé avec un grand nombre de centroïdes et réduisent ensuite un grand nombre de sous-clusters grâce à une technique des proches voisins. Leur but est de traiter des nuages de

³⁴Nos premiers tests ont montré que cette approche n'était pas adaptée à notre problème : il arrivait souvent qu'un des pics de l'histogramme corresponde à une "mauvaise" colonne de paramètres estimés b_{im} .

³⁵Ce n'est pas un point critique pour les méthodes développées ici mais peut le devenir dans le cas d'applications réelles où le nombre de sources est inconnu.

points de grandes cardinalité et dimension (issus de flux cytométriques, c.-à-d. des cellules en suspension dans un fluide qu'il faut compter et analyser), pour lesquelles il n'est pas possible de choisir naturellement de "bons" centroïdes initiaux. Cependant, à cause de la nature de leurs données à traiter, le choix d'arrêter la réduction du nombre de clusters n'est pas dicté par un critère mathématique : cette décision est prise "manuellement" par l'utilisateur.

- D. Pelleg et A. Moore [193] proposent de déterminer le nombre de clusters en ajoutant un critère sur la forme des clusters. Ils supposent avoir une gamme de nombres de clusters à laquelle le véritable k appartient raisonnablement. Ils partent de k_{min} centroïdes obtenus par K-means, k_{min} étant la borne minimale de cette gamme. Grâce à un critère d'information bayésienne (*Bayesian Information Criterion* BIC en anglais) [138, 218], leur approche, appelée X-means, découpe les clusters existant en deux tant qu'un meilleur ajustement des données est obtenu. Cependant, à cause de ce critère, cette méthode a tendance à diviser à tort de "bons" clusters non-sphériques [110].
- G. Hamerly et C. Elkan [110] veulent eux-aussi déterminer le nombre de clusters en divisant des clusters existants. Pour cela, ils partent d'un centroïde et divisent les clusters grâce à un critère de mesure de gaussiannité (d'où le G dans le nom de la méthode : G-means) appelé test d'Anderson-Darling [224]. G-means est plus souvent performant que X-means, en particulier quand les clusters ne sont pas sphériques. Mais il peut aussi diviser des clusters denses en deux sans que cela soit nécessaire en réalité [91]. De plus, la précision de l'estimation du nombre k de clusters décroît exponentiellement avec k^{36} .
- T. Elomaa et H. Koivistoinen [91] proposent d'améliorer G-means sur ce dernier point, en ajoutant une approximation a priori sur le nombre de clusters de G-means. Pour cela, ils
 1. sous-échantillonnent les données
 2. lancent K-means avec différentes valeurs croissant géométriquement de k et évaluent la classification obtenue en utilisant le test d'Anderson-Darling.
 3. arrêtent de faire croître k quand le critère ne montre plus d'amélioration,
 4. choisissent \hat{k} et la classification associée qui a obtenu le meilleur score par le test d'Anderson-Darling.

La première comparaison [91] entre cette méthode et G-means montre une faible amélioration sur l'estimation du nombre réel de clusters lorsque ceux-ci sont nombreux : des tests ont été faits pour 100 à 700 clusters réels. Les deux méthodes trouvent le même nombre de clusters (avec une différence allant de 1 à 14 clusters près en moyenne, soit une erreur différence relative de 1 à 2% en moyenne).

L'algorithme flou C-means

L'algorithme C-means [36, 89] peut-être vu comme une extension *floue* de K-means. En effet, avec K-means, chaque individu appartient à un seul cluster. C'est ce qu'on appelle

³⁶Notez que, dans le cadre de méthodes de SAS basées sur la parcimonie, ce n'est pas un point critique puisque ces méthodes sont naturellement limitées à un nombre relativement restreint de sources.

le *partitionnement dur*. C-means propose que chaque individu x_i appartienne à tous les clusters de centroïde c_j avec une valeur d'appartenance différente $u_{i,j}$ telle que

$$\sum_{j=1}^k u_{i,j} = 1, \quad \forall i \in \{1, \dots, Q\}. \quad (5.2)$$

La fonction de coût à minimiser (5.1) s'écrit alors

$$J = \sum_{j=1}^k \sum_{i=1}^Q u_{i,j}^m \|x_i - c_j\|^2, \quad m \geq 1, \quad (5.3)$$

où l'exposant $m \geq 1$ est le paramètre de distorsion qui permet d'ajuster le flou de la partition. L'algorithme peut être défini de la façon suivante.

1. Initialiser les valeurs $u_{i,j}$.
2. Calculer la position des centroïdes c_j selon la relation

$$c_j = \sum_{i=1}^Q \frac{u_{i,j}^m}{\sum_{i=1}^Q u_{i,j}^m} x_i. \quad (5.4)$$

3. Recalculer les valeurs des paramètres de distorsion selon la relation

$$u_{i,j} = \frac{1}{\sum_{p=1}^k \left(\frac{\|x_i - c_j\|}{\|x_i - c_p\|} \right)^{\frac{2}{m-1}}}. \quad (5.5)$$

4. Répéter les étapes 2 et 3 jusqu'à ce que les valeurs $u_{i,j}$ ne bougent plus.

Nous testerons cette méthode sur nos approches de SAS dans le chapitre 6.

5.1.3 Les méthodes basées sur les mélanges de densités

D'un point de vue probabiliste, on peut supposer que les individus d'un même cluster sont répartis selon une distribution donnée. Les individus de deux clusters différents sont donc générés selon différentes densités (ou la même classe de densités mais avec des paramètres différents). Le nuage de points est alors modélisé comme un *mélange* de ces distributions, c.-à-d. une combinaison linéaire de ces distributions. Supposons connues la probabilité a priori (*probabilité de mélange* ou *mixing probability* en anglais) $\mathbb{P}(C_j)$ du cluster C_j et la densité de probabilité conditionnelle $f(x_i|C_j, \underline{\theta}_j)$ (aussi nommée *densité de la composante* ou *component density* en anglais), où $\underline{\theta}_j$ est le vecteur de paramètres inconnus de la distribution des individus de C_j . Alors la densité de probabilité du nuage de points s'écrit de la façon suivante :

$$f(\underline{x}|\underline{\theta}) = \sum_{j=1}^k f(\underline{x}|C_j, \underline{\theta}_j) \mathbb{P}(C_j), \quad (5.6)$$

où $\underline{\theta} = (\theta_1, \dots, \theta_k)$ et $\sum_{j=1}^k \mathbb{P}(C_j) = 1$. Si le vecteur $\underline{\theta}$ est connu, la probabilité a posteriori pour assigner un individu à un cluster peut facilement être calculée à l'aide du théorème de Bayes. En pratique, les mélanges sont souvent construits à partir de densités gaussiennes, de paramètres différents.

L'approche par maximum de vraisemblance permet d'estimer $\underline{\theta}$, en considérant que la meilleure estimée est celle qui maximise la densité conjointe qui s'écrit, sous réserve d'indépendance des composantes x_i de \underline{x} :

$$f(\underline{x}|\underline{\theta}) = \prod_{i=1}^Q f(x_i|\underline{\theta}), \quad (5.7)$$

ce qui s'écrit aussi sous forme logarithmique :

$$l(\underline{\theta}) \triangleq \ln f(\underline{x}|\underline{\theta}) = \sum_{i=1}^Q \log [f(x_i|\underline{\theta})]. \quad (5.8)$$

La meilleure estimation est obtenue en résolvant les équations de log-vraisemblance :

$$\frac{\partial l(\underline{\theta})}{\partial \theta_j} = 0 \quad \forall j = 1, \dots, k. \quad (5.9)$$

Comme les solutions à cette équation ne peuvent généralement pas être calculées de façon analytique, des approches sous-optimales ont été proposées, notamment une approche basée sur la maximisation de l'espérance (EM) [70]. Cet algorithme considère les données comme incomplètes et divise chaque individu x_i en deux parties comprenant les données observables x_i^g et les données manquantes $x_i^m = \{x_{i1}^m, \dots, x_{ik}^m\}$, avec x_{ij}^m valant 1 ou 0 selon que x_i appartienne au cluster C_j ou pas. En combinant les Eq. (5.6) et (5.8) pour les données complètes $x_i = \{x_i^g, x_i^m\}$, la log-vraisemblance du nuage de points des données complètes s'écrit :

$$l(\underline{\theta}) = \sum_{i=1}^Q \sum_{j=1}^K x_{ij}^m \log [\mathbb{P}(C_j) f(x_i^g|\underline{\theta}_j)]. \quad (5.10)$$

Pour résoudre l'Eq. (5.9) à partir de la log-vraisemblance des données complètes (5.10), l'algorithme standard EM génère une série de paramètres estimés $\{\underline{\theta}^0, \dots, \underline{\theta}^N\}$, où N est ici le nombre d'itérations avant que l'algorithme converge, selon les itérations suivantes :

1. Initialiser $\underline{\theta}^0$ et poser $n = 0$.
2. Etape d'espérance : calculer l'espérance de la log-vraisemblance des données complètes :

$$\mathcal{L}(\underline{\theta}, \underline{\theta}^n) = \mathbb{E} [\log f(\underline{x}^g, \underline{x}^m|\underline{\theta}) | \underline{x}^g, \underline{\theta}^n]. \quad (5.11)$$

3. Etape de maximisation : sélectionner un nouveau paramètre tel que

$$\underline{\theta}^{n+1} = \arg \max_{\underline{\theta}} \mathcal{L}(\underline{\theta}, \underline{\theta}^n). \quad (5.12)$$

4. Poser $n \leftarrow n + 1$ et répéter les étapes 2 et 3 jusqu'à la convergence des paramètres $\underline{\theta}$.

Les principaux inconvénients de l'algorithme EM sont :

- sa sensibilité aux paramètres initiaux,
- l'effet d'une matrice de covariance singulière,
- la possibilité de convergence vers un optimum local
- et une convergence lente.

On peut montrer que les algorithmes EM et K-means sont étroitement liés : en effet, si on suppose que les mélanges sont gaussiens sphériques, alors ces deux algorithmes sont équivalents [52].

5.2 Les algorithmes proposés pour les méthodes de SAS à parcimonie

5.2.1 CBCE : une méthode d'élimination de clusters basée sur la cardinalité

Comme nous l'avons vu au début de ce chapitre, nous désirons éviter les "fausses détections" qui nous empêchent d'estimer correctement la matrice de mélange. En effet, en parcourant la liste ordonnée des zones d'analyse dans l'étape de détection, il arrive qu'une "fausse zone mono-source" soit trouvée et que la colonne estimée de paramètres b_{im} soit gardée, entraînant une chute des performances. Fort heureusement, si on se limite au début de la liste ordonnée, ces "mauvaises" zones d'analyse sont minoritaires par rapport aux "bonnes". Nous proposons donc une première méthode qui permet d'éviter cet effet indésirable en comptant le nombre de zones d'analyse qui correspondent à la même source. Les fausses détections étant peu nombreuses, le nombre de zones d'analyse qui correspondront à une "mauvaise source" sera donc faible par rapport au nombre de zones qui correspondront aux vraies sources. Nous allons voir plus en détail comment fonctionne cette méthode. Pour illustrer ce principe, nous allons utiliser la méthode AD-TIFCORR-CF et plus précisément l'étape de détection et d'identification des colonnes de coefficients d'échelle b_{im} . Notons qu'il est possible d'appliquer cette méthode de clustering à AD-TIFFROM-CF, AD-TICORR-CT ou AD-TIFFROM-CT, en utilisant les paramètres propres à ces approches.

L'idée principale est de former des clusters d'individus correspondant à des colonnes estimées de paramètres b_{im} . Pour cela, nous ne considérons que les zones telles que

$$\overline{|r_{x_1 x_i}(T, \omega_l)|} \geq 1 - \epsilon_4, \quad (5.13)$$

où ϵ_4 est un petit seuil positif défini par l'utilisateur. Nous ne gardons donc que les zones mono-sources qui se trouvent au début de la liste ordonnée dans l'étape de détection³⁷. Nous considérons successivement chacune des zones d'analyse ci-dessus et nous les utilisons d'une façon légèrement différente de la procédure d'identification basique présentée dans la section 4.2.1.

Ici encore, pour chaque zone d'analyse considérée, on prend comme valeur des paramètres b_{im} associés à une colonne de $B(\omega)$ les modules des $I_i(T, \omega_l)$ définis dans l'Eq.

³⁷ Ainsi, nous limitons naturellement le coût de calcul et espérons limiter le nombre de fausses détections.

(4.17). La colonne estimée associée à la première zone dans la liste ordonnée est gardée comme le premier point dans le premier cluster. Chaque colonne estimée dans les autres zones est utilisée comme suit. Nous calculons sa distance par rapport à tous les clusters créés à cette étape, en définissant la distance par rapport à un cluster comme la distance par rapport au premier point qui a été inclus dans celui-ci, c.-à-d. le "meilleur" point selon notre critère mono-source. Si une telle distance est en-dessous d'un seuil ϵ_1 défini par l'utilisateur, cette nouvelle colonne est insérée comme un nouveau point de ce cluster. Sinon, cette nouvelle colonne est gardée comme le premier point d'un nouveau cluster. Ceci est répété pour toutes les zones d'analyse qui vérifient la condition (5.13). Si le seuil ϵ_1 est suffisamment petit, le nombre de clusters ainsi créés est au moins égal au nombre N connu de sources à extraire³⁸. Nous gardons alors les N clusters qui contiennent le plus de points. Pour chaque cluster, nous sélectionnons en pratique un représentant, en prenant le point qui correspond à la plus grande valeur de $|r_{x_1 x_i}(T, \omega_l)|$, et donc, la valeur estimée pour la meilleure zone mono-source³⁹.

Nous obtenons alors les N colonnes d'estimées des paramètres b_{im} . Cette méthode basée sur l'élimination de certains clusters en fonction du nombre d'individus qu'ils contiennent s'appelle CBCE (pour *Cardinality Based Cluster Elimination*). Rappelons encore une fois que CBCE doit être considéré comme une méthode simple, particulière pour estimer ces paramètres de manière plus robuste qu'avec la méthode basique proposée dans le chapitre 4. De plus, l'utilisateur doit encore sélectionner beaucoup de paramètres, de façon à obtenir les meilleures performances. Certains de ces paramètres heuristiques, comme par exemple le seuil qui indique si nous avons trouvé une zone mono-source correspondant à une nouvelle source en calculant la distance entre deux colonnes de paramètres b_{im} , peuvent être évités. En effet, certaines méthodes générales de clustering que nous avons présentées précédemment ou l'algorithme proposé dans la section 5.2.3 permettent d'ajuster *automatiquement* ces seuils, sans aucune connaissance sur les valeurs de b_{im} .

5.2.2 l'algorithme DEMIX de S. Arberet et al. [12, 13]

Dans [12, 13], S. Arberet et al. proposent une nouvelle méthode de clustering qu'ils appellent DEMIX et qui estime le nombre de clusters et la direction des sources, c.-à-d. les colonnes de paramètres de la matrice de mélange⁴⁰. Cette méthode a été développée pour être utilisée avec des méthodes de SAS basées sur la parcimonie et peut être vue comme une version améliorée de la méthode précédente décrite dans la section 5.2.1. Comme cette dernière, elle est originale car elle tire parti de la "qualité" des données à classer. En effet, les auteurs proposent de mesurer la *qualité mono-source* dans le voisinage de chaque fenêtre TF (n, ω) . Ils doivent donc classer des directions θ des sources, couplées à

³⁸En pratique, nous avons constaté que le nombre de clusters obtenu était bien supérieur à N .

³⁹D'autres représentants, tels que le centre de gravité des points dans le cluster considéré, peuvent être utilisés.

⁴⁰DEMIX a d'abord été développée pour les méthodes LI [12] et une version étendue aux mélanges AD, nommée DEMIX Anechoic, a été récemment proposée [13]. Ces méthodes sont très proches puisque dans les deux cas, le critère d'appariement des individus dans le même cluster est la distance entre les paramètres d'atténuation b_{im} . Aussi, dans la suite de cette description, nous nous limiterons aux mélanges LI.

une "mesure de confiance" τ . Cette mesure de confiance est l'équivalent de nos critères mono-sources définis pour les méthodes TIFROM et TIFCORR. L'algorithme peut être résumé de la façon suivante.

- **Initialiser k clusters** : DEMIX crée itérativement k clusters $C_k \subset P$ où P représente le nuage de points à classer, en initialisant $k = 0$ et $P_k = P_0 = P$.
 1. Trouver le point (θ_k, τ_k) dans P_k avec la plus grande mesure de confiance τ_k ,
 2. créer le cluster C_k avec tous les points (θ, τ) dans P "suffisamment proches" du centroïde (θ_k, τ_k) ,
 3. créer un nouveau nuage de points $P_{k+1} \triangleq P_k \setminus C_k$. Si $P_{k+1} = \emptyset$, alors on arrête. Sinon $k \leftarrow k + 1$ et on retourne au point 1.

Remarquons que les clusters C_k peuvent être non-disjoints : en effet, les individus du cluster C_k sont, par définition, les individus du nuage de point initial les plus proches du centroïde (θ_k, τ_k) .

- **Estimation des directions** : puisque les clusters C_k peuvent être non-disjoints, les auteurs proposent de ne considérer que les points "non-biaisés" (θ', τ') qui appartiennent exclusivement à un cluster C_k . Ces points sont supposés indépendants et distribués selon $\theta' = \mathcal{N}(\theta_k^{\text{true}}, \sigma_{\theta'}^2(\tau'))$, où θ_k^{true} est la direction inconnue à estimer, $\sigma_{\theta'}^2(\tau') = \frac{1}{M-1} \frac{\tau}{(\tau-1)^2}$ et M est le nombre de fenêtres TF dans la zone d'analyse où θ' a été estimé. Un sous-ensemble $C'_k \subset C_k$ est ainsi créé et le centroïde estimé $\hat{\theta}_k^c$ est choisi comme l'estimateur non-biaisé à variance minimale de θ_k^{true} .
- **Elimination des clusters** : cette étape consiste à supprimer les "faux" clusters. Si le vrai nombre de clusters est connu, les auteurs proposent de garder les centroïdes à variance minimale. Sinon, ils proposent de supprimer les directions d'un cluster s'il existe un autre cluster tel que la distance entre leurs centroïdes respectifs est sous un seuil défini par l'utilisateur. Ils recalculent ensuite les directions, comme proposé dans les précédentes étapes.

Remarquons que cette méthode nécessite elle-aussi le réglage de plusieurs paramètres heuristiques (seuil pour trouver les individus "suffisamment proches" des centroïdes dans l'étape d'initialisation des k clusters, seuil pour éliminer un cluster dans l'étape d'élimination).

5.2.3 Notre approche modifiée de K-means pour la séparation de sources

Dans la section 5.1.2, nous avons présenté plusieurs méthodes de classification qui ont l'avantage de classer sans paramètre les individus (c.-à-d. les colonnes estimées de b_{im} dans notre problème de SAS) et dans les sections 5.2.1 et 5.2.2, de nouvelles approches qui tiennent compte de la "qualité mono-source" des zones d'analyse des zones TF à classer mais qui nécessitent des paramètres heuristiques. Nous proposons donc de tirer parti de ces deux familles de méthodes.

Ainsi, dans l'étape de détection décrite dans la section 4.2.1, nous avons vu que les coefficients d'atténuation b_{im} sont classés selon les valeurs décroissantes de la moyenne par rapport à i du paramètre $|r_{x_1 x_i}(T, \omega_l)|$ défini dans l'Eq. (4.16). Nous proposons alors

de profiter de cette condition dans une version modifiée de K-means qui nous présentons ici.

Deux modifications qui peuvent être appliquées indépendamment sont alors faites.

- La première concerne la sensibilité de la méthode aux points aberrants : plutôt que d'utiliser la moyenne dans les calculs, nous utilisons la médiane, d'où le nom de K-medians de cette méthode proposée par ailleurs dans la littérature [48]. Nous considérons que la "médiane" des colonnes estimées b_{im} d'un cluster est le vecteur constitué des médianes sur i de ces valeurs b_{im} . Par rapport à l'algorithme original décrit dans la section 5.1.2, la troisième étape de la méthode est modifiée ainsi : on recalcule les centroïdes de telle façon que leurs coordonnées sont les médianes des coordonnées des individus des clusters définis dans l'étape 2.
- La seconde concerne l'initialisation des centroïdes. Nous proposons de les choisir comme les premiers points "distants" de la liste ordonnée définie dans l'Eq. (5.13), c.-à-d. les valeurs obtenues dans les zones TF où $|r_{x_1 x_i}(T, \omega)|$ prend les plus grandes valeurs, comme dans les procédures présentées dans les sections 4.2.1 et 5.2.1. Ainsi, dans l'étape d'initialisation de K-means ou K-medians, le premier centroïde est fixé égal à la première colonne de paramètres b_{im} estimée. Dans la liste ordonnée des zones mono-sources, chacun des $k - 1$ autres centroïdes est rendu égal à une colonne estimée de coefficients b_{im} si la distance entre cette colonne et les centroïdes précédemment trouvés est au-dessus d'un seuil ϵ_1 . Au contraire de l'approche basique ou de notre première version de classification présentée dans la section 5.2.1, ce seuil n'a qu'un rôle mineur. Il permet en effet d'éviter à un centroïde de se focaliser sur une fausse détection et de laisser les $k - 1$ clusters restant partitionner les données.

Nous allons maintenant décrire le formalisme des méthodes de SAS faisant appel à cette version modifiée de K-means.

5.2.4 Description des approches de SAS utilisant notre variante de K-means

Nous venons de voir les modifications que nous proposons à K-means pour améliorer les performances de nos méthodes TF de SAS, c.-à-d. l'utilisation de K-medians et l'initialisation des centroïdes comme les "meilleurs" individus (au sens de notre critère mono-source). En conséquence, les approches AD-TIFCORR-CF et AD-TIFCORR-CT utilisant cette approche⁴¹ peuvent être résumées de la façon suivante :

AD-TIFCORR-CF et K-medians

La méthode présentée ici est sensiblement identique à la version basique proposée dans le chapitre précédent, à l'exception de la détection de zones mono-sources et l'identification de colonnes de paramètres b_{im} qui peuvent être résumées de la façon suivante :

1. Dans un premier temps, nous ne considérons que les zones TF qui vérifient l'Eq. (5.13). Quand le nombre de ces zones dépasse 1000, nous prenons les 1000 zones

⁴¹Bien évidemment, cette méthode de classification peut aussi être utilisée avec les approches AD-TIFFROM-CF et AD-TIFFROM-CT. Cependant, les méthodes en résultant ne seront pas décrites ici pour simplifier la lecture de ce document.

qui maximisent $\overline{|r_{x_1 x_i}(T, \omega_l)|}$. En effet, alors que la méthode CBCE ne conserve que les "meilleures" estimées des b_{im} (au sens de notre critère mono-source), quels que soient le nombre de clusters et leur cardinalité, l'algorithme K-medians (ou K-means) présenté dans ce chapitre tient compte de tous les individus, même s'ils correspondent à des estimées "grossières" de ces coefficients d'échelle. En particulier, quand le nombre de zones qui vérifient l'Eq. (5.13) est grand (typiquement de l'ordre de plusieurs milliers à plusieurs dizaines de milliers), nous avons remarqué que cette méthode de classification identifiait avec une faible précision ces coefficients, parce que le nombre d'estimées "grossières" était plus important que le nombre d'estimées "précises"⁴². Ce résultat est important car de nombreuses méthodes de SAS font appel à des méthodes comme K-means sur l'ensemble des points du domaine d'analyse. Nos résultats expérimentaux suggèrent donc qu'il est possible d'augmenter considérablement les capacités de séparation de ces méthodes si on fait appel à un critère mono-source et si on limite la classification aux "meilleurs" points seulement. Nous validons expérimentalement ce choix dans l'annexe F. Dans ces zones, les colonnes de paramètres b_{im} sont estimées comme les modules des paramètres $I_i(T, \omega_l)$ définis dans l'Eq. (4.17).

2. Nous initialisons les N centroïdes c_j comme expliqué dans la section 5.2.3 ci-dessus⁴³ qui maximisent la moyenne du paramètre (4.16).
3. Lancer K-medians.
4. Estimer la matrice de coefficients d'atténuation comme les centroïdes finaux de l'algorithme de classification.

AD-TIFCORR-CT et K-medians

Comme nous l'avons souligné dans la section 4.3, l'approche AD-TIFCORR-CT est intéressante car elle permet d'estimer dans les mêmes zones mono-sources les paramètres $b_{im'}$ et $\mu_{im'}$ associés. Par conséquent, nous ne devons pas classer ici uniquement des colonnes de coefficients d'atténuation mais des colonnes de "points" $(b_{im'}, \mu_{im'})$. L'approche globale peut être résumée de la façon suivante :

1. Calculer les TFCT des observations.
2. Sélectionner les zones TF vérifiant la condition (5.13) transposée aux zones d'analyse à temps constant. Calculer les couples de colonnes $(b_{im'}, \mu_{im'})$ comme cela a été

⁴²De manière alternative, on pourrait choisir une valeur plus faible pour le seuil ϵ_3 dans l'Eq. (5.13). Néanmoins, nous avons remarqué qu'alors, quand les décalages temporels sont relativement importants, il arrivait que certaines sources n'apparaissent pas comme visible à cause de cette faible valeur de seuil. Le choix de limiter à 1000 zones au plus est donc un compromis acceptable : d'un côté, quand les zones mono-sources sont nombreuses, les 1000 premières suffisent à estimer nos coefficients b_{im} . Quand elles sont peu nombreuses, notre seuil ϵ_3 indique que les zones d'analyse conservées présentent quand même une "qualité mono-source" relativement correcte.

⁴³S'il arrivait que le nombre N de centroïdes initiaux ne soit pas trouvé avec le seuil ϵ_1 , nous proposons d'abaisser la valeur de ce seuil de façon à ce que ce nombre soit atteint. En particulier, dans le cas où $N = 2$, cela revient à chercher la colonne b_{im} la plus distante du centroïde déjà trouvé parmi toutes les colonnes estimées disponibles.

proposé dans la section 4.3.1. Au besoin, limiter à 1000 le nombre de ces zones (en choisissant les zones qui maximisent $\overline{r_{x_1 x_i}(n_{p'}, \Omega)}$, comme discuté ci-dessus).

3. Initialiser les N centroïdes c_j comme expliqué dans la section 5.2.3 ci-dessus.
4. Lancer K-medians.
5. Estimer les paramètres de mélange de la matrice $B(\omega)$ comme les centroïdes finaux de la méthode de clustering.

5.3 Conclusion

Nous avons expliqué les limitations des méthodes présentées dans le chapitre 4 : lorsque les décalages temporels sont importants, il devient difficile d'estimer avec précision tous les paramètres de mélange. C'est pourquoi nous avons présenté dans ce chapitre diverses techniques de classification non-supervisée. Nous avons proposé une méthode simple de classification et vu une extension de cette approche proposée par S. Arberet et al. [12]. Cependant, ces deux approches nécessitent l'utilisation de paramètres heuristiques, ce qu'il vaudrait mieux éviter, afin d'automatiser nos méthodes de SAS. C'est pourquoi nous avons étudié plusieurs méthodes classiques de classification. En particulier, la méthode K-means permet d'éviter l'utilisation de ces paramètres. Nous avons montré ses limitations (sensibilité aux points aberrants et initialisation des centroïdes) et avons proposé des solutions pour chacun de ces problèmes.

Nous allons maintenant montrer les performances de nos méthodes dans le chapitre suivant.

Chapitre 6

Performances des méthodes à atténuations et décalages

Les résultats présentés dans ce chapitre visent à illustrer les performances des méthodes AD de SAS présentées dans les chapitres 4, 5 et l'annexe D. Nous montrerons en particulier l'apport des méthodes de clustering détaillées dans le chapitre 5. Dans la section 6.1, nous présenterons le protocole expérimental (sources, mélange, etc) ainsi que les critères pour mesurer les performances de nos méthodes. Dans la section 6.2, nous étudierons les performances des approches de classification, appliquées à une méthode TF à fréquence constante et à une méthode TF à temps constant. Enfin, nous comparerons dans la section 6.3 les performances des méthodes de SAS, appliquées avec la meilleure des méthodes de classification ci-dessus.

6.1 Conditions des tests et critères de performances

Dans ce chapitre, nous présentons un grand nombre de tests réalisés avec les méthodes présentées dans ce manuscrit, à savoir AD-TIFCORR-CF, AD-TIFCORR-CT, AD-TIFFROM-CF et AD-TIFFROM-CT. Ces méthodes ont été testées dans les conditions suivantes.

- Nous avons utilisé les jeux de signaux sources numéro 2 et 6 présentés dans la section 3.1.1. Nous rappelons que ces signaux sont des enregistrements de parole en anglais, échantillonnés à 20kHz. Les sources du jeu 2 sont extraites de celles du jeu 6, ce dernier contenant des phases de silence temporel en supplément.
- Nous avons créé plusieurs mélanges AD artificiels de ces sources.
- Nous avons traité ces signaux mélangés avec les méthodes AD-TIFCORR-CF, AD-TIFFROM-CF, AD-TIFCORR-CT et AD-TIFFROM-CT resp. définies dans le chapitre 4 et l'annexe D, mais faisant appel aux méthodes de classification vues dans le chapitre 5.

Le critère de performances des méthodes de SAS utilisé ici est toujours le SIR défini dans la section C.2 de l'annexe C. Nous mesurons le SIR^{out} ou le $SIRI$, toujours liés par la relation (C.10), c.-à-d. :

$$(SIRI)_{dB} = (SIR^{out})_{dB} - (SIR^{in})_{dB}. \quad (6.1)$$

Les paramètres de performances sont mesurés pour tous les tests où les méthodes considérées réussissent à identifier N colonnes de $B(\omega)$, où N est le nombre de sources considérées. Comme nous le verrons ci-dessous, les rares cas où l'identification de ces colonnes échoue correspondent aux situations où la largeur temporelle des TFCT est suffisamment grande pour rendre les sources invisibles dans les représentations TF correspondantes. Rappelons aussi que les paramètres estimés $\mu_{im'}$ sont à valeurs entières. Nous pouvons ainsi vérifier si tous les paramètres $\mu_{im'}$ sont *exactement* égaux aux valeurs théoriques. Le pourcentage des cas où cette condition est vérifiée, parmi tous les essais (c.-à-d. comprenant les essais où quelques colonnes de $B(\omega)$ n'ont pas été identifiées), est utilisé comme critère additionnel de performances et est appelé le "taux de succès" ci-dessous. Il est clair néanmoins que même si une méthode de SAS considérée ne "réussit" pas à identifier tous les $\mu_{im'}$ *exactement*, elle peut quand même être capable d'identifier toutes les colonnes de $B(\omega)$ et d'apporter des performances acceptables en termes de SIR^{out} (ou $SIRI$), pourvu que les estimations de $\mu_{im'}$ ne soient pas trop éloignées de leurs valeurs théoriques. Ce "taux de succès" défini ci-dessus est donc un critère plus pessimiste et partiel que le SIR^{out} global.

Dans ce chapitre, nous ne considérons que le cas où $N = 2$, c.-à-d. la configuration correspondant à deux mélanges de deux sources. Tous nos tests ont pour but d'estimer l'influence des coefficients d'échelle et des décalages temporels sur les performances des méthodes proposées. Nous utilisons donc des matrices de mélange symétriques définies par

$$A(\omega) = \begin{bmatrix} 1 & \lambda e^{-j\omega\eta} \\ \lambda e^{-j\omega\eta} & 1 \end{bmatrix}, \quad (6.2)$$

où nous considérons différentes valeurs réelles pour les facteurs d'échelle λ , de façon à faire varier le taux de mélange et donc le SIR^{in} associé aux observations $x_i(n)$. De façon similaire, l'influence des décalages temporels a été étudiée en faisant varier les paramètres entiers η .

Les valeurs que nous avons considérées pour λ et η sont

$$\begin{cases} \lambda = 0.5, 0.9 \\ \eta = 0, 10, 20 \end{cases} \quad (6.3)$$

Les SIR^{in} sont resp. égaux à 6.0 et 0.9 dB pour $\lambda = 0.5$ et 0.9.

Comme expliqué précédemment, les méthodes proposées utilisent des représentations TF des observations $x_i(n)$, obtenues en calculant leurs TFCT $X_i(n, \omega)$. Plus précisément, ce type de représentation est utilisé deux fois dans les méthodes AD-TIFCORR-CF et AD-TIFFROM-CF, c.-à-d. d'abord quand nous considérons les zones d'analyse à fréquence constante, utilisées pour estimer les paramètres b_{im} , et dans les zones à temps constant, utilisées pour estimer les paramètres $b'_{im'}$ et $\mu_{im'}$. Ces deux types de zones d'analyse peuvent amener à des valeurs optimales différentes, autant pour la taille des TFCT que pour le nombre de telles fenêtres TF adjacentes par zone d'analyse. Ainsi, nous considérons indépendamment deux ensembles de paramètres, resp. associés aux deux types de zones d'analyse dans les méthodes AD-TIFCORR-CF et AD-TIFFROM-CF, c.-à-d. :

- nous notons d (resp. d') le nombre d'échantillons des observations $x_i(n)$ dans chaque fenêtre temporelle des TFCT dans les zones d'analyse à fréquence constante (resp. à temps constant),

- comme expliqué précédemment dans les sections 4.2.1 et 4.2.2, M (resp. M') est le nombre de fenêtres adjacentes dans les zones d'analyse à fréquence constante (resp. à temps constant),
- nous notons ρ (resp. ρ') le recouvrement temporel entre les fenêtres dans les TFCT utilisées dans les zones d'analyse à fréquence constante (resp. à temps constant).

Au contraire, les méthodes AD-TIFCORR-CT et AD-TIFROM-CT n'utilisent qu'un seul type de représentation TF et donc qu'un seul jeu de paramètres, c.-à-d. d' , M' et ρ' .

Comme nous l'avons noté dans la section 4.2.1 quand nous avons développé nos méthodes AD, nous avons supposé que les décalages temporels étaient petits par rapport à la largeur temporelle des fenêtres $h(\cdot)$ utilisées dans le calcul des TFCT. Ainsi, dans les tests présentés dans ce chapitre, nous pouvons uniquement supposer que les performances seront bonnes quand $\eta \ll \min\{d, d'\}$ pour les méthodes à fréquence constante et $\eta \ll d'$ pour les méthodes à temps constant. Nous considérons uniquement les situations où cette condition est rigoureusement vérifiée⁴⁴ dans ce chapitre, c.-à-d. où nous avons typiquement $\frac{\eta}{\min\{d, d'\}} \leq \frac{1}{10}$.

Les deux types de TFCT utilisent une fenêtre de Hanning $h(\cdot)$. L'influence des paramètres des zones d'analyse à fréquence constante a été étudiée en détail dans le chapitre 3 pour le cas LI. Aussi, pour des raisons de simplicité, dans les tests que nous détaillons ici, nous ne considérons que les valeurs optimales trouvées dans ce chapitre 3 c.-à-d. :

- $d = 256$,
- $M = 10$,
- $\rho = 75\%$.

Au contraire, nous analysons en détail l'influence des paramètres associés aux zones à temps constant que nous avons introduites dans la section 4.2.2. Cette analyse est en particulier motivée par le fait que les paramètres des TFCT ont une influence sur l'estimation des décalages temporels d'après [28]. Nous avons fait varier ce jeu de paramètres de la manière suivante :

- Le nombre d' d'échantillons dans chaque TFCT a varié géométriquement de 512 à 16384 échantillons, avec un pas de 2.
- Le nombre M' de fenêtres TF par zone d'analyse a été successivement fixé à 4, 8 ou 16 pour $d' = 512$. Cette gamme de valeurs a varié géométriquement avec d' . Ainsi, les largeurs des bandes fréquentielles associées aux domaines Ω des zones d'analyse $(n_{p'}, \Omega)$ ont pris les mêmes valeurs, quel que soit d' . Dans ces tests, ces valeurs étaient 156.25, 312.5 et 625 Hz.
- Le recouvrement temporel ρ' entre deux fenêtres TF a été successivement fixé à 50%, 75% et 90%. Ce paramètre a varié indépendamment de d' et M' .

Les autres paramètres étaient constants dans ces tests, c.-à-d. :

- Pour les méthodes basiques et celles utilisant notre première méthode de classification décrite dans la section 5.2.1, la distance entre deux colonnes identifiées b_{im} de la matrice B était mesurée comme la plus grande valeur absolue entre les éléments de ces vecteurs qui ont le même indice. Comme pour les tests des méthodes LI, le seuil de distance pour accepter une nouvelle colonne était fixé à 0.15.

⁴⁴Il serait intéressant de pousser l'étude à des décalages temporels plus grands, comme nous en discuterons en conclusion de ce chapitre.

- Pour AD-TIFCORR-CF (resp. AD-TIFCORR-CT), les zones d'analyse (T, ω_l) (resp. $(n_{p'}, \Omega)$) ont été détectées si les moyennes des coefficients de corrélation étaient au-dessus d'un seuil $1 - \epsilon_4$, comme montré dans l'Eq. (5.13). Ce seuil a été fixé à $1.5e - 2$. Pour AD-TIFFROM-CF (resp. AD-TIFFROM-CT), les zones d'analyse (T, ω_l) (resp. $(n_{p'}, \Omega)$) ont été détectées si le critère (D.13) (resp. (D.22)) était en-dessous d'un seuil fixé à $1.5e - 2$.

Nous avons testé dans ces zones les méthodes de classification⁴⁵ suivantes :

- la méthode CBCE décrite dans la section 5.2.1,
- la méthode K-means, avec initialisation des centroïdes, comme expliqué dans le chapitre 5,
- l'approche K-medians, avec l'initialisation des centroïdes,
- la méthode floue C-means (nommée *Fuzzy C-means* dans les tableaux) disponible dans [66], en initialisant les centroïdes comme dans les deux approches ci-dessus et en fixant le paramètre de distorsion à $m = 1.5$.
- Pour les méthodes à fréquence constante, le seuil qui définit quelles zones d'analyse $(n_{p'}, \Omega)$ sont gardées (voir le paragraphe "Couplage des paramètres b_{im} et $(b'_{im'}, \mu_{im'})$ " de la section 4.2.3) a été fixé à 0.1.
- La distance entre deux colonnes identifiées b_{im} et $b'_{im'}$, qui permettent de coupler les paramètres b_{im} et $\mu_{im'}$ dans AD-TIFCORR-CF et AD-TIFFROM-CF (comme expliqué dans la section 4.2.3), était mesurée comme la plus grande valeur absolue entre les éléments de ces vecteurs qui ont le même indice. Le seuil de distance pour accepter une nouvelle colonne était fixé à 0.05.

Pour chaque méthode de SAS considérée, 54 tests ont été effectués dans les conditions définies ci-dessus, pour chaque couple de signaux, chaque matrice de mélange et chaque méthode de clustering.

6.2 Comparaison des méthodes de classification

Dans cette section, nous testons les méthodes de classification ci-dessus sur les deux types de méthodes AD que nous avons développées (les méthodes à fréquence constante et celles à temps constant), de façon à trouver quelle technique de clustering permet d'améliorer au mieux leurs performances. Afin d'en simplifier l'étude, nous ne considérons que les approches basées sur la variance, c.-à-d. AD-TIFFROM-CF pour les méthodes à fréquence constante et AD-TIFFROM-CT pour les méthodes à temps constant.

6.2.1 Classification dans les méthodes à fréquence constante

Dans les méthodes à fréquence constante, les estimées b_{im} sont indépendantes des paramètres d' , ρ' et M' que nous avons fait varier dans ces tests. Le tableau 6.1 présente la norme de Frobenius de la différence entre les matrices estimées et théoriques de paramètres b_{im} , en fonction de la méthode de classification utilisée dans AD-TIFFROM-CF et des valeurs de η . Nous avons aussi testé les versions basiques, c.-à-d. sans classification, de nos méthodes de SAS de façon à pouvoir comparer leurs performances avec celles des

⁴⁵Nous avons développé nos propres versions de K-means et K-medians.

TAB. 6.1 – Norme de Frobenius des différences entre les matrices théoriques de paramètres b_{im} et leur estimées obtenues par AD-TIFROM-CF, en fonction de la méthode de classification utilisée, du jeu de sources et des paramètres λ et η de la matrice de mélange définie dans l'Eq. (6.2).

λ	Jeu	Méthode de clustering	η		
			0	10	20
0.5	2	Aucune	0.0006	0.0803	0.0204
		CBCE	0.0006	0.0803	0.0204
		K-means	0.0001	0.0213	0.0119
		K-medians	0.0003	0.0143	0.0135
		Fuzzy C-means	0.0001	0.0210	0.0129
	6	Aucune	0.0005	0.0040	0.0107
		CBCE	0.0005	0.0040	0.0107
		K-means	0.0004	0.0057	0.0075
		K-medians	0.0002	0.0050	0.0123
		Fuzzy C-means	0.0004	0.0061	0.0079
0.9	2	Aucune	0.0001	0.0162	0.0190
		CBCE	0.0001	0.0190	0.0190
		K-means	0.0003	0.0064	2.7859
		K-medians	0.00002	0.0083	0.0060
		Fuzzy C-means	0.0004	0.0085	0.0879
	6	Aucune	0.0001	0.0019	0.0116
		CBCE	0.0001	0.0019	0.0116
		K-means	0.0001	0.0035	0.0025
		K-medians	0.00003	0.0029	0.0066
		Fuzzy C-means	0.0001	0.0045	0.0046

TAB. 6.2 – Moyennes des normes de Frobenius, obtenues par chaque méthode de classification avec AD-TIFROM-CF, en fonction de η et valeurs globales pour $\eta = 0-20$.

Méthode de clustering	η			
	0	10	20	0-20
Aucune	0.0003	0.0256	0.0154	0.0138
CBCE	0.0003	0.0263	0.0154	0.0140
K-means	0.0002	0.0092	0.7019	0.2371
K-medians	0.0001	0.0076	0.0096	0.0058
Fuzzy C-means	0.003	0.0100	0.0283	0.0138

méthodes utilisant la classification. Mis à part un cas, les normes sont faibles, ce qui démontre que nos méthodes identifient les paramètres b_{im} avec une grande précision le plus souvent.

Notons que pour le jeu 2, $\lambda = 0.9$ et $\eta = 20$, K-means ne permet d'estimer précisément les coefficients d'atténuation. En effet, dans ce cas, quelques points aberrants induisent un biais dans le calcul de la moyenne. Notons que dans cette même situation, C-means, elle-aussi basée sur des moyennes (pondérées), y est relativement sensible. Au contraire, l'approche K-medians, n'est pas perturbée par ces valeurs puisqu'elle apporte dans ce cas les meilleures performances.

Etudions maintenant plus en détail les performances de chaque méthode de classification. Remarquons tout d'abord que l'approche basique de AD-TIFROM-CF et son extension utilisant CBCE apportent les mêmes performances (sauf dans un cas). C'est normal puisque CBCE garde comme centroïdes les colonnes "suffisamment distantes" qui sont estimées au début de la liste ordonnée des zones d'analyse, tout comme l'approche basique. Cependant, on note que dans un cas (jeu 2, $\lambda = 0.9$ et $\eta = 10$), l'approche CBCE apporte un norme de Frobenius légèrement plus élevée que l'approche basique. Ceci est peu gênant ici parce que les normes de Frobenius sont proches. Dans les tests décrits dans cette section, nous n'avons pas rencontré la situation où la méthode basique garde une colonne issue d'une "fausse" détection.

Les méthodes "automatisées", c.-à-d. sans seuil heuristique, donnent de meilleures performances que la méthode basique ou CBCE, sauf dans la situation vue précédemment où K-means (et C-means dans une moindre mesure) échoue. Dans ce cas précis, l'approche K-medians permet l'estimation la plus précise parmi toutes les méthodes proposées. Au contraire, dans le cas du jeu 6, $\lambda = 0.5$ et $\eta = 20$, elle propose la moins bonne estimation (mais la norme est relativement basse, ce qui indique que l'estimation des coefficients b_{im} est quand même précise).

Le tableau 6.2 donne les moyennes des normes de Frobenius obtenues avec chaque méthode de classification, pour tous les couples de signaux et les valeurs de λ , en fonction de la valeur de η . A la lecture de ce tableau, il est clair que l'approche qui apporte les meilleures performances est K-medians, suivie de la méthode basique, C-means et CBCE. Les méthodes basées sur la moyenne sont ici pénalisées par le mauvais résultat obtenu lorsque le nuage de points à classer avait des points aberrants qui ont perturbé l'estimation.

TAB. 6.3 – Performances de AD-TIFROM-CT avec les méthodes de classification, pour chaque jeu de sources, chaque valeur de λ , $\eta = 0, 10$ ou 20 et performances moyennes pour $\eta = 0-20$.

Méthode de Clustering	λ	Critère de perf.	η			
			0	10	20	0-20
Aucune	0.5	SIR^{out}	54.9	38.4	31.6	41.7
		% de succès	88.9	73.2	82.4	81.5
	0.9	SIR^{out}	56.6	34.0	24.0	38.3
		% de succès	100	75	60.2	78.4
CBCE	0.5	SIR^{out}	56.2	41.3	36.7	44.7
		% de succès	93.5	85.2	79.7	86.1
	0.9	SIR^{out}	57.2	37.4	30.3	41.5
		% de succès	100	89.8	80.6	90.2
K-means	0.5	SIR^{out}	29.5	28.7	27.1	27.8
		% de succès	52.8	54.7	50.0	52.3
	0.9	SIR^{out}	38.8	34.6	24.7	29.7
		% de succès	76.8	63.9	50.9	31.0
Kmedians	0.5	SIR^{out}	42.2	36.4	30.8	36.5
		% de succès	91.7	78.7	76.0	82.1
	0.9	SIR^{out}	47.4	41.9	32.4	40.5
		% de succès	100	89.8	78.7	89.5
Fuzzy C-means	0.5	SIR^{out}	12.0	26.4	24.5	21.0
		% de succès	58.3	55.6	53.8	55.9
	0.9	SIR^{out}	17.8	33.2	25.8	25.6
		% de succès	56.5	64.8	59.3	60.2

Critères de performances : i) valeurs moyennes des SIR^{out} (en dB) et ii) taux de succès sur toutes valeurs de paramètres de la méthode de SAS considérée.

En conséquence, nous comparerons les méthodes AD-TIFROM-CF et AD-TIFCORR-CF quand elles sont utilisées avec K-medians.

6.2.2 Classification dans les méthodes à temps constant

Nous évaluons à nouveau nos méthodes de classification, mais elles sont cette fois appliquées à AD-TIFROM-CT. Nous rappelons que contrairement à AD-TIFROM-CF qui servait de méthode étalon dans la section 6.2.1, la méthode AD-TIFROM-CT utilise les méthodes de classification pour estimer non seulement les colonnes de coefficients $b_{im'}$ mais aussi les décalages temporels associés $\mu_{im'}$. La norme de Frobenius utilisée jusqu'à présent est donc ici un critère partiel pour mesurer la qualité d'estimation de nos paramètres de mélange. Nous proposons donc de juger la qualité des méthodes de classification considérées en recourant au SIR^{out} et au taux de succès.

Le tableau 6.3 donne les SIR^{out} moyens (sur les jeux de sources et les paramètres des zones à temps constant) et le taux de réussite obtenus par AD-TIFROM-CT avec chaque méthode de classification, en fonction de η . Ce tableau doit cependant être étudié avec attention : dans certains tests, il se peut qu'une colonne de $B(\omega)$ ne puisse être estimée, parce que la source correspondante n'est pas visible. En particulier, si les zones d'analyse sont grandes (grandes valeurs de d' par exemple), cette situation peut arriver. Rappelons qu'avec l'approche basique (resp. CBCE), une colonne de paramètres $b_{im'}$ (resp. un cluster) n'est créée que si sa distance par rapport aux autres colonnes (resp. clusters) est au dessus d'un seuil. Ainsi, si une source n'est pas visible, ces méthodes n'estiment qu'une sous-matrice de $B(\omega)$ et le SIR^{out} correspondant ne peut être calculé. Au contraire, si le nuage de points, c.-à-d. l'ensemble de toutes les colonnes des $b_{im'}$ estimées, n'est pas réduit à un singleton ou à l'ensemble vide, les méthodes "automatisées", c.-à-d. K-means, K-medians et C-means, estiment N colonnes qui ne correspondent pas aux réalités de mélange, entraînant une baisse des SIR^{out} moyens. En fonction du nombre de cas où certaines sources ne sont pas visibles dans ces tests, comparer seulement les SIR^{out} du tableau 6.3 ne nous permet pas de conclure sur le choix de la meilleure méthode de classification. C'est pourquoi nous allons plutôt nous focaliser sur le taux de succès dans un premier temps.

Avec la version basique de AD-TIFCORR-CT, c.-à-d. sans clustering, au moins une des sources n'a pas été visible dans 63 des 648 tests considérés. Ces 63 tests sont comptabilisés dans le "taux d'échec" (c.-à-d. le nombre de cas où les paramètres $\mu_{im'}$ estimés ne se pas exactement égaux aux valeurs théoriques) et représentent 48.5% de ces "échecs". Appliquée avec CBCE, la méthode n'a pas réussi à identifier au moins une colonne de $B(\omega)$ dans 45 des 648 tests considérés (58.4% des cas où AD-TIFCORR-CT n'a pu trouver exactement les décalages temporels). Avec les approches K-means, K-medians et C-means, ces cas sont resp. au nombre de 25, 12 et 6 cas, ce qui correspond resp. à 10.6%, 13.5% et 2.1% des "taux d'échec". Les nombres de cas où les sources sont invisibles avec les méthodes basique et CBCE sont relativement proches, ce qui nous permet de comparer les SIR^{out} obtenus avec celles-ci. Nous remarquons que CBCE apporte de bien meilleurs SIR^{out} et taux de succès que l'approche basique. Pour les mêmes raisons, nous pouvons comparer entre elles les méthodes "automatisées". Nous remarquons que K-medians est l'approche qui permet d'obtenir les meilleures performances, tant en termes de SIR^{out} que de taux de succès, suivie de K-means et C-means. Nous ne pouvons pas par contre comparer leurs performances à celles de CBCE à cette étape.

Comme le paramètre qui varie sur le plus grande gamme de valeurs est le nombre d' d'échantillons dans la TFCT, nous étudions les comportements des techniques de clustering en fonction de d' , en moyennant les performances par rapport aux autres paramètres (sauf λ). Ces informations sont données dans le tableau 6.4. Nous remarquons que les valeurs pour lesquelles toutes ces méthodes fonctionnent le mieux sont les valeurs intermédiaires de d' , c.-à-d. $d' = 1024, 2048$ ou 4096 . En nous limitant à cette gamme de valeurs, nous voyons que K-medians est l'approche qui apporte à nouveau les meilleures performances (son taux de succès est égal à 100 le plus souvent et ses SIR^{out} sont les plus élevés lorsque $d' = 1024$ ou 2048).

Nous appliquerons donc à nouveau K-medians pour montrer les capacités expérimentales de nos deux méthodes de SAS à temps constant.

TAB. 6.4 – Performances globales de AD-TIFROM-CT, pour chaque valeur de λ et chaque méthode de classification, en fonction du nombre d' d'échantillons dans le calcul des TFCT.

Méthode de clustering	λ	Critères de perf.	d'					
			512	1024	2048	4096	8192	16384
Aucune	0.5	SIR^{out}	34.6	42.5	46.9	48.3	40.4	30.7
		% de succès	59.3	83.3	98.1	96.3	70.4	38.9
	0.9	SIR^{out}	28.1	34.5	43.0	46.1	42.9	31.7
		% de succès	55.6	68.5	98.1	100	87.0	61.1
CBCE	0.5	SIR^{out}	41.9	45.6	47.2	50.3	43.5	34.8
		% de succès	90.8	96.3	100	98.2	85.2	46.3
	0.9	SIR^{out}	38.2	41.0	43.5	46.1	45.0	31.7
		% de succès	87.1	100	100	100	88.9	66.1
K-means	0.5	SIR^{out}	36.0	48.0	43.2	27.4	9.2	-4.4
		% de succès	87.1	98.1	84.3	47.2	5.5	0
	0.9	SIR^{out}	44.1	50.1	45.9	24.3	10.7	3.2
		% de succès	83.3	92.6	98.1	53.7	27.8	11.1
K-medians	0.5	SIR^{out}	50.5	56.4	50.2	41.0	26.2	-5.4
		% de succès	83.3	100	100	98.1	79.6	37.0
	0.9	SIR^{out}	46.4	53.2	50.8	44.0	31.2	20.1
		% de succès	83.3	100	100	100	90.7	63.0
Fuzzy C-means	0.5	SIR^{out}	35.6	49.8	40.4	29.6	7.8	-10.3
		% de succès	83.3	100	86.1	58.3	8.3	0
	0.9	SIR^{out}	27.5	43.2	48.5	35.1	17.3	1.2
		% de succès	77.8	94.4	100	75	40.7	5.6

Critères de performances : les mêmes que dans le tableau 6.3.

TAB. 6.5 – Norme de Frobenius des différences entre les matrices théoriques de paramètres b_{im} et leurs estimées obtenues par AD-TIFCORR-CF et AD-TIFFROM-CF, en fonction des paramètres λ et η de la matrice de mélange définie dans l'Eq. (6.2).

Méthode de SAS	λ	η		
		0	10	20
AD-TIFCORR-CF	0.5	0.0007	0.0130	0.0379
	0.9	0.00003	0.0065	0.0142
AD-TIFFROM-CF	0.5	0.0003	0.0097	0.0129
	0.9	0.00003	0.0056	0.0063

6.3 Performances des méthodes de SAS

6.3.1 Comportement des méthodes AD-TIFCORR-CF et AD-TIFFROM-CF

Dans cette section, nous comparons les performances des méthodes de SAS à fréquence constante, c.-à-d. AD-TIFCORR-CF et AD-TIFFROM-CF, quand elles sont toutes les deux utilisées avec la méthode de classification K-medians.

Comme nous l'avons vu précédemment, les coefficients b_{im} sont estimés indépendamment des décalages temporels $\mu_{im'}$. Le tableau 6.5 montre les variations des normes de Frobenius de la différence entre les matrices théoriques et estimées par AD-TIFCORR-CF et AD-TIFFROM-CF, en fonction de la valeur de η . Nous remarquons que ces normes sont basses, ce qui démontre que nos méthodes de SAS sont capables d'estimer les coefficients d'échelle avec une grande précision.

Le tableau 6.6 montre les performances globales de AD-TIFCORR-CF et AD-TIFFROM-CF (les SIR^{out} sont moyennés sur les paramètres des zones à temps constant). Dans ce tableau, le SIR^{out} varie entre 24.9 et 81.3 dB (avec une valeur moyenne sur toutes les valeurs de η de 51.4 dB) pour AD-TIFCORR-CF et entre 33.4 et 79.1 (valeur moyenne : 51.9 dB) pour AD-TIFFROM-CF. Ce tableau montre que les SIR^{out} décroissent de façon significative quand η augmente. Cependant, le "taux de succès" (en fonction de l'estimation *exacte* des décalages $\mu_{im'}$) reste assez haut, avec une valeur globale de 87.0% pour les deux méthodes de SAS. Nous espérons d'aussi bonnes performances parce que toutes les hypothèses faites dans nos méthodes sont vérifiées ici. Remarquons que, contrairement aux méthodes LI, AD-TIFCORR-CF est globalement un peu moins performante que AD-TIFFROM-CF.

Analysons maintenant avec plus d'attention l'influence des jeux de sources considérés sur les performances. De la même façon que pour les b_{im} , on s'attend à ce que le jeu 6 apporte de meilleures estimées des paramètres $\mu_{im'}$ que le jeu 2. Le taux de succès dans le tableau 6.6 confirme ce résultat. De plus, les SIR^{out} montrent le même phénomène.

Les tableaux 6.7-6.9 apportent une analyse plus détaillée de certains aspects des tests ci-dessus : ils contiennent les valeurs globales des critères de performances quand on considère toutes les valeurs de η (c.-à-d. $\eta = 0, 10$ ou 20), mais chacun de ces tableaux

TAB. 6.6 – Performances des méthodes AD-TIFCORR-CF et AD-TIFFROM-CF pour chaque jeu de sources, chaque valeur de λ , $\eta = 0, 10$ ou 20 et performances moyennes pour $\eta = 0-20$.

Méthode de SAS	λ	Jeu	Critère de perf.	η			
				0	10	20	0-20
AD-TIFCORR-CF	0.5	2	SIR^{out}	62.7	36.4	28.0	42.3
			% de succès	83.3	72.2	68.5	74.7
		6	SIR^{out}	81.5	44.8	37.5	54.6
			% de succès	94.4	88.9	85.2	90.7
	0.9	2	SIR^{out}	77.1	33.3	24.9	45.1
			% de succès	100	77.8	84.6	84.6
		6	SIR^{out}	81.3	45.5	37.1	54.6
			% de succès	100	98.1	96.3	98.1
AD-TIFFROM-CF	0.5	2	SIR^{out}	66.6	38.0	35.9	46.8
			% de succès	83.3	72.2	68.5	74.7
		6	SIR^{out}	75.8	49.3	41.9	55.7
			% de succès	94.4	88.9	88.9	90.7
	0.9	2	SIR^{out}	78.6	33.4	34.4	48.8
			% de succès	100	75.9	77.8	84.6
		6	SIR^{out}	79.1	49.8	40.1	56.3
			% de succès	100	98.1	96.3	98.1

Critères de performances : les mêmes que dans le tableau 6.3.

TAB. 6.7 – Performances globales de AD-TIFCORR-CF et AD-TIFFROM-CF pour les deux jeux de sources, $\eta = 0-20$ et chaque valeur de λ , en fonction du recouvrement ρ' entre les fenêtres TF.

Méthode de SAS	λ	Critère de perfs.	ρ'		
			50	75	90
AD-TIFCORR-CF	0.5	SIR^{out}	45.7	50.0	49.7
		% de succès	75.0	87.0	86.1
	0.9	SIR^{out}	50.4	49.6	49.6
		% de succès	92.6	90.7	90.7
AD-TIFFROM-CF	0.5	SIR^{out}	48.0	53.1	52.7
		% de succès	75.0	88.0	86.1
	0.9	SIR^{out}	52.9	52.2	52.5
		% de succès	92.6	90.7	91.7

Critères de performances : les mêmes que dans le tableau 6.3.

détaille les variations de ces critères en fonction d'un des paramètres des méthodes de SAS (en moyennant sur les autres). Le tableau 6.7 montre les variations du SIR^{out} et du taux de succès en fonction du recouvrement ρ' des TFCT. Au vu des résultats, on choisira de préférence la valeur intermédiaire testée, c.-à-d. $\rho' = 75\%$ pour les deux méthodes de SAS.

Le tableau 6.8 contient les variations des SIR^{out} moyens et des taux de succès par rapport au nombre d' de points dans les TFCT. Ce tableau montre que le taux de succès dépend de manière significative de la valeur de d' . Les meilleurs résultats sont obtenus pour les valeurs intermédiaires de d' testées ici, c.-à-d. $d' = 2048$ ou 4096 , qui permettent d'obtenir des taux de succès souvent égaux à 100% (et toujours supérieurs à 98%). Ce phénomène peut être expliqué ainsi : si nous calculons les TFCT avec trop peu de points, alors la précision fréquentielle de nos représentations TF est basse et nous ne sommes pas en mesure d'estimer précisément les décalages $\mu_{im'}$, ce qui est en accord avec [28]. Au contraire, si le nombre d'échantillons est trop important, il est plus difficile de trouver des zones TF mono-sources dans le jeu 2, parce que la largeur temporelle des TFCT est trop grande. Notons que pour ces valeurs de d' , les SIR^{out} moyens (sur les autres paramètres sauf λ) sont toujours au-dessus de 52 dB pour les deux approches de SAS.

Comme expliqué précédemment, le nombre M' de fenêtres dans les zones d'analyse $(n_{p'}, \Omega)$ définit la largeur fréquentielle de ces zones. Le tableau 6.9 montre qu'il vaut mieux fixer ce paramètre aux valeurs intermédiaires ou élevées testées, c.-à-d. $M' = 32$ ou 64 pour $d' = 2048$. Ce résultat était prévisible puisque, quand le recouvrement ρ' augmente, les chances de trouver une zone d'analyse mono-source augmentent elles-aussi. Avec ces paramètres optimaux, sur tous les mélanges considérés, le SIR^{out} moyen vaut 54.8 dB (avec les deux valeurs de M') pour AD-TIFCORR-CF et 55.9 dB (avec les deux valeurs de M') pour AD-TIFFROM-CF.

En conclusion, les méthodes à fréquence constante doivent être utilisées de préférence avec $d' = 2048$ échantillons et $\rho' = 75\%$ (ou 90%) de recouvrement dans le calcul des

TAB. 6.8 – Performances globales de AD-TIFCORR-CF et AD-TIFOMR-CF pour les deux jeux de sources, $\eta = 0-20$ et chaque valeur de λ , en fonction de la largeur temporelle d' des TFCT (en échantillons).

Méthode de SAS	λ	Critères de perf.	d'					
			512	1024	2048	4096	8192	16384
AD-TIFCORR-CF	0.5	SIR^{out}	47.0	53.3	54.2	54.0	49.3	33.1
		% de succès	76.0	94.4	100	98.1	83.3	44.4
	0.9	SIR^{out}	48.4	52.5	53.0	52.4	52.0	40.8
		% de succès	87.0	98.1	100	98.1	96.3	68.6
AD-TIFFROM-CF	0.5	SIR^{out}	49.5	56.9	57.8	57.3	52.4	33.6
		% de succès	75.9	96.3	100	98.1	85.2	42.6
	0.9	SIR^{out}	50.8	55.3	55.9	55.9	55.2	42.9
		% de succès	87.0	98.1	100	100	98.1	68.5

Critères de performances : les mêmes que dans le tableau 6.3.

TAB. 6.9 – Performances globales des méthodes AD-TIFCORR-CF et AD-TIFFROM-CF pour les deux jeux de sources, $\eta = 0-20$ et pour chaque valeur de λ , en fonction de la largeur des zones d'analyse à temps constant (en Hz).

Méthode de SAS	λ	Critère de perf.	Largeur fréquentielle		
			156.25	312.5	625
AD-TIFCORR-CF	0.5	SIR^{out}	46.1	50.2	49.1
		% de succès	72.2	88.0	88.0
	0.9	SIR^{out}	48.0	51.0	50.6
		% de succès	84.3	95.4	94.4
AD-TIFFROM-CF	0.5	SIR^{out}	48.6	53.0	52.2
		% de succès	7.1	87.0	88.9
	0.9	SIR^{out}	50.5	54.0	53.2
		% de succès	85.2	95.4	94.4

Critères de performances : les mêmes que dans le tableau 6.3.

TAB. 6.10 – Performances de AD-TIFCORR-CT et AD-TIFFROM-CT, pour chaque valeur de λ , $\eta = 0, 10$ ou 20 et performances moyennes pour $\eta = 0-20$.

Méthode de SAS	λ	Jeu	Critère de perf.	η			
				0	10	20	0-20
AD-TIFCORR-CT	0.5	2	SIR^{out}	40.7	33.7	30.7	35.0
			% de succès	77.8	70.4	77.8	75.3
		6	SIR^{out}	56.8	43.4	42.0	43.4
			% de succès	96.3	83.3	81.5	87.0
	0.9	2	SIR^{out}	44.8	35.6	29.8	36.7
			% de succès	100	70.4	70.4	80.2
		6	SIR^{out}	55.7	46.8	42.9	48.5
			% de succès	98.1	87.0	88.9	91.4
AD-TIFFROM-CT	0.5	2	SIR^{out}	33.4	32.1	24.0	29.8
			% de succès	87.0	75.9	66.7	76.5
		6	SIR^{out}	51.1	40.7	37.5	43.1
			% de succès	96.3	87.0	81.2	89.5
	0.9	2	SIR^{out}	43.3	41.2	29.8	38.1
			% de succès	100	83.3	70.4	84.6
		6	SIR^{out}	51.5	46.5	39.2	45.7
			% de succès	100	96.3	87.0	94.4

Critères de performances : les mêmes que dans le tableau 6.3.

TFCT. On peut alors fixer la largeur des zones d'analyse à temps constant à $M' = 32$. Enfin, la méthode de classification à privilégier est K-medians qui apporte de bonnes performances dans tous les cas, contrairement aux autres approches qui peuvent être perturbées par les points aberrants (K-means, C-means) ou nécessitent le réglage de certains paramètres heuristiques (CBCE).

6.3.2 Performances des méthodes AD-TIFCORR-CT et AD-TIFFROM-CT et comparaison avec les méthodes à fréquence constante

Dans cette section, nous analysons les performances des méthodes AD-TIFCORR-CT et AD-TIFFROM-CT resp. décrites dans les sections 4.3 et D.2, quand elles sont toutes les deux utilisées avec K-medians. Nous profiterons de cette étude pour comparer leurs performances avec celles des méthodes à fréquence constante testées dans la section 6.3.1.

Les performances globales de AD-TIFCORR-CT et AD-TIFFROM-CT sont données dans le tableau 6.10. Celui-ci montre que le SIR^{out} varie entre 30.7 et 56.3 dB pour AD-TIFCORR-CT (avec une valeur moyenne égale à 40.9 dB) et entre 24.0 et 51.5 dB pour AD-TIFFROM-CT (avec une moyenne égale à 38.5 dB). Le taux de succès sur tous ces tests est de 83.5% pour AD-TIFCORR-CT et 85.8% avec AD-TIFFROM-CT. Comme on pouvait le prévoir, les sources du jeu 6 apportent de meilleures performances que les sources du jeu 2, grâce aux zones de silence. Comme pour les méthodes à fréquence constante, le SIR^{out}

TAB. 6.11 – Performances globales de AD-TIFCORR-CT et AD-TIFFROM-CT pour les deux jeux de sources, $\eta = 0-20$ et chaque valeur de λ , en fonction du recouvrement ρ' entre les fenêtres TF.

Méthode de SAS	λ	Critère de perf.	ρ'		
			50	75	90
AD-TIFCORR-CT	0.5	SIR^{out}	35.3	42.6	45.6
		% de succès	78.7	84.2	80.6
	0.9	SIR^{out}	36.1	43.6	48.1
		% de succès	82.4	87.0	88.0
AD-TIFFROM-CT	0.5	SIR^{out}	32.0	37.3	40.0
		% de succès	81.5	83.3	84.3
	0.9	SIR^{out}	35.5	42.1	44.0
		% de succès	88.0	90.7	89.8

Critères de performances : les mêmes que dans le tableau 6.3.

décroît de façon significative quand η grandit mais les taux de succès restent relativement stables. On remarquera que AD-TIFCORR-CT a des SIR^{out} légèrement supérieurs à ceux de AD-TIFFROM-CT, tout en ayant un taux de succès légèrement inférieur. Cela signifie que, comme dans le cadre des mélanges LI, TIFCORR permet d'estimer de façon un peu plus précise les coefficients d'atténuation que TIFFROM.

En comparant leurs performances à celles des méthodes à fréquence constante dans le tableau 6.6, on remarque que, mis à part pour $\eta = 0$ (ou les méthodes à fréquence constante ont des SIR^{out} bien supérieurs), AD-TIFCORR-CT a des SIR^{out} équivalentes ou meilleures que AD-TIFCORR-CF. À l'inverse, AD-TIFFROM-CT voit ses SIR^{out} inférieurs à ceux de sa variante à fréquence constante.

Les tableaux 6.11-6.13 apportent une analyse détaillée de certains aspects des tests ci-dessus, en utilisant la même approche que dans les tableaux 6.7-6.9.

Le tableau 6.11 montre que les SIR^{out} sont meilleurs pour les valeurs élevées de ρ' alors que les taux de succès sont relativement peu sensibles à ce paramètre. En conséquence, ρ' sera fixé de préférence à 90%.

Le tableau 6.12 contient les variations des SIR^{out} moyens et des taux de succès des deux approches de SAS, en fonction du nombre d' de points dans le calcul des TFCT. Pour les mêmes raisons que dans la section 6.3.1, la valeur de d' sera choisie de préférence pour les valeurs intermédiaires testées : $d' = 1024$ (ou $d' = 2048$). Les SIR^{out} varient alors entre 50.2 et 61.5 dB.

En comparant ces résultats avec ceux du tableau 6.8, on remarque que les méthodes à temps constant sont plus sensibles aux choix de d' que les méthodes à fréquence constante. Par contre, pour leurs valeurs optimales de d' , ces dernières apportent des SIR^{out} comparables à ceux de leurs variantes à fréquence constante.

Le tableau 6.13 détaille les variations des SIR^{out} et du taux de succès en fonction du nombre M' de fenêtres dans les zones d'analyse. Ce tableau montre que M' doit être fixé à la plus petite valeur testée (ou alors la valeur intermédiaire pour maximiser le taux de

TAB. 6.12 – Performances globales de AD-TIFCORR-CT et AD-TIFFROM-CT, pour les deux jeux de sources, chaque valeur de λ , $\eta = 0-20$, en fonction de la taille d' des TFCT (en échantillons).

Méthode de SAS	λ	Critère de perf.	d'					
			512	1024	2048	4096	8192	16384
AD-TIFCORR-CT	0.5	SIR^{out}	58.6	61.4	54.2	42.3	30.9	-0.4
		% de succès	88.9	98.1	100	94.4	79.6	25.9
	0.9	SIR^{out}	55.3	61.5	54.7	44.7	30.1	9.2
		% de succès	87.0	100	100	96.3	83.3	48.1
AD-TIFFROM-CT	0.5	SIR^{out}	50.5	56.4	50.2	41.0	26.2	-5.4
		% de succès	83.3	100	100	98.1	79.6	37.5
	0.9	SIR^{out}	46.4	53.2	50.8	44.0	31.2	20.1
		% de succès	83.3	100	100	100	90.7	63.0

Critères de performances : les mêmes que dans le tableau 6.3.

succès), c.-à-d. $M' = 8$ (ou $M' = 16$) pour $d' = 1024$.

En conséquence, un compromis entre ces paramètres peut être obtenu en utilisant $d' = 1024$ (ou $d' = 2048$) échantillons et $\rho' = 90\%$ de recouvrement dans les calculs des TFCT. On peut ensuite fixer la longueur des zones d'analyse à temps-constant à $M' = 8$ ou $M' = 16$ (ou $M' = 16$ ou $M' = 32$ si $d' = 2048$). Avec ces paramètres optimaux, sur tous les tests, le SIR^{out} moyen sur tous les tests vaut 65.1 dB si $d' = 1024$ et $M' = 8$ (ou 63.5 dB si $M' = 16$) et 68.8 dB si $d' = 1024$ et $M' = 32$ (ou 63.3 si $M' = 64$) pour AD-TIFCORR-CT et 56.5 dB si $d' = 1024$ et $M' = 8$ (ou 57.1 dB si $M' = 16$) et 63.5 dB si $d' = 1024$ et $M' = 32$ (ou 55.9 si $M' = 64$) pour AD-TIFFROM-CT.

6.4 Conclusion

Dans ce chapitre, nous avons présenté les performances des méthodes TF AD décrites dans le chapitre 4 et l'annexe D, utilisées avec des méthodes de classification décrites dans le chapitre 5.

Nous avons tout d'abord comparé les performances des méthodes de classification, en les appliquant aux versions à fréquence constante et à temps constant de AD-TIFFROM. Nous avons montré que notre version de K-medians, initialisée avec les "meilleurs" individus (au sens de notre critère mono-source) du nuage de points était plus performante que K-means, C-means mais aussi CBCE. De plus, cette méthode nous permet de nous passer de paramètres tels la distance minimale entre deux colonnes de la matrice de mélange.

Nous avons ensuite appliqué cette méthode de classification à nos méthodes AD-TIFCORR-CF et AD-TIFFROM-CF d'une part, et AD-TIFCORR-CT et AD-TIFFROM-CT. Ces quatre méthodes ont montré de grandes performances : quand on choisit leurs paramètres optimaux, leurs SIR^{out} moyens varient entre 54.8 et 63.5 dB et sont en presque toujours supérieurs à 40 dB pour les deux familles de méthodes.

TAB. 6.13 – Performances globales des méthodes AD-TIFCORR-CT et AD-TIFFROM-CT pour les deux jeux de sources, $\eta = 0-20$ et pour chaque valeur de λ , en fonction de la largeur des zones d'analyse à temps constant (en Hz).

Méthode de SAS	λ	Critère de perf.	Largeur fréquentielle		
			156.25	312.5	625
AD-TIFCORR-CF	0.5	SIR^{out}	45.0	41.2	37.3
		% de succès	79.7	81.5	82.4
	0.9	SIR^{out}	44.6	45.1	38.0
		% de succès	77.8	93.5	86.1
AD-TIFFROM-CF	0.5	SIR^{out}	42.8	37.3	29.2
		% de succès	78.7	85.2	85.2
	0.9	SIR^{out}	44.2	40.6	36.9
		% de succès	86.1	95.6	89.8

Critères de performances : les mêmes que dans le tableau 6.3.

Cependant, les résultats expérimentaux ne donnent qu'un aperçu de ces méthodes. Il serait souhaitable de tester ces approches sur d'autres sources et sur plus de matrices de mélanges, tenant notamment compte de grands décalages temporels.

Chapitre 7

Extension des méthodes parcimonieuses aux mélanges convolutifs généraux

Dans les chapitres précédents, nous avons vu des méthodes à parcimonie, applicables aux mélanges LI et AD. Bien que leurs performances soient très intéressantes, la nature des modèles considérés limite ces approches à des mélanges simples et ne permet généralement⁴⁶ pas de traiter des mélanges acoustiques réels de "cocktail party".

Nous proposons ici une extension théorique de nos méthodes aux mélanges convolutifs généraux. En particulier, une extension de TIFCORR-NC servira à illustrer les concepts proposés dans ce chapitre, mais les méthodes TIFCORR-C et TIFROM peuvent aussi être étendues de la même façon aux mélanges convolutifs. Dans la section 7.1, nous poserons le problème de la séparation de mélanges convolutifs. Nous proposerons ensuite dans la section 7.2 notre extension aux mélanges convolutifs. Enfin, nous discuterons dans la section 7.3 du protocole expérimental à mettre en place pour valider cette méthode.

7.1 Position du problème

Dans ce chapitre, nous considérons la configuration de N sources et $P = N$ observations. Dans le cas AD que nous avons vu précédemment, les observations $x_i(n)$ s'écrivaient en fonction des sources $s_j(n)$ selon la relation (4.2).

Dans ce chapitre, nous considérons des mélanges plus généraux, qui tiennent compte du fait que chaque observation contienne plusieurs répliques atténuées et décalées des sources. Le modèle étendu aux mélanges convolutifs qui en résulte s'écrit :

$$x_i(n) = \sum_{j=1}^N \sum_{k=0}^{K-1} a_{ij,k} s_j(n - n_{ij,k}) \quad i = 1, \dots, N, \quad (7.1)$$

$$= \sum_{j=1}^N a_{ij}(n) * s_j(n) \quad i = 1, \dots, N. \quad (7.2)$$

⁴⁶Dans la littérature, on considère souvent que les méthodes AD permettent de séparer convenablement des mélanges en environnement (très) peu réverbérant.

Le problème décrit dans l'Eq. (7.1) et (7.2) correspond donc à la version discrétisée du problème convolutif général décrit dans les Eq. (3) et (4) (en supposant que les filtres de mélange inconnus des Eq. (3) et (4) puissent être modélisés par des filtres RIF causaux, ce qui est généralement supposé pour les méthodes convolutives de SAS) avec $P = N$.

La transformée de Fourier de l'Eq. (7.2) s'écrit :

$$X_i(\omega) = \sum_{j=1}^N A_{ij}(\omega) S_j(\omega) \quad i = 1, \dots, N, \quad (7.3)$$

où $A_{ij}(\omega)$ est la réponse fréquentielle du filtre de propagation de la source j sur l'observation i et s'écrit :

$$A_{ij}(\omega) = \sum_{k=0}^{K-1} a_{ij,k} e^{-j\omega(n+n_{ij,k})}, \quad i, j \in \{1, \dots, N\}. \quad (7.4)$$

Nous supposons que, pour toute pulsation ω , les coefficients $A_{1j}(\omega)$ définis dans l'Eq. (7.4) sont des nombres complexes non-nuls⁴⁷.

Mis à part que les filtres à estimer sont plus complexes, l'expression matricielle de l'Eq. (7.3) est de la même forme que celle pour les mélanges AD, donnée dans l'Eq. (4.4) et que nous rappelons ici :

$$\underline{X}(\omega) = A(\omega) \underline{S}(\omega), \quad (7.5)$$

avec $\underline{S}(\omega) = [S_1(\omega) \cdots S_N(\omega)]^T$, $\underline{X}(\omega) = [X_1(\omega) \cdots X_N(\omega)]^T$ et avec ici

$$A(\omega) = \begin{bmatrix} A_{11}(\omega) & \cdots & A_{1N}(\omega) \\ A_{21}(\omega) & \cdots & A_{2N}(\omega) \\ \vdots & & \vdots \\ A_{N1}(\omega) & \cdots & A_{NN}(\omega) \end{bmatrix}. \quad (7.6)$$

Les Eq. (7.3) et (7.5) prennent la même forme dans le domaine TF, si tous les décalages temporels $n_{ij,k}$ des répliques de la source j dans l'observation i sont suffisamment petits par rapport à la largeur temporelle de la transformée TF utilisée (c.-à-d. la TFCT) :

$$X_i(n, \omega) = \sum_{j=1}^N A_{ij}(\omega) S_j(n, \omega) \quad i = 1, \dots, N, \quad (7.7)$$

$$\underline{X}(n, \omega) = A(\omega) \underline{S}(n, \omega), \quad (7.8)$$

avec $\underline{S}(n, \omega) = [S_1(n, \omega) \cdots S_N(n, \omega)]^T$, $\underline{X}(n, \omega) = [X_1(n, \omega) \cdots X_N(n, \omega)]^T$ et $A(\omega)$ définie dans l'Eq. (7.6).

⁴⁷Rappelons que ce genre d'hypothèse sur les coefficients de mélange a aussi été fait pour les versions LI et AD de TIFCORR, ainsi que pour les autres méthodes présentées dans ce manuscrit. Comme les approches LI-TIFROM-S1 et LI-TIFROM-S2 présentées dans ce document nécessitent que tous les coefficients de mélanges soient non-nuls, leurs extensions aux mélanges convolutifs nécessitent que tous les coefficients $A_{ij}(\omega)$ soient non-nuls, pour toute pulsation ω .

Comme cela a été proposé pour les méthodes LI et AD dans les chapitres précédents et pour les méthodes convolutives dans [38–40], nous fixons les indéterminations de filtre en étendant l’approche [74] de Y. Deville aux mélanges considérés ici. Nous écrivons les observations en fonction des contributions des sources dans la première observation, à une permutation $\sigma(\cdot)$ près :

$$X_i(n, \omega) = \sum_{j=1}^N B_{ij}(\omega) S'_j(n, \omega) \quad i = 1, \dots, N, \quad (7.9)$$

avec

$$\begin{cases} S'_j(n, \omega) = a_{1, \sigma(j)}(\omega) S_{\sigma(j)}(n, \omega) \\ B_{ij}(\omega) = \frac{A_{i, \sigma(j)}(\omega)}{A_{1, \sigma(j)}(\omega)} \end{cases} \quad (7.10)$$

L’Eq. (7.9) peut alors s’écrire sous la forme matricielle :

$$\underline{X}(n, \omega) = B(\omega) \underline{S}'(n, \omega). \quad (7.11)$$

Nous allons maintenant introduire des méthodes pour estimer $B(\omega)$. Les signaux de sortie peuvent par exemple être obtenus dans le domaine fréquentiel (en interpolant la réponse fréquentielle des filtres) en calculant

$$\underline{Y}(\omega) = B^{-1}(\omega) \underline{X}(\omega), \quad (7.12)$$

où $\underline{Y}(\omega) = [Y_1(\omega) \cdots Y_N(\omega)]^T$ est le vecteur de transformées de Fourier des signaux de sorties. Les versions temporelles de ces signaux sont ensuite obtenues en appliquant une transformée de Fourier inverse à $\underline{Y}(\omega)$.

7.2 Extension de TIFCORR-NC aux mélanges convolutifs

Nous proposons dans cette section l’extension de TIFCORR-NC aux mélanges convolutifs. Dans [5–7], B. Albouy et Y. Deville avaient proposé des méthodes convolutives, basées sur une segmentation temporelle des observations, afin de trouver des zones d’analyse mono-sources et d’y estimer les filtres de mélange correspondants. Ces méthodes sont intéressantes car elles évitent le problème d’indétermination des permutations à chaque bande fréquentielle, discuté dans l’introduction générale. Néanmoins, elles imposent l’existence de zones temporelles mono-sources, ce qui est une hypothèse de parcimonie relativement forte.

Pour relâcher cette hypothèse, nous proposons une approche TF afin de revenir à plusieurs problèmes LI sur chaque bande de fréquence. En effet, pour chaque pulsation ω , le problème (7.11) est un problème LI. Nous pouvons alors appliquer la méthode temporelle LI-TEMPCORR-NC⁴⁸, transposée dans le domaine TF à la pulsation ω considérée. Bien

⁴⁸La méthode LI-TEMPCORR-NC proposée dans le chapitre 2 a été décrite dans un formalisme statistique pour des signaux réels (mais elle est applicable aux signaux complexes, comme cela est expliqué dans [74]). La transposition de cette approche dans le cas de nos mélanges convolutifs est ici réduite à un formalisme déterministe, pour des signaux à valeurs complexes.

sûr, des méthodes de classification décrites dans le chapitre 5 peuvent être appliquées afin d'étendre l'approche basique présentée ci-dessous.

En transposant cette approche temporelle au domaine TF, nous devons redéfinir nos hypothèses de parcimonie :

Hypothèse 11 *A chaque pulsation ω_l du domaine TF, les sources sont visibles dans le domaine temporel, c.-à-d. qu'il existe, pour chaque source et à chaque pulsation, une zone temporelle où celle-ci est isolée (au sens de la définition 8).*

Ces hypothèses de parcimonie sont beaucoup plus fortes que les hypothèses faites dans les chapitres précédents et notre premier but sera de vérifier si en pratique, de telles hypothèses sont vérifiées par des signaux réels "courts", c.-à-d. sans zone temporelle de silence. En pratique, si une source n'est pas visible à une pulsation ω_l donnée, la colonne de coefficients $B_{im}(\omega_l)$ correspondante sera fixée égale à 0 (sauf pour $i = 1 : B_{1m}(\omega_l) = 1$ par construction).

La méthode proposée, notée C-TIFCORR-NC⁴⁹, peut être résumée ainsi :

1. Une pré-étape consiste à calculer les transformées TF des observations.
2. Pour chaque pulsation ω_l , l'approche basique LI-TEPCORR-NC (ou une extension faisant appel aux techniques de clustering) est appliquée. La matrice de mélange $B(\omega_l)$ correspondante est calculée.
3. Une approche pour résoudre les indéterminations de permutation bande à bande est utilisée : entre deux pulsations adjacentes ω_l et ω_{l+1} , nous permutons éventuellement les colonnes de $B_{im}(\omega_{l+1})$ de façon à minimiser ce que leur distance par rapport à celles de $B_{im}(\omega_l)$. Une fois la matrice $B(\omega)$ entièrement estimée et les indéterminations de permutation résolues, nous avons alors estimé la réponse fréquentielle de nos filtres inconnus.

7.3 Conclusion et discussion

Nous avons présenté dans ce chapitre une méthode TF permettant de résoudre le problème de SAS pour des mélanges convolutifs. L'approche considérée consiste à calculer les transformées TF des observations, à considérer successivement chaque ω_l afin de calculer la matrice de mélange $B(\omega_l)$ correspondante à l'aide de l'une des méthodes temporelles de SAS présentées dans le chapitre 2 de ce document. Une approche basique est ensuite proposée pour résoudre le problème de permutation bande à bande de cette méthode. Pour cela, nous cherchons à minimiser la distance entre les colonnes de $B(\omega)$ entre deux pulsations adjacentes ω_l et ω_{l+1} . Une étude poussée des techniques existantes, comme par exemple [156, 215], serait nécessaire.

Les premiers tests devront montrer la validation expérimentale de l'hypothèse 11. Il serait intéressant d'étudier les techniques permettant de combler les trous dans la réponse fréquentielle.

⁴⁹Les méthodes C-TIFCORR-C, C-TIFFROM-S1 et C-TIFFROM-S2 ont les mêmes étapes. La différence est qu'à chaque bande fréquentielle, l'approche temporelle LI-TEPCORR-NC est resp. remplacée par LI-TEPCORR-C, LI-TEMPROM-S1 ou LI-TEMPROM-S2.

Conclusion générale

Cette thèse a abordé le problème de la séparation aveugle de sources (SAS) parcimonieuses. Plusieurs méthodes ont été proposées, en fonction du modèle de mélange choisi. Leur point commun est qu'elles cherchent dans le domaine d'analyse des zones où une seule source est active et estiment dans ces petites zones les paramètres de mélange. En dehors de ces zones, les sources peuvent être recouvrantes. En conséquence, ces méthodes nécessitent de faibles hypothèses de parcimonie des sources en comparaison d'autres méthodes parcimonieuses de SAS proposées dans la littérature, comme discuté dans l'état de l'art du chapitre 1.

Les premières études, que nous avons développées dans la partie I de ce document, considèrent des mélanges linéaires instantanés. Dans le chapitre 2, nous avons présenté les méthodes proposées précédemment par notre équipe qui ont servi de point de départ à nos travaux et proposé des premières extensions. Les deux méthodes temporelles et les deux méthodes temps-fréquence (TF) décrites présentent des structures semblables mais utilisent des critères complètement différents.

1. Les méthodes LI-TEMPROM et LI-TIFROM sont basées sur des rapports d'observations. A l'aide de la variance de ces rapports dans des *zones d'analyse*, c.-à-d. des petites zones du domaine d'étude, nous trouvons les zones *mono-sources*. Les colonnes de la matrice de mélange sont fixées égales aux moyennes des rapports d'observations dans ces zones. Nous avons mis en exergue le côté non-symétrique de la variance de ces rapports d'observations et avons ensuite proposé de résoudre ce problème en étudiant à la fois les précédents rapports d'observations mais aussi les rapports inverses correspondants. Cette symétrisation a lieu d'une part dans l'étape de détection des zones mono-sources (d'où les noms respectifs LI-TEMPROM-S1 et LI-TIFROM-S1 de ces variantes) et d'autre part dans l'étape d'identification, pour savoir quelles colonnes de la matrice de mélange estimée garder. Ces secondes extensions s'appellent resp. LI-TEMPROM-S2 et LI-TIFROM-S2.
2. Les méthodes LI-TEPCORR-C et LI-TIFCORR-C sont basées sur des rapports de coefficients de corrélation qui permettent, sous une certaine forme, de trouver les zones mono-sources et sous une autre forme d'identifier les colonnes de la matrice de mélange. Des variantes faisant appel à des versions non-centrées de ces coefficients de corrélation ont aussi été étudiées, d'où les noms respectifs des méthodes correspondantes : LI-TEPCORR-NC et LI-TIFCORR-NC.

Toutes ces méthodes peuvent aussi traiter le cas de sources cachées et/ou le cas de mélanges sous-déterminés. Leur principal souci commun est qu'elles nécessitent, comme la

majorité des méthodes à parcimonie, le réglage d'un nombre important de paramètres heuristiques. Néanmoins, nous avons montré dans le chapitre 3 que les performances étaient globalement très bonnes et que, une fois ces paramètres "bien réglés", les capacités de séparation étaient excellentes : les SIR^{out} obtenus étaient alors supérieurs à 60 dB (et même proches de 70 dB) pour $N = 2$ sources. En comparaison, les méthodes statistiques classiques disponibles dans le logiciel ICALAB ne permettaient d'obtenir que des SIR^{out} compris entre 0 et 40 dB. Quand $N = 4$ sources, nos méthodes ont montré de très bonnes performances puisque les SIR^{out} approchaient alors 50 dB tandis que le nombre et la qualité des zones d'analyse diminuaient. Nous avons aussi testé nos méthodes sur des signaux issus de données astrophysiques. Bien que ces données soient mal adaptées à nos méthodes (signaux très courts limitant les gammes de valeurs des paramètres heuristiques, peu de parcimonie), nos approches ont montré qu'elles étaient capables d'extraire des données intéressantes à partir des observations.

Bien que les méthodes ci-dessus soient très performantes, la simplicité du modèle de mélange limite naturellement leurs champs d'application, notamment dans le cadre de mélanges de signaux de parole. C'est pourquoi nous avons proposé dans la partie II de cette thèse des extensions de ces approches aux mélanges convolutifs.

Dans le chapitre 4, nous avons proposé deux extensions de LI-TIFCORR à des mélanges convolutifs simples : les mélanges à atténuations et décalages temporels (AD). Ces mélanges correspondent au cas où chaque observation est une combinaison linéaire de sources décalées temporellement et atténuées (mélange en chambre anéchoïque). La première extension de LI-TIFCORR fait appel à des zones d'analyse à fréquence constante (d'où son nom AD-TIFCORR-CF) pour estimer les coefficients d'échelle b_{im} et utilise ensuite des zones d'analyse à temps constant pour estimer les décalages temporels associés $\mu_{im'}$, à l'aide des droites des moindres carrés. L'utilisation de telles droites permet d'estimer une large gamme de décalages temporels, contrairement à d'autres méthodes AD, basées elles aussi sur la parcimonie et proposées par ailleurs dans la littérature. La seconde extension de LI-TIFCORR aux mélanges AD n'utilise que les zones à temps constant pour estimer tous les paramètres de mélange, d'où son nom AD-TIFCORR-CT. Ces deux extensions sont aussi applicables à TIFROM. Cependant, si on applique la méthodologie utilisée pour les mélanges LI, c.-à-d. si on ne considère qu'une seule zone mono-source par source afin d'extraire les paramètres de mélange, il se peut que certaines "fausses zones mono-sources" soient trouvées et gardées, engendrant de fortes baisses des performances.

Nous avons donc proposé d'étudier plus de zones mono-sources par source et avons montré l'apport de techniques de classification non-supervisée à ces estimations dans le chapitre 5. Nous avons notamment étudié la méthode K-means qui a été très utilisée en SAS. Nous avons montré ses limitations (importance de l'initialisation des centroïdes, sensibilité aux points aberrants) et les avons résolues : en initialisant les centroïdes comme des valeurs estimées des paramètres de mélanges dans les zones mono-sources, nous réduisons très fortement les risques de convergence vers un optimum local. De plus, en utilisant la médiane plutôt que la moyenne dans le calcul du déplacement des centroïdes, nous rendons la méthode K-medians résultante moins sensible aux points aberrants. Enfin, nous avons montré que, lorsque la taille du nuage de points était importante (plusieurs milliers de points), ne considérer qu'un sous-ensemble de ce nuage correspondant aux "meilleures"

zones mono-sources (au sens du critère de la méthode de SAS considérée) permettait d'améliorer grandement la qualité d'estimation des paramètres de mélange.

Nous avons testé nos méthodes AD et les techniques de clustering dans le chapitre 6 pour deux types de signaux de parole continue et discontinue, pour plusieurs matrices de mélange comprenant plusieurs valeurs de décalages temporels. Nos approches ont montré qu'elles apportaient de bonnes performances dans un nombre varié de cas : avec $N = 2$ et l'utilisation de la technique de classification K-medians, les SIR^{out} des méthodes étaient compris entre 24 et plus de 80 dB pour une gamme de décalages temporels compris entre 0 (mélanges LI) et 20 échantillons.

Enfin, dans le chapitre 7, nous avons montré comment étendre nos méthodes de SAS aux mélanges convolutifs généraux. Nous considérons que, pour chaque pulsation du domaine TF, les mélanges sont des mélanges LI et que nous pouvons les séparer en appliquant nos méthodes temporelles précédentes, transposées dans le domaine d'analyse considéré. Il reste cependant à vérifier que les nouvelles hypothèses de parcimonie faites sont satisfaites dans le cadre de mélanges de "cocktail party".

L'évolution de nos travaux a été dictée par le désir de pouvoir traiter les mélanges les plus généraux possibles. Nous avons donc proposé dans ce document plusieurs méthodes qui permettent de traiter des mélanges de plus en plus réalistes. Dans les parties expérimentales correspondantes, nous avons montré que les performances atteintes étaient très bonnes. Toutefois, les études à venir et les extensions possibles sont encore nombreuses, dont voici une liste non-exhaustive.

1. Nous avons présenté quelques résultats expérimentaux assez prometteurs pour nos méthodes AD. Nous devons étudier leurs comportements pour un plus grand nombre de sources et une plus grande gamme de décalages temporels.
2. L'extension aux mélanges convolutifs des approches ci-dessus nécessite des hypothèses de parcimonie beaucoup plus fortes que celles nécessaires pour traiter les mélanges LI ou AD. Une validation expérimentale de ces hypothèses serait à envisager. En particulier, si elles ne sont pas entièrement validées (cela signifie qu'à certaines fréquences, il se peut qu'au moins une colonne de la matrice de mélange ne soit pas estimée), il faudrait trouver un moyen de "combler les trous" obtenus dans l'estimation des réponses fréquentielles des filtres de mélange. Enfin, une caractérisation des performances de cette approche sera à fournir.
3. Nous pourrions aussi envisager l'extension des méthodes décrites dans ce document à d'autres signaux parcimonieux, comme les images. Il serait notamment intéressant d'étendre les méthodes AD à ces signaux 2D (dans ce cas, le décalage deviendrait spatial).
4. Récemment, les méthodes temporelles de SAS décrites dans ce document ont été étendues au cas de mélanges linéaires-quadratiques [81]. Il serait intéressant de développer d'autres approches pour des mélanges non-linéaires plus complexes.

Annexe A

Influence d'un faible niveau de bruit sur l'étape de détection

Dans les chapitres 2, 4 ou 7, nous avons considéré la configuration théorique sans bruit de méthodes à parcimonie développées resp. pour les mélanges LI, AD ou convolutifs.

Par exemple, pour la méthode LI-TEPCORR-C⁵⁰, les dénominateurs des paramètres de détection et d'identification (2.65) et (2.68) dépendent des variances des observations. Si toutes les sources sont de variance nulle, ces dénominateurs sont nuls aussi et les paramètres (2.65) et (2.68) ne sont pas définis. Donc, si nous ne devons que proposer une présentation formelle de la configuration non-bruitée, dans le but d'éviter cette singularité, nous devons imposer que dans toutes les aires temporelles considérées, au moins une source soit de variance non-nulle.

Considérons maintenant des signaux pratiques, qui proviennent typiquement d'un ensemble de capteurs. On peut alors raisonnablement supposer que cette nécessité sur les sources est rencontrée, parce que les signaux sources n'ont pas de variance strictement nulle. Mais de toute façon, il n'y a pas besoin d'avoir des signaux pratiques qui remplissent cette condition, parce que le bruit de capteur contenu dans des enregistrements réels évite cette singularité et permet à nos méthodes de SAS de fonctionner, comme nous allons le montrer. Considérons d'abord des aires temporelles où au moins une source a une variance significativement plus grande que le bruit de capteur. Dans de telles zones, l'influence du bruit est négligeable, de telle façon que la méthode fonctionne normalement, comme expliqué dans la section 2.3.1. Maintenant, dans les aires temporelles où toutes les sources ont des variances significativement plus petites que le bruit, chaque signal $x_i(t)$ est réduit à la contribution du bruit mesuré dans le capteur considéré. Dans des situations classiques, ces bruits de capteurs sont mutuellement décorrélés. Les coefficients de corrélation correspondants $\rho_{x_1 x_i}(t)$ sont donc faibles. En conséquence, les zones correspondantes ne sont pas insérées au début de la liste ordonnée dans l'étape de détection de la méthode de SAS proposée, et ne sont donc pas utilisées pour identifier une colonne de la matrice de mélange. Donc, nos méthodes de SAS manipulent correctement ces zones temporelles où toutes les sources ont une variance négligeable par rapport au bruit, et même une variance

⁵⁰Dans un but de clarté, nous nous limitons ici à l'influence du bruit sur la méthode LI-TEPCORR-C mais ce raisonnement peut être appliqué à toutes les méthodes décrites dans ce manuscrit.

presque nulle.

Nous avons validé expérimentalement cette analyse théorique comme suit. En partant des deux sources de 50000 échantillons du jeu numéro 2 défini dans la section 3.1.1, nous avons ajouté à la fin de chaque source 50000 échantillons qui étaient égaux à zéro. Nous avons mélangé ces signaux sources avec la matrice A_1 défini dans l'Eq. (3.2). Nous avons ensuite ajouté des composantes de bruit à ces mélanges de sources. Ces composantes de bruits étaient des signaux centrés gaussiens iid mutuellement indépendants. Leurs variances étaient sélectionnées de telle façon qu'elles plus basses de 40 dB par rapport à la plus faible variance des sources. Avec ce bruit de faible niveau, nous pouvions espérer que la méthode LI-TEPCORR-C soit toujours capable de détecter les zones temporelles mono-sources qu'elle détectait dans le cas sans bruit que nous avons étudié dans la section 3.1.5. Nous voulons ici vérifier que :

1. elle réussit à détecter ces zones temporelles, c.-à-d. elle les met toujours dans la liste ordonnée de zones mono-sources (éventuellement dans un ordre sensiblement différent que dans le cas non-bruité, parce que les petites composantes de bruit dans ces zones peut modifier légèrement les valeurs du paramètre de détection dans ces zones)
2. et elle gère correctement les nombreuses aires temporelles où les deux sources ont une variance bien plus faible que les composantes de bruit, c.-à-d. toutes les zones correspondant à la seconde partie des signaux source, où ces signaux ont une variance strictement nulle. Gérer ces zones correctement signifie que la méthode LI-TEPCORR-C ne les mette pas avant les zones mono-sources dans la liste ordonnée.

Les résultats à déduire de ces tests concernent donc la détection des zones temporelles mono-sources. Dans le but de comparer plus facilement ces résultats à ceux du cas non-bruité, nous avons utilisé les zones temporelles que nous avons détecté de la façon suivante. Nous avons de nouveau considéré les signaux source de 100000 échantillons et nous les avons à nouveau mélangé selon la matrice A_1 de l'Eq. (3.2), mais sans ajouter le bruit. En partant de ces observations, nous avons ensuite identifié les colonnes de la matrice pondérée permutée dans les zones temporelles que nous avons détecté ci-dessus. Nous avons ensuite déduit les signaux de sortie de notre système de SAS et le SIR^{out} associé. De façon répétitive, cette procédure était appliquée aux mêmes valeurs de paramètres que notre système de SAS dans la section 3.1.5. Le SIR^{out} moyen résultant est égal à 51.8 dB, son écart type à 17.5 dB et ses valeurs minimales et maximales sont resp. égales à 16.1 et 70.3 dB. Ces résultats doivent être comparés à ceux du jeu 2 dans la table 3.7, qui concerne le cas non-bruité étudié dans la section 3.1.5. Ils sont assez similaires (considérant le nombre limité de configurations étudiées dans ces tests), ce qui confirme que la méthode de SAS proposée gère correctement les aires temporelles où les sources ont des variances bien plus faibles que les composantes de bruit. Alors que nous n'avons considéré que la méthode LI-TEPCORR-C ci-dessus, le même type de commentaires s'applique aux extensions TF des méthodes de SAS proposées dans ce manuscrit.

Annexe B

Démonstration du critère mono-source de TEMPCORR et TIFCORR

Dans cette annexe, nous allons démontrer la validité du critère employé dans les diverses versions de TEMPCORR et TIFCORR pour trouver les zones où une source est isolée.

B.1 Démonstration du critère dans le cadre de mélanges LI

B.1.1 Démonstration du critère pour LI-TEMPCORR-C

Nous allons démontrer ici la validité de la condition nécessaire et suffisante (2.66). Nous étudions ici les propriétés des coefficients d'intercorrélation $\rho(x_1(t)x_i(t))$, définis dans l'Eq. (2.65), suivant le nombre de sources actives à l'instant t considéré. Nous supposons qu'au moins une source soit active. Sinon, ces coefficients $\rho(x_1(t)x_i(t))$ ne sont pas définis. Dans la suite de cette démonstration, dans un but de simplicité, les notations $s(t)$ et $x(t)$ que nous allons introduire ci-dessous font référence aux versions centrées des signaux.

Si seule une source est active, disons $s_l(t)$, alors il peut être montré que

$$\mathbb{E} \{x_i(t)x_k(t)\} = a_{il}a_{kl}\mathbb{E} \{s_k(t)^2\} \quad \forall i, k = 1, \dots, N. \quad (\text{B.1})$$

L'Eq. (2.65) permet alors d'obtenir l'égalité suivante :

$$|\rho_{x_1x_i}(t)| = 1 \quad \forall i = 2, \dots, N. \quad (\text{B.2})$$

Inversement, supposons que l'égalité (B.2) est validée. Nous avons alors (voir [185], Ed. 1965, p. 210)

$$\forall i = 2, \dots, N, \quad \exists \alpha_i / x_i(t) = \alpha_i x_1(t). \quad (\text{B.3})$$

Nous déduisons des Eq. (2.18) et (B.3) la relation suivante

$$\sum_{j=1}^N (a_{ij} - \alpha_i a_{1j})s_j(t) = 0 \quad \forall i = 2, \dots, N. \quad (\text{B.4})$$

En notant $\mathcal{A}(t)$ l'ensemble des indices des sources actives à l'instant t , on déduit de l'Eq. (B.4) que

$$\mathbb{E} \left\{ \left(\sum_{j \in \mathcal{A}(t)} (a_{ij} - \alpha_i a_{1j}) s_j(t) \right)^2 \right\} = 0 \quad \forall i = 2, \dots, N. \quad (\text{B.5})$$

L'hypothèse 7 implique alors que, pour tout $1 \leq i \leq N$,

$$a_{ij} - \alpha_i a_{1j} = 0, \quad \forall j \in \mathcal{A}(t). \quad (\text{B.6})$$

On peut montrer à partir de l'expression (B.6) que si plusieurs sources sont actives à l'instant t , alors les colonnes de A correspondantes sont proportionnelles. Cela est en contradiction avec l'hypothèse que la matrice soit de rang plein (hypothèse admise par toutes les méthodes de SAS).

En conséquence, l'Eq. (B.2) est vérifiée si et seulement si une seule source est active.

B.1.2 Démonstration du critère pour LI-TIFCORR-C

Nous allons démontrer maintenant la validité de la condition nécessaire et suffisante (2.73). En calculant les transformées TF des observations dans l'Eq. (2.18), en les centrant sur les zones d'analyse et en les écrivant sous forme vectorielle, nous obtenons l'expression suivante :

$$V_{x_i}(T, \omega) = \sum_{j=1}^N a_{ij} V_{s_j}(T, \omega) \quad i = 1, \dots, N. \quad (\text{B.7})$$

Les coefficients de corrélation $c_{x_i x_1}(t, \omega)$ des observations, définis dans l'Eq. (2.71), peuvent alors s'écrire en fonction des vecteurs $V_{x_i}(T, \omega)$:

$$c_{x_i x_1}(t, \omega) = \frac{\langle V_{x_i}(T, \omega), V_{x_1}(T, \omega) \rangle}{\|V_{x_i}(T, \omega)\| \cdot \|V_{x_1}(T, \omega)\|}, \quad (\text{B.8})$$

où les notations $\langle \cdot, \cdot \rangle$ et $\|\cdot\|$ correspondent resp. au produit scalaire et à la norme vectorielle. En appliquant la relation de Cauchy-Schwarz à l'Eq. (B.8), nous obtenons l'inégalité suivante :

$$|c_{x_i x_1}(T, \omega)| \leq 1 \quad \forall i = 1, \dots, N, \quad (\text{B.9})$$

avec l'égalité si et seulement si les vecteurs $V_{x_i}(T, \omega)$ et $V_{x_1}(T, \omega)$ sont linéairement dépendants.

Analysons maintenant cette condition, dans une zone d'analyse (T, ω) donnée, en fonction du nombre de vecteurs non-nuls $V_{s_j}(T, \omega)$, c.-à-d. le nombre de sources actives dans la zone d'analyse. D'après l'hypothèse 8, au moins un de ces vecteurs $V_{s_j}(T, \omega)$ n'est pas égal à zéro partout sur cette zone (T, ω) .

- Il est aisé de montrer que si un seul des vecteurs $V_{s_j}(T, \omega)$ est non-nul dans la zone d'analyse (T, ω) considérée, comme les coefficients a_{1j} sont supposés non-nuls, alors d'après l'Eq. (B.7), $V_{x_1}(T, \omega)$ est non-nul et les vecteurs $V_{x_i}(T, \omega)$ sont colinéaires, pour tout $i \in \{2, \dots, N\}$. Ainsi, la relation (2.73) est vérifiée.

- La difficulté réside donc à considérer au moins deux vecteurs $V_{s_j}(T, \omega)$ non-nuls. On peut montrer facilement que si les vecteurs $V_{x_i}(T, \omega)$ et $V_{x_1}(T, \omega)$ sont linéairement dépendants pour tout i , avec $2 \leq i \leq N$, alors grâce à l'hypothèse 8, toutes les colonnes de la matrice de mélange A avec les indices égaux aux indices j des sources non-nulles seraient colinéaires. Cela n'est pas possible puisque A est supposée inversible. Ainsi, au moins une paire de vecteurs $(V_{x_i}(T, \omega), V_{x_1}(T, \omega))$ ne consiste pas en deux vecteurs colinéaires, de telle façon que $|c_{x_i x_1}(T, \omega)| < 1$ et que la relation (2.73) n'est pas vérifiée.

En conséquence, la relation (2.73) est vérifiée si et seulement si exactement un des vecteurs $V_{x_i}(T, \omega)$ n'est pas égal à zéro dans la zone d'analyse considérée, c.-à-d. si une source est isolée dans cette zone.

B.1.3 Démonstration du critère pour LI-TIFCORR-NC

En utilisant le raisonnement de la section B.1.2 ci-dessus et en se basant sur l'hypothèse 9, nous pouvons facilement montrer que la relation (2.73) est vérifiée si et seulement si une source est isolée.

B.2 Démonstration du critère pour AD-TIFCORR-CT

Considérons les modules des observations $X_i(n, \omega_l)$ comme des vecteurs de dimension M' :

$$V_{|x_i|}(n, \Omega) = [|X_i(n, \omega_1)|, \dots, |X_i(n, \omega_{M'})|]^T. \quad (\text{B.10})$$

Les paramètres de corrélation $R'_{x_i x_q}(n, \Omega)$ peuvent alors être exprimés en fonction d'un produit scalaire :

$$R'_{x_i x_q}(n, \Omega) = \frac{1}{M'} \sum_{l'=1}^{M'} |X_i(n, \omega_{l'})| \cdot |X_q(n, \omega_{l'})| \quad (\text{B.11})$$

$$= \frac{1}{M'} \langle V_{|x_i|}(n, \Omega), V_{|x_q|}(n, \Omega) \rangle. \quad (\text{B.12})$$

Les coefficients de corrélation $r'_{x_i x_1}(n, \Omega)$ peuvent alors s'écrire sous la forme

$$r'_{x_i x_1}(n, \Omega) = \frac{\langle V_{|x_i|}(n, \Omega), V_{|x_1|}(n, \Omega) \rangle}{\|V_{|x_i|}(n, \Omega)\| \cdot \|V_{|x_1|}(n, \Omega)\|}. \quad (\text{B.13})$$

En appliquant l'inégalité de Cauchy-Schwarz à l'Eq. (B.13), on obtient :

$$|r'_{x_i x_1}(n, \Omega)| \leq 1, \quad (\text{B.14})$$

avec l'égalité si et seulement si les vecteurs $V_{|x_i|}(n, \Omega)$ et $V_{|x_1|}(n, \Omega)$ sont linéairement dépendants, c.-à-d. (puisque $V_{|x_1|}(n, \Omega) \neq 0$) grâce aux hypothèses 4 et 10 si et seulement si il existe un réel positif constant α tel que

$$V_{|x_i|}(n, \Omega) = \alpha V_{|x_1|}(n, \Omega), \quad (\text{B.15})$$

c.-à-d.

$$\forall l' \in \Omega, \quad |X_i(n, \omega_{l'})| = \alpha |X_1(n, \omega_{l'})|, \quad (\text{B.16})$$

c.-à-d.

$$\forall l' \in \Omega, \quad \frac{|X_i(n, \omega_{l'})|}{|X_1(n, \omega_{l'})|} = \alpha. \quad (\text{B.17})$$

Analysons cette condition, en fonction du nombre de sources actives dans la zone d'analyse considéré (n, Ω) . Grâce à l'hypothèse 4, il existe au moins un vecteur $V_{|s_j|}(n, \Omega)$ non-nul par zone d'analyse. Analysons le cas où au moins deux sources sont actives. Si l'Eq. (B.17) est vérifiée pour tous les indices i , alors les modules des rapports correspondants d'observations sont constants sur cette zone d'analyse, ce qui n'est pas possible grâce à l'hypothèse 10.

En conséquence, si $|r'_{x_i x_1}(n, \Omega)| = 1 \quad \forall i = 2, \dots, N$, alors une seule source est active dans cette zone d'analyse.

Annexe C

SIR des signaux mélangés et critères de performance des méthodes de SAS

Dans cette annexe, nous décrivons les critères de performances que nous avons utilisés dans ce document. Notons qu'il existe d'autres critères pour les mélanges LI ou convolutifs [162, 164].

C.1 SIR pour les mélanges LI

Nous définissons le rapport signal sur interférence (SIR - *signal to interference ratio* en anglais) associé aux observations qui sont utilisées par nos méthodes LI de SAS et les paramètres principaux utilisés pour mesurer les performances de ces méthodes dans les tests reportés dans ce manuscrit.

Considérons d'abord une source seule, avec un indice k donné⁵¹. Nous définissons le SIR d'entrée de notre système de SAS associé à une source k en utilisant l'approche en deux étapes suivante. Dans une première étape, nous considérons un seul capteur d'indice i du système de SAS, qui reçoit le signal de mélange $x_i(n)$. Ce signal consiste en :

1. Une contribution de la source d'indice k . Cette contribution est considérée comme le signal d'intérêt contenu par l'entrée i de notre système de SAS et est égale à $a_{ik}s_k(t)$.
2. Les contributions de toutes les autres sources d'indices $j \neq k$. Ces sources sont considérées comme des signaux d'interférence ou du "bruit" contenu dans l'entrée i de notre système de SAS. Leur contribution globale dans $x_i(t)$ est égale à $x_i(t) - a_{ik}s_k(t)$.

Le SIR d'entrée élémentaire de toutes les méthodes LI de SAS, associées à ses entrées i et à la source k , est alors défini comme le rapport des puissances des contributions des

⁵¹L'indice de chaque source est connue, quand nous testons nos méthodes de SAS avec des signaux sources donnés.

signaux et bruits définis ci-dessus⁵², i.e.

$$SIR_k^{in}(i) = \frac{\mathbb{E} \{|a_{ik}s_k(t)|^2\}}{\mathbb{E} \{|x_i(t) - a_{ik}s_k(t)|^2\}}. \quad (C.1)$$

Dans une deuxième étape, nous définissons le SIR d'entrée global du système de SAS considéré, associé à la source k , c.-à-d. tenant compte de toutes les observations $x_i(t)$ utilisées comme les entrées de ce système. Ce SIR global est défini comme

$$SIR_k^{in} = \max_{i=1\dots N} (SIR_k^{in}(i)), \quad (C.2)$$

c.-à-d., pour la source k considérée, nous tenons compte des observations où cette source a le plus grand SIR.

Nous utilisons ensuite la même approche pour définir le SIR de sortie du système de SAS considéré. Ainsi, nous considérons d'abord la source d'indice k et la sortie i de ce système, qui nous donne le signal $y_i(t)$. Ce signal consiste en :

1. La contribution voulue associée à la sortie i du système de SAS. Cette contribution est définie comme la valeur idéale de la sortie $y_i(t)$ quand la source extraite de cette sortie est la source k . A cause du principe des méthodes de SAS considérées dans ce manuscrit, cette sortie idéale est égale à la contribution de la source k dans le premier mélange, c.-à-d. $a_{1k}s_k(t)$ comme vu dans l'Eq. (2.7).
2. De la même façon que pour les signaux d'entrée, la contribution de bruit dans la sortie i est alors définie comme les résidus de $y_i(t)$, c.-à-d. égal à $y_i(t) - a_{1k}s_k(t)$.

Le SIR de sortie élémentaire du système de SAS considéré, associé à sa sortie i et à la source k , est alors définie comme le rapport des puissance de ces contributions de bruit et signal ci-dessus, c.-à-d.

$$SIR_k^{out}(i) = \frac{\mathbb{E} \{|a_{1k}s_k(t)|^2\}}{\mathbb{E} \{|y_i(t) - a_{1k}s_k(t)|^2\}}. \quad (C.3)$$

Nous définissons le SIR de sortie global du système de SAS considéré, associé à la source k comme

$$SIR_k^{out} = \max_{i=1\dots N} (SIR_k^{out}(i)), \quad (C.4)$$

Nous définissons l'amélioration de SIR (*SIRI* - *SIR Improvement* en anglais) effectuée par le système de SAS par rapport à la source k comme

$$SIRI_k = \frac{SIRI_k^{out}}{SIRI_k^{in}}. \quad (C.5)$$

Tous les paramètres ci-dessus réfèrent uniquement à une seule source. Les paramètres globaux correspondants sont effectivement définis comme les moyennes géométriques sur

⁵²Dans nos tests, nous centrons d'abord chaque série temporelle globale définissant un signal source.

toutes les sources k de chaque paramètre considéré, c.-à-d. comme la moyenne arithmétique de ce paramètre exprimé en dB . Cela s'écrit explicitement :

$$SIR^{in} = \left(\prod_{k=1}^N SIR_k^{in} \right)^{\frac{1}{N}} \quad \text{et} \quad (SIR^{in})_{dB} = \frac{1}{N} \sum_{k=1}^N (SIR_k^{in})_{dB} \quad (\text{C.6})$$

$$SIR^{out} = \left(\prod_{k=1}^N SIR_k^{out} \right)^{\frac{1}{N}} \quad \text{et} \quad (SIR^{out})_{dB} = \frac{1}{N} \sum_{k=1}^N (SIR_k^{out})_{dB} \quad (\text{C.7})$$

$$SIRI = \left(\prod_{k=1}^N SIRI_k \right)^{\frac{1}{N}} \quad \text{et} \quad (SIRI)_{dB} = \frac{1}{N} \sum_{k=1}^N (SIRI_k)_{dB} \quad (\text{C.8})$$

Notez que ceci nécessite également :

$$\frac{SIR^{out}}{SIR^{in}} = \left(\prod_{k=1}^N \left[\frac{SIR_k^{out}}{SIR_k^{in}} \right] \right)^{\frac{1}{N}} = \left(\prod_{k=1}^N SIRI_k \right)^{\frac{1}{N}} = SIRI \quad (\text{C.9})$$

et donc

$$(SIRI)_{dB} = (SIR^{out})_{dB} - (SIR^{in})_{dB}. \quad (\text{C.10})$$

Les paramètres SIR^{out} et/ou $SIRI$ sont utilisés pour mesurer les performances des systèmes de SAS considérés, où SIR^{in} indique à quel niveau les signaux utilisés par ce système sont mélangés.

Remarquez que SIR^{in} a une simple expression dans la configuration comprenant $N = 2$ mélanges de $N = 2$ sources, comme nous allons le voir. Considérons d'abord la source d'indice $k = 1$. L'Eq. (C.1) devient :

$$SIR_1^{in}(i) = \frac{\mathbb{E} \{ |a_{i1} s_1(t)|^2 \}}{\mathbb{E} \{ |a_{i2} s_2(t)|^2 \}}. \quad (\text{C.11})$$

En notant m la valeur de l'entrée d'indice i qui correspond au plus grand $SIR_1^{in}(i)$, l'Eq. (C.2) s'écrit

$$SIR_1^{in} = \frac{\mathbb{E} \{ |a_{m1} s_1(t)|^2 \}}{\mathbb{E} \{ |a_{m2} s_2(t)|^2 \}}. \quad (\text{C.12})$$

De même, pour la source d'indice $k = 2$, l'Eq. (C.1) devient

$$SIR_2^{in}(i) = \frac{\mathbb{E} \{ |a_{i2} s_2(t)|^2 \}}{\mathbb{E} \{ |a_{i1} s_1(t)|^2 \}}. \quad (\text{C.13})$$

En notant n la valeur de l'entrée d'indice i qui correspond au plus grand $SIR_2^{in}(i)$, l'Eq. (C.2) s'écrit

$$SIR_2^{in} = \frac{\mathbb{E} \{ |a_{n2} s_2(t)|^2 \}}{\mathbb{E} \{ |a_{n1} s_1(t)|^2 \}}. \quad (\text{C.14})$$

Ainsi

$$SIR^{in} = \sqrt{SIR_1^{in} SIR_2^{in}} \quad (C.15)$$

$$= \left| \frac{a_{m1} a_{n2}}{a_{m2} a_{n1}} \right| \quad (C.16)$$

En particulier, dans la situation standard où la source 1 (resp. 2) est prédominante dans l'entrée 1 (resp. 2), nous avons : $m = 1$ et $n = 2$. L'Eq. (C.15) s'écrit alors

$$SIR^{in} = \left| \frac{a_{11} a_{22}}{a_{12} a_{21}} \right|. \quad (C.17)$$

C.2 SIR pour les mélanges AD

Nous définissons maintenant le SIR associé aux observations qui sont utilisées par nos méthodes AD de SAS et les paramètres principaux utilisés pour mesurer les performances des méthodes correspondantes. Les paramètres sont très proches de ceux que nous venons de développer.

Considérons d'abord une source seule, avec un indice k donné. Nous définissons le SIR d'entrée de notre système de SAS associé à une source k en utilisant l'approche en deux étapes suivante. Dans une première étape, nous considérons un seul capteur d'indice i du système de SAS, qui reçoit le signal de mélange $x_i(n)$. Ce signal consiste en :

1. Une contribution de la source d'indice k égale à $a_{ik} s_k(n - n_{ik})$.
2. Les contributions de toutes les autres sources d'indices $j \neq k$. Leur contribution globale dans $x_i(n)$ est égale à $x_i(n) - a_{ik} s_k(n - n_{ik})$.

Le SIR d'entrée élémentaire, défini dans l'Eq. (C.1) pour les mélanges LI, s'écrit ici :

$$SIR_k^{in}(i) = \frac{\mathbb{E} \{ |a_{ik} s_k(n - n_{ik})|^2 \}}{\mathbb{E} \{ |x_i(n) - a_{ik} s_k(n - n_{ik})|^2 \}}. \quad (C.18)$$

le SIR d'entrée global du système, associé à la source k , est toujours défini par :

$$SIR_k^{in} = \max_{i=1 \dots N} (SIR_k^{in}(i)). \quad (C.19)$$

Considérons maintenant la source d'indice k et la sortie i de ce système, qui nous donne le signal $y_i(n)$. Ce signal consiste en :

1. La contribution voulue associée à la sortie i du système de SAS. Cette contribution est définie comme la valeur idéale de la sortie $y_i(n)$ quand la source extraite de cette sortie est la source k . A cause du principe des méthodes de SAS considérées dans ce manuscrit, cette sortie idéale est égale à la contribution de la source k dans le premier mélange, c.-à-d. $a_{1k} s_k(n - n_{1k})$ comme vu dans l'Eq. (4.6).
2. De la même façon que pour les signaux d'entrée, la contribution de bruit dans la sortie i est alors définie comme les résidus de $y_i(n)$, c.-à-d. égal à $y_i(n) - a_{1k} s_k(n - n_{1k})$.

Le SIR de sortie élémentaire, défini dans l'Eq. (C.3) pour les mélanges LI, s'écrit :

$$SIR_k^{out}(i) = \frac{\mathbb{E} \{ |a_{1k}s_k(n - n_{1k})|^2 \}}{\mathbb{E} \{ |y_i(n) - a_{1k}s_k(n - n_{1k})|^2 \}}. \quad (C.20)$$

Le SIR de sortie global du système, associé à la sortie k , est toujours défini par l'Eq. (C.4) :

$$SIR_k^{out} = \max_{i=1\dots N} (SIR_k^{out}(i)). \quad (C.21)$$

Le SIRI du système de SAS par rapport à la source k s'écrit alors :

$$SIRI_k = \frac{SIR_k^{out}}{SIR_k^{in}}. \quad (C.22)$$

Les paramètres globaux de performance de la méthode restent inchangés et s'écrivent :

$$SIR^{in} = \left(\prod_{k=1}^N SIR_k^{in} \right)^{\frac{1}{N}} \quad \text{et} \quad (SIR^{in})_{dB} = \frac{1}{N} \sum_{k=1}^N (SIR_k^{in})_{dB} \quad (C.23)$$

$$SIR^{out} = \left(\prod_{k=1}^N SIR_k^{out} \right)^{\frac{1}{N}} \quad \text{et} \quad (SIR^{out})_{dB} = \frac{1}{N} \sum_{k=1}^N (SIR_k^{out})_{dB} \quad (C.24)$$

$$SIRI = \left(\prod_{k=1}^N SIRI_k \right)^{\frac{1}{N}} \quad \text{et} \quad (SIRI)_{dB} = \frac{1}{N} \sum_{k=1}^N (SIRI_k)_{dB} \quad (C.25)$$

Notez que les Eq. (C.9) et (C.10) sont toujours vérifiées dans le cadre des mélanges AD, c.-à-d. :

$$\frac{SIR^{out}}{SIR^{in}} = \left(\prod_{k=1}^N \left[\frac{SIR_k^{out}}{SIR_k^{in}} \right] \right)^{\frac{1}{N}} = \left(\prod_{k=1}^N SIRI_k \right)^{\frac{1}{N}} = SIRI \quad (C.26)$$

et

$$(SIRI)_{dB} = (SIR^{out})_{dB} - (SIR^{in})_{dB}. \quad (C.27)$$

C.3 SIR pour les mélanges convolutifs

Nous définissons maintenant le SIR associé aux observations qui sont utilisées par nos méthodes convolutives de SAS et les paramètres principaux utilisés pour mesurer les performances des méthodes correspondantes. Les paramètres sont très proches de ceux que nous venons de développer.

Considérons d'abord une source seule, avec un indice k donné. Nous définissons le SIR d'entrée de notre système de SAS associé à une source k en utilisant l'approche en deux étapes suivante. Dans une première étape, nous considérons un seul capteur d'indice i du système de SAS, qui reçoit le signal de mélange $x_i(n)$. Ce signal consiste en :

1. Une contribution de la source d'indice k égale à $a_{ik}(n) * s_k(n)$.

2. Les contributions de toutes les autres sources d'indices $j \neq k$. Leur contribution globale dans $x_i(n)$ est égale à $x_i(n) - a_{ik}(n) * s_k(n)$.

Le SIR d'entrée élémentaire, défini dans l'Eq. (C.1) pour les mélanges LI, s'écrit ici :

$$SIR_k^{in}(i) = \frac{\mathbb{E} \{ |a_{ik}(n) * s_k(n)|^2 \}}{\mathbb{E} \{ |x_i(n) - a_{ik}(n) * s_k(n)|^2 \}}. \quad (C.28)$$

le SIR d'entrée global du système, associé à la source k , est toujours défini par :

$$SIR_k^{in} = \max_{i=1 \dots N} (SIR_k^{in}(i)). \quad (C.29)$$

Considérons maintenant la source d'indice k et la sortie i de ce système, qui nous donne le signal $y_i(n)$. Ce signal consiste en :

1. La contribution voulue associée à la sortie i du système de SAS. Cette contribution est définie comme la valeur idéale de la sortie $y_i(n)$ quand la source extraite de cette sortie est la source k . A cause du principe des méthodes de SAS considérées dans ce manuscrit, cette sortie idéale est égale à la contribution de la source k dans le premier mélange, c.-à-d. $a_{1k}(n) * s_k(n)$.
2. De la même façon que pour les signaux d'entrée, la contribution de bruit dans la sortie i est alors définie comme les résidus de $y_i(n)$, c.-à-d. égal à $y_i(n) - a_{1k}(n) * s_k(n)$.

Le SIR de sortie élémentaire, défini dans l'Eq. (C.3) pour les mélanges LI, s'écrit :

$$SIR_k^{out}(i) = \frac{\mathbb{E} \{ |a_{1k}(n) * s_k(n)|^2 \}}{\mathbb{E} \{ |y_i(n) - a_{1k}(n) * s_k(n)|^2 \}}. \quad (C.30)$$

Le SIR de sortie global du système, associé à la sortie k , est toujours défini par l'Eq. (C.4). Le SIRI du système de SAS par rapport à la source k s'écrit alors :

$$SIRI_k = \frac{SIR_k^{out}}{SIR_k^{in}}. \quad (C.31)$$

Les paramètres globaux de performance de la méthode restent inchangés et s'écrivent :

$$SIR^{in} = \left(\prod_{k=1}^N SIR_k^{in} \right)^{\frac{1}{N}} \quad \text{et} \quad (SIR^{in})_{dB} = \frac{1}{N} \sum_{k=1}^N (SIR_k^{in})_{dB} \quad (C.32)$$

$$SIR^{out} = \left(\prod_{k=1}^N SIR_k^{out} \right)^{\frac{1}{N}} \quad \text{et} \quad (SIR^{out})_{dB} = \frac{1}{N} \sum_{k=1}^N (SIR_k^{out})_{dB} \quad (C.33)$$

$$SIRI = \left(\prod_{k=1}^N SIRI_k \right)^{\frac{1}{N}} \quad \text{et} \quad (SIRI)_{dB} = \frac{1}{N} \sum_{k=1}^N (SIRI_k)_{dB} \quad (C.34)$$

Notez que les Eq. (C.9) et (C.10) sont toujours vérifiées dans le cadre des mélanges convolutifs, c.-à-d. :

$$\frac{SIR^{out}}{SIR^{in}} = \left(\prod_{k=1}^N \left[\frac{SIR_k^{out}}{SIR_k^{in}} \right] \right)^{\frac{1}{N}} = \left(\prod_{k=1}^N SIRI_k \right)^{\frac{1}{N}} = SIRI \quad (C.35)$$

et

$$(SIRI)_{dB} = (SIR^{out})_{dB} - (SIR^{in})_{dB}. \quad (\text{C.36})$$

Annexe D

Extension de LI-TIFROM aux mélanges atténués et retardés

Tout comme la méthode LI-TIFCORR, nous pouvons étendre LI-TIFROM aux cas des mélanges atténués et décalés temporellement. La position du problème est exactement la même que celle vue dans la section 4.1, c.-à-d. que nous devons estimer les éléments de la matrice $B(\omega)$ définis par l'expression (4.9). Dans un but de clarté et de simplicité, nous n'allons détailler que l'extension de LI-TIFROM-S2, détaillée dans la section 2.1.3, aux mélanges AD. En effet, parmi toutes les variantes de LI-TIFROM, celle-ci nous a donné les résultats meilleurs dans nos tests (voir chapitre 3) et l'extension aux mélanges AD de LI-TIFROM-S1 ou la version originale de LI-TIFROM se fait de façon similaire à la version proposée ici. Dans la suite de ce chapitre, pour plus de clarté, nous n'indiquerons plus le $S2$ de la méthode.

Enfin, tout comme pour AD-TIFCORR, nous proposons aussi deux approches :

- l'une découplant l'identification des paramètres b_{im} et μ_{im} , et utilisant des zones d'analyse à fréquence constante pour estimer les coefficients d'échelle (d'où le nom AD-TIFROM-CF de la méthode),
- l'autre n'utilisant que les zones d'analyse mono-sources à temps constant pour estimer les paramètres (d'où le nom AD-TIFROM-CT de celle-ci).

D.1 Description de l'approche AD-TIFROM-CF

Comme nous l'avons vu dans la première partie de ce manuscrit, LI-TIFROM et LI-TIFCORR ont des structures très semblables tout en utilisant des paramètres différents dans les étapes de détection et d'identification. Il en est de même pour AD-TIFROM-CF et AD-TIFCORR-CF. Ainsi, la méthode AD-TIFROM-CF est composée des étapes suivantes.

- la méthode que nous proposons pour détecter les zones mono-sources (à fréquence constante) et pour identifier les paramètres b_{im} à l'intérieur de celles-ci (voir la sous-section D.1.1 ci-dessous),
- la méthode que nous proposons pour détecter les zones mono-source (à temps constant) et pour identifier les paramètres $\mu_{im'}$ à l'intérieur de celles-ci (voir la

- sous-section 4.2.2),
- la procédure pour coupler ces paramètres identifiés ci-dessus b_{im} et μ_{im} et qui correspond à l'étape décrite dans la section 4.2.3,
 - et la méthode globale qui résulte de ces principes (voir la sous-section D.1.2 ci-dessous).

D.1.1 Etapes de détection basique et identification pour les paramètres b_{im}

Comme pour le cas des mélanges LI, la méthode de SAS que nous introduisons ici inclut d'abord une étape de détection pour trouver les zones TF mono-sources. En partant des équations de mélanges AD (4.2), nous avons déduit dans la section 4.1 les équations de mélange dans le domaine fréquentiel (4.3). Ici encore, nous supposons que les décalages temporels sont suffisamment petits en comparaison de la largeur temporelle des fenêtres $h(\cdot)$ utilisées dans le calcul des TFCT et donc que l'expression (4.13) est vérifiée. Nous supposons aussi que l'hypothèse 10 (voir page 110) est vérifiée.

Comme dans le cas de mélanges LI, nous étudions les rapports $\alpha_i(n, \omega)$ et $\beta_i(n, \omega)$ définis par les équations (2.13) et (2.35), qui ici s'écrivent

$$\alpha_i(n, \omega) = \frac{\sum_{j=1}^N a_{ij} e^{-j\omega n_{ij}} S_j(n, \omega)}{\sum_{j=1}^N a_{1j} e^{-j\omega n_{1j}} S_j(n, \omega)} \quad (\text{D.1})$$

$$\beta_i(n, \omega) = \frac{\sum_{j=1}^N a_{1j} e^{-j\omega n_{1j}} S_j(n, \omega)}{\sum_{j=1}^N a_{ij} e^{-j\omega n_{ij}} S_j(n, \omega)} \quad (\text{D.2})$$

Si une source $S_k(n, \omega)$ est isolée dans une fenêtre TF (n_p, ω_l) alors, l'Eq. (4.21) nous permet d'obtenir :

$$\alpha_i(n_p, \omega_l) = b_{im} e^{-j\omega_l \mu_{im}}, \quad (\text{D.3})$$

avec b_{im} et μ_{im} définis par l'Eq. (4.9) et $k = \sigma(m)$. Comme nous avons supposé tous les coefficients d'atténuation a_{ik} réels et positifs, tous les coefficients d'échelle b_{im} le sont aussi. Le module du paramètre fourni dans l'expression (4.21) est encore égal à :

$$|\alpha_i(n_p, \omega_l)| = b_{im} \quad (\text{D.4})$$

Ainsi, si une source $S_k(n, \omega)$ est isolée dans une zone d'analyse (T, ω_l) , alors $|\alpha_i(n_p, \omega_l)|$ est constant dans cette zone et sa moyenne, définie par

$$\overline{|\alpha_i|}(T, \omega_l) = \frac{1}{M} \sum_{p=1}^M |\alpha_i(n_p, \omega_l)|, \quad (\text{D.5})$$

se lit ensuite

$$\overline{|\alpha_i|}(T, \omega_l) = \frac{a_{ik}}{a_{1k}} = b_{im}. \quad (\text{D.6})$$

De la même façon, $|\beta_i(n_p, \omega_l)|$ est constant sur cette zone d'analyse (T, ω_l) et $\overline{|\beta_i|}(T, \omega_l)$, défini par

$$\overline{|\beta_i|}(T, \omega_l) = \frac{1}{M} \sum_{p=1}^M |\beta_i(n_p, \omega_l)|, \quad (\text{D.7})$$

se lit ensuite

$$\overline{|\beta_i|}(T, \omega_l) = \frac{a_{1k}}{a_{ik}} = \frac{1}{b_{im}}. \quad (\text{D.8})$$

Une approche naturelle peut alors être proposée pour détecter les zones d'analyse et identifier les paramètres b_{im} . Cela consiste à utiliser le même type d'approche que dans LI-TIFROM, sauf que les paramètres b_{im} sont ici définis par $|\alpha_i(n, \omega)|$ et $|\beta_i(n, \omega)|$. Donc nous calculons les variances de $|\alpha_i(n, \omega)|$ et $|\beta_i(n, \omega)|$:

$$\text{var} [|\alpha_i|] (T, \omega_l) = \frac{1}{M} \sum_{p=1}^M \left| |\alpha_i(n_p, \omega_l)| - \overline{|\alpha_i|}(T, \omega_l) \right|^2, \quad (\text{D.9})$$

$$\text{var} [|\beta_i|] (T, \omega_l) = \frac{1}{M} \sum_{p=1}^M \left| |\beta_i(n_p, \omega_l)| - \overline{|\beta_i|}(T, \omega_l) \right|^2 \quad (\text{D.10})$$

et les moyennes sur i de ces variances

$$MVAR [|\alpha|] (T, \omega_l) = \frac{1}{N-1} \sum_{i=2}^N \text{var} [|\alpha_i|] (T, \omega_l), \quad (\text{D.11})$$

$$MVAR [|\beta|] (T, \omega_l) = \frac{1}{N-1} \sum_{i=2}^N \text{var} [|\beta_i|] (T, \omega_l). \quad (\text{D.12})$$

Nous classons alors les zones d'analyse TF selon les valeurs croissantes du paramètre

$$\min \{ MVAR [|\alpha|] (T, \omega_l), MVAR [|\beta|] (T, \omega_l) \}. \quad (\text{D.13})$$

Les plus petites valeurs de ce paramètre sont encore obtenues dans les "meilleures" zones mono-sources.

L'étape d'identification des coefficients d'échelle b_{im} consiste donc, soit à étudier successivement les premières zones d'analyse de la liste ordonnée décrite plus haut, soit à utiliser une des méthodes de classification proposées dans le chapitre 5. Pour chaque zone de celle-ci, les estimées des paramètres b_{im} associés à une colonne de $B(\omega)$ sont affectées aux valeurs de $\overline{|\alpha_i|}(T, \omega_l)$ ou de $1/\overline{|\beta_i|}(T, \omega_l)$, suivant lequel de ces rapports minimise le critère (D.13).

D.1.2 Résumé global de la méthode AD-TIFROM-CF

En conséquence de la description ci-dessus, la méthode AD-TIFROM-CF que nous proposons ici contient les étapes suivantes :

1. L'étape de prétraitement consiste à calculer les TFCT $X_i(n, \omega)$ des signaux de mélange, selon l'expression (2.2).
2. Nous considérons ensuite uniquement les paramètres b_{im} . Nous détectons d'abord les zones d'analyse mono-source à fréquence constante (T, ω_l) et nous identifions ensuite les paramètres b_{im} dans ces zones, utilisant la procédure décrite dans la sous-section D.1.1.
3. Nous considérons ensuite les paramètres $b'_{im'}$ et $\mu_{im'}$. Nous calculons d'abord les droites de régression et leurs erreurs quadratiques moyennes associées sur les zones d'analyse TF à temps constant $(n_{p'}, \Omega)$ définies dans la section 4.2.2. Nous détectons ensuite les zones mono-sources et identifions les paramètres $b'_{im'}$ et $\mu_{im'}$ dans ces zones, en utilisant la procédure décrite dans les sous-sections 4.2.2 et 4.2.3.
4. On couple ensuite les paramètres b_{im} et $\mu_{im'}$, comme expliqué dans la sous-section 4.2.3. Tous les paramètres, associés à la matrice $B(\omega)$, qui sont requis sont disponibles à cette étape.
5. En utilisant ces paramètres, nous obtenons en pratique les sources extraites des observations, selon l'équation (4.12).

D.2 Description de l'approche AD-TIFROM-CT

Comme pour l'approche AD-TIFCORR, nous avons introduit une méthode qui exploite seulement les zones TF à temps constant $(n_{p'}, \Omega)$ pour détecter les zones mono-sources et estimer tous les paramètres de $B(\omega)$.

Si $S_{k'}(n, \omega)$ est isolée dans une zone d'analyse $(n_{p'}, \Omega)$, alors les modules de $\alpha_i(n_{p'}, \omega_{l'})$ et $\beta_i(n_{p'}, \omega_{l'})$ resp. exprimés dans les équations (D.1) and (D.2), sont constants. Leur moyennes, resp. définies par

$$\overline{|\alpha_i|}(n_{p'}, \Omega) = \frac{1}{M'} \sum_{l'=1}^{M'} |\alpha_i(n_{p'}, \omega_{l'})| \quad (\text{D.14})$$

et

$$\overline{|\beta_i|}(n_{p'}, \Omega) = \frac{1}{M'} \sum_{l'=1}^{M'} |\beta_i(n_{p'}, \omega_{l'})|, \quad (\text{D.15})$$

s'écrivent alors

$$\overline{|\alpha_i|}(n_{p'}, \Omega) = \frac{a_{ik'}}{a_{1k'}} = b_{im'} \quad (\text{D.16})$$

$$\overline{|\beta_i|}(n_{p'}, \Omega) = \frac{a_{1k'}}{a_{ik'}} = \frac{1}{b_{im'}} \quad (\text{D.17})$$

avec $k' = \sigma(m')$. Nous calculons les variances de $|\alpha_i(n_{p'}, \omega_{l'})|$ et de $|\beta_i(n_{p'}, \omega_{l'})|$:

$$\text{var} [|\alpha_i|](n_{p'}, \Omega) = \frac{1}{M'} \sum_{l'=1}^{M'} \left| |\alpha_i(n_{p'}, \omega_{l'})| - \overline{|\alpha_i|}(n_{p'}, \Omega) \right|^2, \quad (\text{D.18})$$

$$var [|\beta_i|] (n_{p'}, \Omega) = \frac{1}{M} \sum_{l'=1}^{M'} \left| |\beta_i(n_{p'}, \omega_{l'})| - \overline{|\beta_i|}(n_{p'}, \Omega) \right|^2 \quad (D.19)$$

et les moyennes sur i de ces variances

$$MVAR [|\alpha|] (n_{p'}, \Omega) = \frac{1}{N-1} \sum_{i=2}^N var [|\alpha_i|] (n_{p'}, \Omega), \quad (D.20)$$

$$MVAR [|\beta|] (n_{p'}, \Omega) = \frac{1}{N-1} \sum_{i=2}^N var [|\beta_i|] (n_{p'}, \Omega). \quad (D.21)$$

Les paramètres de mélanges $b'_{im'}$ sont ensuite identifiés, soit en utilisant la méthode basique définie dans la sous-section 4.2.1, soit en faisant appel aux techniques de classification décrites dans le chapitre 5, à l'exception du fait que la zone d'analyse (T, ω_l) est remplacée par la zone $(n_{p'}, \Omega)$. Ainsi :

1. les meilleures zones d'analyse sont celles qui minimisent le critère

$$\min \{MVAR [|\alpha|] (n_{p'}, \Omega), MVAR [|\beta|] (n_{p'}, \Omega)\} \quad (D.22)$$

au lieu de celui décrit dans l'Eq. (D.13).

2. Les paramètres utilisés ici pour identifier les paramètres $b'_{im'}$ sont $\overline{|\alpha_i|}(n_{p'}, \Omega)$ et $\overline{|\beta_i|}(n_{p'}, \Omega)$ au lieu de $\overline{|\alpha_i|}(T, \omega_l)$ et de $\overline{|\beta_i|}(T, \omega_l)$.

L'identification des paramètres $\mu_{im'}$ est appliquée de la façon suivante. Dans un premier temps, en considérant le cas de la méthode basique (adaptée de la sous-section 4.2.1), l'identification des paramètres $b'_{im'}$ est effectuée. Ensuite, dans chaque zone mono-source sélectionnée où une colonne de paramètres $b'_{im'}$ est identifiée (et gardée), la colonne correspondante de paramètres $\mu_{im'}$ est simultanément identifiée, en utilisant de nouveau la méthode basée sur les droites de régression décrite dans la sous-section 4.2.2. Maintenant, dans une méthode faisant appel aux techniques de classification, nous formons d'abord N clusters en considérant uniquement les paramètres $b'_{im'}$, en utilisant une des approches choisies dans le chapitre 5. Pour chaque point de ces clusters, nous identifions aussi les paramètres associés $\mu_{im'}$, de nouveau au moyen des droites de régression. Nous déduisons ensuite une seule colonne de $\mu_{im'}$ de chaque cluster en utilisant le même type d'approches que celles de la sous-section 4.2.3, c.-à-d. : pour chaque indice i indépendamment, nous étudions toutes les valeurs $\mu_{im'}$ correspondant au même cluster et nous gardons la valeur qui a le plus grand nombre d'occurrences.

Annexe E

Atténuations négatives dans les méthodes AD

Comme nous l'avons noté dans la section 4.1, nous supposons que les coefficients d'atténuation a_{ij} sont positifs lorsque le mélange considéré est AD. Physiquement, cela n'est pas vraiment une limitation puisque dans le cas d'une propagation mono-trajet d'une onde sonore, l'atténuation ne peut être que positive. Néanmoins, il peut être intéressant de traiter aussi le cas d'atténuations négatives.

En réécrivant les observations en fonction des contributions des sources dans la première observation, l'Eq. (4.8) peut s'écrire dans le domaine TF sous la forme suivante

$$X_i(n, \omega) = \sum_{j=1}^N |b_{ij}| e^{-j\omega\mu_{ij} - j\pi\mathbb{1}_{ij}} S'_j(n, \omega) \quad i = 1 \dots N, \quad (\text{E.1})$$

où b_{ij} et μ_{ij} sont définis dans l'Eq. (4.9) et où

$$\mathbb{1}_{ij} = \begin{cases} 1 & \text{si } b_{ij} < 0 \\ 0 & \text{si } b_{ij} \geq 0 \end{cases}. \quad (\text{E.2})$$

En pratique, grâce aux versions basiques ou améliorées (c.-à-d. faisant appel aux techniques de clustering) de AD-TIFCORR et AD-TIFFROM décrites dans les chapitres 4, 5 et l'annexe D, il est facile de voir que nous pouvons toujours estimer les modules des colonnes de paramètres b_{im} et leurs décalages temporels μ_{im} associés. Il nous reste donc à estimer les coefficients $\mathbb{1}_{im}$, définis dans l'Eq. (E.2), et associés aux coefficients d'échelle b_{im} . Nous proposons d'étudier le cas particulier où la pulsation est nulle dans l'Eq. (E.1). Dans cette situation particulière, les observations s'écrivent :

$$X_i(n, 0) = \sum_{j=1}^N |b_{ij}| e^{-j\pi\mathbb{1}_{ij}} S'_j(n, 0) \quad i = 1, \dots, N, \quad (\text{E.3})$$

$$= \sum_{j=1}^N b_{ij} S'_j(n, 0) \quad i = 1, \dots, N. \quad (\text{E.4})$$

En particulier, supposons qu'il existe une fenêtre TF $(n_p, 0)$ tel que la source $S_k(n, \omega)$ est isolée, alors

$$X_i(n_p, 0) = b_{im} S'_m(n_p, 0) \quad i = 1, \dots, N, \quad (\text{E.5})$$

avec $k = \sigma(m)$. Nous remarquons alors que dans une telle fenêtre, il est possible d'estimer le signe de b_{im} .

Nous proposons alors d'estimer le signe des coefficients d'échelle de la façon suivante.

1. Dans une première étape, nous calculons les coefficients $|b_{im}|$, comme proposé dans les versions basiques ou améliorées des méthodes AD-TIFCORR ou AD-TIFROM.
2. Dans une seconde étape, nous ne gardons que les zones d'analyse qui correspondent *effectivement*⁵³ aux positions TF permettant l'estimation de ces paramètres.
3. Dans une troisième étape, nous créons de nouvelles zones d'analyse correspondant à la concaténation de chaque point TF $(n_p, 0)$ avec chacune des zones TF sélectionnées dans l'étape précédente.
4. Grâce aux critères mono-sources développés dans AD-TIFCORR ou AD-TIFROM, nous pouvons trouver quelles sont les nouvelles zones créées à l'étape précédente qui sont toujours mono-sources, c.-à-d. quel quel point TF $(n_p, 0)$ mono-source correspond à une colonne de paramètres $|b_{im}|$ estimés. Il suffit alors de voir par exemple le signe des rapports $\alpha_i(n_p, 0)$ pour estimer le signe des paramètres b_{im} correspondants.

⁵³Dans le cas des versions faisant appel aux méthodes de classification, nous choisissons d'utiliser les zones TF qui correspondent aux positions TF des centroïdes des clusters, c.-à-d.

- le "premier" point de chaque cluster dans la méthode CBCE, décrite dans la section 5.2.1,
- la médiane des clusters dans K-medians (voir section 5.2.4).
- l'individu d'un cluster le plus proche de son centroïde pour les approches basées sur la moyenne.

Annexe F

Limiter numériquement le nombre de zones d'analyse mono-sources : validation expérimentale

Dans cette annexe, nous montrons expérimentalement pourquoi il vaut mieux limiter numériquement le nombre de zones d'analyse mono-sources quand on applique une méthode de classification telle K-means, K-medians ou C-means.

Nous proposons de comparer les estimées des coefficients b_{im} obtenues avec l'approche AD-TIFROM-CF décrite dans la section 6.2.1, c.-à-d. quand on a limité à 1000 le nombre de zones mono-sources, avec ces mêmes estimées quand on ne limite pas le nombre de zones mono-sources.

Le tableau F.1 donne les normes de Frobenius entre les valeurs estimées et théoriques des colonnes de b_{im} . Dans ce tableau, nous n'avons pas limité numériquement le nombre de zones mono-sources. Par rapport aux normes obtenues dans le tableau 6.1, les normes que nous obtenons ici sont souvent bien plus importantes, apportant une faible qualité d'estimation des paramètres b_{im} . Nous expliquons cette différence de comportement de la manière suivante : quand nous limitons le nombre d'individus à classer, nous ne prenons que les premières colonnes de paramètres b_{im} estimées, issues des "meilleures" zones mono-sources et apportant une excellente qualité d'estimation (sauf en cas de "fausse" détection, comme nous l'avons vu précédemment). Cependant, lorsqu'on étudie tous les individus du nuage de points, la plupart des individus sont issus d'estimations moins précises que les précédentes, induisant un biais dans l'estimation de nos colonnes de coefficients d'échelle. Nous montrons dans la figure F.1 les histogrammes de largeur $1e-2$ des coefficients b_{im} obtenus par AD-TIFROM-CF, pour le jeu 2 de sources de parole, testé dans le chapitre 6, et la matrice de mélange (6.2) avec les paramètres $\lambda = 0.5$ et $\eta = 0$. Dans la figure du haut, nous avons conservé toutes les zones telles que le critère (D.13) soit en-dessous du seuil fixé dans le chapitre 6. Au contraire, dans la figure en-dessous, nous n'avons conservé que les mille "meilleures" zones (au sens de notre critère mono-source). Nous voyons alors que les points dans le deuxième cas sont beaucoup moins dispersés autour des valeurs théoriques (0.5 et 2) que dans le premier cas. Cette différence entraîne la différence de performances que l'on peut observer en comparant les normes de Frobenius données dans les tableaux 6.1 et F.1.

TAB. F.1 – Norme de Frobenius des différences entre les matrices théoriques de paramètres b_{im} et leur estimées obtenues par AD-TIFROM-CF, en fonction de la méthode de classification utilisée, du jeu de sources et des paramètres λ et η de la matrice de mélange définie dans l'Eq. (6.2). Le nombre de zones mono-sources n'a pas été limité.

λ	Jeu	Méthode de clustering	η		
			0	10	20
0.5	2	K-means	0.0434	0.0289	0.0148
		K-medians	0.0206	0.0201	0.0205
		Fuzzy C-means	0.0435	0.0292	0.0156
	6	K-means	0.0339	0.0191	0.0118
		K-medians	0.0142	0.0185	0.0181
		Fuzzy C-means	0.0341	0.0194	0.0124
0.9	2	K-means	0.0383	0.0101	2.7859
		K-medians	0.0253	0.0132	0.0060
		Fuzzy C-means	0.0394	0.0125	0.0879
	6	K-means	0.0372	0.00072	0.0025
		K-medians	0.0213	0.0060	0.0066
		Fuzzy C-means	0.00382	0.0089	0.0046

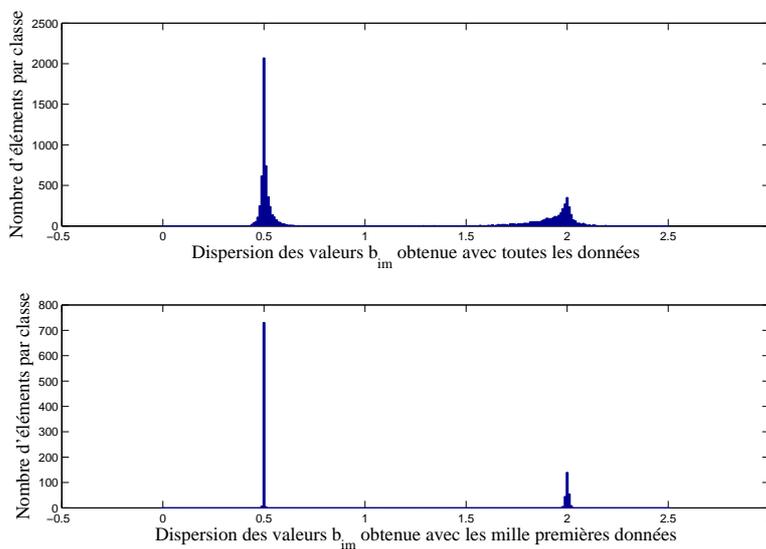


FIG. F.1 – Histogrammes de dispersion des paramètres de mélanges estimés avec AD-TIFROM-CF, en fonction du nombre de points utilisés. *En haut* : tous les points qui vérifient l'Eq. (5.13) sont utilisés. *En bas* : Seuls les mille points qui maximisent ce critère (5.13) le sont.

En conclusion, ce résultat nous semble très important car de nombreuses méthodes à parcimonie de SAS étudient l'ensemble des données du domaine d'analyse et estiment les paramètres de mélanges à l'aide d'une méthode de classification classique, comme K-means. Nous montrons ici qu'il semble plus pertinent de chercher un sous-ensemble de points de "très bonne qualité" (au sens de la "qualité mono-source") et d'effectuer la classification sur ces points. En outre, cette limitation a l'avantage d'alléger considérablement le coût de calcul.

Bibliographie

- [1] F. Abrard: *Méthodes de séparation aveugle de sources et applications : des statistiques d'ordre supérieur à l'analyse temps-fréquence*. Thèse de doctorat, Université Paul Sabatier Toulouse 3, 2003.
- [2] F. Abrard and Y. Deville: *A time-frequency blind signal separation method applicable to underdetermined mixtures of dependent sources*. *Signal Processing*, 85(7):1389–1403, July 2005.
- [3] F. Abrard, Y. Deville, and P. White: *From blind source separation to blind source cancellation in the underdetermined case: a new approach based on time-frequency analysis*. In *Proceedings of the 3rd International Symposium on Independent Component Analysis and Blind Signal Separation (ICA 2001)*, San Diego, California, USA, December 9–13, 2001.
- [4] F. Abrard, Y. Deville, and P. White: *A new source separation approach based on time-frequency analysis for instantaneous mixtures*. In *Proceedings of the 5th International Workshop on Electronics, Control, Modelling, Measurement and Signals (ECM2S 2001)*, pages 259–267, Toulouse, France, May 30 – June 1, 2001.
- [5] B. Albouy: *Méthodes de séparation aveugle de sources applicables à des signaux de parole*. Thèse de doctorat, Université Paul Sabatier Toulouse 3, 2004.
- [6] B. Albouy and Y. Deville: *Alternative structures and power spectrum criteria for blind segmentation and separation of convolutive speech mixtures*. In *Proceedings of the Fourth International Symposium on Independent Component Analysis and Blind Signal Separation (ICA2003)*, pages 361–366, Nara, Japan, April 1–4, 2003.
- [7] B. Albouy and Y. Deville: *Segmentation and separation of speech and/or noise signals, using coherence functions and power spectra*. In *Proceedings of the third workshop on Physics in Signal and Image Processing (PSIP 2003)*, pages 9–12, Grenoble, France, January 29–31, 2003.
- [8] B. Albouy and Y. Deville: *A time-frequency blind source separation method based on segmented coherence function*. In *Proceedings of the 7th International Workshop on Artificial and Natural Neural Networks (IWANN 2003)*, volume 2, pages 289–296, Mao, Menorca, Spain, June 3–6, 2003. J. Mira and J. R. Alvarez eds (Springer).
- [9] J. Anemüller and B. Kollmeier: *Adaptive separation of acoustic sources for anechoic conditions: A constrained frequency domain approach*. *Speech Communications*, 39(1–2):79–95, January 2003.

- [10] P. Anttilaa, P. Paatero, U. Tapperb, and O. Järvinen: *Source identification of bulk wet deposition in finland by positive matrix factorization*. Atmospheric Environment, 29(14):1705–1718, 1995.
- [11] S. Araki, H. Sawada, R. Mukai, and S. Makino: *Underdetermined blind sparse source separation for arbitrarily arranged multiple sensors*. Signal Processing, 87(8):1833–1847, August 2007.
- [12] S. Arberet, R. Gribonval, and F. Bimbot: *A robust method to count and locate audio sources in a stereophonic linear instantaneous mixture*. In *Proceedings 6th International Conference on Independent Component Analysis and Blind Source Separation (ICA 06)*, pages 536–543, Charleston, South Carolina, USA, March 2006.
- [13] S. Arberet, R. Gribonval, and F. Bimbot: *A robust method to count and locate audio sources in a stereophonic linear anechoic mixture*. In *Proceedings of the 2007 IEEE International Conference on Acoustics, Sound and Signal Processing (ICASSP 2007)*, volume 3, pages 745–748, Honolulu, Hawai'i, USA, April 15–20, 2007.
- [14] A. Aïssa-El-Bey: *Séparation aveugle de sources audio*. Thèse de doctorat, Ecole nationale supérieure des télécommunications, Paris, 2007.
- [15] A. Aïssa-El-Bey, K. Abed-Meraim et Y. Grenier: *Séparation aveugle sous-déterminée de sources audio par la méthode EMD (Empirical Modal Decomposition)*. Dans *Actes du 20^{ème} Colloque National GRETSI 2005*, tome 2, pages 1233–1236, Louvain-la-neuve, Belgique, septembre 2005.
- [16] A. Aïssa-El-Bey, K. Abed-Meraim, and Y. Grenier: *Underdetermined blind separation of audio sources in time-frequency domain*. In *Proceedings of the Workshop on Signal Processing with Adaptive Sparse Structured Representations (SPARS'05)*, Rennes, France, November 16–18, 2005.
- [17] A. Aïssa-El-Bey, K. Abed-Meraim, and Y. Grenier: *Underdetermined blind audio source separation using modal decomposition*. EURASIP Journal on Audio, Speech, and Music Processing, 2007, March 2007.
- [18] A. Aïssa-El-Bey, K. Abed-Meraim, and Y. Grenier: *Underdetermined blind separation of audio sources from the time-frequency representation of their convolutive mixtures*. In *Proceedings of the 2007 IEEE International Conference on Acoustic, Speech and Signal Processing (ICASSP 2007)*, volume 1, pages 153–156, Hawaii, U.S.A, April 15–20, 2007.
- [19] A. Aïssa-El-Bey, N. Linh-Trung, K. Abed-Meraim, A. Belouchrani, and Yves Grenier: *Underdetermined blind separation of nondisjoint sources in the time-frequency domain*. IEEE Transactions on Signal Processing, 55(3):897–907, March 2007.
- [20] M. Babaie-Zadeh and C. Jutten: *A general approach for mutual information minimization and its application to blind source separation*. Signal Processing, 85(5):975–995, May 2005.
- [21] M. Babaie-Zadeha, C. Jutten, and A. Mansour: *Sparse ICA via cluster-wise PCA*. Neurocomputing, 69(13–15):1458–1466, August 2006.

-
- [22] A. D. Back and A. S. Weigend: *A first application of independent component analysis to extracting structure from stock returns*. International journal of neural systems, 8(5):473–484, October 1997.
- [23] M. Baeck and U. Zölder: *Performance analysis of a source separation algorithm*. In *Proceedings of the 5th International Conference on Digital Audio Effects (DAFx-02)*, pages 207–210, Hamburg, Germany, September 26–28, 2002.
- [24] M. Baeck and U. Zölder: *Real-time implementation of a source separation algorithm*. In *Proceedings of the 6th International Conference on Digital Audio Effects (DAFx-03)*, London, United Kingdom, September 8–11, 2003.
- [25] T. C. Bakker Schut, B. G. De Groot, and J. Greve: *Cluster analysis of flow cytometric list mode data on a personal computer*. Cytometry, 14:649–659, 1993.
- [26] R. Balan and J. Rosca: *Source separation using sparse discrete prior models*. In *Proceedings of the 2006 IEEE International Conference on Acoustics, Sound and Signal Processing (ICASSP 2006)*, volume 4, pages 1113–1116, Toulouse, France, May 14–19, 2006.
- [27] R. Balan, J. Rosca, and S. Rickard: *Non-square blind source separation under coherent noise by beamforming and time-frequency masking*. In *Proceedings of the 4th International Symposium on Independent Component Analysis and Blind Signal Separation (ICA 2003)*, pages 313–318, Nara, Japan, April 2003.
- [28] R. Balan, J. Rosca, S. Rickard, and J. O’Ruanaidh: *The influence of windowing on time delay estimation*. In *Proceedings of the 35th Annual Conference on Information Sciences and Systems (CISS 2000)*, volume 1, pages WP1(15–17), Princeton, NJ, March 2000.
- [29] A. J. Bell and T. J. Sejnowski: *An information-maximisation approach to blind separation and blind deconvolution*. Neural Computation, 7:1129–1159, 1995.
- [30] A. Belouchrani, K. Abed-Meraim, M. G. Amin, and A. M. Zoubir: *Joint anti-diagonalisation for blind source separation*. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2001)*, volume 5, pages 2789–2792, Salt Lake City, Utah, USA, May 7–11, 2001.
- [31] A. Belouchrani, K. Abed-Meraim, J. F. Cardoso, and E. Moulines: *A blind source separation technique using second-order statistics*. IEEE Transactions on Signal Processing, 45(2):434–444, February 1997.
- [32] A. Belouchrani and M. G. Amin: *Blind source separation based on time-frequency signal representations*. IEEE Transactions on Signal Processing, 46(11):2888–2897, November 1998.
- [33] A. Belouchrani and J. F. Cardoso: *Maximum likelihood source separation for discrete sources*. In *Proceedings of the European Signal Processing Conference (EU-SIPCO’94)*, volume 2, pages 768–771, Edinburgh, United Kingdom, September 1994.
- [34] O. Berné, C. Joblin, Y. Deville, J. D. Smith, M. Rapaccioli, J. P. Bernard, J. Thomas, W. Reach, and A. Abergel: *Analysis of the emission of very small dust*

- particles from spitzer spectro-imagery data using blind signal separation methods.* Astronomy & Astrophysics, 469(2):575–586, July 2007.
- [35] M. W. Berry, M. Browne, A. N. Langville, V. P. Pauca, and R. J. Plemmons: *Algorithms and applications for approximate nonnegative matrix factorization.* Computational Statistics & Data Analysis, 52(1):155–173, September 2007.
- [36] J. C. Bezdek: *Pattern Recognition with Fuzzy Objective Function Algorithms.* Plenum Press, New York, 1981.
- [37] D. Bissessur: *Méthodes de séparation aveugle de sources basées sur l'analyse en ondelettes : Application aux signaux de parole et aux images astrophysiques.* Rapport de Master 2 Recherche Signal, Image, Acoustique et Optimisation, LATT, Toulouse, 2005.
- [38] A. Blin, S. Araki, and S. Makino: *Blind source separation when speech signals outnumber sensors using a sparseness-mixing matrix estimation (SMME).* In *Proceedings of the International Workshop on Acoustic Echo and Noise Control (IWAENC 2003)*, pages 211–214, Kyoto, Japan, September 2003.
- [39] A. Blin, S. Araki, and S. Makino: *A sparseness-mixing matrix estimation (SMME) solving the underdetermined bss for convolutive mixtures.* In *Proceedings of the 2004 IEEE International Conference on Acoustics, Sound and Signal Processing (ICASSP 2004)*, volume 4, pages 85–88, Montreal, Quebec, Canada, May 17–21, 2004.
- [40] A. Blin, S. Araki, and S. Makino: *Underdetermined blind separation of convolutive mixtures of speech using time-frequency mask and mixing matrix estimation.* IE-ICE Transactions Fundamentals, E88–A (Special Section on Multi-channel Acoustic Signal Processing)(7):1693–1700, July 2005.
- [41] P. Bofill: *Underdetermined blind separation of delayed sound sources in the frequency domain.* Neurocomputing, 16(6):627–641, 2003.
- [42] P. Bofill and E. Monte: *Underdetermined convoluted source reconstruction using LP and SOCP, and a neural approximator of the optimizer.* In Springer-Verlag (editor): *Proceedings of the 6th International Conference on Independent Component Analysis and Blind Signal Separation (ICA 2006)*, volume LNCS 3889, pages 569–576, Charleston, South Carolina, USA, March 5–8, 2006.
- [43] P. Bofill and M. Zibulevsky: *Blind separation of more sources than mixtures using sparsity of their short-time Fourier transform.* In *Proceedings of the 2nd International Workshop on Independent Component Analysis and Blind Signal Separation (ICA 2000)*, pages 87–92, Helsinki, Finland, June 19–22, 2000.
- [44] P. Bofill and M. Zibulevsky: *Underdetermined blind source separation using sparse representations.* Signal Processing, 81:2353–2362, 2001.
- [45] H. Le Borgne: *Analyse de scènes naturelles par composantes indépendantes.* Thèse de doctorat, Institut National Polytechnique de Grenoble, 2004.
- [46] H. Boumaraf: *Séparation aveugle de mélanges convolutifs de sources.* Thèse de doctorat, Université Joseph Fourier et Institut National Polytechnique de Grenoble, 2005.

-
- [47] H. Bousbia-Salah, A. Belouchrani, and K. Abed-Meraim: *Jacobi-like algorithm for blind signal separation of convolutive mixtures*. *Electronic Letters*, 37:1049–1050, August 2001.
- [48] P. S. Bradley, O. L. Mangasarian, and W. N. Street: *Clustering via concave minimization*. In *Proceedings of the Conference on Advances in Neural Information Processing Systems (NIPS 1996)*, volume 9, pages 368–374, Denver, Colorado, USA, December 2–5 1996.
- [49] H. Buchner, R. Aichner, and W. Kellermann: *A generalization of blind source separation algorithms for convolutive mixtures based on second-order statistics*. *IEEE Transactions on Speech and Audio Processing*, 13(1):120–134, January 2005.
- [50] J. F. Cardoso: *Blind beamforming for non-Gaussian signals*. *IEE Proceedings F Radar and Signal Processing*, 140(6):362–370, December 1993.
- [51] J. F. Cardoso: *Infomax and maximum likelihood for blind source separation*. *IEEE Signal Processing Letters*, 4:112–114, April 1997.
- [52] G. Celeux and G. Govaert: *A classification EM algorithm for clustering and two stochastic versions*. *Computational Statistics & Data Analysis*, 14(3):315–332, October 1992. Special issue on optimization techniques in statistics.
- [53] C. Choi: *Real-time binaural blind source separation*. In *Proceedings of the 4th International Symposium on Independent Component Analysis and Blind Signal Separation (ICA 2003)*, pages 567–572, Nara, Japan, April 1–4, 2003.
- [54] S. Choi, A. Cichocki, and A. Belouchrani: *Second order non-stationary source separation*. *Journal of VLSI Signal Processing*, 32(1–2):93–104, August 2002.
- [55] S. Choi, A. Cichocki, H. M. Park, and S. Y. Lee: *Blind source separation and independent component analysis: a review*. *Neural Information Processing – Letters and Reviews*, 6(1), January 2005. Review.
- [56] A. Cichocki and S. Amari: *Adaptive Blind Signal and Image Processing: Learning Algorithms and Applications*. Wiley-Interscience, New York, April 2002.
- [57] A. Cichocki, S. Amari, K. Siwek, T. Tanaka, and al.: *ICALAB Toolboxes*. <http://www.bsp.brain.riken.jp/ICALAB>.
- [58] L. Cohen: *Time-frequency analysis*. Prentice hall PTR, Englewood, Cliffs, New Jersey, 1995.
- [59] M. Coli, R. Di Nisio, and L. Ippoliti: *Exploratory analysis of financial time series using independent component analysis*. In *Proceedings of the 27th International Conference on Information Technology Interfaces*, Cavtat, Croatia, June 20–23, 2005.
- [60] P. Comon: *Separation of sources using higher-order cumulants*. In *Proceedings of the SPIE Conference on Advanced Algorithms and Architectures for Signal Processing*, volume 1152, pages 170–181, San Diego, California, August 8–10, 1989.
- [61] P. Comon: *Independant component analysis, a new concept?* *Signal Processing*, 36(3):287–314, April 1994.

- [62] P. Comon: *Contrasts for multichannel blind deconvolution*. IEEE Signal Processing Letters, 3(7):209–211, July 1996.
- [63] P. Comon: *Blind identification and source separation in 2×3 under-determined mixtures*. IEEE Transactions on Signal Processing, 52(1):11–22, January 2004.
- [64] P. Comon and B. Emile: *Estimation of time delays in the blind mixture problem*. In *Proceedings of the 7th European Signal Processing Conference (EUSIPCO 1994)*, pages 482–485, Edinburgh, Scotland, September 13-16, 1994.
- [65] P. Comon, C. Jutten et al.: *Séparation de Sources*, tome 1 (*concepts de base et analyse en composantes indépendantes*). Collection *Traité IC2, Information – Commande – Communication*, Hermès - Lavoisier, Paris, France, 2007.
- [66] D. Corney: *My clustering toolbox*. <http://www.lottolab.org/David%20Corney/ClusteringToolbox.zip>.
- [67] N. Delfosse and P. Loubaton: *Adaptive blind separation of independent sources: a deflation approach*. Signal Processing, 45(1):59–83, July 1995.
- [68] S. Deligne and R. Gopinath: *An EM algorithm for convolutive independent component analysis*. Neurocomputing, 49(1):187–211, December 2002.
- [69] C. Demars: *Représentations bidimensionnelles d'un signal de parole : éléments de monographie*, 2005. <http://www.limsi.fr/Individu/chrd/ftpindex2004.html>.
- [70] A. P. Dempster, N. M. Laird, and D. B. Rubin: *Maximum likelihood from incomplete data via the EM algorithm*. Journal of the Royal Statistical Society. Series B (Methodological), 39(1):1–38, 1977.
- [71] Y. Deville: *Traitement d'antenne et séparation de sources*. Notes de cours du Master 2 Signal, Image, Acoustique et Optimisation de Toulouse.
- [72] Y. Deville: *Towards industrial applications of blind source separation and independent component analysis*. In *Proceedings of the International Conference on Independent Component Analysis and Signal Separation (ICA 1999)*, pages 19–24, Aussois, France, January 11–15, 1999.
- [73] Y. Deville: *Panorama des applications biomédicales des méthodes de séparation aveugle de sources*. Dans *Actes du 19^{ème} Colloque National GRETSI 2003*, tome 1, pages 31–34, Paris, France, 8–11 septembre 2003.
- [74] Y. Deville: *Temporal and time-frequency correlation-based blind source separation methods*. In *Proceedings of the 4th International Symposium on Independent Component Analysis and Blind Signal Separation (ICA 2003)*, pages 1059–1064, Nara, Japan, April 1–4, 2003.
- [75] Y. Deville: *Applications des méthodes de séparation aveugle de sources*, chapitre 13 du livre multi-auteurs *Séparation de Sources*, tome 2, sous la direction de C. Jutten et P. Comon. Collection *Traité IC2, Information – Commande – Communication*, Hermès - Lavoisier, Paris, France, 2007.
- [76] Y. Deville, M. Benali, and F. Abrard: *Differential source separation for underdetermined instantaneous or convolutive mixtures : concept and algorithms*. Signal Processing, 84(10):1759–1776, October 2004.

-
- [77] Y. Deville, D. Bissessur, M. Puigt, S. Hosseini, and H. Carfantan: *A time-scale correlation-based blind separation method applicable to correlated sources*. In *Proceedings of the European Symposium on Artificial Neural Networks (ESANN 2006)*, pages 337–344, Bruges, Belgium, April 25–28, 2006.
- [78] Y. Deville, J. Chappuis, S. Hosseini, and J. Thomas: *Differential fast fixed-point BSS for underdetermined linear instantaneous mixtures*. In Springer-Verlag (editor): *Proceedings of the 6th International Conference on Independent Component Analysis and Blind Signal Separation (ICA 2006)*, volume LNCS 3889, pages 48–56, Charleston, South Carolina, USA, March 5–8, 2006.
- [79] Y. Deville, J. Damour, and N. Charkani: *Multi-tag radio-frequency identification systems based on new blind source separation neural networks*. *Neurocomputing*, 49:369–388, 2002.
- [80] Y. Deville and S. Hosseini: *Blind identification and separation methods for linear-quadratic mixtures and/or linearly independent non-stationary signals*. In *Proceedings of the 9th International Symposium on Signal Processing and its Applications (ISSPA 2007)*, Sharjah, United Arab Emirates, February 12–15, 2007. Invited paper in special session.
- [81] Y. Deville and S. Hosseini: *Blind identification and separation methods for linear-quadratic mixtures and/or linearly independent non-stationary signals*. In *Proceedings of the 9th IEEE International Symposium on Signal Processing and its Applications (ISSPA 2007)*, Sharjah, United Arab Emirates, February 12–15, 2007.
- [82] Y. Deville and M. Puigt: *Temporal and time-frequency correlation-based blind source separation methods. part 1: linear instantaneous mixtures*. *Signal Processing*, 87(3):374–407, March 2007.
- [83] Y. Deville, M. Puigt, and B. Albouy: *Time-frequency blind signal separation: extended methods, performance evaluation for speech sources*. In *Proceedings of the International Joint Conference on Neural Networks (IJCNN 2004)*, volume 1, pages 255–260, Budapest, Hungary, July 25–29, 2004.
- [84] D. Donoho: *On minimum entropy deconvolution*. In *Proceedings of the Applied Time series Analysis II*, pages 556–608, Tulsa, Oklahoma, March 1980.
- [85] D. Donoho: *Sparse components of images and optimal atomic decompositions*. *Constructive approximation*, 17(3):353–382, 2001.
- [86] D. Donoho and M. Elad: *Optimally sparse representation in general (nonorthogonal) dictionaries via ℓ^1 minimization*. *Proceedings of the National Academy of Sciences of the United States of America (PNAS)*, 100(5):2197–2202, 2003.
- [87] S. Douglas, H. Sawada, and S. Makino: *A spatio-temporal fastica algorithm for separating convolutive mixtures*. In *Proceedings of the 2005 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2005)*, volume 5, pages 165–168, Philadelphia, Pennsylvania, March 2005.
- [88] F. X. Désert, F. Boulanger, and J. L. Puget: *Interstellar dust models for extinction and emission*. *Astronomy & Astrophysics*, 237(1):215–236, October 1990.

- [89] J. C. Dunn: *A fuzzy relative of the ISODATA process and its use in detecting compact, well-separated clusters*. Journal of Cybernetics, 3(3):32–57, 1973.
- [90] K. Eggert and E. Korner: *Sparse coding and nmf*. In *Proceedings of the 2004 IEEE International Joint Conference on Neural Networks (IJCNN 2004)*, volume 4, pages 2529–2533, July 25–29, 2004.
- [91] T. Elomaa and H. Koivistoinen: *On autonomous k-means clustering*. In *Proceedings of the 15th International Symposium on Methodologies for Intelligent Systems (ISMIS 2005)*, pages 228–236, Saratoga Springs, New York, USA, May 25-28, 2005.
- [92] B. Emile and P. Comon: *Estimation of time delays between unknown colored signals*. Signal Processing, 69(1):93–100, August 1998.
- [93] B. Emile, P. Comon, and J. Le Roux: *Estimation of time delays with fewer sensors than sources*. IEEE Transactions on Signal Processing, 46(7):2012–2015, July 1998.
- [94] D. Erdogmus, K. E. Hild, Y. N. Rao, and J. C. Principe: *Minimax mutual information approach for independent component analysis*. Neural Computation, 16(6):1235–1252, June 2004.
- [95] E. M. Fadaili, N. Thirion-Moreau, and E. Moreau: *Joint zero-diagonalization for blind separation of instantaneous and convolutive mixtures of white correlated sources*. In *Proceedings of the Second International Symposium on Communications, Control and Signal Processing (ISCCSP 2006)*, Marrakech, Morocco, March 13–15, 2006.
- [96] A. Favrot, M. Erne, and C. Faller: *Improved cocktail-party processing*. In *Proceedings of the 9th International Conference on Digital Audio Effects (DAFx-06)*, pages 227–232, Montreal, Canada, September 18–20, 2006.
- [97] P. Flandrin, G. Rilling, and P. Goncalvés: *Empirical mode decomposition as a filter bank*. IEEE Signal Processing Letters, 11(2):112–114, 2004.
- [98] J. Foote: *Visualizing music and audio using self-similarity*. In *Proceedings of the seventh ACM international conference on Multimedia*, pages 77–80, Orlando, Florida, USA, October 1999.
- [99] C. Févotte: *Approche temps-fréquence pour la séparation aveugle de sources non-stationnaires*. Thèse de doctorat, École Centrale de Nantes et Université de Nantes, 2003.
- [100] C. Févotte and C. Doncarli: *Two contributions to blind source separation using time-frequency distributions*. IEEE Signal Processing Letters, 11(3):386–389, March 2004.
- [101] M. Gaeta et J. L. Lacoume: *Estimateurs du maximum de vraisemblance étendus à la séparation de sources non gaussiennes*. Traitement du Signal, 7(5) :419–434, 1990.
- [102] L. Giulieri, H. Ghennioui, N. Thirion-Moreau, and E. Moreau: *Nonorthogonal joint diagonalisation of spatial quadratic time-frequency matrices for source separation*. IEEE Signal Processing Letters, 12(5):415–418, May 2005.
- [103] I. F. Gorodnitsky and B. D. Rao: *Sparse signal reconstruction from limited data using FOCUSS: a re-weighted minimum norm algorithm*. IEEE Transactions on Signal Processing, 45(3):600–616, March 1997.

-
- [104] A. Gorokhov and P. Loubaton: *Subspace-based techniques for blind separation of convolutive mixtures with temporally correlated sources*. IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications, 44(9):813–820, September 1997.
- [105] R. Gribonval and S. Lesage: *A survey of sparse component analysis for source separation : principles, perspectives, and new challenges*. In *Proceedings of the 14th European Symposium On Artificial Neural Networks (ESANN 2006)*, pages 323–330, Bruges, Belgium, April 26–28, 2006.
- [106] R. Guidara: *Séparation aveugle de sources markoviennes, bidimensionnelles et autocorrélées – Application à l’Astrophysique*. Rapport de Master 2 Recherche Signal, Image, Acoustique et Optimisation, LATT, Toulouse, 2005.
- [107] R. Guidara, S. Hosseini, and Y. Deville: *Blind separation of non-stationary images using markov models*. In Springer-Verlag (editor): *Proceedings of the 7th International Conference on Independent Component Analysis and Blind Source Separation (ICA 2007)*, volume LNCS 4666, pages 722–729, London, United Kingdom, September 9–12, 2007.
- [108] R. Guidara, S. Hosseini, and Y. Deville: *Markovian blind separation of non-stationary temporally correlated sources*. In *Proceedings of the 15th European Symposium on Artificial Neural Networks (ESANN 2007)*, Bruges, Belgium, April 25–27, 2007.
- [109] M. Halkidi, Y. Batistakis, and M. Vazirgiannis: *On clustering validation techniques*. Journal of Intelligent Information Systems, 17(2–3):107–145, 2001.
- [110] G. Hamerly and C. Elkan: *Learning the k in k -means*. In *Proceedings of the 17th annual conference on Neural Information Processing Systems (NIPS 2003)*, pages 281–288, Whistler, British Columbia, Canada, December 8–13, 2003.
- [111] A. Hardy: *On the number of clusters*. Computational Statistics & Data Analysis, 23(1):83–96, November 1996.
- [112] T. Hastie, R. Tibshirani, and G. Walther: *Estimating the number of clusters in a dataset via the gap statistic*. Technical Report 208, Department of Statistics, Stanford University, Stanford, California, USA, 2000. www-stat.stanford.edu/~tibs/ftp/gap.ps.
- [113] K. E. Hild, D. Erdogmus, and J. C. Principe: *Blind source separation using renyi’s mutual information*. IEEE Signal Processing Letters, 8(6):174–176, June 2001.
- [114] K. E. Hild, D. Erdogmus, and J. C. Principe: *On-line minimum mutual information method for time-varying blind source separation*. In *Proceedings of the 3rd International Symposium on Independent Component Analysis and Signal Separation (ICA 2001)*, pages 126–131, San Diego, California, December 9–13, 2001.
- [115] K. E. Hild, D. Pinto, D. Erdogmus, and J. C. Principe: *Convolutional blind source separation by minimizing mutual information between segments of signals*. IEEE Transactions on Circuits and Systems: I Regular Papers, 52(10):2188–2196, October 2005.

- [116] F. Hlawatsch and G.F. Boudreaux-Bartels: *Linear and quadratic time-frequency signal representations*. IEEE SP Magazine, pages 21–67, April 1992.
- [117] A. Holobar, C. Févotte, C. Doncarli, and D. Zazula: *Single autoterms selection for blind source separation in time-frequency plane*. In *Proceedings of the 11th European Signal Processing Conference (EUSIPCO 2002)*, Toulouse, France, September 3–6, 2002.
- [118] S. Hosseini and Y. Deville: *Blind separation of nonstationary sources by spectral decorrelation*. In Springer-Verlag (editor): *Proceedings of the Fifth International Conference on Independent Component Analysis and Blind Signal Separation (ICA 2004)*, volume LNCS 3195, pages 279–286, Granada, Spain, September 22–24, 2004.
- [119] S. Hosseini, Y. Deville, and H. Saylani: *From time-domain separation of stationary temporally correlated sources to frequency-domain separation of nonstationary sources*. In *Proceedings of the Fourth International Conference on Physics in Signal and Image Processing (PSIP 2005)*, pages 55–59, Toulouse, France, January 31 – February 2, 2005.
- [120] S. Hosseini, R. Guidara, Y. Deville, and C. Jutten: *Markovian blind image separation*. In Springer-Verlag (editor): *Proceedings of the 6th International Conference on Independent Component Analysis and Blind Signal Separation (ICA 2006)*, volume LNCS 3889, pages 106–114, Charleston, South Carolina, USA, March 5–8, 2006.
- [121] S. Hosseini, R. Guidara, Y. Deville, and C. Jutten: *Maximum likelihood separation of spatially autocorrelated images using a markov model*. In *Proceedings of the 26th International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering (MaxEnt 2006)*, pages 105–112, Paris, France, July 8–13 2006. Invited paper.
- [122] J. Hérault, C. Jutten et B. Ans: *Détection de grandeurs primitives dans un message composite par une architecture de calcul neuromimétique en apprentissage non-supervisé*. Dans *Actes du 10^{ème} colloque GRESTI*, pages 1017–1022, Nice, France, mai 1985.
- [123] N. E. Huang, Z. Shen, S. R. Long, M.C. Wu, E. H. Shih, Q. Zheng, C.C. Tung, and H.H. Liu: *The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis*. *Proceedings of the Royal Society London A*, 454:903–995, 1998.
- [124] A. Hyvärinen: *New approximations of differential entropy for independent component analysis and projection pursuit*. In *Proceedings of the Conference on Advances in Neural Information Processing Systems (NIPS 1997)*, volume 10, pages 273–279, Denver, Colorado, December 1997.
- [125] A. Hyvärinen: *Fast and Robust Fixed-Point Algorithms for Independent Component Analysis*. *IEEE Transactions on Neural Networks*, 10(3):626–634, May 1999.
- [126] A. Hyvärinen: *Survey on independent component analysis*. *Neural Computing Surveys*, 2:94–128, 1999.
- [127] A. Hyvärinen, J. Karhunen, and E. Oja: *Independent Component Analysis*. Wiley-Interscience, New York, 2001.

-
- [128] A. Hyvärinen and E. Oja: *A Fast Fixed-Point Algorithm for Independent Component Analysis*. *Neural Computation*, 9(7):1483–1492, October 1997.
- [129] A. Hyvärinen and E. Oja: *Independent component analysis: Algorithms and applications*. *Neural Networks*, 13(4–5):411–430, June 2000.
- [130] P. Jallon and A. Chevreuril: *Second-order based cyclic frequency estimates: the case of digital communication signals*. In *Proceedings of the 2006 IEEE International Conference on Acoustics, Sound and Signal Processing (ICASSP 2006)*, volume 4, pages 389–392, Toulouse, France, May 14–19, 2006.
- [131] S. C. Johnson: *Hierarchical clustering schemes*. *Psychometrika*, 32(3):241–254, September 1967.
- [132] A. Jourjine, S. Rickard, and O. Yilmaz: *Blind separation of disjoint orthogonal signals: demixing N sources from 2 mixtures*. In *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2000)*, volume 5, pages 2985–2988, Istanbul, Turkey, June 5–9, 2000. IEEE Press.
- [133] C. Jutten and M. Babaie-Zadeh: *Source separation: principles, current advances and applications*. In *Proceedings of the 4th Workshop on Advanced Control and Diagnosis (ACD 2006)*, Nancy, France, November 16–17, 2006.
- [134] C. Jutten, P. Comon et al.: *Séparation de Sources*, tome 2 (*au-delà de l’aveugle et applications*). Collection *Traité IC2, Information – Commande – Communication*, Hermès - Lavoisier, Paris, France, 2007.
- [135] C. Jutten and J. Héroult: *Blind separation of sources, part 1: an adaptive algorithm based on neuromimetic architecture*. *Signal Processing*, 24:1–10, 1991.
- [136] C. Jutten and A. Taleb: *Source separation: from dusk till dawn*. In *Proceedings of the 2nd International Workshop on Independent Component Analysis and Blind Source Separation (ICA 2000)*, pages 15–26, Helsinki, Finland, June 2000. Invited paper.
- [137] A. Kachenoura, L. Albera et L. Senhadja: *Séparation aveugle de sources en ingénierie biomédicale*. *ITBM–RBM*, 28(1) :20–34, mars 2007.
- [138] R. E. Kass and L. Wasserman: *A reference Bayesian test for nested hypotheses and its relationship to the Schwarz criterion*. *Journal of the American Statistical Association*, 90(431):928–934, September 1995.
- [139] M. Kawamoto, A. K. Barros, A. Mansour, K. Matsuoka, and N. Ohnishi: *Blind separation of convolutive mixtures of non-stationary signals*. In *Proceedings of the 5th International Conference on Neural Information Processing (ICONIP 1998)*, pages 743–746, Kitakyushu, Japan, October 21–23, 1998.
- [140] M. Kawamoto, K. Matsuoka, and N. Ohnishi: *A method of blind separation for convolved non-stationary signals*. *Neurocomputing*, 22(1–3):157–171, November 1998.
- [141] M. Knaak, S. Araki, and S. Makino: *Geometrically constraint ICA for convolutive mixtures of sound*. In *Proceedings of the 2003 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2003)*, volume 2, pages 725–728, April 6–10, 2003.

- [142] C. H. Knapp and G. C. Carter: *The generalized correlation method for estimation of time delay*. IEEE Transactions on Acoustics, Speech, and Signal Processing, 24(4), August 1976.
- [143] D. D. Lee and H. S. Seung: *Learning the parts of objects by non-negative matrix factorization*. Nature, 401:788–791, 1999.
- [144] D. D. Lee and H. S. Seung: *Algorithms for non-negative matrix factorization*. In *Proceedings of the 14th annual conference on Neural Information Processing Systems (NIPS 2000)*, pages 556–562, 2000.
- [145] T. W. Lee: *Independent Component Analysis: Theory and Applications*. Kluwer Academic Publishers, Boston, September 1998.
- [146] T. W. Lee, M. Girolami, and T. J. Sejnowski: *Independent component analysis using an extended infomax algorithm for mixed sub-gaussian and super-gaussian sources*. Neural Computation, 11(2):609–633, 1999.
- [147] A. Léger, L. d’Hendecourt, and D. Défourneau: *Physics of IR emission by interstellar PAH molecules*. Astronomy & Astrophysics, 216(1–2):148–164, June 1989.
- [148] R. Li and M. Xiao: *Estimation of delays and attenuations for underdetermined BSS in frequency domain*. In Springer-Verlag (editor): *Proceedings of the 3rd International Symposium on Neural Networks (ISNN 2006), Part 1*, volume LNCS 3971, pages 1183–1188, Chengdu, China, May 28 – June 1, 2006.
- [149] Y. Li, S. I. Amari, A. Cichocki, D. W. C. Ho, and S. Xie: *Underdetermined blind source separation based on sparse representation*. IEEE Transactions on Signal Processing, 54(2):423–437, 2006.
- [150] Y. Li, A. Cichocki, and S. Amari: *Analysis of sparse representation and blind source separation*. Neural Computation, 16(6):1193–1234, 2004.
- [151] Y. Li, A. Cichocki, S. I. Amari, and C. Guan: *Analysis of source sparsity and recoverability for SCA based blind source separation*. In Springer-Verlag (editor): *Proceedings of the 6th International Conference on Independent Component Analysis and Blind Signal Separation (ICA 2006)*, volume LNCS 3889, pages 831–837, Charleston, South Carolina, USA, March 5–8, 2006.
- [152] Y. Luo, S. Lambotharan, and J. A. Chambers: *A new block based time-frequency approach of underdetermined blind source separation*. In *Proceedings of the 2004 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2004)*, volume 5, pages 537–540, Montreal, Quebec, Canada, May 17–21, 2004.
- [153] Y. Luo, W. Wang, J. A. Chambers, S. Lambotharan, and I. Proudler: *Exploitation of source nonstationarity in underdetermined blind source separation with advanced clustering techniques*. IEEE Transactions on Signal Processing, 54(6), June 2006.
- [154] Z. Q. Luo and J. Lu: *On blind source separation using mutual information criterion*. Mathematical programming, 97(3):587–603, August 2003.
- [155] D. J. C. MacKay: *Information Theory, Inference, and Learning Algorithms*. Cambridge University Press, Cambridge, 2003. <http://www.inference.phy.cam.ac.uk/mackay/itila/>, Free on-line version.

-
- [156] S. Makino, H. Sawada, R. Mukai, and S. Araki: *Blind source separation of convolutive mixtures of speech in frequency domain*. IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences, E88-A(7):1640–1655, July 2005.
- [157] S. G. Mallat and Z. Zhang: *Matching pursuit in a time-frequency dictionary*. IEEE Transactions on Signal Processing, 41(12):3397–3415, December 1993.
- [158] A. Mansour: *A mutually referenced blind multiuser separation of convolutive mixture algorithm*. Signal Processing, 81(11):2253–2266, November 2001.
- [159] A. Mansour: *Batch mutually referenced separation algorithm for MIMO convolutive mixtures*. In Springer-Verlag (editor): *Proceedings on the 5th International Symposium on Independent Component Analysis and Blind Signal Separation (ICA 2004)*, volume LNCS 3195, Granada, Spain, September 22–24, 2004.
- [160] A. Mansour, A. K. Barros, and N. Ohnishi: *Blind separation of sources: Methods, assumptions and applications*. EICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences, E83–A(8):1498–1512, August 2000.
- [161] A. Mansour, N. Bencheikroun, and C. Gervaise: *Blind separation of underwater acoustic signals*. In Springer-Verlag (editor): *Proceedings of the 6th International Conference on Independent Component Analysis and Blind Signal Separation (ICA 2006)*, volume LNCS 3889, pages 181–188, Charleston, South Carolina, USA, March 5–8, 2006.
- [162] A. Mansour and A. Al Falou: *Performance indices of bss for real-world applications*. In *Proceedings of the 14th European Signal Processing Conference (EUSIPCO 2006)*, Florence, Italy, September 4–8, 2006.
- [163] A. Mansour and C. Jutten: *What should we say about the kurtosis?* IEEE Signal Processing Letters, 6(12):321–322, December 1999.
- [164] A. Mansour, M. Kawamoto, and N. Ohnishi: *A survey of the performance indexes of ica algorithms*. In *Proceedings of the 21st IASTED International Conference on Modelling, Identification, and Control (MIC 2002)*, pages 660–666, Innsbruck, Austria, February 18–21, 2002.
- [165] A. S. Master: *Sound source separation of N sources from stereo signals via fitting to N models each lacking one source*. Technical report, Center for Computer Research in Music and Acoustics, Stanford University, 2003. <http://www-ccrma.stanford.edu/asmaster/>.
- [166] A. S. Master: *Bayesian two source modeling for separation of N sources from stereo signals*. In *Proceedings of the 2004 IEEE International Conference on Acoustics, Sound and Signal Processing (ICASSP 2004)*, volume 4, pages 281–284, Montreal, Quebec, Canada, May 17–21, 2004.
- [167] K. Matsuoka and M. Kawamoto: *Blind signal separation based on mutual information criterion*. In *Proceedings of the International Symposium on Nonlinear Theory and its Applications (NOLTA 1995)*, pages 85–90, Las Vegas, Nevada, USA, December 10–14, 1995.

- [168] K. Matsuoka, M. Ohoya, and M. Kawamoto: *A neural net for blind separation of nonstationary signals*. *Neural Networks*, 8(3):411–419, 1995.
- [169] J. McQueen: *Some methods for classification and analysis of multivariate observations*. In *Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability*, volume 1, pages 281–297, 1967.
- [170] T. Melia and S. Rickard: *Extending the DUET blind source separation technique*. In *Proceedings of the Workshop on Signal Processing with Adaptive Sparse Structured Representations (SPARS'05)*, Rennes, France, November 16–18, 2005.
- [171] T. Melia and S. Rickard: *Underdetermined blind source separation in echoic environments using DESPRIT*. *EURASIP Journal on Advances in Signal Processing*, 2007:Article ID 86484, 19 pages, 2007.
- [172] A. Mertins and I. Russel: *An extended ACDC algorithm for the blind estimation of convolutive mixing systems*. In *Proceedings of the Seventh International Symposium on Signal Processing and Its Applications (ISSPA 2003)*, volume 2, pages 527–530, July 1–4, 2003.
- [173] N. Mitianoudis and M. E. Davies: *Audio source separation of convolutive mixtures*. *IEEE Transactions on Speech and Audio Processing*, 11(5):489–497, September 2003.
- [174] E. Moreau, J. C. Pesquet, and N. Thirion-Moreau: *Convolutive blind signal separation based on asymmetrical contrast functions*. *IEEE Transactions on Signal Processing*, 55(1):356–371, January 2007.
- [175] E. Moreau and N. Thirion-Moreau: *Nonsymmetrical contrasts for sources separation*. *IEEE Transactions on Signal Processing*, 47(8):2241–2252, August 1999.
- [176] J. Mouba and S. Marchand: *A source localization/separation/respatialization system based on unsupervised classification of interaural cues*. In *Proceedings of the 9th International Conference on Digital Audio Effects (DAFx-06)*, pages 233–238, Montreal, Canada, September 18–20, 2006.
- [177] E. Moulines, J. F. Cardoso, and E. Gassiat: *Maximum likelihood for blind separation and deconvolution of noisy signals using mixture models*. In *Proceedings of the 1997 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 1997)*, volume 5, pages 3617–3620, Munich, Germany, April 21–24, 1997.
- [178] L. T. Nguyen, A. Belouchrani, K. Abed-Meraim, and B. Boashash: *Separating more sources than sensors using time-frequency distributions*. In *Proceedings of the Sixth International Symposium on Signal Processing and its Applications (ISSPA 2001)*, volume 2, pages 583–286, Kuala Lumpur, Malaysia, August 13–16, 2001.
- [179] P. D. O'Grady and B. A. Pearlmutter: *Convolutive non-negative matrix factorisation with a sparseness constraint*. In *Proceedings of the 2006 16th IEEE Signal Processing Society Workshop on Machine Learning for Signal Processing*, pages 427–432, September 2006.
- [180] P. D. O'Grady, B. A. Pearlmutter, and S. Rickard: *Survey of sparse and non-sparse methods in source separation*. *International Journal of Imaging Systems and Technology (IJIST)*, 15:18–33, 2005.

-
- [181] M. Ohata, K. Matsuoka, and T. Mukai: *An adaptive blind separation method using para-hermitian whitening filter for convolutively mixed signals*. *Signal Processing*, 87(1):33–50, January 2007. <http://www.sciencedirect.com/science/article/B6V18-4K48JFV-2/2/7e0dc378c818b35b45980445e20faaa5>.
- [182] M. Ohata, K. Matsuoka, and T. Mukai: *An adaptive blind separation method using para-hermitian whitening filter for convolutively mixed signals*. *Signal Processing*, 87(1):33–50, January 2007.
- [183] E. Oja, K. Kiviluoto, and S. Malaroiu: *Independent component analysis for financial time series*. In *Proceedings of the 2000 IEEE Symposium on Adaptive Systems for Signal Processing, Communications, and Control (AS-SPCC 2000)*, pages 111–116, Lake Louise, Canada, October 1–4, 2000.
- [184] P. Paatero and U. Tapper: *Positive matrix factorization: A non-negative factor model with optimal utilization of error estimates of data values*. *Environmetrics*, 5:111–126, 1994.
- [185] A. Papoulis and S. U. Pillai: *Probability, random variables, and stochastic processes*. McGraw-Hill, New York, 1965 and 2002.
- [186] L. Parra and C. Spence: *Convolutive blind separation of non-stationary sources*. *IEEE Transactions on Speech and Audio Processing*, 8(3):320–327, May 2000.
- [187] L. Parra, C. Spence, and B. De Vries: *Convolutive source separation and signal modeling with ML*. In *Proceedings of the International Symposium on Intelligent Systems (ISIS 1997)*, Reggio Calabria, Italy, 1997. <http://newton.bme.columbia.edu/~lparra/publish/isis97.pdf>.
- [188] R. M. Parry and I. Essa: *Blind source separation using repetitive structure*. In *Proceedings of the 8th International Conference on Digital Audio Effects (DAFx-05)*, pages 143–148, Madrid, Spain, September 2005.
- [189] R. M. Parry and I. Essa: *Source detection using repetitive structure*. In *Proceedings of the 2006 IEEE International Conference on Acoustics, Sound and Signal Processing (ICASSP 2006)*, volume 4, pages 1093–1096, Toulouse, France, May 14–19, 2006.
- [190] G. Patanchon, H. Snoussi, J.F. Cardoso, and J. Delabrouille: *Component separation for cosmic microwave background data : a blind approach based on spectral diversity*. In *Proceedings of the 3rd Workshop on Physics in Signal and Image Processing (PSIP 2003)*, Grenoble, France, January 29–31, 2003.
- [191] B. Pearlmutter and L. Parra: *A context-sensitive generalization of independent component analysis*. In *Proceedings of the 1996 International Conference on Neural Information Processing (ICONIP 1996)*, Hong Kong, September 24–27, 1996. <http://newton.bme.columbia.edu/~lparra/publish/iconip96.pdf>.
- [192] M. S. Pedersen, J. Larsen, U. Kjems, and L. C. Parra: *A survey of convolutive blind source separation methods*. In *Springer Handbook of Speech (to appear)*. Springer Press, September 2007. <http://www2.imm.dtu.dk/pubdb/p.php?4924>.
- [193] D. Pelleg and A. Moore: *X-means: Extending K-means with efficient estimation of the number of clusters*. In *Proceedings of the 17th International Conference on Machine Learning*, pages 727–734, San Francisco, 2000. Morgan Kaufmann.

- [194] D. T. Pham: *Mutual information approach to blind separation for stationary sources*. In *Proceedings of the International Conference on Independent Component Analysis and Signal Separation (ICA 1999)*, pages 215–220, Aussois, France, January 11–15, 1999.
- [195] D. T. Pham: *Blind separation of instantaneous mixture of sources via the gaussian mutual information criterion*. *Signal Processing*, 81(4):855–870, April 2001.
- [196] D. T. Pham: *Mutual information approach to blind separation of stationary sources*. *IEEE Transactions on Information Theory*, 48(7):1935–1946, July 2002.
- [197] D. T. Pham: *Mutual information approach to blind separation-deconvolution*. In *Proceedings of the European Signal Processing Conference (EUSIPCO 2005)*, Antalya, Turkey, September 2005.
- [198] D. T. Pham and J.F. Cardoso: *Blind separation of instantaneous mixtures of nonstationary sources*. *IEEE Transactions on Signal Processing*, 49(9):1837–1848, September 2001.
- [199] D. T. Pham and P. Garat: *Blind separation of mixtures of independent sources through a quasi maximum likelihood approach*. *IEEE Transactions on Signal Processing*, 45(7):1712–1725, July 1997.
- [200] D. T. Pham, P. Garat, and C. Jutten: *Separation of mixture of independent sources through a maximum likelihood approach*. In *Signal Processing VI, Proceedings of the 6th European Signal Processing Conference (EUSIPCO 1992)*, pages 7716–774, Brussels, Belgium, August 1992.
- [201] D. T. Pham, C. Servière, and H. Boumaraf: *Blind separation of speech mixtures based on nonstationarity*. In *Proceedings of the 7th International Symposium on Signal Processing and Its Applications (ISSPA 2003)*, volume 2, pages 73–76, Paris, France, July 1–4, 2003.
- [202] M. D. Plumbley: *Dictionary learning for l_1 -exact sparse coding*. In *Proceedings of the 7th International Conference on Independent Component Analysis and Signal Separation (ICA 2007)*, page to appear, London, UK, September 9–12, 2007. <http://www.elec.qmul.ac.uk/people/markp/2007/Plumbley-ica07.pdf>.
- [203] M. Puigt: *Séparation aveugle de sources pour les mélanges à atténuations et retards – Application aux signaux de parole et de musique*. Rapport de DEA Signal, Image et Acoustique, LAMI, Toulouse, 2003.
- [204] C.G. Puntonet, A. Mansour, C. Bauer, and E. Lang: *Separation of sources using simulated annealing and competitive learning*. *Neurocomputing*, 49(1–4):39–60, December 2002.
- [205] M. El Rhabi, G. Gelle, H. Fenniri, and G. Delaunay: *A penalized mutual information criterion for blind separation of convolutive mixtures*. *Signal Processing*, 84(10):1979–1984, October 2004.
- [206] S. Rickard, R. Balan, and J. Rosca: *Real-time time-frequency based blind source separation*. In *Proceedings of the 3rd International Symposium on Independent Component Analysis and Blind Signal Separation (ICA 2001)*, pages 651–656, San Diego, USA, December 9–13, 2001.

-
- [207] S. Rickard and M. Fallon: *The gini index of speech*. In *Proceedings of the 39th Annual Conference on Information Sciences and Systems (CISS2004)*, Princeton, New Jersey, USA, March 2004.
- [208] S. Rickard, T. Melia, and C. Fearon: *DESPRIT – hitogram based blind source separation of more sources than sensors using subspace methods*. In *Proceedings of the 2005 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, New Paltz, New York, USA, October 16–19, 2005.
- [209] S. Rickard and O. Yilmaz: *On the approximate W-disjoint orthogonality of speech*. In *Proceedings of International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2002)*, Orlando, USA, May 13–17, 2002.
- [210] R. Roy and T. Kailath: *ESPRIT – estimation of signal parameters via rotational invariance techniques*. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 37(7):954–995, 1989.
- [211] R. Saab, Ö. Yilmaz, M. J. McKeown, and R. Abugharbieh: *Underdetermined sparse blind source separation with delays*. In *Proceedings of the Workshop on Signal Processing with Adaptive Sparse Structured Representations (SPARS’05)*, Rennes, France, November 16–18, 2005.
- [212] R. Saab, Ö. Yilmaz, M. J. McKeown, and R. Abugharbieh: *Underdetermined anechoic blind source separation via ℓ^q -basis-pursuit with $q < 1$* . *IEEE Transactions on Signal Processing*, 55(8):4004–4017, August 2007.
- [213] S. Samadi, M. Babaie-Zadeh, C. Jutten, and K. Nayebi: *Blind source separation by adaptive estimation of score function difference*. In Springer-Verlag (editor): *Proceedings on the 5th International Symposium on Independent Component Analysis and Blind Signal Separation (ICA 2004)*, volume LNCS 3195, Granada, Spain, September 22–24, 2004.
- [214] G. Sanguinetti, J. Laidler, and N. D. Lawrence: *Automatic determination of the number of clusters using spectral algorithms*. In *Proceedings of the 2005 IEEE Workshop on Machine Learning for Signal Processing*, pages 55–60, Mystic, Connecticut, USA, September 28–30 2005.
- [215] H. Sawada, S. Araki, R. Mukai, and S. Makino: *Grouping separated frequency components by estimating propagation model parameters in frequency-domain blind source separation*. *IEEE Transactions on Audio, Speech and Language Processing*, 15(5):1592–1604, July 2007.
- [216] H. Saylani, Y. Deville, S. Hosseini, and M. Habibi: *A multi-tag radio-frequency identification system using a new blind source separation method based on spectral decorrelation*. In *Proceedings of the Second International Symposium on Communications, Control and Signal Processing (ISCCSP 2006)*, Marrakech, Morocco, March 13–15, 2006.
- [217] H. Saylani, S. Hosseini et Y. Deville: *Une nouvelle approche fréquentielle pour la séparation aveugle de signaux non-stationnaires et autocorrélés*. Dans *Actes du 21^{ème} Colloque National GRETSI 2007*, Troyes, France, 11–14 septembre 2007.

- [218] G. Schwarz: *Estimating the dimension of a model*. Annals of Statistics, 6(2):461–464, March 1978.
- [219] P. Smaragdis: *Convolutional speech bases and their application to supervised speech separation*. IEEE Transactions on Audio, Speech and Language Processing, 15(1):1–12, January 2007.
- [220] D. Smith, J. Lukasiak, and I. Burnett: *A block-adaptative audio separation technique based upon time-frequency information*. In *Proceedings of the 12th European Signal Processing Conference (EUSIPCO 2004)*, pages 393–396, Vienna, Austria, September 6–10, 2004.
- [221] D. Smith, J. Lukasiak, and I. Burnett: *Two channel, block adaptative audio separation using the cross correlation of time frequency information*. In Springer-Verlag (editor): *Proceedings on the 5th International Symposium on Independent Component Analysis and Blind Signal Separation (ICA 2004)*, volume LNCS 3195, Granada, Spain, September 22–24, 2004.
- [222] D. Smith, J. Lukasiak, and I. Burnett: *A sequential approach to sparse component analysis*. In *Proceedings of the 2005 IEEE 7th Workshop on Multimedia Signal Processing (MMSP'05)*, pages 129–132, Shangai, China, October 30 – November 2, 2005.
- [223] A. Souselias: *Blind source detection and separation using second-order non-stationarity*. In *Proceedings of the IEEE 1995 International Conference on Acoustics, Speech, and Signal Processing (ICASSP 1995)*, volume 3, pages 1912–1915, Detroit, Michigan, USA, May 9–12, 1995.
- [224] M. A. Stephens: *EDF statistics for goodness of fit and some comparisons*. Journal of the American Statistical Association, 69(347):730–737, September 1974.
- [225] I. Takigawa, M. Kudo, and J. Toyama: *Performance analysis of minimum ℓ_1 -norm solutions for underdetermined source separation*. IEEE Transactions on Signal Processing, 52(3):582–591, March 2004.
- [226] A. Taleb: *Séparation des mélanges non-linéaires*. Thèse de doctorat, Institut National Polytechnique de Grenoble, 1999.
- [227] A. R. Teixeira, A. M. Tomé, E. W. Lang, and K. Stadlthanner: *Delayed AMUSE – a tool for blind source separation and denoising*. In Springer-Verlag (editor): *Proceedings on the 5th International Symposium on Independent Component Analysis and Blind Signal Separation (ICA 2004)*, volume LNCS 3195, Granada, Spain, September 22–24, 2004.
- [228] F. J. Theis, C. G. Puntonet, and E. W. Lang: *Median-based clustering for underdetermined blind signal processing*. IEEE Signal Processing Letters, 13(2), February 2006.
- [229] J. Thomas, Y. Deville, and S. Hosseini: *Time-domain fast fixed-point algorithms for convolutional ICA*. IEEE Signal Processing Letters, 13(4):228–231, April 2006.
- [230] J. Thomas, Y. Deville, and S. Hosseini: *Differential fast fixed-point algorithms for underdetermined instantaneous and convolutional partial blind source separation*. IEEE Transactions on Signal Processing, 55(7):3717–3729, July 2007.

-
- [231] L. Tong, R. W. Liu, V. C. Soon, and Y. F. Huang: *Indeterminacy and identifiability of blind identification*. IEEE Transactions on Circuits and Systems, 38(5):499–509, May 1991.
- [232] L. Tong, V. C. Soon, R. Liu, and Y. Huang: *AMUSE: a new blind identification algorithm*. In *Proceedings of the IEEE International Symposium on Circuits and Systems (ISCAS)*, volume 3, pages 1784–1787, New Orleans, Louisiana, USA, May 1–3, 1990.
- [233] K. Torkkola: *Blind separation of convolved sources based on information maximization*. In *Proceedings of the IEEE Workshop on Neural Networks for Signal Processing (NNSP 1996)*, Kyoto, Japan, September 4–6, 1996.
- [234] K. Torkkola: *Blind separation for audio signals – are we there yet?* In *Proceedings of the Workshop on Independent Component Analysis and Blind Signal Separation*, Aussois, France, January 11–15 1999.
- [235] J. Tugnait: *Identification and deconvolution of multichannel linear non-gaussian processes using higher order statistics and inverse filter criteria*. IEEE Transactions on Signal Processing, 45(3):658–672, March 1997.
- [236] H. Viste: *Binaural Localization and Separation Techniques*. PhD thesis, Ecole polytechnique fédérale de Lausanne, 2004.
- [237] H. Viste and G. Evangelista: *An extension for source separation techniques avoiding beats*. In *Proceedings of the 5th International Conference on Digital Audio Effects (DAFx-02)*, pages 71–75, Hamburg, Germany, September 26–28, 2002.
- [238] B. Wang and M. D. Plumbley: *Musical audio stream separation by non-negative matrix factorization*. In *Proceedings of the DMRN Summer Conference*, Glasgow, Scotland, July 23–24, 2005. <http://www.elec.qmul.ac.uk/people/markp/2005/WangPlumbley05-dmrn.pdf>.
- [239] B. Wang and M. D. Plumbley: *Investigating single-channel audio source separation methods based on non-negative matrix factorization*. In *Proceedings of the 2006 ICA Research Network International Workshop (ICArn 2006)*, pages 17–20, Liverpool, England, September 18–19, 2006. <http://www.elec.qmul.ac.uk/people/markp/2006/WangPlumbley06-ICArnWS.pdf>.
- [240] R. A. Wiggins: *On minimum entropy deconvolution*. Geoplotting, 16:21–35, January 1978.
- [241] S. Winter, W. Kellermann, H. Sawada, and S. Makino: *MAP-based underdetermined blind source separation of convolutive mixtures by hierarchical clustering and ℓ^1 -norm minimization*. EURASIP Journal on Advances in Signal Processing, 2007:Article ID 24717, 12 pages, 2007.
- [242] M. Xiao, S. Xie, and Y. Fu: *A novel approach for underdetermined blind sources separation in frequency domain*. In Springer-Verlag (editor): *Proceedings of the International Symposium on Neural Networks (ISNN 2005)*, volume LNCS 3497, pages 484–489, Chongqing, China, May 30 – June 1, 2005.
- [243] M. Xiao, S. Xie, and Y. Fu: *A statistically sparse decomposition principle for underdetermined blind source separation*. In *Proceedings of the 2005 IEEE International*

- Symposium on Intelligent Signal Processing and Communication Systems (ISPACS 2005)*, pages 165–168, Hong Kong, China, December 13–16, 2005.
- [244] R. Xui and D. Wunsch II: *Survey of clustering algorithms*. IEEE Transactions on Neural Networks, 16(3):645–678, May 2005.
- [245] A. Yeredor: *Blind source separation with pure delays mixture*. In *Proceedings of the third International Workshop on Independent Component Analysis and Blind Source Separation (ICA2001)*, San-Diego, USA, December 2001.
- [246] A. Yeredor: *Non-orthogonal joint diagonalization in the least-squares sense with application in blind source separation*. IEEE Transactions on Signal Processing, 50(7):1545–1551, July 2002.
- [247] A. Yeredor: *Time-delay estimation in mixtures*. In *Proceedings of the 2003 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2003)*, volume 5, pages 237–240, Hong Kong, China, April 6–10, 2003.
- [248] Ö. Yilmaz and S. Rickard: *Blind separation of speech mixtures via time-frequency masking*. IEEE Transactions on Signal Processing, 52(7):1830–1847, July 2004.
- [249] Y. J. Zhang and Z. Q. Liu: *Self-splitting competitive learning: a new on-line clustering paradigm*. IEEE Transactions on Neural Networks, 13(2):369–380, March 2002.

Publications scientifiques de l'auteur

Articles de revue

Y. Deville et M. Puigt : *Temporal and time-frequency correlation-based blind source separation methods. Part I : Determined and underdetermined linear instantaneous mixtures*, revue Signal Processing, Volume 87, Issue 3, pp. 374-407, mars 2007.

M. Puigt et Y. Deville : *Time-frequency ratio-based blind separation methods for attenuated and time-delayed sources*, revue Mechanical System and Signal Processing (MSSP), Volume 19, Issue 6, pp. 1348-1379, novembre 2005.

Actes de conférence internationales

Conférences à comité de lecture

M. Puigt et Y. Deville : *A time-frequency correlation-based blind source separation method for time-delayed mixtures*, Actes de "the 2006 IEEE International Conference on Acoustics, Sound and Signal Processing (ICASSP 2006)", ISBN : 1-4244-0469-X, pp. V-853 - V-856, Toulouse, France, 14-19 mai 2006.

Y. Deville, D. Bissessur, M. Puigt, S. Hosseini et H. Carfantan : *A time-scale correlation-based blind separation method applicable to correlated sources*, Actes de "the European Symposium on Artificial Neural Networks (ESANN 2006)", pp. 337-344, Bruges, Belgique, 25-28 avril 2006.

M. Puigt et Y. Deville : *Two time-frequency ratio-based blind source separation methods for time-delayed mixtures*, Actes de "the 6th International Conference on Independent Component Analysis and Blind Source Separation (ICA 2006)", Charleston, Caroline du Sud, USA, 5-8 mars 2006.

Y. Deville, M. Puigt et B. Albouy : *Time-frequency blind signal separation : extended methods, performance evaluation for speech sources*, Actes de "the International Joint Conference on Neural Networks (IJCNN 2004)", numéro de catalogue IEEE : 04CH37541C, ISBN : 0-7803-8360-5, pp. 255-260, Budapest, Hongrie, 25-29 juillet 2004.

Conférences sans comité de lecture

M. Puigt et Y. Deville : *A new time-frequency correlation-based source separation method for attenuated and time shifted mixtures*, Actes de "the 8th International Workshop on Electronics, Control, Modelling, Measurement and Signals (ECMS 2007)", Liberec, République Tchèque, 21-23 mai 2007.

M. Puigt et Y. Deville : *A time-frequency ratio-based source separation method for time-delayed mixtures exploiting constant-time zones*, Actes de "the 7th International Workshop on Electronics, Control, Modelling, Measurement and Signals (ECMS 2005)", Toulouse, France, 17-20 mai 2005.

Colloques en français

M. Puigt et Y. Deville : *Une méthode temps-fréquence de séparation aveugle de sources, pour mélanges à retard, exploitant les zones à temps constant*, Colloque de l'Ecole Doctorale Informatique et Télécommunications (EDIT 2006), pp. 143-147, Toulouse, France, 22-23 mai 2006.

Blind source separation methods based on time-frequency transforms. Application to speech signals

Matthieu PUIGT

Thesis supervisor : Yannick DEVILLE

Abstract

Several time-frequency (TF) blind source separation (BSS) methods have been proposed in this thesis. In the systems output that have been used, a contribution of each source is estimated, using only mixed signals. All the methods proposed in this manuscript find tiny TF zones where only one source is active and estimate the mixing parameters in these zones. These approaches are particularly well suited for non-stationary sources (speech, music).

We first studied and improved linear instantaneous methods based on variance or correlation criteria, that have been previously proposed by our team. They yield excellent performance for speech signals and can also separate spectra from astrophysical data. However, the nature of the mixtures that they can process limits their application fields.

We have extended these approaches to more realistic mixtures. The first extensions consider attenuated and delayed mixtures of sources, which corresponds to mixtures in anechoic chamber. They require less restrictive sparsity assumptions than some approaches previously proposed in the literature, while addressing the same type of mixtures. We have studied the contribution of clustering techniques to our approaches and have achieved good performance for mixtures of speech signals.

Lastly, a theoretical extension of these methods to general convolutive mixtures is described. It needs strong sparsity hypotheses and we have to solve classical indeterminacies of frequency-domain BSS methods.

Keywords: blind source separation, linear instantaneous mixtures, attenuated and delayed mixtures, convolutive mixtures, sparsity, time-frequency analysis, non-stationary sources, clustering.

Additional keywords: short time Fourier transform, correlation, variance, speech, astrophysic.

Méthodes de séparation aveugle de sources fondées sur des transformées temps-fréquence. Application à des signaux de parole

Matthieu PUIGT

Directeur de thèse : Yannick DEVILLE

Thèse (discipline : **Traitement du Signal**) soutenue le **13 décembre 2007**, en **salle de conférences du CESR** (Toulouse).

Résumé

Plusieurs méthodes de séparation aveugle de source (SAS), fondées sur des transformées temps-fréquence (TF), ont été proposées au cours de cette thèse. En sortie des systèmes utilisés, une contribution de chaque source est estimée, uniquement à l'aide des signaux mélangés. Toutes les méthodes étudiées dans ce manuscrit trouvent des petites zones du plan TF où une seule source est présente et estiment dans ces zones les paramètres de mélange. Ces approches sont particulièrement adaptées aux sources non-stationnaires.

Nous avons tout d'abord étudié et amélioré des méthodes proposées précédemment par l'équipe, basées sur des critères de variance ou de corrélation, pour des mélanges linéaires instantanés. Elles apportent d'excellentes performances pour des signaux de parole et peuvent aussi séparer des spectres issus de données astrophysiques. Cependant, la nature des mélanges qu'elles peuvent traiter limite leur champ d'application.

Nous avons donc étendu ces approches à des mélanges plus réalistes. Les premières extensions considèrent des mélanges de sources atténuées et décalées temporellement, ce qui correspond physiquement aux mélanges en chambre anéchoïque. Elles nécessitent des hypothèses de parcimonie beaucoup moins fortes que certaines approches de la littérature, tout en traitant le même type de mélanges. Nous avons étudié l'apport de méthodes de classification non-supervisée sur nos approches et avons obtenu de bonnes performances pour des mélanges de signaux de parole.

Enfin, une extension théorique aux mélanges convolutifs généraux est décrite mais nécessite de fortes hypothèses de parcimonie et le réglage d'indéterminations propres aux méthodes fréquentielles.

Mots-clés : séparation aveugle de sources, mélanges linéaires instantanés, mélanges atténués et décalés temporellement, mélanges convolutifs, parcimonie, analyse temps-fréquence, sources non-stationnaires, classification non-supervisée.

Mots-clés complémentaires : transformée de Fourier à court terme, corrélation, variance, parole, astrophysique.

Laboratoire d'Astrophysique de Toulouse-Tarbes,
Université de Toulouse, CNRS,
Observatoire Midi-Pyrénées,
14 Av. Edouard Belin, 31400 Toulouse,
FRANCE