Prise de son distante par système multimicrophone. Application à la communication parlée en environnement bruyant

Ludovick Lepauloux

▶ **To cite this version:**

Ludovick Lepauloux. Prise de son distante par système multimicrophone. Application à la communication parlée en environnement bruyant. Traitement du signal et de l'image [eess.SP]. Université Rennes 1, 2010. Français. NNT : . tel-00636256

HAL Id: tel-00636256

https://theses.hal.science/tel-00636256

Submitted on 27 Oct 2011

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.

N° d'ordre : 4156 ANNÉE 2010

THÈSE / UNIVERSITÉ DE RENNES 1
sous le sceau de l'Université Européenne de Bretagne

pour le grade de
DOCTEUR DE L'UNIVERSITÉ DE RENNES 1

Mention : Traitement du signal et télécommunications

Ecole doctorale MATISSE

présentée par

Ludovick Lepauloux

préparée à l'unité de recherche (6074 IRISA UMR)
(Institut de Recherche en Informatique et Systèmes Aléatoires)

Prise de son distante par système multimicrophone. Application à la communication parlée en environnement bruyant

Thèse soutenue à Lannion
le 8 décembre 2010

devant le jury composé de :

Maurice CHARBIT
Professeur des universités
Télécom ParisTech / président

Gaël MAHE
Maître de conférence
Université Paris Descartes / rapporteur

Iain MCCOWAN
Assistant Professeur
Queensland University of Technology / rapporteur

Dominique PASTOR
Professeur des universités
Télécom Bretagne / examinateur

Pascal SCALART
Professeur des universités
ENSSAT / directeur de thèse

Claude MARRO
Ingénieur R&D
Orange Labs / co-directeur de thèse

"Rien n'est plus pratique qu'une bonne théorie"
Kurt Lewin
(1890-1947)

Résumé : Dans notre société de plus en plus mobile, nous constatons que les systèmes de communication mains libres sont devenus une norme, que ce soit pour les interactions homme-machine ou homme-homme. Cependant, il est difficile de garantir la qualité du signal de parole lors d'une prise de son distante car les microphones captent en plus du signal désiré le bruit de fond, la réverbération, le signal d'écho et les autres sources de parole concurrentes. Les antennes acoustiques associées aux techniques de formation de voies sont une solution efficace et aujourd'hui technologiquement réalisable. Grâce à leur directivité, elles améliorent le rapport signal sur bruit de même que, en exploitant la corrélation spatiale du champ sonore, elles permettent de réduire la réverbération, l'écho et le bruit. Les antennes acoustiques jouent un rôle croissant dans la téléphonie mains libres, la reconnaissance de locuteur, les systèmes à commande vocale, l'aide aux malentendants ou la surveillance audio. Dans cette thèse, pour s'adapter aux environnements variants dans le temps, aux non-stationnarités des signaux et aux éventuels mouvements du locuteur, nous considérons les antennes adaptatives de type GSC (*Generalized Sidelobe Canceller*) pour leur implémentation avantageuse. Cependant, l'un des défauts majeurs de cette antenne, couramment utilisée, réside dans la fuite du signal utile dans l'étage d'estimation du bruit provoquant de fortes dégradations voire l'auto-annulation de ce signal en sortie. Pour résoudre ce problème, nous proposons de mettre à profit la structure récursive spécifiquement dédiée à la gestion de la diaphonie dans la technique d'annulation adaptative de bruit bicapteur, nommée CTRANC (*Crosstalk-Resistant Adaptive Noise Canceller*). Ainsi, avant de proposer une extension de cette méthode au cas du GSC et donc de définir une nouvelle architecture récursive pour celui-ci, nous donnons dans un premier temps une analyse complète du CTRANC et proposons de nouveaux algorithmes adaptatifs dans le domaine fréquentiel. Par ailleurs, notre étude fournit également de nouveaux résultats sur les propriétés de convergence et l'existence d'un point d'équilibre pour le CTRANC. Enfin, des résultats expérimentaux confortent l'analyse et montrent l'efficacité des solutions proposées pour la suppression du signal de fuite et l'amélioration des performances.

Mots-clés : Traitement de la parole - Antenne acoustique - Filtrage adaptatif - Structure récursive - Formation de voies - Annulation adaptative de bruit - Déréverbération

Abstract : In our increasingly mobile society we find that hands-free operation of communication systems is becoming more and more the norm, whether we talk to other humans or to machine. There are significant challenges in enabling high quality speech communication in adverse acoustic environments where the recorded microphone signals are corrupted by background noise, room reverberation, far-end echo signals and competing speech sources. Microphone arrays and more specifically beamforming methods are enabling technology for hands-free communication that is now viable and cost effective. By offering directional gain to improve the signal-to-noise ratio and taking the spatial correlation of sound field into account to dereverberate the desired speech signal and to reduce noise and acoustic echoes, microphone arrays techniques play an important role in hands-free mobile telephony, distant-talker speech recognition, voice-controlled systems, hearing aids, or audio monitoring. In this thesis, to tackle time-varying environments with both nonstationary signal characteristics and potentially moving sources, we consider the Generalized Sidelobe Canceller (GSC) which is an efficient implementation of adaptive beamformers. However, one of the main drawback of this well used antenna are the self-cancellation phenomena of the desired signal caused by the signal leakage into the noise reference. To cope with this problem, we propose to take benefit of the ability of the crosstalk-resistant adaptive noise canceller (CTRANC) to deal with crosstalk problem that, in fact, is the same as the signal leakage problem in the GSC. Thus, before describing the new feedback structure for the GSC, we first give a complete analysis of the CTRANC and propose new adaptive algorithms in the frequency-domain. In addition, our study establish new results about the convergence properties and the existence of an equilibrium point for this feedback structure. Finally, results show that feedback GSC is an effective solution to solve the leakage problem and to improve the performance.

Keywords : Speech processing - Microphone array - Adaptive filtering - Feedback structure - Beamforming - Adaptive noise cancellation - Dereverberation

Sommaire

Introduction **1**

1 Débruitage de la parole par antenne acoustique **15**

1.1 De l'antenne au postfiltre 16

1.1.1 Spécificités de l'antenne acoustique 19

1.1.2 Influence des perturbations et des incertitudes 27

1.2 Techniques pour l'amélioration de la robustesse 31

1.2.1 Formation de voies adaptative robuste 34

1.3 Performances du *Generalized Sidelobe Canceller* (GSC) 48

1.4 Synthèse du chapitre 49

2 Vers la définition d'une cellule élémentaire **53**

2.1 Constat d'une analogie 54

2.2 Annulation de bruit adaptative robuste à la diaphonie 56

2.3 Modèle de mélange, conditions de stabilité et positions des sources . . . 60

2.3.1 Modélisation des canaux acoustiques 60

2.3.2 Simplification du modèle de mélange 63

2.3.3 Notions de stabilité 66

2.3.4 Influence de l'hypothèse de causalité sur la position des sources et les performances 68

2.4 Minimisation de l'Erreur Quadratique Moyenne (EQM) ou décorrélation ? . 74

2.4.1 Minimisation de l'erreur quadratique moyenne 75

2.4.2 Illustration sur un cas simple 77

2.5 Synthèse du chapitre 81

3 Solution architecturale : le DCTRANC **83**

3.1 Présentation du DCTRANC 84

3.1.1 Description générale 84

3.1.2 Découplage simple - Filtre adaptatif $\mathbf{w_1}$ . . . 85
3.1.3 Version hybride - Filtre adaptatif $\mathbf{w_2}$ . . . 86
3.1.4 Gestion du bruit du gradient - Contrainte algorithmique . . . 87
3.2 Approches dédiées à l'amélioration de la gestion des pas . . . 89
3.2.1 Règles algorithmiques pour la gestion des pas . . . 90
3.2.2 Bilan des méthodes existantes . . . 93
3.2.3 Méthodes proposées . . . 93
3.3 Etude des solutions optimales . . . 104
3.3.1 Solutions analytiques théoriques . . . 104
3.3.2 Réflexion sur le DCTRANC . . . 108
3.4 Implémentation fréquentielle et performances en environnement réel . . 108
3.4.1 Influence de la sous-modélisation . . . 110
3.4.2 Prise de son binaurale . . . 111
3.5 Synthèse du chapitre . . . 112

4 Solution algorithmique optimisée : le LCCTRANC 115
4.1 Algorithme optimal de Zinser, Mirchandani et Evans . . . 116
4.1.1 Analyse de l'erreur quadratique moyenne . . . 116
4.1.2 Filtres optimaux . . . 120
4.1.3 Algorithme des moindres carrés transverses . . . 122
4.1.4 Evaluation de la complexité . . . 124
4.2 Réduction de la complexité de l'algorithme optimal . . . 126
4.2.1 Concept à la base de nos simplifications algorithmiques . . . 126
4.2.2 Obtention du nouvel algorithme . . . 127
4.2.3 Evaluation de la complexité du LCCTRANC . . . 129
4.3 Comparaison des performances . . . 131
4.3.1 Mélange de signaux synthétiques . . . 131
4.3.2 Mélange de signaux réels . . . 133
4.3.3 Gestion des instabilités - Réduction de l'ordre . . . 135
4.4 Implémentation fréquentielle et performances en environnement réel . . 139
4.4.1 Influence de la sous-modélisation . . . 139
4.4.2 Prise de son binaurale . . . 141
4.5 Synthèse du chapitre . . . 142

5 Analyse de la convergence et de la stabilité de la structure récursive 143

5.1 Données et points bloquants du problème 144
5.1.1 Description du problème 144
5.1.2 Absence de solution analytique facilement exploitable 146
5.2 Recherche de l'état d'équilibre 149
5.2.1 Analyse de l'erreur résiduelle sur les signaux estimés 149
5.2.2 Conservation de l'énergie de l'erreur résiduelle 150
5.2.3 Analyse des Rapports Signaux à Artefacts (RSA) 152
5.2.4 Analyse de l'équilibre à partir du principe d'égalité 152
5.3 Résultats expérimentaux avec des signaux synthétiques 155
5.3.1 Convergence initiale et asymptotique 155
5.3.2 RSA sur chaque voie 156
5.4 Résultats expérimentaux avec des signaux de parole 160
5.4.1 Conservation de l'énergie de l'erreur résiduelle - Principe d'égalité . . . 160
5.4.2 RSA sur chaque voie - Relation d'équilibre 161
5.5 Etude des résultats expérimentaux dans le cas dissymétrique 166
5.5.1 Cas des sources de même puissance 168
5.5.2 Cas des sources de puissance différente 168
5.6 Synthèse du chapitre 173

6 Extension au cas multicapteur : vers la définition d'un GSC récursif 177
6.1 Implémentations récursives du GSC 178
6.1.1 Rappel : GSC robuste (RGSC) 178
6.1.2 GSC robustes récursifs 179
6.1.3 Inconvénients de ces méthodes 181
6.2 GSC récursif à faible coût 182
6.3 Impact du chemin direct sur les estimées produites par le LCCTRANC . . 183
6.4 Synchronisation dans les antennes récursives 185
6.5 Evaluation en conditions réelles 192
6.5.1 Description de l'environnement acoustique et de la prise de son 192
6.5.2 Description du corpus 193
6.5.3 Résultats expérimentaux 194
6.6 Synthèse du chapitre 198

Conclusion et perspectives 201

Références bibliographiques **224**

Annexes **225**

A Generalized Sidelobe Canceller **227**

A.1 GSC généralisé 227

A.1.1 Solution optimale sous contraintes - Frost 228

A.1.2 Solution optimale non contrainte - GSC 231

A.1.3 Construction de la matrice de blocage 234

A.1.4 Pointage de la source - *Matched Beamformer* 235

A.1.5 Annuleur de bruit multicanal 236

A.2 Estimation des Fonctions de Transfert FTA et FTR 237

A.2.1 Méthode du sous-espace signal 237

A.2.2 Méthode utilisant la non-stationnarité du signal 238

B Réduction de bruit par atténuation spectrale à court terme **241**

B.1 Introduction 242

B.2 Estimation du bruit et calcul de RSB 242

B.2.1 Estimation du bruit avec DAV 243

B.2.2 Estimation continue 243

B.2.3 Calcul du RSB par l'approche décision-dirigée (DD) 244

B.3 Règles d'atténuation spectrale 246

B.3.1 Approches non statistiques 246

B.3.2 Approches statistiques 248

C Evaluation objective **251**

C.1 Description de la méthode d'évaluation 252

C.2 Critères utilisés 252

D Roomsim **255**

D.1 Présentation des possibilités de *Roomsim* 256

D.2 Scène synthétique n°1 256

D.3 Scène synthétique n°2 257

E Prise de son binaurale **265**

E.1 Introduction 266

E.2 Description de l'environnement acoustique 266

Abréviations et acronymes

ABF *Adaptive Blocking Filter*

ABM *Adaptive Blocking Matrix*

ACF *Adaptive Canceling Filter*

AIC *Adaptive Interference Canceller*

ALU Antenne Linéaire Uniforme

ANC *Adaptive Noise Canceller*

APSM *Adaptive Projected Subgradient Method*

CTRANC *Crosstalk Resistant Adaptive Noise Canceller*

CTSI Centre des Technologies de la Sécurité Intérieure

DAV Détection d'Activité Vocale

DCTRANC *Decoupled Crosstalk Resistant Adaptive Noise Canceller*

DD *Decision-Directed*

DI *Directivity Index*

DSP Densité Spectrale de Puissance

EQM Erreur Quadratique Moyenne

EQMM Erreur Quadratique Moyenne Minimale

EUSIPCO *European Signal Processing Conference*

FBF *Fixed Beamformer*

FFT *Fast Fourier Transform*

FLMS *Frequency domain Least Mean Square*

FTA Fonctions de Transferts Acoustiques

FTR Fonctions de Transferts Relatives

FWM Filtrage de Wiener Multicanal

GMDF *Generalized Multidelay Filter*

GSC *Generalized Sidelobe Canceller*

GSVD Generalized Singular Value Decomposition

HRNR *Harmonic Regeneration Noise Reduction*

ITU *International Telecommunication Union*

IWAENC *International Workshop on Acoustic Echo and Noise Control*

LCCTRANC *Low Complexity Crosstalk Resistant Adaptive Noise Canceller*

LCMV *Linearly Constraint Minimum Variance*

LMS *Least Mean Square*

MVDR *Minimum Variance Distortionless Response*

MSD *Mean Square Deviation*

NLMS *Normalized Least Mean Square*

ODE *Ordinary Differential Equation*

OLA *Overlap-Add*

OLS *Overlap-Save*

RGSC *Robust Generalized Sidelobe Canceller*

RIF Réponse Impulsionnelle Finie

RII Réponse Impulsionnelle Infinie

RMVB *Robust Minimum Variance Beamformer*

RSA Rapport Signal à Artefact

RSB Rapport Signal à Bruit

RSI Rapport Signal à Interférent

SAD *Symmetric Adaptive Decorrelation*

SDR-GSC *Speech Distortion Regularized-Generalized Sidelobe Canceller*

SOCP *Second-Order Cone Program*

SRP *Steered Response Power*

SSPD *Sensor Signal Processing for Defence*

TDOA *Time Difference of Arrival*

TFCT Transformée de Fourier à Court Terme

TFD Transformée de Fourier Discrète

TFDI Transformée de Fourier Discrète Inverse

TPS Traitement de la Parole et du Son

TSNR *Two-Step Noise Reduction*

WNG *White Noise Gain*

WOLA *Weighted Overlap-Add*

Introduction

Si depuis près de quarante ans la communauté scientifique travaille à l'utilisation des antennes acoustiques couplées aux méthodes de formation de voies pour l'enrichissement de la parole, cela n'a pas été toujours suivi dans la pratique. Ce n'est que depuis une vingtaine d'années que naît l'engouement pour les systèmes multimicrophones et ceux-ci commencent à apparaître dans les systèmes de visioconférences, les kits mains libres, les interfaces hommes/machines, les implants cochléaires... Depuis environ deux ans, une nouvelle génération de téléphones portables équipés d'au moins deux microphones tel que le LG SH400 ou Nokia 6720 relancent les activités de recherche des équipementiers du secteur des télécommunications qui voient poindre le besoin de tels systèmes notamment pour la conception des environnements dit "intelligents".

Longtemps jugées trop coûteuses et encombrantes, les techniques de formation de voies étaient aussi victime d'une incapacité technologique. En effet, il a fallu attendre l'arrivée de processeurs de traitement du signal performants et la maîtrise de la miniaturisation des microphones (notamment sur silicium) et des circuits pour que le coût de calcul et l'encombrement de l'antenne acoustique ne soient plus des facteurs limitants au déploiement de ces méthodes, notamment dans les systèmes embarqués. Aujourd'hui, les verrous technologiques sont levés au prix de contraintes d'utilisations plus diverses et plus adverses. L'utilisateur souhaite par exemple pouvoir téléphoner dans les lieux publics ou au milieu d'une foule sans que son correspondant ne soit gêné par le bruit ambiant. Ces conditions de bruit sont encore plus pénalisantes dès lors que l'on souhaite effectuer la prise de son en mains libres. Cette volonté dynamise les travaux de recherche sur la prise de son par antenne acoustique et, parallèlement, nous voyons apparaître de nouvelles méthodes en réponse à ces nouveaux cas d'utilisation.

Parmi elles, les techniques de séparation de sources qui connaissent un véritable intérêt depuis plus de dix ans. A l'origine, celles-ci se démarquent des méthodes de traitement par antennes acoustiques par le fait que les microphones sont censés être disposés de manière aléatoire et inconnue dans l'espace. Mais la différence, si tant est qu'elle existe, est aujourd'hui infime et la frontière est parfois très floue puisque

certains systèmes fusionnent les deux approches. Néanmoins, les méthodes de séparation de sources connaissent les difficultés initiales rencontrées par les antennes acoustiques. En effet, hormis celles clairement prévues pour travailler hors ligne, les plus évoluées restent encore trop coûteuses pour effectuer le traitement en temps réel.
Les travaux de thèse présentés dans ce mémoire, réalisés au laboratoire de Traitement de la Parole et du Son (TPS) au sein des *Orange Labs* (Lannion) et financés par le Centre des Technologies de la Sécurité Intérieure (CTSI), concernent plus spécifiquement la prise de son distante dématérialisée en environnement bruyant par système multimicrophone. Le but de ce travail est de concevoir et proposer un nouvel algorithme de réduction de bruit et d'amélioration de la qualité vocale pour une prise de son réalisée à l'aide d'une antenne acoustique. L'aspect dématérialisé signifie que les locuteurs sont dépourvus de tout équipement audio ou aidant à la capture du signal qu'ils émettent.

Formulation du problème

La Figure 1 illustre une communication parlée mains libres dans un environnement typique. Contrairement aux systèmes de communication classiques (par exemple le téléphone portable), les locuteurs s_1 et s_2 que l'on prendra comme sources utiles, peuvent circuler librement dans la pièce sans porter un équipement audio doté d'un microphone. L'objectif de l'antenne microphonique, généralement située à une position fixe, est de capter les signaux de parole prononcés par s_1 et s_2 en leur soustrayant tous les autres signaux perturbateurs (musique, ventilation et réverbération). Ceux-ci sont à l'origine de plusieurs types de dégradation sur les signaux de parole. De manière générale, en raison de la grande distance entre les sources et le réseau de microphones, le bruit de fond est également bien présent au niveau des microphones, y compris le signal des locuteurs provenant des réflexions sur les murs, les planchers, le plafond et les objets présents dans les salles d'enregistrement (réverbération). A cela, nous pourrions ajouter l'écho acoustique qui, par exemple dans le cas de la visioconférence, résulte d'un couplage acoustique entre hauts-parleurs et microphones. Le bruit de fond peut être très divers et diffère en fonction du lieu de la prise de son. Il peut se composer d'autres sources de parole provenant d'autres intervenants dans la salle (par exemple nous pourrions voir la source s_2 comme perturbatrice). Un restaurant est un lieu typique de ce genre de conditions de bruit où il faudra de plus ajouter les bruits impulsionnels liés aux claquement des portes, des verres et des couverts. C'est ce que l'on appelle le bruit de cocktail ou bien de foule qui constitue un environnement critique de prise de son où l'intelligibilité du signal est très dégradée. Bien que la plupart des sources de bruit correspondent à des signaux

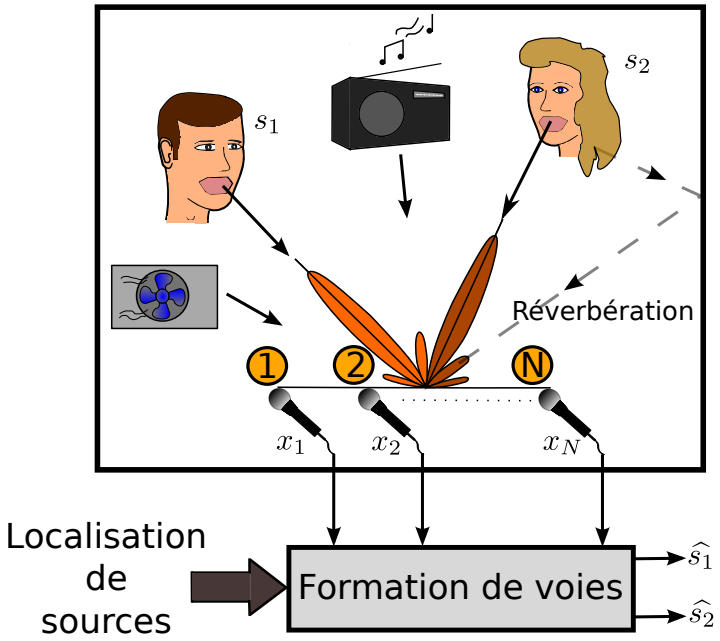


Figure 1 – Environnement acoustique typique d'une prise de son distante

inconnus, il est parfois possible d'obtenir un signal de référence à leur image à l'aide d'un microphone supplémentaire. Par exemple, un microphone dédié à l'enregistrement du bruit du moteur pourra aider l'antenne placée dans l'habitacle d'une voiture à la réduction de bruit. De même que l'utilisation d'un signal synthétique de bruit de roulement guidera l'algorithme de débruitage dans le cas d'une prise de son dans une voiture ou train. Ces cas correspondent à des techniques de rehaussement du signal spécifiques. Dans cette thèse, nous nous concentrerons sur les sources de bruit inconnues pour lequelles aucun signal synthétique de référence n'est disponible.

Réverbération

L'environnement sonore joue un rôle important dans le contexte de la prise de son dématérialisée. Les ondes acoustiques en provenance du locuteur se propagent dans l'air et sont réfléchies par les murs, le plancher, le plafond, et par tout objet présent dans la salle avant d'être captées par les micros. Il en résulte un étalement temporel du signal et une distorsion spectrale, appelée réverbération. Bien que celle-ci puisse être agréable et quelquefois ajoutée ou préservée, elle contribue la plupart du temps à une diminution de l'intelligibilité du signal. De plus, les performances des techniques de localisation sont sérieusement dégradées par la réverbération ambiante, voire même plus que par le bruit de fond. L'ensemble des méthodes mises en œuvre pour la réduction ou la suppression de la réverbération sont appelées techniques de déréverbération ou techniques de déconvolution aveugles. La réverbération

est donc une grandeur extrêmement importante pour la conception des systèmes de prise de son, elle est quantifiée par le temps de réverbération défini comme le temps nécessaire pour que l'énergie de la source sonore présente dans la pièce décroisse de 60 dB. Ce temps est fonction de la dimension de la pièce et des matériaux utilisés pour couvrir les murs, le sol et les plafonds. En ce sens, chacun des matériaux est caractérisé par un coefficient appelé coefficient d'absorption. Celui-ci varie en fonction de la fréquence et de l'angle d'incidence de la source et traduit le comportement des matériaux en réponse à l'excitation par l'onde acoustique. Ces coefficients sont tabulés et utilisés notamment dans l'outil de simulation d'environnement acoustique que nous utiliserons (Annexe D) pour la création des scènes sonores synthétiques avec un temps de réverbération que nous choisirons. L'outil en question tient compte également des autres phénomènes observés dans la pratique, à savoir la diffraction, la diffusion et la dépendance par rapport à la température des grandeurs physiques liées à la propagation. Enfin, il existe une distance limite entre l'antenne acoustique et la source liée à la réverbération et caractéristique de toute prise de son réalisée en environnement clos. Celle-ci, appelée distance critique [69] correspond à la distance où l'énergie émise par une source ponctuelle omnidirectionnelle et provenant du chemin direct est égale à celle issue de la somme des énergies provenant des chemins réfléchis. Au delà de cette distance, il y a davatange d'énergie contenue dans les chemins réverbérants que dans le chemin direct, ce qui se produit pour un environnement classique dès lors que la distance source-antenne acoustique est de l'ordre du mètre. La notion de distance de la source par rapport à l'antenne sera reconsidérée dans le document au Chapitre 2 pour différencier les notions de champ proche et champ lointain.

Par rapport à ce qui vient d'être exposé, il convient de remarquer que le système auditif humain est remarquablement robuste dans la plupart des situations défavorables. En effet, nous sommes en mesure de nous concentrer sur une source de parole même lorsque les conditions de bruit sont difficiles et dans des environnements extrêmement réverbérants. C'est grâce à l'association de notre propre système de prise de son binaural (oreilles) et au traitement adaptatif non-linéaire effectué au niveau de l'oreille interne et de notre cerveau qu'aujourd'hui encore aucun système n'est capable d'égaler les capacités du système auditif humain, en particulier vis-à-vis de sa sélection spatiale. De plus, la vision apporte une aide essentielle pour se focaliser sur une source.

Caractérisation des signaux de parole et de bruits

Avant de nous intéresser à une quelconque méthode de traitement pour le débruitage, nous rappelons tout d'abord les informations importantes sur les signaux d'intérêts. Ceci est primordial puisqu'elles définissent la manière de construire ces techniques autant du point de vue logiciel que matériel. De plus, ces caractéristiques vont nous permettre de comprendre pourquoi une méthode est plus efficace qu'une autre et pourquoi certains phénomènes apparaissent en fonction des méthodes choisies. Nous donnons ici uniquement les propriétés utiles à nos développements.

Signal de parole

Il s'agit d'un signal large bande, de bande utile [50 10000] Hz avec une plage spécifique de [100 3400] Hz qui couvre la zone où la parole est considérée comme intelligible. Cette première caractéristique porte déjà à conséquences. Les antennes à considérer ainsi que les méthodes numériques mises en oeuvre devront être adaptées, c'est-à-dire qu'il faudra concevoir les antennes afin qu'elles soient capables de récupérer le signal sur toute sa bande utile ou alors procéder à un traitement par composantes fréquentielles du signal, en particulier si l'on souhaite utiliser directement les méthodes de traitement d'antenne dites à bande étroite.
Dans sa structure, le signal de parole est aussi marqué au niveau temporel par :

- une quasi-périodicité du signal qui atteste de la prononciation d'un son voisé ;
- une variation du signal semblable à un bruit coloré typique de la présence de fricative ;
- une impulsion lorsqu'une plosive se trouve dans le mot prononcé ;
- et enfin par des périodes de silence qui représentent environ 50% du temps, ceci rapporté sur la durée totale d'une conversation.

Une autre particularité et non des moindres concerne ses propiétés statistiques. Le signal vocal est non-stationnaire ou plus précisément non-stationnaire à l'ordre 2 au sens strict puisque néanmoins de moyenne nulle. De manière générale la stationnarité est supposée vérifiée sur des trames de 20 à 30 ms ce qui permet de garantir la viabilité des estimateurs statistiques utilisés dans les algorithmes au moins sur cette durée. Néanmoins, cela n'est pas forcément nécessaire et pertinent pour les méthodes multimicrophones adaptatives.

Signal de bruit

Par définition, il s'agit d'un signal difficile à classifier, même si certains sont bien connus et très utilisés dans le domaine vaste du traitement du signal, il est rare de voir ces modèles de bruit tels le bruit blanc, le bruit rose, le bruit de foule, le bruit diffus,... être en adéquation parfaite avec la réalité. Ceci s'explique par

le fait qu'il est d'une part difficile d'obtenir des modèles précis de tous les bruits d'environnement existants dans lesquels le système est plongé. D'autre part, quand bien même cela serait possible, la complexité de ces modèles, ainsi que la difficulté d'anticiper la présence ou non d'un ou de plusieurs types de bruit au même instant, met en défaut l'idée d'une recherche exhaustive des modèles pour le débruitage. Typiquement, pour une antenne acoustique nous pouvons dénombrer les différentes sources de bruit suivantes :

- Bruit thermique : intrinsèque aux microphones et dû à l'échauffement des différents composants le constituant ;
- Bruit ambiant : regroupe toutes les sources de signal autre que celles liées à la parole utile. Par exemple, le bruit généré par les pâles d'un ventilateur, du vent, un ordinateur, une foule, un véhicule,...
- Bruit de mesures : lié aux incertitudes sur la position des microphones et leurs propres imperfections.

En résumé, le bruit peut être à bande étroite ou large bande, stationnaire ou non-stationnaire, intermittent ou persistant et peut avoir les mêmes caractéristiques spectrales et/ou d'angle d'arrivée que la source d'intérêt. Mais le facteur le plus important et déterminant pour les méthodes de formation de voies, concerne le degré de cohérence du bruit entre les microphones qui constituent l'antenne acoustique. En ce sens, on parlera de champ cohérent, incohérent ou diffus. Les mélanges réels étant formés le plus souvent de ces trois types de champs de bruit, nous verrons que certaines précautions sont nécessaires pour la conception d'une antenne acoustique afin qu'elle soit en mesure de supprimer ces différents bruits.

Techniques pour l'amélioration de la qualité vocale

Pour chaque forme de dégradation du signal (bruit, écho, réverbération), il existe des traitements que l'on peut distinguer dans un premier temps selon qu'ils utilisent un ou plusieurs microphones. Les traitements monocapteurs, classés suivant si ils utilisent une modélisation paramétrique ou non, sont bien connus de la communauté. Ils ont été largement exploités et bien qu'aujourd'hui de nouvelles méthodes de séparation de sources monocapteurs voient le jour [18, 76, 159], à l'évidence, le manque d'information disponible et discriminante demeure pénalisant. En effet, en ne disposant que d'une seule observation, seul le plan temps-fréquence peut être mis à profit pour les traitements. Or, dans un mélange réel de signaux qui se recouvrent spectralement et dont l'information pertinente peut être contenue dans les mêmes bandes de fréquence, il devient très difficile de concevoir un traitement efficace, peu complexe et capable d'opérer dans des conditions très adverses. A titre d'exemple, la Figure 2 illustre le pire cas. Les sources de paroles sont actives aux mêmes instants,

le recouvrement spectral rend les sources visuellement indissociables. En imaginant maintenant que la scène se tienne dans un véhicule, le bruit qui en résulte aggrave encore un peu plus la situation. Aujourd'hui, la communauté s'accorde à dire que ce type de méthodes se situe à un stade où les gains en performances sont faibles en regard de l'investissement réalisé dans le raffinement des outils du traitement du signal. Le gain en réduction de bruit repose sur la qualité de l'estimation de la Densité Spectrale de Puissance (DSP) du bruit. Plus nous seront capables d'améliorer son estimation et plus ce type de méthodes sera efficace, mais la plus value reste faible. En revanche, l'information de phase est un paramètre très peu exploité dans ce type de méthodes et aucun algorithme efficace, à notre connaissance, n'a encore été proposé. Des études discutent également des modèles de lois les plus appropriés pour représenter le signal de parole et proposent des modélisations plus ou moins complexes pour lesquelles le choix des paramètres reste une tâche plus ou moins subjective. De plus, si l'approximation par une loi gaussienne n'est pas des plus précises, la complexité de la modélisation induite par l'utilisation des distributions de Laplace, de Weibull [174] ou du modèle gaussien généralisé par rapport à la faible amélioration apportée n'encourage pas forcément au développement de ces méthodes dans la pratique. En effet, la force des modèles plus complexes repose sur la capacité de modifier les paramètres en temps réel afin d'ajuster au mieux la distribution. Notre affirmation s'appuie sur les travaux récents de Sehr et al. qui montrent dans [160, 161] que la réverbération change la forme, la moyenne et la variance de la distribution du signal de parole en fonction de l'énergie du vecteur caractéristique de la trame précédente. Or, il est parfois difficile d'envisager une modification en temps réel de la distribution puisque la plupart du temps celle-ci est tabulée et stockée en mémoire. Par conséquent, il faut avoir une capacité mémoire plus importante pour exploiter pleinement la modularité des distributions paramétriques dans les systèmes embarqués.

Par opposition, en combinant le filtrage temporel et spatial des méthodes de traitement d'antenne linéaire [1], les méthodes multimicrophones ajoutent une dimension supplémentaire qui permet de simplifier dans un premier temps les traitements. En effet, l'objectif est de combiner de façon linéaire les signaux disponibles sur chacun des micros de manière à ce que l'analyse soit focalisée sur un signal provenant d'une direction privilégiée. Cette opération de pointage électronique est connue sous le nom de formation de voies ou *beamforming*. L'objectif escompté est de parvenir à séparer des signaux qui se recouvrent fréquentiellement mais provenant, on l'espère, de lieux différents. Ce filtrage spatial réalisé par le formateur de voies ou *beamformer*,

1. Il existe des méthodes de traitement d'antenne non linéaire basées notamment sur les filtres de Volterra

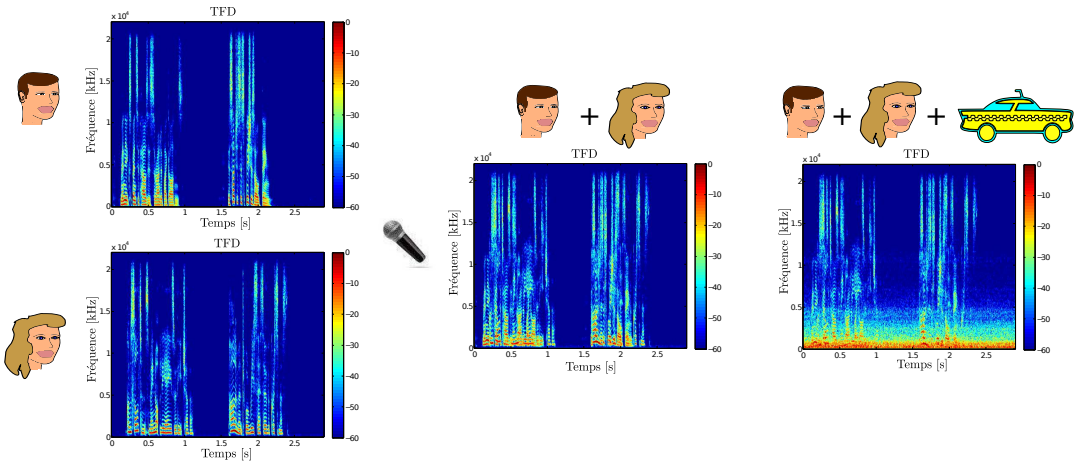


Figure 2 – Enregistrement monocapteur de signaux réels qui se recouvrent en temps et en fréquence.

nom attribué au processeur qui réalise l'échantillonnage spatial discret, réussit bien souvent là où un filtrage temporel classique échoue. En associant la réponse spatiale de chacun des microphones utilisés pour construire l'antenne au traitement logiciel, nous sommes en mesure de former la réponse spatiale souhaitée, Figure 3, encore appelée directivité de l'antenne. Nous profitons donc de la richesse et de la souplesse du traitement du signal en regard de la directivité physique imposée par le microphone. En ce sens, en profitant des traitements les plus performants développés pour le cas d'une prise de son avec un seul microphone et de leurs dérivés, les techniques multimicrophones seront nécessairement plus efficaces.

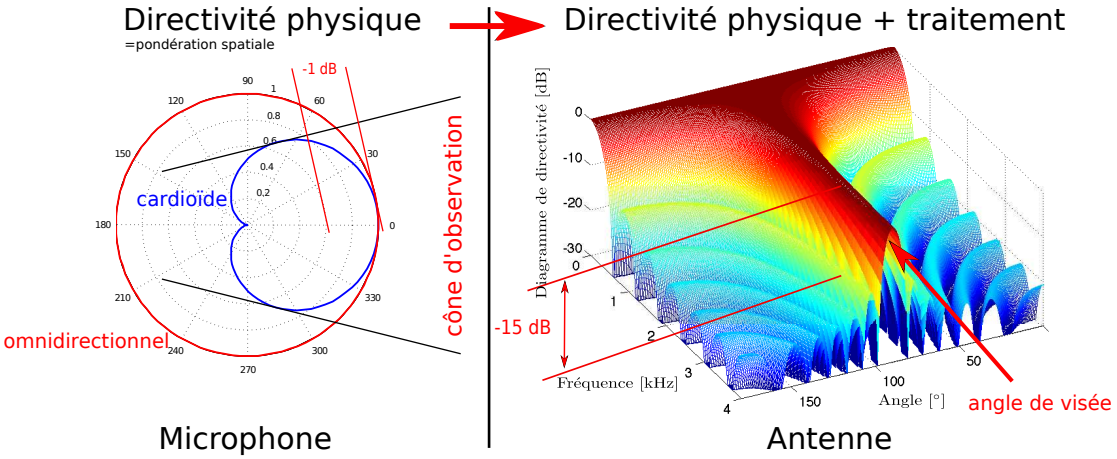


Figure 3 – Principe du traitement d'antenne.

Les méthodes de formation de voies associées aux antennes acoustiques peuvent

être distinguées selon deux classes [46] selon qu'elles s'adaptent automatiquement à la scène sonore ou pas.

Formations fixe et adaptative de voies

Les techniques de formation de voies fixes filtrent le signal microphonique à l'aide de filtres invariants dans le temps et sont donc indépendantes des données. La réponse spatiale est élaborée sous l'hypothèse que les statistiques temporelles et spatiales du signal et du bruit soient statiques. Ces méthodes incluent les antennes de type retard-somme, différentielle [54], superdirective [19], à directivité constante [186, 187] et sphérique à décomposition modale [55]. Cependant, avec ce type de formations de voies, il n'est en général pas possible de concevoir des directivités arbitraires pour une configuration donnée des microphones. Ceci est rendu possible avec une formulation plus générale de l'antenne retard-somme où l'ajout de filtres sur chaque voie permet d'approcher le gabarit voulu, d'où son appellation d'antenne filtrage-sommation. Pour définir les coefficients des filtres, divers critères d'optimalité peuvent être utilisés, comme par exemple la minimisation de la norme L_2.

Les techniques de formation adaptative de voies [189, 191] combinent la focalisation spatiale des techniques fixes à l'annulation adaptative de bruit. L'adaptation est faite sous contraintes en fonction de l'évolution de l'environnement et des statistiques des données. Parmi ces techniques, la solution de Frost [61] (*Minimum Variance Distortionless Response* (MVDR)) qui appartient à la classe des formateurs de voies linéairement contraint (*Linearly Constraint Minimum Variance* (LCMV)), calcule les coefficients optimaux des filtres de voies suivant le principe de minimisation de la puissance du signal en sortie d'antenne sous la contrainte que le signal provenant de la direction pointée soit préservé. Une alternative à cette solution qui se définit comme la résolution d'un problème d'optimisation sous contrainte est l'antenne appelée GSC qui reporte la complexité algorithmique sur l'architecture. Le problème de l'identification des coefficients est transformé en un problème d'optimisation non contraint. Ce formateur de voies, initialement proposé par Griffiths et Jim [73], est composé d'un formateur de voie fixe qui crée une référence du signal utile ; d'une matrice de blocage qui crée une référence de bruit et d'un annuleur de bruit multicanal qui supprime les composantes de bruit contenues dans le signal utile de référence et corrélées avec le signal de bruit de référence. Du fait de l'existence de réverbération, des incertitudes sur la position des microphones et des erreurs sur la position réelle de la source de parole utile, celle-ci peut fuir dans l'étage d'estimation du bruit et conduire à l'annulation du signal utile et l'introduction de distorsion, inévitables pour le GSC classique. Des méthodes robustes pour pallier ce problème

seront présentées au Chapitre 1.

Les techniques adaptatives offrent en général de meilleures performances en termes de réduction de bruit puisqu'elles sont capables de s'adapter aux évolutions de l'environnement. Néanmoins, elles sont plus sensibles aux erreurs de modélisation et donc moins robustes que les solutions fixes, souvent utilisées dans les milieux très réverbérants ou lorsque les conditions de prise de son et les positions des sources sont connues et invariantes.

Filtrage de Wiener multicanal

Le filtrage de Wiener multicanal constitue la seconde classe des techniques de réduction de bruit multimicrophones. Basé sur la théorie du filtrage optimal non contraint, ces techniques calculent les différents filtres selon le critère de l'Erreur Quadratique Moyenne Minimale (EQMM). Cependant, la frontière entre la classe précédente et cette classe de méthode est fine puisque Simmer et al. démontrent dans [166] que la solution optimale au sens de l'EQMM peut être décomposée en la solution optimale MVDR suivie d'un postfiltre de Wiener monocanal, ceci dans le cas où les sources de parole et de bruit sont décorrélées. Doclo et Moonen [46, 49] proposent une implémentation efficace du filtre de Wiener basée sur la décomposition généralisée en valeurs singulières (Generalized Singular Value Decomposition (GSVD)) de la matrice des signaux reçus sur les micros. Ceci conduit à une estimation optimale au sens de l'EQMM des composantes du signal désiré sur l'un des microphones. Une structure efficace est également proposée pour la mise à jour récursive de la GSVD. De plus, en ajoutant une opération de postfiltrage optionnelle, une estimée du canal bruité peut être obtenue pour améliorer la qualité du signal de parole. Spriet et al. proposent une implémentation en sous-bandes de la structure fondée sur la GSVD [169] tandis que Rambouts et Moonen proposent d'utiliser la décomposition QR [152]. Dans [171], Spriet et al. étudient la robustesse du GSC et du Filtrage de Wiener Multicanal (FWM) pour une utilisation à l'aide aux malentendants. Il en ressort que, si le GSC est capable de maintenir de bonnes performances dans un contexte de bruit non-stationnaire par rapport au FWM, ce dernier est quant à lui indifférent aux erreurs de pointage, de positionnement et de gains des microphones ainsi qu'aux erreurs dues au modèle de signal utilisé puisqu'il ne nécessite pas de connaissance *a priori* sachant qu'il ne s'appuie que sur les estimées des statistiques d'ordre deux. Néanmoins, dans le cas d'un bruit non-stationnaire, le FWM préservera son avantage sur le GSC lorsque la DSP du bruit varie lentement dans le temps. Cependant, le FWM ne peut garantir un signal non distordu en sortie et ses performances sont très fortement liées à l'efficacité de la Détection d'Activité Vocale (DAV). Pour pallier au

défaut de distorsion, une modification du critère du minimum de l'EQM ainsi que l'inclusion des Fonctions de Transferts Acoustiques (FTA) reliant la source de parole aux microphones est proposée dans [50, 168]. L'estimation des FTA pour remplacer le modèle trop réducteur d'une propagation selon de simple retard sera discutée au Chapitre 1.

Objectifs et contributions de la thèse

Les conditions de bruit des environnements considérés peuvent être très défavorables, par conséquent, nous considérons des méthodes multimicrophones qui, en exploitant à la fois le filtrage temporel et spatial, permettent une meilleure réduction de bruit comparées aux méthodes monomicrophones, et sont également capables de supprimer efficacement des interférents grâce à la localisation et la focalisation spatiale. Enfin, l'opération de pointage inhérente à ce type de méthodes permet de supprimer une partie de la réverbération.

Ces travaux de thèse visent à étudier et proposer un nouvel algorithme et une nouvelle architecture pour l'antenne GSC afin de corriger le problème de la fuite du signal pour en améliorer ses performances. Le choix de cette antenne s'appuie sur le fait qu'il s'agit d'une méthode adaptative, ce qui constitue un premier critère discriminant dans le sens où nous cherchons à définir un système autonome fonctionnant dans différents types d'environnements. De plus, nous visons une solution qui soit compatible avec une implémentation temps réel tout en garantissant la qualité du signal de parole offerte par cette antenne.

Pour parvenir à la solution proposée, nous étudierons dans un premier temps les méthodes bicapteurs, en particulier les méthodes récursives qui sont à la base de la nouvelle architecture que nous proposons. En ce sens, nous allons tout d'abord présenter une analyse du problème à résoudre en établissant un parallèle entre les méthodes multicapteurs et bicapteurs. Ensuite, nous donnerons une analyse des performances de la structure récursive choisie qui permettra de statuer sur ses cas d'utilisations, sa stabilité et ses propriétés de convergence. Nous montrerons également que les algorithmes les plus souvent utilisés avec cette structure sont sous-optimaux, ce que nous tenterons d'améliorer au travers de deux solutions, l'une architecturale, l'autre algorithmique. Enfin, à partir des résultats et des observations du cas bicapteur, nous étendrons la solution proposée au cas du GSC.

Organisation du mémoire

Ce rapport est organisé en six chapitres. Le Chapitre 1 présente tout d'abord un tour d'horizon des principales méthodes pour l'amélioration de la robustesse des antennes adapatives utilisées dans le traitement de la parole. Après avoir identifié les principaux défauts et sources de dégradation des méthodes existantes, nous présenterons les méthodes qui ont trait à l'amélioration des performances du GSC, structure sur laquelle nous concentrons nos travaux.

Dans le Chapitre 2, nous montrerons que les méthodes développées dans le cadre du contrôle de la diaphonie fournissent des réponses intéressantes au problème de la fuite du signal rencontré dans le GSC. Nous concentrerons plus spécifiquement notre étude sur la structure récursive, véritable pierre angulaire de ce travail de thèse. Ainsi, après une analyse des spécifités des réponses impulsionnelles acoustiques ou FTA et des différents contextes de prise de son, nous étudions un premier aspect de la stabilité de la structure récursive. Nous analysons également son champ d'application en donnant les conditions requises à son bon fonctionnement. En dernier lieu, nous montrons que l'algorithme de décorrélation couramment utilisé pour mettre à jour les coefficients est inadéquat et proposons en conséquence deux nouvelles solutions.

Les Chapitres 3 et 4 sont consacrés à l'étude des deux solutions proposées. Dans le Chapitre 3, nous décrivons celle basée sur une modification de l'architecture permettant l'utilisation des algorithmes d'annulation de bruit classiques. Le Chapitre 4 décrit la seconde solution proposée et fondée sur les travaux de Zinser et Mirchandani [202]. A partir de l'algorithme optimal proposé par ces derniers pour la structure récursive, nous déclinons un nouvel algorithme adapté au débruitage de la parole.

Le Chapitre 5 présente nos analyses et résultats concernant l'étude de la convergence et de l'existence d'un point d'équilibre stable pour la structure récursive. Ce travail fait suite à l'étude de la stabilité. La force et la pertinence de notre nouvelle approche pour traiter ce problème délicat du filtrage adaptatif tient dans le fait que notre analyse s'appuie uniquement sur les relations liant les signaux estimés et leurs artefacts. Nous obtenons donc des résultats indépendants du type d'algorithme mis en œuvre.

Finalement, le Chapitre 6 est consacré à l'extension au cas multimicrophone et plus particulièrement du GSC, de la méthode bicapteur analysée dans les trois chapitres précédents. Ce dernier chapitre expose tout d'abord les deux seuls et uniques

travaux à notre connaissance qui envisagent une implémentation récursive du GSC. C'est aussi l'occasion de décliner une nouvelle architecture récursive moins coûteuse que les deux précédentes. Ensuite, nous discutons de l'influence des chemins acoustiques sur la structure récursive en lien avec les problèmes de synchronisation dans les antennes. Enfin, nous donnons les résultats d'évaluation en environnement réel.

Chapitre 1

Débruitage de la parole par antenne acoustique

1.1 De l'antenne au postfiltre 16

1.1.1 Spécificités de l'antenne acoustique 19

1.1.2 Influence des perturbations et des incertitudes 27

1.2 Techniques pour l'amélioration de la robustesse 31

1.2.1 Formation de voies adaptative robuste 34

1.3 Performances du GSC 48

1.4 Synthèse du chapitre 49

Le domaine du traitement d'antenne, y compris acoustique, est un sujet très vaste (cf. frise chronologique en fin de chapitre). Il est donc difficile, voir même prétentieux, de prétendre le maîtriser dans son ensemble et surtout réussir à le présenter complètement, en particulier dans un mémoire de thèse. Par conséquent, notre objectif ici n'est pas de redéfinir une fois de plus les grands principes qui lui sont associés, ni même de faire l'inventaire des techniques les plus connues pour le traitement d'antenne acoustique avec leurs avantages et leurs inconvénients. De nombreuses références très pertinentes existent déjà sur ce sujet. Pour une description complète de l'ensemble des méthodes de formation de voies depuis leur création, nous renvoyons le lecteur aux références [96, 104, 108, 178, 182]. Pour une lecture plus spécifique dédiée aux antennes acoustiques, c'est-à-dire tenant compte du caractère large bande du signal de parole, nous pourrons nous reporter aux références [23, 46, 55, 68] et [126] où une synthèse comparative des différentes méthodes est donnée. Il ressort de ces lectures une absence de solution "universelle" ; aujourd'hui, il n'existe pas de méthode capable d'offrir une réduction de bruit satisfaisante avec la garantie d'un signal de bonne qualité quelles que soient les conditions d'utilisations. Ces conditions diffèrent par le type de bruit caractérisant l'environnement du système (voiture, foule, métro, salon, etc.).

Remarque 1 : *Notons qu'il est possible d'étendre les techniques développées pour les signaux à bande étroite aux antennes acoustiques en opérant par bande de fréquence, d'où l'intérêt des premières lectures référencées.*

1.1 De l'antenne au postfiltre

La conception d'une antenne acoustique est une chose délicate qui requiert en réalité la conception de trois éléments inter-dépendants, Figure 1.1.

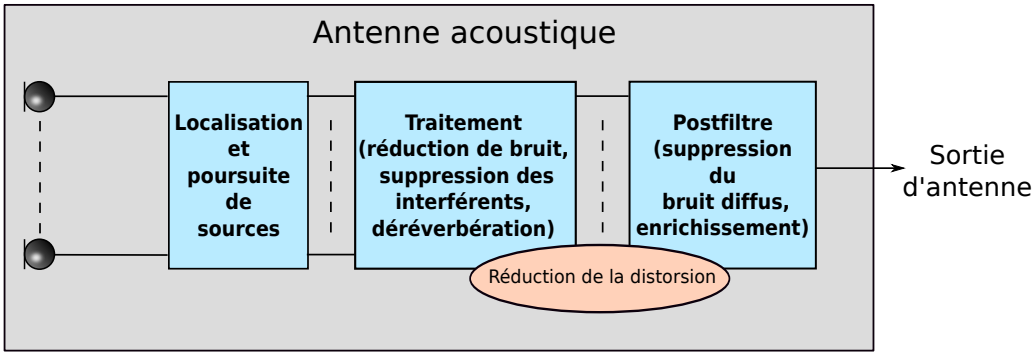


Figure 1.1 – Les trois blocs constitutifs d'une antenne acoustique.

Chacun de ces éléments couvre à lui seul un domaine de recherche. En effet, pour obtenir un système de prise de son performant et autonome les trois blocs suivants sont nécessaires :

1. Localisation et poursuite de sources : ce premier bloc conditionne les performances du système, en particulier pour les systèmes que nous appellerons autonomes où la position des sources est inconnue. Nous les opposerons aux systèmes fixes conçus pour fonctionner dans des environnements connus à l'avance et supposés invariants. Par exemple, le mur de télé-présence utilisé pour la visioconférence fait parti de ces systèmes. La conception est faite sur-mesure dans le sens où la position des locuteurs est fixée ainsi que les conditions acoustiques (sources de bruit présentes et leurs positions, temps de réverbération, etc.). Dans ce cas, ce bloc est inutile et est remplacé par une simple opération de pointage invariante dans le temps. Les méthodes de localisation de sources peuvent être classées en trois grandes familles suivant le procédé d'estimation : celles utilisant les temps d'arrivée (*Time Difference of Arrival* (TDOA)), celles basées sur la recherche du maximum de puissance reçue par une antenne effectuant un balayage de l'espace d'observation (*Steered Response Power* (SRP)), et les méthodes dites haute résolution qui s'appuient sur la matrice de corrélation des signaux reçus pour déterminer certaines caractéristiques du signal comme par exemple sa direction d'arrivée. Pour de plus amples détails, nous pourrons nous reporter aux références [36, 44, 87]. Les techniques de poursuite de sources sont quant à elles principalement fondées sur le filtrage de Kalman [97].

2. Traitement : ce bloc réalise le filtrage des signaux après pointage dans le but de réduire le bruit, supprimer les interférents et améliorer la qualité du signal vocal extrait du mélange. Il réalise le filtrage spatio-temporel (formation de voies) à partir de l'information fournie par le bloc de localisation et poursuite de sources. Les systèmes adaptatifs sont ceux pour lesquels ce bloc de traitement utilise les techniques de filtrage adaptatif qui ont l'avantage de s'adapter aux variations de l'environnement au cours du temps. Ils s'appuient donc sur l'évolution des statistiques des signaux captés, par opposition aux systèmes fixes où l'algorithme de filtrage est indépendant des données (cf. le chapitre Introduction).

3. Postfiltrage : ce bloc en fin de chaîne est dédié à la suppression des bruits non localisés que l'antenne est incapable de supprimer seule. Plus spécifiquement, il permet de supprimer le bruit diffus que l'on associe à la réverbération et les bruits incohérents. En plus de raffiner le débruitage, il peut également être conçu de sorte à améliorer la qualité du signal en limitant la distorsion

introduite par le bloc de traitement. Par exemple, nous pourrions imaginer l'utilisation de l'algorithme HRNR [143] pour la régénération des harmoniques perdues par le filtrage de Wiener. Une première utilisation d'un postfiltre en sortie d'antenne est décrite par Zelinski dans [199] où il utilise un filtrage de Wiener accompagné d'un post-traitement sur la fonction d'intercorrélation afin de diminuer les erreurs d'estimation. Une analyse des décroissances de la fonction d'autocorrélation permet quant à elle de distinguer le signal utile du bruit. Le lien entre les performances de l'antenne et ce type de postfiltre est analysé par Marro et al. dans [123, 124] où l'incapacité de ce postfiltre à améliorer les performances là où l'antenne est elle-même inefficace (basses fréquences) y est démontrée. Dans [94], Bitzer et al. proposent l'association du GSC et d'un postfiltre adaptatif. Le GSC utilisé est quelque peu spécifique puisqu'il est conçu comme une antenne superdirective avec des filtres fixes pour l'annulation de bruit. La méthode ainsi obtenue améliore la réduction de bruit, diminue l'annulation du signal utile et réduit le bruit musical. Dans cet article, les auteurs proposent également une méthode d'évaluation que nous utiliserons et décrirons au Chapitre 6 pour évaluer la réduction du bruit et la qualité vocale. Dans [129, 130], McCowan et Bourlard améliorent le postfiltre proposé par Zelinski en tenant compte de la corrélation existant entre les micros, le rendant ainsi plus en adéquation avec la réalité. L'expression plus générale obtenue permet d'adapter la méthode aux divers types de champ de bruit dont la fonction de cohérence peut être modélisée. Ce modèle plus précis améliore par ailleurs les performances en basses fréquences. Cependant, ce postfiltre sera lui-même amélioré par Lefkimmiatis et Maragos [111] qui vont tenir compte de la réduction de bruit apportée par le formateur de voies, contrairement aux travaux de McCowan qui surestime le dénominateur du postfiltre. De plus leur méthode gagne encore en généralité puisqu'elle traite la conception de postfiltre linéaire ou non. Avec les travaux de Cohen, Berdugo et Gannot [37, 39, 67], le postfiltre est agrémenté d'un modèle probabiliste pour la détection de l'absence de parole. A l'aide d'une fonction de vraisemblance globale de présence de parole, ils traitent la réduction des bruits transitoires et non-stationnaires. Dans le même sens mais en utilisant la fonction de cohérence pour construire la probabilité d'absence du signal, on trouve la méthode hybride de Li et Akagi [99, 100] qui combine les méthodes d'estimation multicanal et monocanal. Enfin, une généralisation des postfiltres de Zelinski, Marro, McCowan et Lefkimmiatis a été récemment proposée par Wolff et Burck dans [192] où ils établissent les liens entre ces différents postfiltres et exposent la manière d'en construire de nouveaux adaptés aux critères choisis. La déclinaison adaptative de ce postfiltre généralisé est également donnée avec les résultats qui prouvent sa supériorité

sur les autres méthodes. Ces travaux, qui s'inscrivent dans une démarche de conception de postfiltre adaptatif adéquat avec les antennes adaptatives linéairement contraintes, sont repris dans [193] pour mettre en avant l'influence de la matrice de blocage qui doit être de préférence adaptative afin d'améliorer la qualité du signal de parole fourni par l'antenne. Cette observation a son importance, car elle correspond à une justification supplémentaire de l'implémentation du GSC choisie dans ce mémoire.

Les contributions décrites dans ce mémoire concernent plus spécifiquement le bloc de traitement et visent les techniques adaptatives pour leur capacité d'adaptation aux variations de l'environnement. Bien que nos travaux émanent en partie d'un constat vis-à-vis des faiblesses des méthodes de localisation et de poursuite de source, nous tentons de résoudre le problème qui en découle au niveau du bloc de traitement. Dans une certaine mesure, les méthodes proposées peuvent néanmoins être vues comme une opération de postfiltrage, mais ce n'est pas sous cette forme que nous les présenterons dans la suite du document. Dans le paragraphe suivant, nous décrivons les notions nécessaires du traitement d'antenne qui vont nous permettre au paragraphe 1.1.2 de montrer l'influence des incertitudes sur les performances du filtrage spatial réalisé par l'antenne.

1.1.1 Spécificités de l'antenne acoustique

Une manière simple d'aborder les antennes acoustiques en faisant abstraction des fondements théoriques à leurs origines est de considérer leur conception comme celle d'un filtre à Réponse Impulsionnelle Finie (RIF). En établissant ce parallèle, Figure 1.2, nous pouvons définir les grandeurs significatives associées aux antennes acoustiques et mettre en évidence les aspects qui ont motivé les travaux décrits dans la suite de ce mémoire.
L'analogie entre le filtrage temporel à l'aide d'un RIF et le filtrage spatial à l'aide d'une antenne sous-entend une certaine dualité entre les grandeurs caractéristiques. En effet, nous allons montrer qu'il y a équivalence entre l'ordre d'un filtre RIF et le nombre de microphones utilisés. De même que nous allons définir une fréquence d'échantillonnage spatiale, et distinguer différents types d'antennes.

Type d'antenne, géométrie et sélectivité

Au même titre qu'un filtre temporel est complètement défini par son gain et sa phase, une antenne sera complètement définie par sa fonction de transfert, notée $H(f, \theta, \phi)$, qui est la réponse harmonique de l'antenne à l'onde provenant de la direction (θ, ϕ) dans le système de coordonnées sphériques et de fréquence f. Typiquement, en prenant le cas très simple de l'antenne linéaire de la Figure 1.2 dont

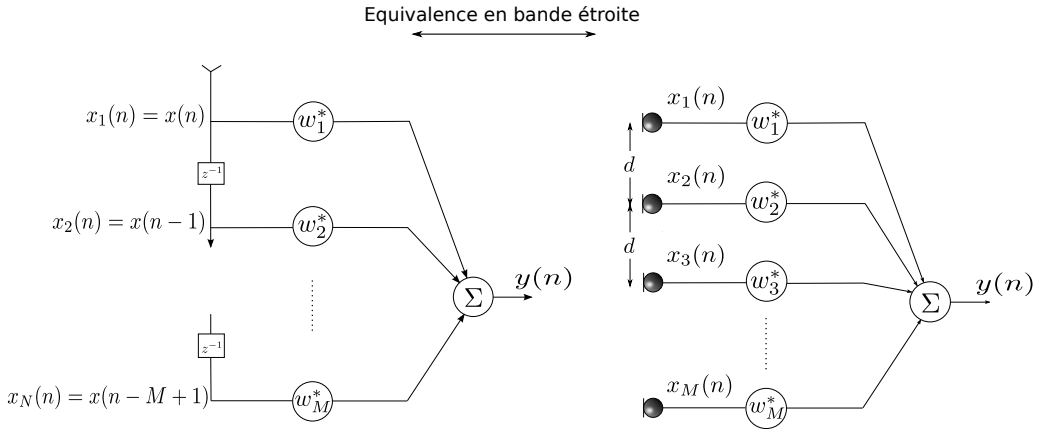


Figure 1.2 – Analogie entre un filtrage RIF mono canal d'ordre M et une ALU composée de M microphones omnidirectionnels.

les microphones sont équi-espacés de la distance d (communément nommée Antenne Linéaire Uniforme (ALU)) et en supposant que les coefficients de pondération w_i^* soient indépendants de la fréquence, nous pouvons tracer cette fonction de transfert dans le plan azimutal pour différentes valeurs de f. La Figure 1.3 illustre une telle réponse en tenant compte également de la caractéristique omnidirectionnelle supposée des microphones.

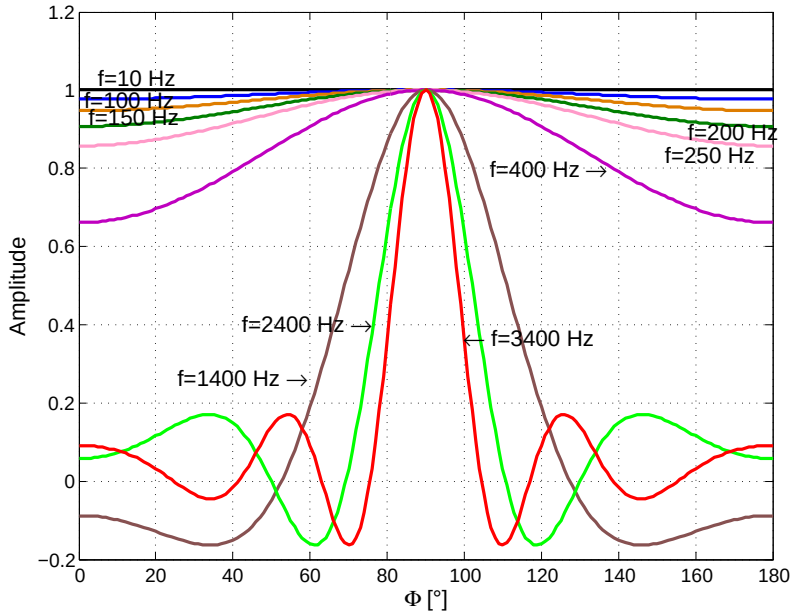


Figure 1.3 – Fonction de transfert de l'antenne ALU pour différentes valeurs de fréquence f.

Le maximum observé en $\phi = 90°$ signifie que l'antenne pointe dans cette direction de l'espace. L'antenne considérée étant composée de microphones placés sur une droite, il n'y a pas de dépendance par rapport à l'angle d'élévation θ. En prenant pour axe la ligne imaginaire passant par les micros, la direction de pointage obtenue ici se situe selon la perpendiculaire à cet axe et correspond à une visée dite en *broadside*. Si ce maximum est atteint en $\phi = 0°$, la visée est dite en *endfire* et peut être obtenue en fixant pour chaque micro i, $w_i^* = \frac{1}{M} e^{j\frac{-2\pi f}{c}(i-1)dcos\phi'} = \frac{1}{M} e^{j\frac{-2\pi f(i-1)d}{c}}$ avec $i = 1 \dots M$ et $\phi' = 0$. Ceci revient à appliquer le délai $\delta_i = \frac{(i-1)d}{c}$ sur chaque voie de sorte à pointer l'antenne dans la direction $\phi = \phi' = 0°$ (le micro indicé 1 à la Figure 1.4 sert de référence à l'application des retards). Par extension, le cas représenté à la Figure 1.3 correspond au cas où $w_i^* = \frac{1}{M}$.

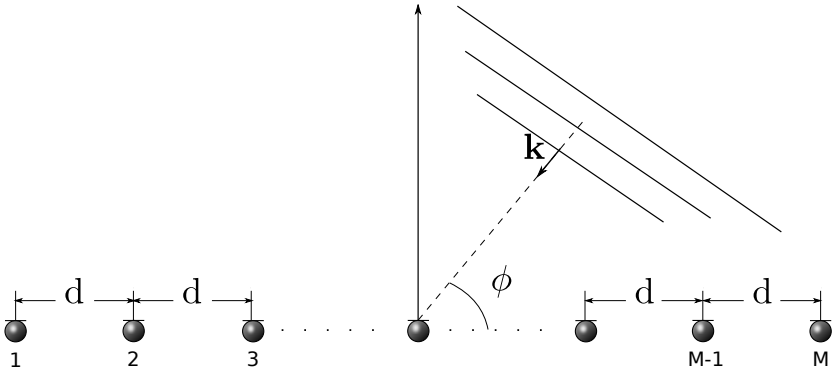


Figure 1.4 – Vue schématique de l'antenne ALU.

Une description plus globale des caractéristiques de directivité et de sélectivité de l'antenne s'obtient grâce au diagramme de directivité défini par $|H(f, \theta, \phi)|^2$. Celui-ci permet de visualiser la réponse de l'antenne en trois dimensions et est particulièrement utile pour l'analyse du comportement spatial de l'antenne sur une bande de fréquence donnée. En l'occurrence, nous traçons la directivité de l'antenne considérée (ALU) à la Figure 1.5 où nous faisons varier le nombre de microphone utilisés. A partir des résultats obtenus, nous constatons que l'antenne sera d'autant plus sélective spatialement que le nombre de microphones augmente. Ainsi, le nombre de microphones est assimilable au degré du filtre RIF, puisque plus le degré d'un filtre RIF est élevé et plus celui-ci est sélectif.
Pour visualiser cette propriété en trois dimensions et apprécier le potentiel d'une antenne acoustique, nous donnons à la Figure 1.6 un exemple de directivité créée à l'aide d'une sphère microphonique. L'avantage d'une telle antenne réside dans la

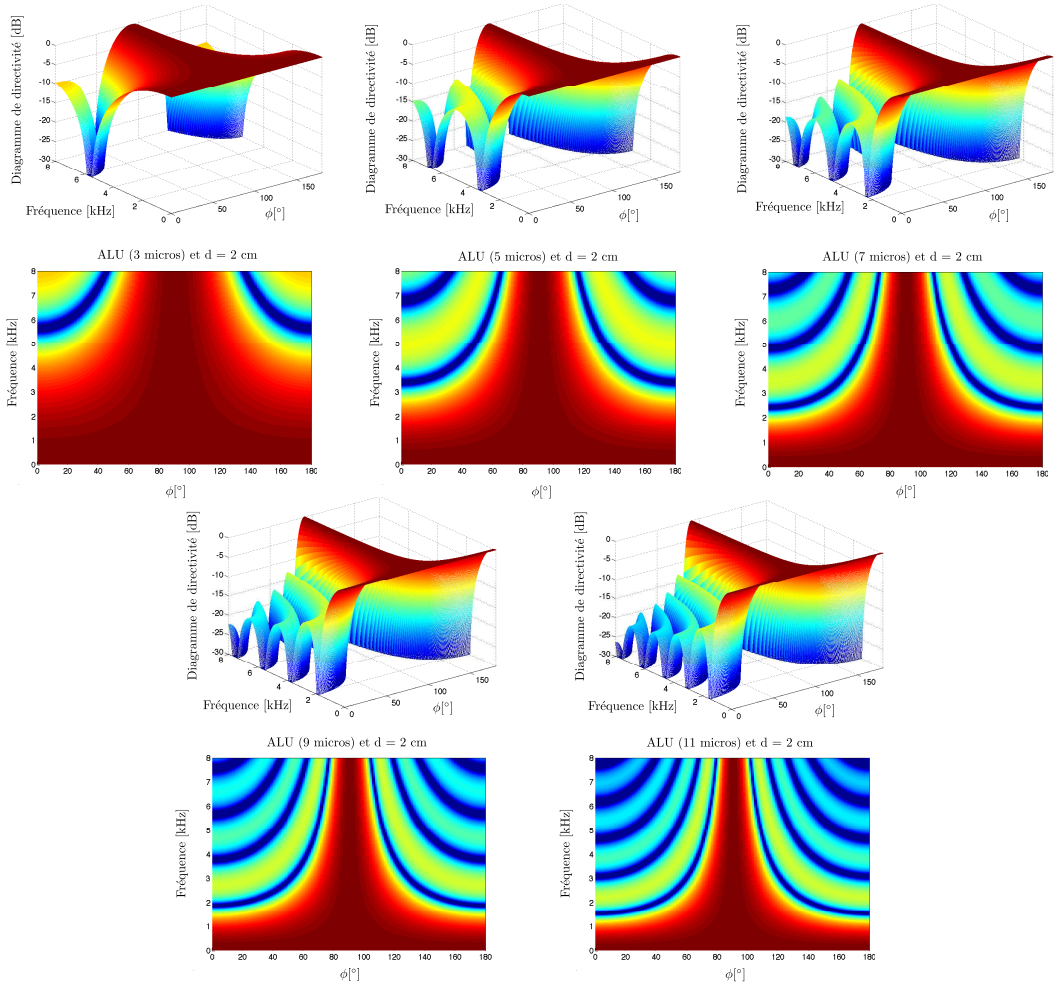


Figure 1.5 – Diagramme de directivité de l'antenne ALU sur la bande de fréquence [0 8000] Hz en fonction du nombre de microphones.

finesse des lobes qu'il est possible de créer. L'espace est discrétisé selon trois dimensions et la sélection spatiale pourra être faite sur des zones d'autant plus restreintes que le nombre de microphones est important. Pour une lecture plus approfondie sur les performances des méthodes de formation de voies utilisant des antennes sphériques nous pourrons nous référer à [12, 52, 146–148]. La géométrie et le nombre de microphones sont donc deux caractéristiques qui définissent une antenne acoustique et influent sur les performances du système final. En l'occurrence, l'analyse de l'impact de la géométrie sur la réverbération est donnée dans [139, 163]. En effet, en considérant un modèle d'observation additif classique et en supposant que le bruit est de type diffus, la matrice de covariance des signaux d'observation est la somme de la matrice de covariance du signal dans la direction de pointage avec celle du bruit provenant de toutes les directions. Pour supprimer le bruit diffus et

estimer la DSP du signal utile, les auteurs proposent de diagonaliser la matrice de covariance du bruit afin de supprimer les termes hors la diagonale. Le point fort de la méthode provient du caractère "aveugle" de la décorrélation réalisée, c'est-à-dire ne nécessitant pas le calcul, la connaissance de la matrice de covariance du bruit nécessaire aux méthodes de projection sur des sous-espaces propres au signal utile et au bruit. Pour y parvenir, les auteurs utilisent des antennes à géométries régulières identiques aux structures cristallines (cf. Figure 1.7) et la Transformée de Fourier Discrète (TFD). Plus précisément, c'est la matrice de la TFD dont l'ordre est fixé en fonction du nombre de microphones utilisés qui joue le rôle de matrice unitaire pour la diagonalisation de la matrice de covariance du bruit. Cette méthode est appliquée à la définition d'un postfiltre dans [90].

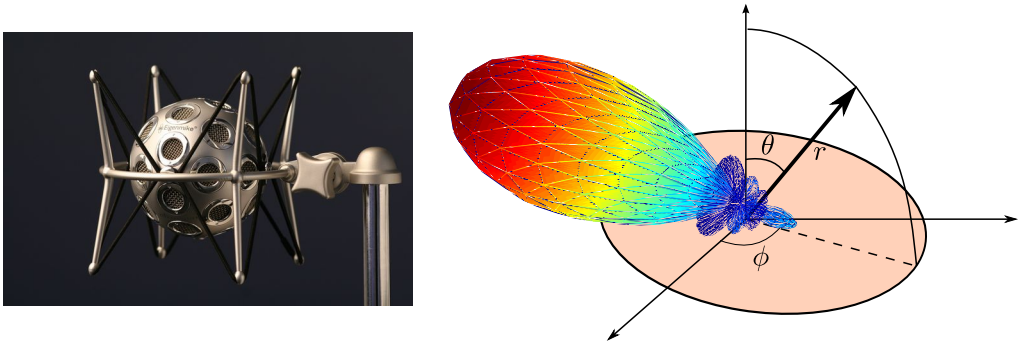


Figure 1.6 – Exemple d'antenne sphérique (*Eigenmike*) avec son diagramme de directivité $|H(f,\theta,\phi)|^2$ pointant dans la direction (θ_0,ϕ_0).

Aliasing spatial

Nous venons de voir que le nombre et la disposition des microphones créent une géométrie particulière qui influence le diagramme de directivité de l'antenne. Cependant, ceux-ci ne sont pas, en général, placés au hasard dans l'espace. En effet, même si il existe des études sur une distribution aléatoire des microphones pour l'analyse et la synthèse de champ sonore [74], le plus souvent, le placement permettant de réaliser une géométrie d'antenne donnée est fait sous la contrainte du respect du théorème d'échantillonnage spatial. En effet, le réseau de microphones réalise un échantillonnage spatial de fronts d'ondes qui se propagent selon des directions différentes et selon leurs propres longueurs d'ondes. Ainsi, de la même manière que le théorème d'échantillonnage dans le domaine temporel, ou théorème de Nyquist-Shannon, assure une acquisition des données discrètes sans perte d'information due au repliement spectral, la distance inter-microphone est liée à une fréquence spatiale dépendante de la longueur d'onde et correspond donc à une période spatiale

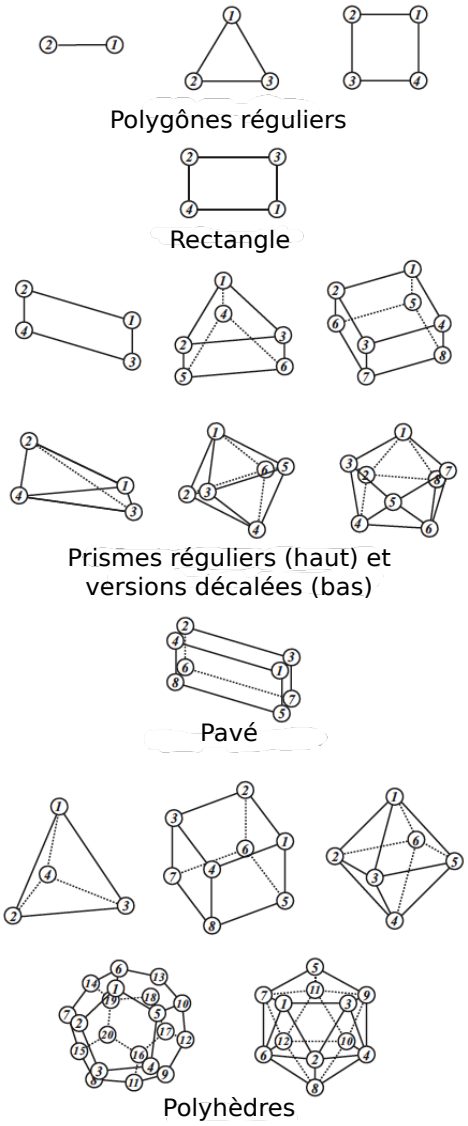


Figure 1.7 – Cinq classes d'antennes cristallines [139].

contrainte par une période d'échantillonnage spatiale. Il existe différentes manières pour le démontrer [126], [123] et [55], ici, nous proposons un autre raisonnement qui rejoint celui décrit dans [109].
A partir de l'analogie donnée à la Figure 1.2, nous pouvons déduire pour l'antenne ALU la définition de la fréquence d'échantillonnage spatiale ν_s telle que :

$$\nu_s = \frac{1}{d} \tag{1.1}$$

où la période d'échantillonnage spatiale est déterminée par l'espacement inter-microphone d exprimé en mètre. De façon analogue à l'échantillonnage uniforme dans le domaine temporel pour lequel la différence de phase entre deux séquences à temps discrets $x(n)$

et $x(n-1)$ correspond à $e^{j2\pi F}$ où F désigne la fréquence normalisée, nous pouvons définir une fréquence spatiale normalisée par la quantité suivante :

$$U = \frac{\nu}{\nu_s} = \frac{dcos(\phi)}{\lambda} \tag{1.2}$$

Cette définition s'explique facilement à partir du calcul de la réponse de l'antenne ALU, (1.3), à une onde plane (hypothèse de champ lointain) de fréquence f, de longueur d'onde $\lambda = c/f$ et provenant de la direction ϕ (cf. Figure 1.4). La fonction de transfert ainsi obtenue correspond à l'écriture d'une TFD à la fréquence U de la suite des gains $w_i(f)$.

$$H(f,\phi) = \sum_{i=1}^{M} w_i(f) e^{j\frac{2\pi f}{c}(i-1)dcos(\phi)} = \sum_{i=1}^{M} w_i(f) e^{j\frac{2\pi}{\lambda}(i-1)dcos(\phi)} \tag{1.3}$$

Etablissons l'analogie avec le domaine temporel :

$$F_e \geq 2F \Rightarrow U_e \geq 2\frac{dcos\phi}{\lambda} \tag{1.4}$$

Sachant que les définitions angulaires sont non ambiguës pour $\phi \in [0\ \pi]$. D'où

$$0 \leq \phi \leq \pi$$

$$\iff \quad -1 \leq cos\phi \leq 1$$

$$\iff \quad -\tfrac{2d}{\lambda} \leq \tfrac{2d}{\lambda} cos\phi \leq \tfrac{2d}{\lambda}$$

$$\frac{-2d}{\lambda} \leq 2U \leq \frac{2d}{\lambda} \tag{1.5}$$

Par conséquent, l'analyse donnée dans [109] affirme à juste titre que si $d/\lambda > 1/2$ ou encore $d > \lambda/2$ alors il existe un repliement des fréquences angulaires. En effet, U est compris entre $-d/\lambda$ et d/λ et peut devenir supérieur à $1/2$, donc se replier puisque la réponse de l'antenne ALU $H(f,\phi)$ est périodique en U de période 1. Par contre si $d/\lambda < 1/2$ soit $d < \lambda/2$ la transformée de Fourier discrète des coefficients $w_i(f)$ n'est pas entièrement parcourue lorsque ϕ croît de $[0\ 2\pi]$ puisque dans ce cas U croît de 0 à $d/\lambda < 1/2$. La conséquence est que l'antenne est utilisée de manière sous-optimale, la largeur du lobe principal pourrait être réduite. Logiquement, nous en déduisons que la valeur optimale qui garantit le lobe principal le plus fin en fonction de la fréquence et sans repliement spatial est $d = \lambda/2$. Néanmoins, Il existe une antenne particulière de type *endfire* et appelée antenne superdirective qui utilise un espacement entre les microphones inférieur à $\lambda/2$. Ses spécificités seront données au paragraphe 1.1.2.

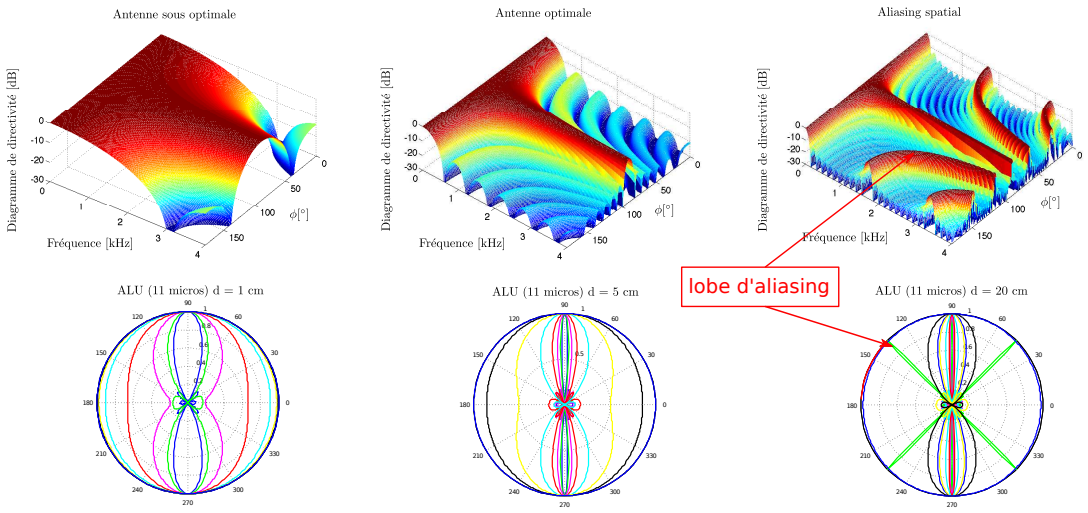


Figure 1.8 – Diagramme de directivité de l'antenne ALU composée de 11 microphones pour d valant (de gauche à droite) 1, 5 et 20 ?cm. La distance $d = 5\ cm$ (milieu) correspond à l'antenne optimale pour l'onde de fréquence $f = 3400\ Hz$. A gauche, nous illustrons le cas d'une antenne sous-optimale, à droite le phénomène de repliement spatial. Le code de couleur utilisé pour les directivités dans le plan n'a pas de lien avec les directivités 3D, il sert seulement à la lisibilité.

Au regard de la Figure 1.8, le repliement spatial se traduit par l'apparition de lobes parasites supplémentaires qui impliquent une confusion sur les signaux captés dans une direction donnée. En effet, la directivité est maximale en plusieurs directions différentes.

Antenne large bande

Nous venons de mettre en évidence l'importance du rapport d/λ qui intervient pour l'apparition de l'aliasing spatial et pour la sélectivité. Les lobes d'aliasing surviennent en effet lorsque la fréquence augmente, alors que la sélectivité spatiale, en l'occurrence l'angle d'ouverture du lobe principal, est d'autant meilleure que la fréquence est grande. En regard au caractère large bande du signal de parole que nous souhaitons extraire grâce aux antennes acoustiques, il nous faut donc conserver des caractéristiques homogènes avec la fréquence et empêcher l'apparition du phénomène d'aliasing. Le rapport d/λ doit donc rester constant. Pour ce faire, différentes approches existent et l'une d'entre elle consiste à utiliser sur une même antenne un espacement différent entre les microphones et ajouter un filtre sur chaque voie [123]. L'antenne se trouve ainsi constituée de plusieurs sous-antennes et permet entre autre de créer une directivité telle que celle donnée à la Figure 1.9. D'autres approches, basées sur l'antenne filtrage-sommation décrite à la section 1.2 et à laquelle nous avons déjà fait allusion en Introduction, sont données dans [102, 185–187] et [48].

Dans cette dernière, la conception tient compte du type de propagation lié à l'utilisation de l'antenne : champ proche ou champ lointain. Par ailleurs, en plus des fonctions de coût habituelles (moindres carrés pondérés, maximisation du rapport d'énergie entre la bande passante et la bande atténuée, minimisation de l'erreur entre l'amplitude du diagramme de directivité réel et celui désiré), les auteurs proposent deux nouvelles méthodes moins coûteuses que celles utilisées classiquement, basées sur les vecteurs propres [32, 103], et où les filtres optimaux correspondent aux valeurs propres minimales de la matrice du problème mis sous forme quadratique.

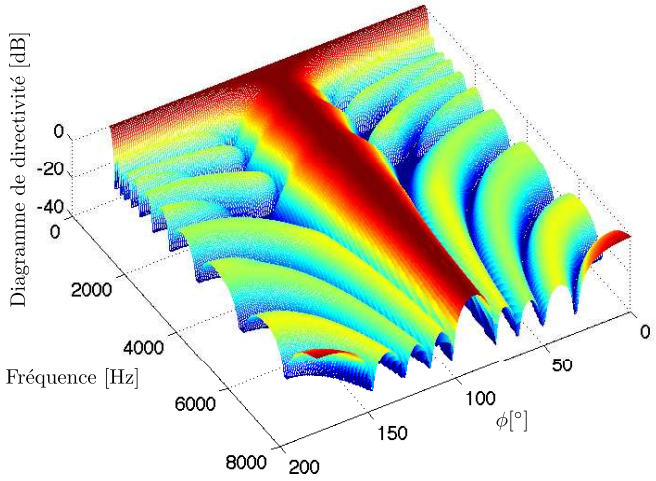


Figure 1.9 – Antenne à directivité contrôlée. La réponse $|H(f,\phi)|^2$ tracée correspond à celle d'une antenne rectiligne logarithmique composée de 11 microphones cardioïdes groupés en 4 sous-antennes. Le pointage est de type *broadside*.

1.1.2 Influence des perturbations et des incertitudes

En tenant compte de la première contrainte qui nous est imposée, à savoir la discrétion du système proposé, alors le choix d'une antenne superdirective est approprié car ces antennes ont la particularité d'être très directives et de petites tailles [19]. Cependant, elles sont très sensibles aux bruits des microphones, aux erreurs de pointage et de positionnement des microphones ainsi qu'aux variations des caractéristiques de ces derniers. En effet, les gains des microphones peuvent être très différents d'un microphone à l'autre et dépasser le gain unité, créant ainsi des phénomènes d'amplification pour les erreurs. Les erreurs de positionnement peuvent être considérées indépendantes d'un capteur à l'autre mais elles affectent le formateur de voies de la même manière. L'évaluation de la robustesse d'une antenne vis-à-vis des erreurs précitées se fait à l'aide du gain en bruit blanc noté WNG (pour *White Noise*

Gain (WNG)) qui mesure la réduction des bruits incohérents ou spatialement blancs, ce qui corrobore l'hypothèse que les erreurs entre deux microphones sont supposées non corrélées. En prenant pour modèle des signaux d'observations et de sortie d'antenne les relations (1.6) et (1.7), la définition du WNG est donnée en (1.11) à partir des définitions du gain de l'antenne et de la directivité.

$$\mathbf{x}(n) = s(n)\mathbf{d}(\phi_s) + \mathbf{b}(n) \tag{1.6}$$

$$y(n) = \mathbf{w}^H \mathbf{x}(n) \tag{1.7}$$

où s désigne le signal utile, $\mathbf{b}$ le vecteur du signal de bruit capté sur chaque microphone, $\mathbf{x}$ le vecteur d'observations, y le signal en sortie d'antenne pondéré par le vecteur des coefficients $\mathbf{w}^H$ et $\mathbf{d}$ la réponse de l'antenne dont l'expression est la suivante :

$$\mathbf{d}^T(\phi_s) = \left[1 \dots e^{j\frac{-2\pi f}{c}(i-1)dcos\phi_s} \dots e^{j\frac{-2\pi f}{c}(M-1)dcos\phi_s}\right] \tag{1.8}$$

Les exposants ^{T} et ^{H} désignent respectivement l'opérateur de transposition et celui de transposition-conjugaison (Hermitien).

- **Gain de l'antenne**

Il quantifie l'amélioration apportée par le formateur de voies en termes de Rapport Signal à Bruit (RSB).

$$G = \frac{RSB_{antenne}}{RSB_{micro}} = \frac{|\mathbf{w}^H\mathbf{d}|^2}{\mathbf{w}^H\mathbf{Q}\mathbf{w}} \tag{1.9}$$

avec $\mathbf{Q}$ la matrice de corrélation spatiale du bruit. Notons que le terme au numérateur de $\mathbf{G}$ n'est autre que le diagramme de directivité de l'antenne défini par $|H(f,\phi)|^2 = |\mathbf{w}^H\mathbf{d}|^2$ pour lequel l'angle ϕ du vecteur de propagation balaie tout l'espace d'observation.

- **Directivité**

Elle exprime la capacité de l'antenne à supprimer le bruit diffus. La grandeur indicative utilisée est l'Indice de Directivité (*Directivity Index* (DI)) donné par :

$$DI = 10 \times log_{10}\left(\frac{|\mathbf{w}^H\mathbf{d}|^2}{\mathbf{w}^H\mathbf{\Gamma}\mathbf{w}}\right) \tag{1.10}$$

où $\mathbf{\Gamma}$ est la matrice de corrélation spatiale du bruit diffus dont le terme général est donné par la fonction de cohérence $\Gamma_{mn} = \frac{sin(2\pi f r_{mn}/c)}{2\pi f r_{mn}/c}$, avec r_{mn} la distance entre les microphones m et n.

• **Gain en bruit blanc** (WNG)

Il correspond au gain de l'antenne lorsque les bruits présents sur les microphones possèdent une matrice de corrélation spatiale diagonale. Ainsi, sachant que celle-ci vaut $\mathbf{Q} = \mathbf{I}$ alors :

$$WNG = \frac{|\mathbf{w}^H \mathbf{d}|^2}{\mathbf{w}^H \mathbf{w}} \tag{1.11}$$

Le gain en bruit blanc est maximal pour l'antenne retard-somme, ce qui en fait intrinsèquement l'antenne la plus robuste vis-à-vis des erreurs énoncées plus haut, contrairement à l'antenne superdirective où en général $WNG < 10^{-3}$. Celle-ci est donc très sensible aux erreurs et bien qu'elle offre un DI de M^2 contre M pour l'antenne retard-somme, nous ne l'utiliserons pas pour les travaux présentés dans ce mémoire. La raison est, qu'en plus de sa forte sensiblité aux erreurs, elle nécessite un pointage précis de la source utile du fait de sa forte sélectivité.
Afin de montrer l'influence des incertitudes sur le diagramme de directivité et donc le filtrage spatial réalisé par l'antenne, nous nous appuyons sur les travaux de Gilbert et Morgan [72] ainsi que ceux de Yan et Ma [195] afin d'inclure ces erreurs dans le formalisme descriptif du filtrage réalisé par l'antenne. Prenons pour expression nominale des coefficients $w_n = g_n e^{-j\Phi_n}$. La modélisation des différentes erreurs identifiées peut être faite comme suit : $\tilde{\mathbf{p}}_n = \mathbf{p}_n + \Delta\mathbf{p}_n$ avec $\mathbf{p}_n$ le vecteur nominal des coordonnées du $n^{\text{ième}}$ microphone et $\tilde{w_n} = g_n(1 + \Delta g_n)e^{-j(\Phi_n + \Delta\Phi_n)}$ où Δg_n, $\Delta\Phi_n$ et $\Delta\mathbf{p}_n$ sont des variables aléatoires indépendantes, gaussiennes et de moyennes nulles. En posant σ_g^2 et σ_Φ^2 les variances de l'erreur sur le gain et la phase ; $\sigma_\lambda = 2\pi\sigma_p/\lambda$ avec λ la longueur d'onde et σ_p^2 la variance de l'erreur sur la position, il est possible de démontrer que la moyenne d'ensemble du diagramme de directivité $|H(f,\theta,\phi)|^2$ dans le cas d'un bruit diffus a pour expression

$$E\left[|H(f,\phi)|^2\right] = |H(f,\phi)|^2 e^{-(\sigma_\Phi^2 + \sigma_\lambda^2)} + ||\mathbf{w}||^2(\sigma_g^2 + \sigma_\Phi^2 + \sigma_\lambda^2) \tag{1.12}$$

A partir de la relation (1.12), nous constatons que les variations aléatoires ont deux effets sur le diagramme de directivité. Tout d'abord, elles atténuent de façon exponentielle la réponse de l'antenne dans toute les directions de l'espace. La deuxième incidence, plus critique, est liée au second terme qui engendre une remontée des lobes secondaires, ce qui conduit à une diminution de la sélectivité de l'antenne dans les directions des interférents. Ceci est mis en image aux Figures 1.10 et 1.11 sur un cas particulier, mais le lecteur intéressé pourra trouver de plus amples détails dans [195] où une analyse du gain de l'antenne montre que la sensibilité de la méthode MVDR est plus forte en basse fréquence.
Notons que si la norme des coefficients diminue, alors le WNG augmente, la sensibilité du formateur de voies aux erreurs est donc réduite. Cela explique la robustesse

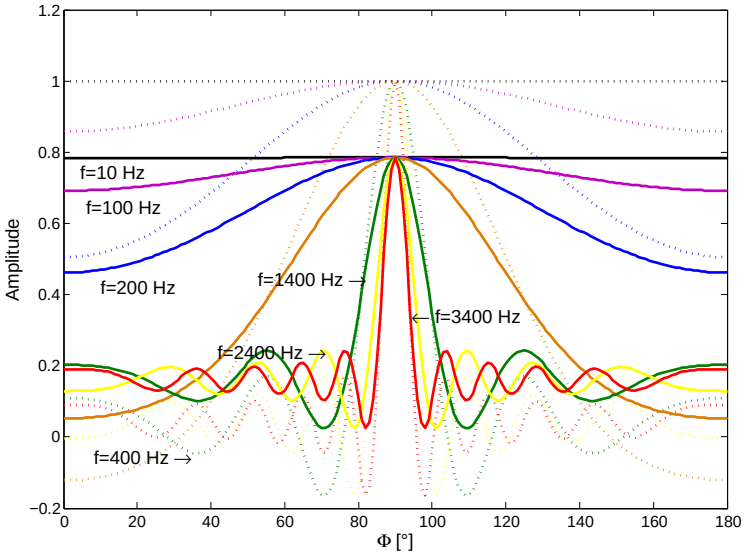


Figure 1.10 – Fonction de transfert de l'antenne ALU dans le cas d'incertitude sur la position des micros, la valeur des gains et la direction de la source utile. En pointillé la fonction de transfert dans le cas idéal, en trait continu le résultat pour $\sigma_g = 0,1$, $\sigma_\Phi = 10°$ et $\sigma_p = 0,01\ m$.

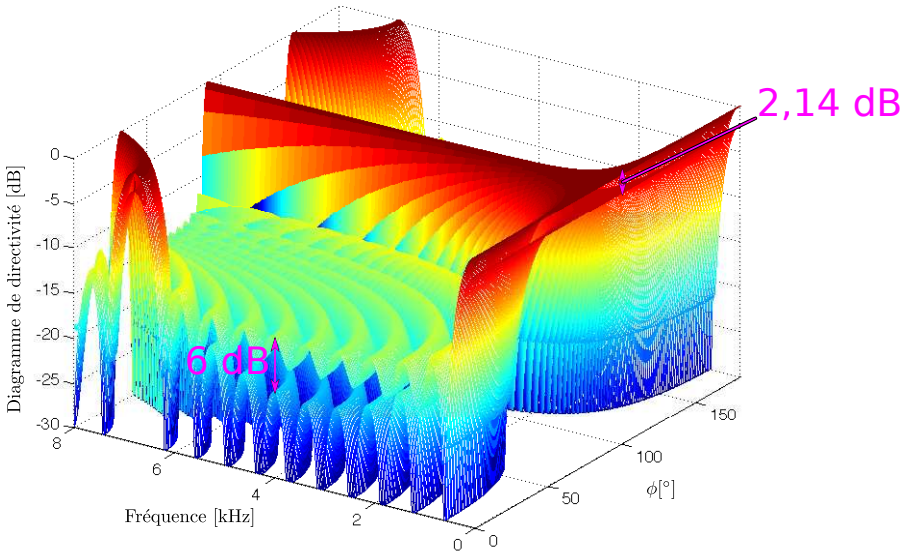


Figure 1.11 – Directivité de l'antenne ALU dans le cas d'incertitude sur la position des micros, la valeur des gains et la direction de la source utile ($\sigma_g = 0,1$, $\sigma_\Phi = 10°$ et $\sigma_p = 0,01\ m$).

de l'antenne retard-somme qui réalise une pondération uniforme sur chacune des voies $(1/M)$ et impose ainsi la condition d'une norme minimale pour les coefficients même si comme nous le montre les résultats des Figures 1.10 et 1.11, les erreurs ont une incidence sur son diagramme directivité. Les lobes secondaires dans la di-

rection $\phi = 0°$ à la fréquence 4 kHz remontent de 6 dB et le lobe principal perd 2,14 dB à la fréquence 10 Hz dans la direction $\phi = 90°$. Pour résoudre ces problèmes de modélisation qui conduisent à l'amplification du bruit blanc en l'absence supposée d'erreurs, l'idée fondamentale consiste à ajouter un terme de régularisation sur la diagonale de la matrice de covariance du bruit [27], en particulier pour les antennes superdirectives. La formalisation de la méthode pour les antennes adaptatives a été donnée par Cox [42] qui introduit une contrainte sur le WNG lors de la formulation du problème d'optimisation posé dans la méthode MVDR. La résolution de ce problème par la méthode de Lagrange fait apparaître le multiplicateur de Lagrange comme un terme de régularisation appelé facteur de *diagonal loading*. Il s'agit donc de rajouter une contrainte pour pallier les erreurs de modélisation, ceci en complément de celles déjà utilisées pour définir la réponse souhaitée en fonction de la position de la source utile et des interférents. Ceci est discuté en détail ci-après.

1.2 Techniques pour l'amélioration de la robustesse

En lien avec ce que nous venons de voir précédemment, il s'agit dans cette partie de présenter les techniques utilisées pour s'adapter et se prémunir des erreurs liées à la modélisation et au mauvais pointage dues d'une part aux imperfections du frontal acoustique et d'autre part à la réverbération qui met en défaut les techniques de localisation et de poursuite de sources. Tout d'abord, la prise en compte de la réverbération et du caractère large bande des signaux nous oblige à considérer l'antenne filtrage-sommation décrite à la Figure 1.12 (le postfiltre est mis à titre indicatif). Grâce au filtrage du signal de parole sur toute la bande utile, cette antenne sera en mesure d'atténuer les versions "réverbérées" du signal utile de manière identique [14], voire de supprimer la réverbération à l'aide de la technique du filtrage adapté proposée par Jan et Flanagan [92, 93] et Rabinkin et al. [145]. Cette méthode consiste à calculer le filtre inverse de la réponse impulsionnelle reliant la source à chaque microphone et de l'appliquer sur les observations. Elle nécessite donc l'invariance et la connaissance de ces réponses impulsionnelles afin de pouvoir procéder à la déconvolution par filtrage RIF réalisé au niveau de l'antenne. Une analyse et une évaluation des méthodes classiques de déréverbération sont données dans [123]. Des travaux plus récents sur une approche de déréverbération multicanal et un algorithme d'estimation du temps de réverbération se trouvent dans [119, 177]. Par ailleurs, l'antenne filtrage-sommation sera capable de créer des zéros dans la direction des sources interférentes connues, ce qui ne peut être fait avec la simple antenne retard-somme. Les techniques que nous allons décrire concernent plus spécifiquement l'amélioration des antennes adaptatives linéairement contraintes (LCMV) dont fait partie celle proposée par Frost [61] en 1972 et sa variante proposée par Griffiths et Jim [73] dix

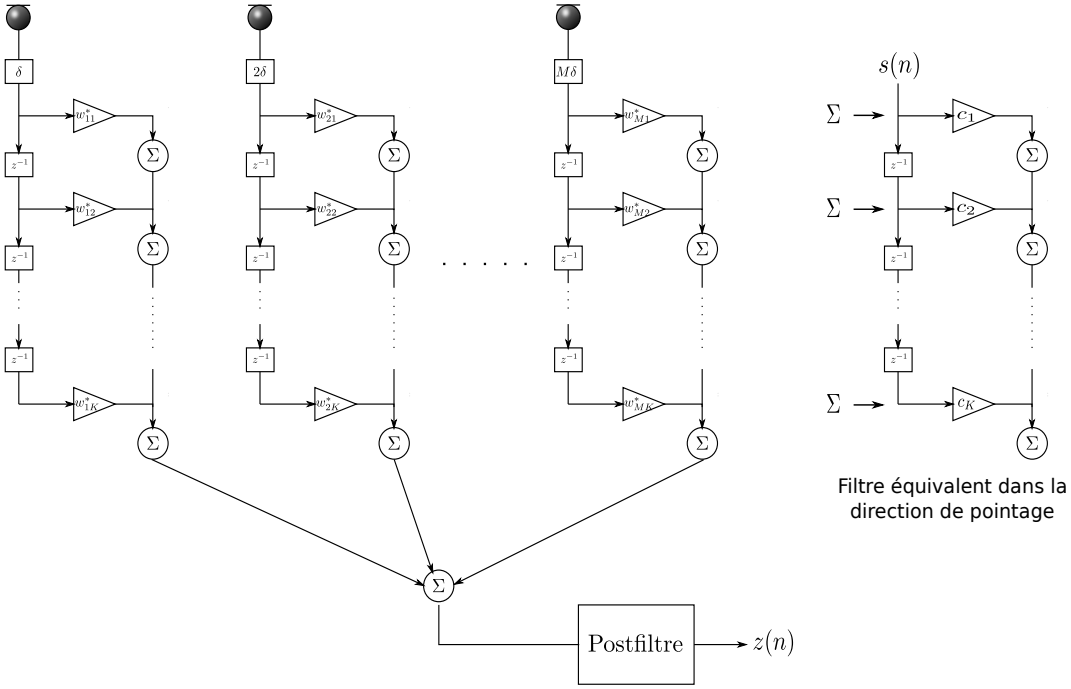


Figure 1.12 – Antenne large bande filtrage-sommation. Le premier bloc placé derrière chaque microphone assure la remise en phase des signaux par l'application d'un délai à définir en fonction de la position de la source cible. Un filtre RIF assure ensuite la conception de la réponse souhaitée par application des contraintes sur les coefficients que l'on obtient grâce à la résolution d'un problème d'optimisation suivant une métrique choisie.

années plus tard. Frost propose un algorithme adaptatif pour la détermination des coefficients optimaux ($\mathbf{w_{opt}}$) de l'antenne filtrage-sommation de la Figure 1.12. Cet algorithme correspond en fait à une mise en œuvre adaptative de l'estimateur de Capon ou MVDR qui minimise la puissance du signal en sortie (*Minimum Variance*) tout en préservant le signal utile provenant de la direction pointée (*Distortionless Response*). Le problème mathématique associé s'écrit comme suit :

$$\begin{aligned} &\min \mathbf{w^H R w} \\ &\text{s.c.q } \mathbf{w}^H\mathbf{d} = 1 \end{aligned}$$

La contrainte se justifie par deux aspects. D'une part, la matrice de corrélation du bruit et des interférents $\mathbf{Q}$ est inconnue, dans la pratique nous disposons uniquement de la matrice de corrélation des observations $\mathbf{R}$. Par conséquent, pour ne pas supprimer le signal utile provenant de la direction donnée par $\mathbf{d}$, en particulier lorsque celui-ci prédomine, il est important d'ajouter cette contrainte de non distorsion. Par ailleurs, cette contrainte permet d'éviter la solution évidente $\mathbf{w} = \mathbf{0}$. En ce sens, l'algorithme décrit dans [61] est un algorithme du gradient stochastique contraint dit *Constrained Least Mean Square* pour lequel il est nécessaire d'avoir une

connaissance *a priori* sur la direction d'arrivée du signal d'intérêt. Par rapport à cette méthode, l'antenne GSC transforme le problème d'optimisation contraint en un problème d'annulation de bruit non contraint par le report de la complexité sur l'architecture. Le schéma bloc de l'antenne GSC est donné à la Figure 1.13 où nous retrouvons l'antenne filtrage-sommation dans le bloc formateur de voies fixe. Avec une réalisation algorithmique plus simple, cette structure offre donc la possibilité de contraindre indépendamment la réponse de l'antenne pour l'extraction de la source utile et/ou la suppression des interférents dont la position est connue. Pour cela, il suffira de contraindre la conception du formateur de voies fixe fournissant une estimée du signal utile et de la matrice de blocage dont le but est de créer une référence des signaux de bruit par blocage de la source utile. Ceux-ci seront ensuite supprimés grâce au schéma classique de l'annulation de bruit adaptative qui dans ce cas est multivoie. Ces caractéristiques nous ont conduit à retenir, dans le cadre de cette thèse, la technique du GSC afin de réaliser l'implémentation de l'antenne LCMV plutôt que celle de Frost car, comme nous allons le voir, pour être efficace, la méthode de Frost nécessite d'utiliser un grand nombre de microphones pour profiter des différentes contraintes existantes et destinées à la rendre plus robuste, ou bien alors de contraindre la directivité de l'antenne par décomposition de la matrice de corrélation. Ces différents points feront l'objet des paragraphes ci-après.

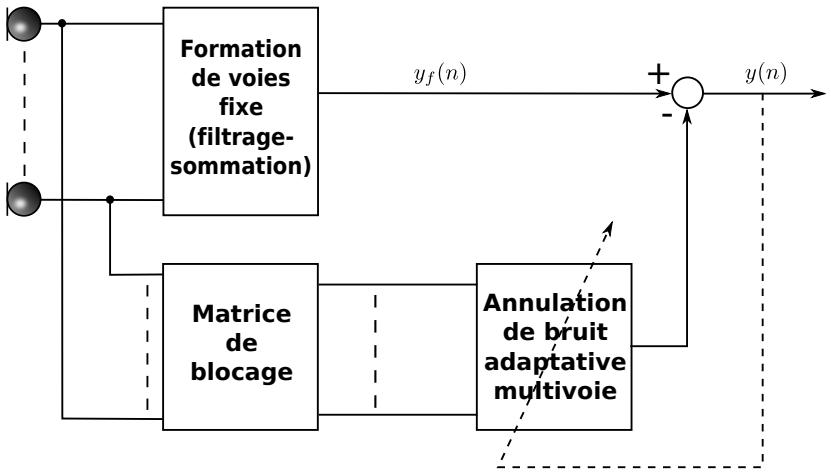


Figure 1.13 – Schéma bloc de l'antenne GSC. L'antenne fixe de type filtrage-sommation réalise le pointage dans la direction de la source cible et permet un premier traitement pour l'enrichissement du signal. La matrice de blocage fournit une référence de bruit en créant un zéro spatial dans la direction du signal cible. L'annuleur de bruit adaptatif supprime par soustraction le bruit résiduel contenu en sortie de formation de voies.

Néanmoins, outre les erreurs de modélisation, le problème critique à résoudre pour l'antenne GSC est la fuite du signal utile dans l'étage d'estimation du bruit qui cause l'annulation en sortie du signal d'intérêt estimé par le formateur de voies fixe ($y_f(n)$). Ce phénomène d'auto-annulation du signal crée des distorsions sur le signal de parole estimé ($y(n)$). Par conséquent, chacun des blocs constitutifs du GSC doit être rendu plus robuste. Dans le paragraphe suivant nous présentons les méthodes proposées pour améliorer la conception du frontal acoustique, de la matrice de blocage et de l'annuleur de bruit multivoie.

1.2.1 Formation de voies adaptative robuste

Dans [85], Hoshuyama et Sugiyama classent les méthodes mises en œuvre pour pallier la fuite du signal selon qu'elles traitent le problème au niveau de la matrice de blocage ou bien de l'annuleur de bruit multivoie. Dans le premier cas, cela s'apparente à une approche directe, alors que pour le second, il s'agit plutôt de minimiser l'influence du signal de fuite en fin de chaîne, soit donc une méthode indirecte. Néanmoins, nous pouvons déjà affirmer que la solution optimale se trouve dans la combinaison des deux approches car les facteurs aggravants sont inévitables. Ainsi, nous allons dans un premier temps décrire les différentes approches qui ont trait à l'amélioration de la robustesse du frontal acoustique par l'ajout d'un ensemble de contraintes. Cette description est construite autour du problème MVDR donné précédemment. Dans un second temps, nous nous attachons à la conception de la matrice de blocage et de l'annuleur de bruit multivoie sous les problèmes liés aux incertitudes.

Contraintes sur le formateur de voies fixe (frontal acoustique)

Le rôle du formateur de voies fixe est d'extraire le signal cible en maximisant sa réponse dans cette direction tout en créant des zéros dans la direction des interférents. Les contraintes qui lui sont imposées ont pour but de garantir un filtrage des signaux en "parfaite" cohérence de phase, y compris en présence d'erreurs sur la position des microphones. De plus, lorsque ce formateur de voies est conçu comme une antenne superdirective pour pallier aux faibles performances du GSC dans un bruit diffus [20, 31, 95], la robustesse vis-à-vis du bruit blanc est un facteur déterminant. Une telle implémentation est proposée notamment par Bitzer et al. dans [94] et McCowan et al. dans [128].
Pour améliorer la robustesse de cet élément, l'idée consiste a ajouter des contraintes sur les filtres qui réalisent le filtrage spatial. Ces contraintes sont de quatres types.

1. Ponctuelles [61, 179] : elles contraignent la réponse de l'antenne $H(f, \phi)$ a prendre certaines valeurs pour un angle et une fréquence (ϕ_0, f_0) donnés ;

2. Lignes [46] : elles contraignent la réponse fréquentielle de $H(f,\phi)$ pour une direction angulaire ϕ_0 choisie ;
3. Dérivatives [25, 58, 59, 180] : elles permettent de lisser la réponse de l'antenne $H(f,\phi)$. Typiquement, il sera possible d'aplanir $H(f,\phi)$ pour une direction et une fréquence donnée en procédant comme suit :

$$\frac{\partial H(f,\phi)}{\partial \phi}\Big|_{(f=f_0,\phi=\phi_0)} = 0 \qquad \frac{\partial H(f,\phi)}{\partial f}\Big|_{(f=f_0,\phi=\phi_0)} = 0 \tag{1.13}$$

4. De fuites [118] : elles s'appuient sur un développement en série de Taylor de la matrice de corrélation du signal utile. En ne considérant que les dérivées à l'ordre 1 et 2, cette approche contraint le produit du vecteur des coefficients des filtres par les dérivées obtenues et la matrice de blocage à être nul dans la direction de la source utile. Du point de vue conceptuel, elles sont identiques aux contraintes dérivatives mais, selon les auteurs, permettent de diminuer le nombre de contraintes utilisées pour ajuster les lobes.

L'introduction de multiples contraintes linéaires a été proposé initialement par Booker et Ong [22], puis repris par Frost dans son algorithme adaptatif. L'utilisation de contraintes spatiales et fréquentielles dans le cas du filtrage de Wiener en sous-bandes est également donnée par Nordholm et al. dans [138]. Dans le cas du problème MVDR, ces contraintes peuvent toujours être écrites sous la forme

$$\mathbf{C}^H\mathbf{w} = \mathbf{g} \tag{1.14}$$

où $\mathbf{C}$ est la matrice $M \times J$ des contraintes dans les J directions de l'espace et $\mathbf{g}$ le vecteur de taille $J \times 1$ spécifiant la réponse désirée dans ces J directions. La solution au problème MVDR sous ces multiples contraintes est :

$$\mathbf{w_{opt}} = \mathbf{R}^{-1}\mathbf{C}(\mathbf{C}^H\mathbf{R}^{-1}\mathbf{C})^{-1}\mathbf{g} \tag{1.15}$$

Cependant, le nombre de contraintes applicables sera limité par le nombre M de microphones, soit $J \leq M$, ce qui signifie que le nombre de degrés de liberté peut devenir rapidement problématique, en particulier pour des antennes de petites tailles. Un des moyens de pallier ce problème, sans augmenter le nombre de microphones, consiste à utiliser une décomposition en valeurs singulières "tronquée" [105] de l'équation des contraintes (1.16). En effet, la meilleure approximation de rang k de $\mathbf{C}$ au sens des moindres carrés est donnée par $\mathbf{U}\mathbf{\Sigma}\mathbf{V}^H$ où $\mathbf{\Sigma}$ est la matrice diagonale contenant les k plus grandes valeurs singulières, $\mathbf{U}$ la matrice des vecteurs singuliers à gauche et $\mathbf{V}$ la matrice des vecteurs singuliers à droite. L'équation des contraintes "tronquée" sera donc de la forme

$$\mathbf{V}\mathbf{\Sigma}^T\mathbf{U}^H\mathbf{w} = \mathbf{g} \iff \mathbf{U}^H\mathbf{w} = \mathbf{\Sigma}^\dagger\mathbf{V}^H\mathbf{g} \tag{1.16}$$

où † désigne la pseudo-inverse de Moore-Penrose. La solution du problème sous ces contraintes particulières est dite à minimum de variance sous contraintes environnementales [105]. Son expression est la suivante :

$$\mathbf{w_{ce}} = \mathbf{R}^{-1}\mathbf{U}(\mathbf{U}^H\mathbf{R}^{-1}\mathbf{U})^{-1}\mathbf{\Sigma}^\dagger\mathbf{V}^H\mathbf{g} \tag{1.17}$$

En ce sens, certaines méthodes proposent également de travailler sur les sous-espaces après décomposition en valeurs propres de la matrice de corrélation [110, 167, 200]. En effet, pour obtenir une estimée plus précise du vecteur de pointage inconnu, Yu et Er [198] définissent l'ensemble des possibles pour celui-ci par la projection sur le sous-espace bruit plus interférents. Ce sous-espace, extrait de la matrice de corrélation, permet selon les auteurs de ne pas avoir à faire de compromis entre la suppression efficace des interférents ou du bruit, à la différence des méthodes d'optimisation convexe décrites plus loin dans ce paragraphe. Une fois l'espace interférents plus bruit délimité, les coefficients optimaux sont ceux qui maximisent la puissance en sortie du formateur de voies de Capon à l'aide des vecteurs de pointage définis auparavant.
Supposons que K signaux arrivent sur l'antenne. La matrice de corrélation spatiale est décomposée telle que $\mathbf{R} = \mathbf{U\Gamma U}^H$, avec $\mathbf{U}$ la matrice des vecteurs propres et $\mathbf{\Gamma}$ la matrice des valeurs propres. La base orthogonale $\mathbf{U_d}$ des vecteurs de pointage, formée des vecteurs propres issus des K plus grandes valeurs propres de $\mathbf{U}$, définit l'espace O ci-après.

$$O = \{\mathbf{d}| \quad \mathbf{d} = \mathbf{U_d c}, \quad \mathbf{c} \in C^K\}$$

En projetant le vecteur de pointage sur cet espace, l'erreur commise entre le vecteur de pointage estimé $\mathbf{d_{estimé}} = \mathbf{U_d c_{est}}$ avec $\mathbf{c_{est}} = \arg_c min \ ||\mathbf{U_d c} - \mathbf{d_0}||^2$ et celui pointant la vraie direction de la source est minimisée puisque celui-ci est pris dans un sous-espace orthogonal. Le vecteur $\mathbf{d_0}$ représente le vecteur de pointage avant projection. Ainsi, pour obtenir une solution robuste à l'aide du formateur de voies de Capon, la contrainte $||\mathbf{d} - \mathbf{d_{estimé}}||^2 \leq \epsilon$ est ajoutée.

$$\begin{aligned} \min \mathbf{w^H R w} \quad \text{s.c.q} \qquad & \mathbf{d^H w} = 1 \\ & ||\mathbf{d} - \mathbf{d_{estimé}}||^2 \leq \epsilon \end{aligned}$$

Dans la même idée, Slavakis et al. proposent dans [167] d'utiliser un algorithme de projection adaptatif (*Adaptive Projected Subgradient Method* (APSM)). Cependant, cette méthode est moins réaliste car elle suppose à l'avance connues les erreurs de pointage ainsi que le rayon de la boule contenant le vecteur de pointage recherché. Nous trouverons aussi dans la littérature des approches bayesiennes [13, 197] où un modèle de densité de probabilité est utilisé pour estimer la direction d'arrivée de la source utile.

Jusqu'ici, les contraintes sont plutôt destinées à résoudre le problème des incertitudes sur la position des sources et des microphones sur le frontal acoustique. Pour tenir compte de la variabilité temporelle des caractéristiques (gain et phase) des microphones, Doclo et al. proposent dans [47] d'inclure, dans les méthodes de conception d'antenne large bande décrites dans [48], un modèle de densité de probabilité à l'image des variations des caractéristiques des microphones au cours du temps. La réponse de l'antenne $H(f, \phi)$ est donc agrémentée de la réponse nominale des microphones, et la fonction de coût répondant à un critère des moindres carrés est pondérée par la densité de probabilité. Cette méthode, qui optimise les performances en moyenne, nécessite donc d'avoir une connaissance *a priori* de la densité de probabilité des caractéristiques. Par opposition, les auteurs proposent une autre technique qui optimise le pire cas selon une approche *minimax*.

Si maintenant nous nous intéressons au problème de l'amplification des bruits spatialement blancs, en particulier pour les antennes superdirectives, il nous faut commencer par étudier la méthode du *diagonal loading* vue au paragraphe 1.1.2 et initiée par Carlson en 1988 [27]. La méthode consiste à imposer une contrainte sur le gain en bruit blanc (WNG) [42], soit $WNG = \delta^2 \leq M$, ce qui revient à contraindre la norme au carré des filtres comme suit $\mathbf{w}^H\mathbf{w} \leq \delta^{-2}$. La résolution du problème MVDR sous cette contrainte supplémentaire à l'aide de la méthode de Lagrange conduit à la solution suivante :

$$\mathbf{w_{dl}} = \frac{(\mathbf{R} + \epsilon\mathbf{I})^{-1}\mathbf{d}}{\mathbf{d}^H(\mathbf{R} + \epsilon\mathbf{I})^{-1}\mathbf{d}} \tag{1.18}$$

Il s'agit donc d'une méthode de régularisation qui ajoute la quantité ϵ (multiplicateur de Lagrange) sur la diagonale de la matrice de covariance, soit donc sur les valeurs propres, sans modifier les vecteurs propres. Ce terme $\epsilon > 0$ appelé facteur de *diagonal loading* peut être vu comme un terme qui pénalise les fortes valeurs de $\mathbf{w}$ et doit être ajusté pour satisfaire $WNG \leq M$. Une méthode relativement proche du *diagonal loading* et donnant les mêmes performances, consiste à seuiller les valeurs propres de la matrice de covariance [78] de sorte à ce que la plus petite valeur propre ne soit pas inférieure à ϵ fois la plus grande. Ceci peut être interprété comme l'injection d'un bruit blanc contrôlé par le facteur de *diagonal loading*. Si ces méthodes se révèlent efficaces, le choix de ce facteur de *diagonal loading* est délicat à régler car la relation qui le lie à δ^2 est difficilement exploitable [195]. Jusqu'en 2003 et avant l'introduction dans le traitement d'antenne des problèmes d'optimisation convexes par Vorobyov [183], il était déterminé à partir d'une méthode itérative qui requiert un coût de calcul non négligeable. Grâce à l'utilisation de l'optimisation convexe et plus particulièrement du *Second-Order Cone Program* (SOCP), le problème MVDR va être résolu en un temps polynomial avec l'ensemble des contraintes linéaires précé-

dentes plus celles de type conique associées notamment à la contrainte sur le WNG. Le SOCP est utilisable uniquement sur des fonctions de coût linéaires convexes, par conséquent, il faut reformuler le problème qui nous intéresse sous la forme générale décrite ci-après.

$$\begin{aligned} & \min \mathbf{c^T x} \\ \text{s.c.q} \quad & \mathbf{a_i^T x} \leq \mathbf{b_i} \quad \forall \mathbf{a_i} \in \varepsilon_i \quad i = 1 \dots m \end{aligned}$$

où $\varepsilon_i = \{\mathbf{a_i} + \mathbf{P_i u} |\ ||\mathbf{u}||^2 \leq 1\}$ représente une ellipse qui contraint le domaine de variation des $\mathbf{a_i}$, d'où le nom de la méthode. La matrice $\mathbf{P_i}$ est quant à elle à l'image de cette incertitude, c'est elle qui définit la forme de l'ellipse (plate ou non). Lorsque $\mathbf{P_i}$ est réduite à la matrice nulle alors cela signifie qu'aucune incertitude n'existe sur les $\mathbf{a_i}$. Ici, nous donnons un cas particulier où seuls les $\mathbf{a_i}$ varient dans le temps mais on peut aussi établir les mêmes contraintes sur $\mathbf{c}$ et $\mathbf{b_i}$.

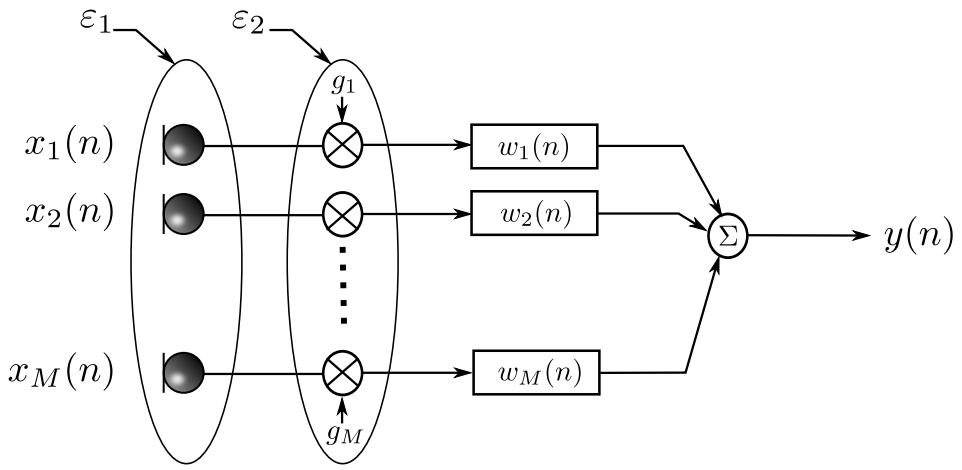


Figure 1.14 – Les variations possibles de la réponse de l'antenne sont contenues dans l'ellipse ε_1. Celles du gain des micros dans ε_2.

La mise en œuvre la plus simple et la plus explicite de cet algorithme se trouve dans [195] pour le cas d'une antenne circulaire. La fonction de coût quadratique du problème MVDR est transformée en une fonction de coût linéaire à l'aide d'une décomposition de Cholesky de la matrice de covariance. Une autre implémentation plus complexe est faite par Lorenz et Boyd dans [120] pour pallier aux erreurs de pointage. Ils obtiennent une antenne optimale dans le sens où, pour une matrice de covariance $\mathbf{R}$ donnée et une réponse de l'antenne comprise dans l'ellipse de contrainte, Figure 1.14, aucun autre vecteur de pondération ne parvient à des puissances pondérées aussi faibles tout en maintenant la partie réelle de la réponse plus grande que l'unité dans la direction du signal utile. L'objectif est donc de garantir le respect de

la contrainte de non distorsion sur toute la zone correspondant à la fenêtre d'incertitude. Cette solution appelée *Robust Minimum Variance Beamformer* (RMVB) est obtenue en se ramenant à l'étude du problème suivant :

$$\begin{aligned} &\min \mathbf{x^T R x} \\ \text{s.c.q} \quad &||\mathbf{A^T x}|| \leq \mathbf{c^T x} - 1 \end{aligned}$$

où $\mathbf{A}$ est la matrice qui conditionne l'ellipse (ses demi-axes), $\varepsilon = \{\mathbf{Au} + \mathbf{c} \mid ||\mathbf{u}|| \leq 1\}$, $\mathbf{c}$ est le centre de l'ellipse et $\mathbf{x} = [Re(\mathbf{w})\ Im(\mathbf{w})]^T$ le vecteur des coefficients. Nos résultats de simulation donnés à la Figure 1.15 montrent que, autour de la position du signal utile situé à 45°, une zone plane est obtenue en fonction de l'angle d'ouverture souhaité. Pour chacune de ces zones, nous constatons que la réponse de l'antenne est égale à 1 (soit 0 dB sur le graphique) tandis que les interférents situés aux positions 30° et 75° sont supprimés. Ainsi, grâce à ces zones planes de largeur variable, on pare aux erreurs de pointage en tolérant une variation de la position de la source à l'intérieur de cette fenêtre d'incertitude. Néanmoins, lorsque l'on s'attache à l'importance du lobe secondaire obtenu en $\phi = 20°$ pour une fenêtre de 12°, nous nous apercevons que, dans l'hypothèse où les sources interférentes sont nombreuses, alors il se peut que l'une d'elle située à cette position ne soit plus filtrée spatialement à cause de ce type de comportement. En offrant une réduction d'au moins 10 dB sur tous les lobes secondaires, la solution de Capon pourrait donc paraître suffisante mais ce n'est pas le cas. Li et al montrent dans [116] que la solution de Capon aura un comportement encore plus critique si nous ajoutons aux erreurs de pointage un nombre d'échantillons insuffisant pour le calcul de l'estimée de la matrice de covariance. Nous venons simplement de mettre en image une limite du RMVB lorsque l'ellipse des contraintes est mal dimensionnée et ne couvre pas l'ensemble des variations possibles de la position de la source, soit ici dans une fenêtre supérieure à 10°. Ceci rejoint le travail de Vorobyov et al. qui se sont intéressés à l'optimisation du pire cas, [183], et au même titre que la méthode précédente, ont utilisé un algorithme SOCP afin d'obtenir une réponse sans distorsion en fixant la valeur du facteur de *diagonal loading* à la valeur maximale que peut prendre la norme du vecteur de pointage $\mathbf{d}$ lorsque celui-ci est biaisé. Une implémentation temps réel de cette méthode basée sur le filtrage de Kalman a été proposée dans [53]. Comparée à la méthode de projection sur les sous-espaces qui sont les plus efficaces pour des RSB favorables, cette méthode présente l'avantage de ne pas nécessiter une connaissance précise de la dimension du sous-espace signal plus interférent qui doit de plus être de faible dimension. Certes, la réjection des interférents est moins importante lorsque les RSB sont défavorables, mais suffisante comparée aux méthodes basées sur les sous-espaces qui deviennent inefficaces. Des exemples d'utilisation du SOCP pour la conception d'antenne acoustique robuste sont donnés dans [196] ainsi que [121, 122]

où les auteurs traitent de la conception d'une antenne superdirective.

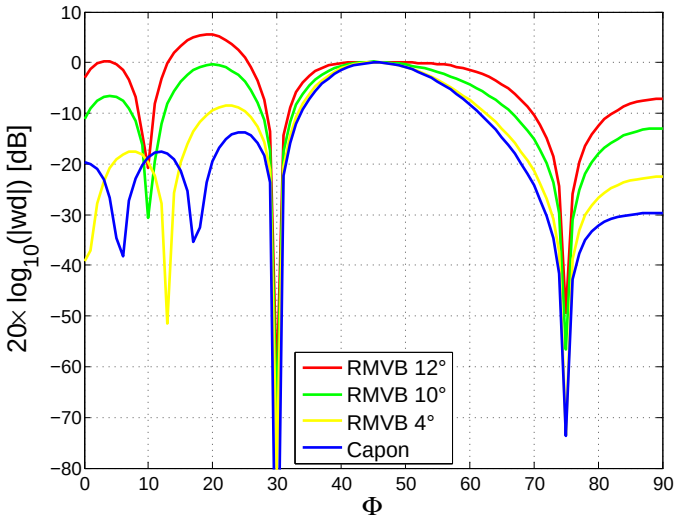


Figure 1.15 – Réponse de l'antenne robuste MVDR (RMVB) pour différentes fenêtres d'incertitudes autour de la position du signal utile placé à $45°$. Les interférents situés en $30°$ et $75°$ ont des puissances supérieures à la source utile de 20 dB et 10 dB respectivement.

Enfin, nous pourrons compléter les méthodes précédentes par les méthodes de poursuite de sources [4, 5, 8] qui en l'absence de celles-ci peuvent conduire au pointage sur un interférent de même type (parole). Par ailleurs, si celles fondées sur le filtrage de Kalman peuvent avoir un coût de calcul incompatible avec un traitement en temps réel, nous noterons que celles utilisant les TDOA pâtissent d'un modèle de propagation trop réducteur. Ceci sera discuté dans le paragraphe suivant.

Lorsque cela est possible, la prise en compte des imperfections peut se faire à l'aide de techniques de calibration qui vont permettre d'étalonner le système avant son utilisation. Dans [62] Fugde et al. utilisent un signal test qui va permettre de pré-calculer, à l'aide d'une décomposition en valeurs singulières, la matrice de blocage et la matrice de passage associées au frontal acoustique. Une procédure d'auto-calibration est également analysée par Nordholm et al. dans [137]. A contrario, McCowan et al. [127] évitent l'utilisation du signal de calibration en tirant profit du lien entre la fonction de cohérence dans les conditions d'un bruit diffus et la position des micros. De plus, cette méthode laisse la possibilité de pouvoir recalibrer le système pendant son fonctionnement.

Contraintes sur la matrice de blocage et l'annuleur de bruit multivoie

Les contraintes imposées à la matrice de blocage et l'annuleur de bruit multivoie sont destinées à supprimer le signal de fuite. En fait, la réponse de la matrice de blocage correspond à un zéro spatial dans la direction de la source utile. En admettant comme précédemment que la position de la source varie dans une fenêtre d'amplitude donnée, nous pourrons ré-appliquer les contraintes dérivatives pour la matrice de blocage [9, 63, 64] afin d'élargir le "zéro" pour qu'il englobe la fenêtre d'incertitude. Il s'agit finalement de créer une antenne complémentaire [154, 155], sachant qu'il est plus facile, dans le domaine du filtrage en général, de créer des zéros précis que des maxima. Claesson et Nordholm [35] proposent de concevoir la matrice de blocage à partir de filtres spatiaux passe-hauts fondés sur une connaissance *a priori* de la position et de la bande utile de la source cible. Si, dans le cas de la parole il est simple de fixer la bande utile, la fenêtre d'observation angulaire autour de la position de la source cible peut quant à elle évoluer au cours du temps.
Mais comme l'affirme Gannot et al. dans [65, 66, 68], le principal défaut du GSC proposé par Griffiths et Jim réside dans le modèle de propagation des signaux qui est supposé correspondre à un simple retard. En effet, les travaux de Affes et Grenier [3] montrent que de bons résultats ne peuvent être obtenus sans procéder à l'estimation précise des réponses impulsionnelles pour le pointage ou à leurs inversions (filtre adapté). Cependant, nous verrons au Chapitre 2 paragraphe 2.3.1, que les méthodes d'inversion se heurtent aux propriétés de phase non minimale des réponses impulsionnelles acoustiques. L'objectif est néanmoins de remplacer la synchronisation des signaux sur l'antenne faite avec de simples retards par un modèle plus précis et réaliste. Avant la méthode d'estimation des Fonctions de Transfert Acoustiques (FTA) proposée par Affes et Grenier [6], les méthodes de la littérature avaient pour modèle une propagation idéale que l'on suppose en ondes planes et selon le seul chemin direct. Ce modèle est trop réducteur et en ce sens Gannot et al. généralisent le GSC au cas d'une propagation échoïque en incorporant les FTA dans le calcul de chacun des blocs constitutifs. Le détail des calculs est donné en Annexe A et le schéma résultant est présenté à la Figure 1.16 en supposant le modèle de signal mis sous forme vectorielle de l'équation (1.19).

$$\mathbf{X}(k,l) = \mathbf{A}^{S_1}(k)S_1(k,l) + \mathbf{A}^{S_2}(k)S_2(k,l) + \mathbf{B}(k,l) \tag{1.19}$$

avec l l'indice de trame, k l'indice des composantes fréquentielles et les quantités

– $\mathbf{X}(k,l)$: vecteur des observations ;

$$\mathbf{X}(k,l) = [X_1(k,l)\ X_2(k,l) \cdots X_M(k,l)]$$

– $\mathbf{A}^{S_1}(k)$: vecteur des réponses fréquentielles entre le signal source et chacun des micros ;

$$\mathbf{A}^{S_1}(k) = \left[A_1^{S_1}(k)\ A_2^{S_1}(k) \cdots A_M^{S_1}(k)\right]$$

– $\mathbf{A}^{S_2}(k)$: vecteur des réponses fréquentielles entre le signal interférent et chacun des micros ;

$$\mathbf{A}^{S_2}(k) = \left[A_1^{S_2}(k)\ A_2^{S_2}(k) \cdots A_M^{S_2}(k)\right]$$

– $\mathbf{B}(k,l)$: vecteur du bruit reçu sur chacun des micros.

$$\mathbf{B}(k,l) = [B_1(k,l)\ B_2(k,l) \cdots B_M(k,l)]$$

En fonction de ces quantités, le formateur de voies fixe et la matrice de blocage ont pour expressions :

$$\mathbf{W_0}(k) = \frac{\frac{\left\|\mathbf{A}^{S_2}(k)\right\|^2}{\left|A_1^{S_2}(k)\right|}\frac{\mathbf{A}^{S_1}(k)}{A_1^{S_1}(k)} - \frac{\mathbf{A}^{S_2}(k)}{A_1^{S_2}(k)}\frac{\mathbf{A}^{S_2^H}(k)}{A_1^{S_2}(k)}\frac{\mathbf{A}^{S_1}(k)}{A_1^{S_1}(k)}}{\frac{\left\|\mathbf{A}^{S_1}(k)\right\|^2}{\left|A_1^{S_1}(k)\right|^2}\frac{\left\|\mathbf{A}^{S_2}(k)\right\|^2}{\left|A_1^{S_2}(k)\right|^2} - \frac{\mathbf{A}^{S_1^H}(k)}{A_1^{S_1}(k)}\frac{\mathbf{A}^{S_2}(k)}{A_1^{S_2}(k)}\frac{\mathbf{A}^{S_2^H}(k)}{A_1^{S_2}(k)}\frac{\mathbf{A}^{S_1}(k)}{A_1^{S_1}(k)}}\mathcal{F}(k) \tag{1.20}$$

$$\mathcal{H}(k) = \begin{bmatrix} Q_3(k) & Q_4(k) & \dots & Q_M(k) \\ L_3(k) & L_4(k) & \dots & L_M(k) \\ 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \\ 0 & 0 & \dots & 1 \end{bmatrix}$$

Avec

$$L_m(k) = -\frac{\frac{A_m^{S_1^*}(k)}{A_1^{S_1^*}(k)} - \frac{A_m^{S_2^*}(k)}{A_1^{S_2^*}(k)}}{\frac{A_2^{S_1^*}(k)}{A_1^{S_1^*}(k)} - \frac{A_2^{S_2^*}(k)}{A_1^{S_2^*}(k)}}$$

$$Q_m(e^{j\omega}) = -\frac{\frac{A_2^{S_1^*(k)}}{A_1^{S_1^*(k)}} \cdot \frac{A_m^{S_2^*(k)}}{A_1^{S_2^*(k)}} - \frac{A_2^{S_2^*(k)}}{A_1^{S_2^*(k)}} \cdot \frac{A_m^{S_1^*(k)}}{A_1^{S_1^*(k)}}}{\frac{A_2^{S_1^*(k)}}{A_1^{S_1^*(k)}} - \frac{A_2^{S_2^*(k)}}{A_1^{S_2^*(k)}}}$$

et $\mathcal{F}(k)$ la réponse d'un filtre arbitrairement choisi. La matrice de blocage ainsi définie permet de bloquer les deux sources localisées S_1 et S_2. Les filtres $G_m(k,l)$ de la Figure 1.16 sont les filtres adaptatifs de l'annuleur de bruit multivoie que l'on peut regrouper dans une seule et même matrice notée $\mathbf{G}(k,l)$. Rappelons que, dans sa version de base et dans le cas où la source S_2 est absente, Griffiths et Jim définissent le formateur de voies fixe et la matrice de blocage du GSC comme suit :

$$\mathbf{W_0}(k) = \left[1 \quad e^{j\frac{2\pi}{K}\tau_2} \quad e^{j\frac{2\pi}{K}\tau_3} \dots \quad e^{j\frac{2\pi}{K}\tau_M}\right] \tag{1.21}$$

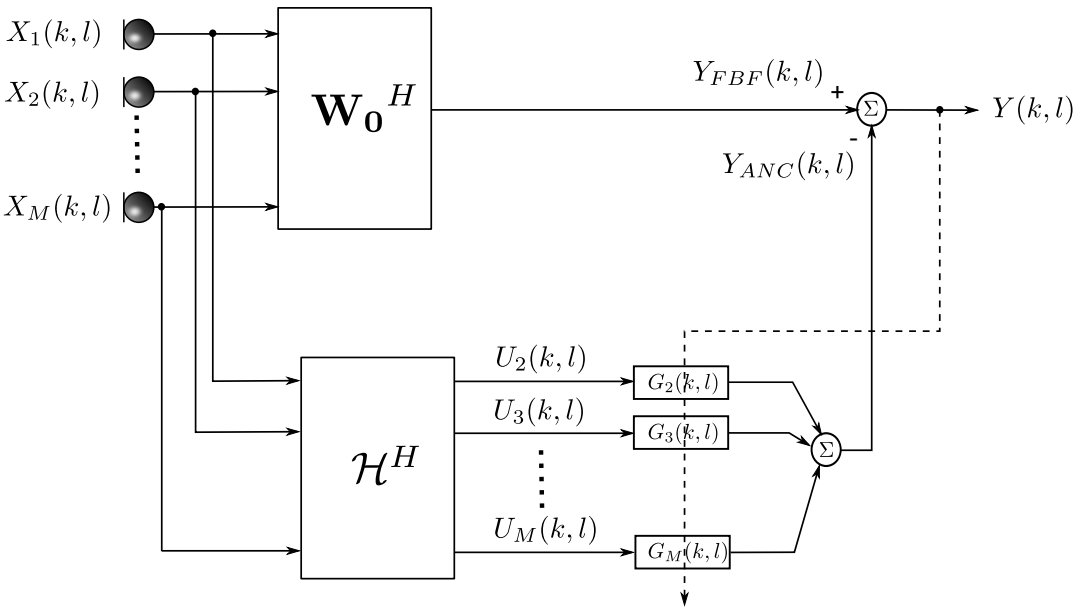


Figure 1.16 – GSC généralisé.

ce qui permet effectivement de compenser le retard sur chaque voie lorsque l'on suppose une propagation selon de simples retards (1.22) et où $\tau_2, \ldots, \tau_M$ sont les retards relatifs entre les microphones et la référence.

$$\mathbf{A}^{S_1}(k) = \begin{bmatrix} 1 & e^{-j\frac{2\pi}{K}\tau_2} & e^{-j\frac{2\pi}{K}\tau_3} \ldots & e^{-j\frac{2\pi}{K}\tau_M} \end{bmatrix} \tag{1.22}$$

Remarque 2 : *Le premier microphone de la Figure 1.16 est pris pour référence.*

$$\mathcal{H}(k) = \begin{bmatrix} -e^{j\frac{2\pi}{K}\tau_2} & -e^{j\frac{2\pi}{K}\tau_2} & \ldots & -e^{j\frac{2\pi}{K}\tau_2} \\ 1 & 0 & \ldots & 0 \\ 0 & 1 & \ldots & 0 \\ \vdots & \vdots & \ddots & \\ 0 & 0 & \ldots & 1 \end{bmatrix}$$

La prise en compte des FTA dans chacun des blocs du GSC est donc une solution au signal de fuite puisque le pointage est maintenant réalisé en tenant compte de l'effet de salle. Néanmoins, pour être exploitable, encore faut-il être capable d'estimer ces FTA, ou plutôt les rapports $\frac{\mathbf{A}^{S_1(k)}}{A_1^{S_1}(k)}$ et $\frac{\mathbf{A}^{S_2(k)}}{A_1^{S_2}(k)}$ qui apparaissent dans les expressions précédentes et sont appelés Fonctions de Transferts Relatives (FTR). Une référence incontournable à ce sujet écrite par Gannot et Cohen se trouve dans l'ouvrage *Handbook of Speech Processing* [68]. Alors qu'une description adaptée au cas du modèle

de signal donné plus haut est proposée en Annexe A, nous allons maintenant faire état des méthodes adaptatives à partir de la description donnée dans cette même référence [68].

Nous avons vu au paragraphe précédent qu'il fallait contraindre la norme des coefficients pour rendre l'antenne robuste. A partir de la définition suivante de la sensibilité de l'antenne aux variations sur les FTA, $J(k,l) = \mathbf{W}^H(k)\mathbf{W}(k)$, et de la décomposition du vecteur des filtres $\mathbf{W}(k)$ dans l'antenne GSC décrite en Annexe A, les auteurs font le constat suivant :

$$\mathbf{W}(k) = \mathbf{W_0}(k) - \mathbf{V}(k) = \mathbf{W_0}(k) - \mathcal{H}(k)\mathbf{G}(k,l) \tag{1.23}$$

Or $\mathbf{W_0}$ étant fixe, la contrainte se reporte sur le second membre, il suffira donc de contraindre la matrice des filtres adaptatifs.

$$\mathbf{G}^H(k,l)\mathbf{G}(k,l) = ||\mathbf{G}^H(k,l)||^2 \leq \Omega(k,l) \tag{1.24}$$

Ainsi, en reportant cette contrainte sur l'équation de mise à jour des filtres $\mathbf{G}(k,l)$ associés au GSC décrit en Annexe A, il vient :

$$\widetilde{\mathbf{G}}'(k,l+1) = \widetilde{\mathbf{G}}(k,l) + \mu\frac{U_m(k,l)Y^*(k,l)}{P_{est}(k,l)}$$

$$\widetilde{\mathbf{G}}(k,l+1) = \begin{cases} \widetilde{\mathbf{G}}'(k,l+1) & \text{si } ||\widetilde{\mathbf{G}}'(k,l+1)||^2 \leq \Omega(k,l+1) \\ \frac{\sqrt{\Omega(k,l+1)}}{||\widetilde{\mathbf{G}}'(k,l+1)||}\mathbf{G}'(k,l+1) & \text{sinon} \end{cases}$$

Cette méthode, qui consiste à contraindre la norme des coefficients des filtres adaptatifs dans le GSC, a été pleinement exploitée par Hoshuyama et Sugiyama [85, 86] en remplaçant la matrice de blocage usuelle par une matrice de blocage adaptative où les filtres seront également contraints. Ceci permet un contrôle adaptatif de la direction de pointage avec un nombre de contraintes réduit. Il n'est donc pas nécessaire d'augmenter le nombre de microphones. Le GSC robuste issu de ces travaux, décrit initialement dans le domaine temporel et dont la structure est donnée à la Figure 1.17, a ensuite été transposé dans le domaine fréquentiel par Herbordt et Kellermann [80, 81]. Les filtres ACF sont équivalents aux filtres G_m de la Figure 1.16. Pour plus de lisibilité et de cohérence avec les calculs présentés dans ce chapitre et l'Annexe A, nous noterons H_m les filtres adaptatifs de la matrice de blocage, ceux-ci sont équivalents aux filtres ABF [1].

1. La cohérence des notations est maintenue pour ce chapitre avec l'objectif de faire le lien avec les développements du Chapitre 6, d'où l'introduction des filtres ABF et ACF.

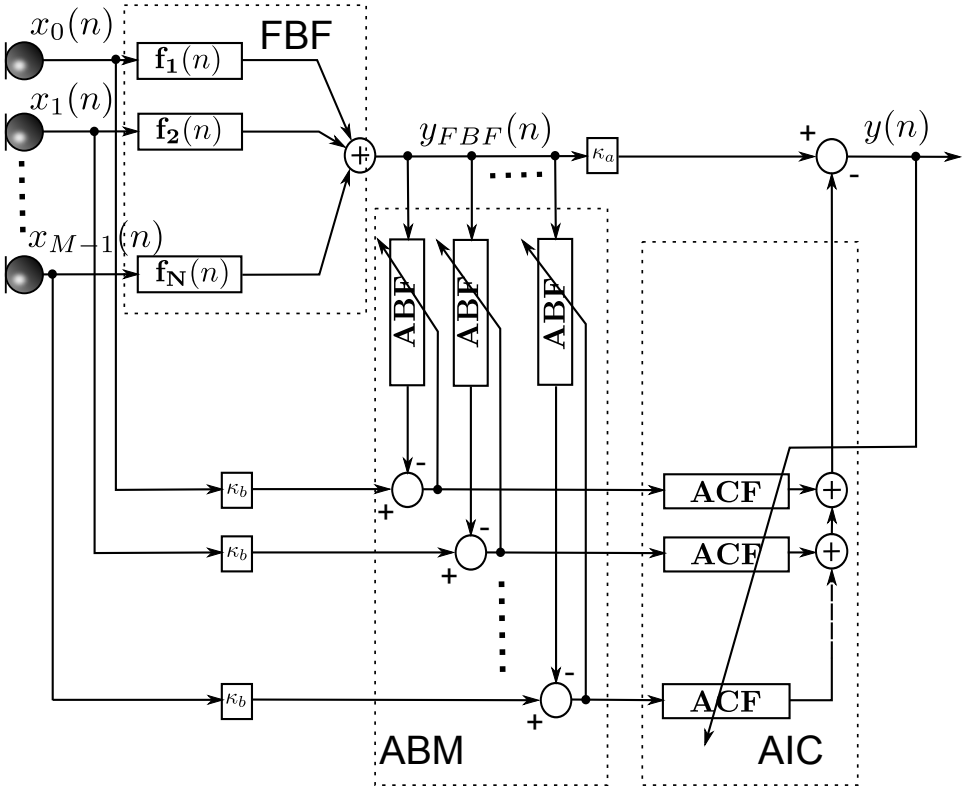


Figure 1.17 – Description du GSC robuste dans le domaine temporel. Le bloc FBF correspond au formateur de voies fixe, ABM à la matrice de blocage adaptative constituée des filtres adaptatifs noté ABF et l'annuleur de bruit multivoie AIC fait des filtres ACF.

En adaptant pour la Figure 1.17 le formalisme utilisé à la Figure 1.16, les estimées des signaux de bruit en sortie de la matrice de blocage adaptative (*Adaptive Blocking Matrix* (ABM)) et du signal de parole $y(n)$ sont :

$$\begin{aligned} U_m(k,l) &= X_m(k,l) - H_m^*(k,l)Y_{FBF}(k,l) \\ Y(k,l) &= Y_{FBF}(k,l) - \mathbf{G}^H(k,l)\mathbf{U}(k,l) \end{aligned}$$

En ce qui concerne la mise à jour des filtres adaptatifs H_m contenus dans la matrice de blocage, Hoshuyama et Sugiyama proposent deux versions. La première emploie l'algorithme *Leaky LMS* (1.25), déjà proposé par Claesson pour l'annuleur de bruit multivoie [35]. Cet algorithme introduit un bruit blanc sur les coefficients, ce qui est équivalent à contraindre leurs normes. Notons que l'injection d'un bruit blanc artificiel sur les coefficients des filtres de l'annuleur de bruit multivoie a été proposé dans [91]. L'ouverture de la fenêtre angulaire est contrôlée par le facteur de fuite α et le pas d'adaptation μ, bien qu'il n'existe pas de relation simple entre ces paramètres

et l'ouverture du lobe.

$$H_m(k,l+1) = (1-\alpha)H_m(k,l) + \frac{\mu}{|Y_{FBF}(k,l)|^2}U_m(k,l)Y^*_{FBF}(k,l) \tag{1.25}$$

En ce sens, les auteurs proposent une seconde approche qui utilise un masque prédéfini, servant de gabarit à la réponse temporelle des filtres calculés. L'équation associée est donnée en (1.26). Cette fois le pointage est contrôlé à travers les quantités ϕ_{low} et ϕ_{up} qui peuvent être ajustées pour couvrir une plage angulaire choisie autour de la source utile. Nous mettons ceci en image à la Figure 1.18.

$$H^{'}_m(k,l+1) = H_m(k,l) + \frac{\mu}{|Y_{FBF}(k,l)|^2}U_m(k,l)Y^*_{FBF}(k,l) \tag{1.26}$$

$$H_m(k,l+1) = \begin{cases} \phi_{low}(k,l+1) & H^{'}_m(k,l+1) \leq \phi_{low}(k,l+1) \\ \phi_{up}(k,l+1) & H^{'}_m(k,l+1) \geq \phi_{up}(k,l+1) \\ H^{'}_m(k,l+1) & \text{sinon} \end{cases}$$

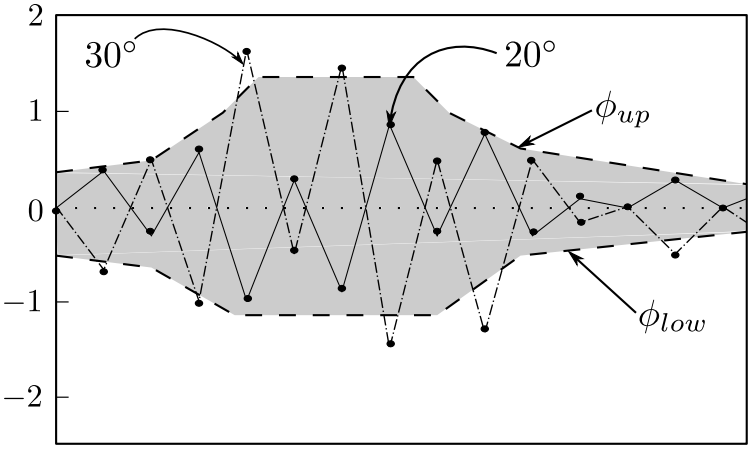


Figure 1.18 – Exemple de masque pour une ouverture à $20°$. Il s'agit d'une mise en image n'ayant aucun lien avec des résultats de simulation.

Par la suite, une mise en œuvre complètement adaptative a également été proposée par Gannot [65] en utilisant le critère de décorrélation introduit par Weinstein et al. [188]. En effet, à partir de la Figure 1.16 et des relations de l'Annexe A, nous pouvons exprimer les signaux de bruit et de parole estimés dans le domaine temporel suivant :

$$u_m(n) = x_m(n) - \sum_{k=-q_L}^{q_R} h_m(k)x_1(n-k)$$

$$y(n) = y_{FBF}(n) - \sum_{m=2}^{M}\sum_{k=-p_L}^{p_R} g_m(k)u_m(n-k)$$

Les filtres sont contraints à une structure de RIF. Avec une écriture vectorielle de ces différentes quantités,

$$\mathbf{h}_m = [h_m(-q_L) \dots h_m(q_R)]^T$$

$$\mathbf{g}_m = [g_m(-p_L) \dots g_m(p_R)]^T$$

$$\mathbf{y}(n) = [y(n+q_L) \dots y(n) \dots y(n-q_R)]^T$$

$$\mathbf{u}_m = [u_m(n+p_L) \dots u_m(n) \dots u_m(n-p_R)]^T$$

$$\mathbf{x_1}(n) = [x_1(n+q_L) \dots x_1(n) \dots x_1(n-q_R)]^T$$

nous obtenons les relations (1.27) et (1.28) suivantes :

$$u_m(n) = x_m(n) - \mathbf{x_1}^T(n)\mathbf{h}_m(n) \tag{1.27}$$

$$y(n) = y_{FBF}(n) - \sum_{m=2}^{M} \mathbf{g}_m^T(n)\mathbf{u}_m(n) \tag{1.28}$$

Or, l'objectif est d'ajuster $\mathbf{g}_m(n)$ pour éliminer les composantes de bruit de $y(n)$, et déterminer les filtres $\mathbf{h}_m(n)$ pour supprimer les composantes du signal désiré (signal de fuite) de la référence de bruit. En supposant que les sources de bruit et de parole sont décorrélées, cela revient à trouver $\mathbf{g}_m(n)$ et $\mathbf{h}_m(n)$ de sorte que les signaux $y(n)$ et $u_m(n)$ soient décorrélés, soit donc $E\left[y(n)u_m(n+\tau)\right] = 0, \quad m = 2, \dots, M \; \forall \, \tau$. Par conséquent, il est possible de déterminer chacun des filtres en appliquant le critère de décorrélation de Weinstein sur lequel nous reviendrons au Chapitre 2. Les deux fonctions de coût associées sont :

$$E\left[u_m(n)\mathbf{y}(n)\right] = 0 \quad \Longrightarrow \quad \mathbf{h}_m \quad m = 2, \dots, M \tag{1.29}$$

$$E\left[y(n)\mathbf{u}_m(n)\right] = 0 \quad \Longrightarrow \quad \mathbf{g}_m \quad m = 2, \dots, M \tag{1.30}$$

Celles-ci pourront, par exemple, être mises en œuvre à l'aide de l'algorithme des moindres carrés normalisé (NLMS) comme suit :

$$\begin{aligned} \mathbf{h}_m(n) &= \mathbf{h}_m(n-1) + \frac{\mu}{||\mathbf{y}(n)||^2}\mathbf{y}(n)u_m(n) \quad m = 2, \dots, M \\ \mathbf{g}_m(n) &= \mathbf{g}_m(n-1) + \frac{\mu}{||\mathbf{u}_m(n)||^2}\mathbf{u}_m(n)y(n) \quad m = 2, \dots, M \end{aligned}$$

Pour l'ensemble des algorithmes décrits dans ce paragraphe, l'adaptation sera contrôlée en pratique par l'ajout d'une Détection d'Activité Vocale (DAV) de sorte à garantir l'adaptation des filtres aux moments opportuns. En d'autres termes, l'adaptation de $\mathbf{h}_m(n)$ sera réalisée durant les instants de présence du signal de parole alors que $\mathbf{g}_m(n)$ sera mis à jour durant les instants d'inactivité vocale. En ce sens, Hoshuyama et Sugiyama agrémentent leur GSC robuste d'un bloc nommé (*Adaptation-Mode Controller*) [84, 85, 173] qui réalise cette tâche de contrôle à l'aide d'une

détection du signal de parole basée sur l'estimation du RSB calculé entre la sortie du formateur de voies fixe et les sorties de la matrice de blocage.

Enfin, Spriet et al. adoptent une approche différente pour la gestion du signal de fuite en modifiant le critère d'adaptation pour les filtres de l'annuleur de bruit multivoie de sorte à ce qu'il prenne en compte de manière distincte la distorsion introduite par la fuite du signal et la réduction de bruit [168, 170]. Le critère modifié s'écrit :

$$E\left[||Y^s_{FBF}(k,l) - \mathbf{G}^H(k,l)(\mathbf{U}^s(k,l) + \mathbf{U}^b(k,l)||^2\right] \tag{1.31}$$

Les exposants ^{s} et ^{b} désignent respectivement les composantes de signal et de bruit contenues dans le signal ainsi estampillé. Sous l'hypothèse que le bruit et la parole sont décorrélés, (1.31) devient :

$$\underbrace{E\left[||\mathbf{G}^H(k,l)\mathbf{U}^b(k,l)||^2\right]}_{Bruit} + \underbrace{E\left[||Y^s_{FBF}(k,l) - \mathbf{G}^H(k,l)\mathbf{U}^s(k,l)||^2\right]}_{Distorsion} \tag{1.32}$$

Le critère à l'origine du filtrage de Wiener peut donc être généralisé afin de permettre un compromis entre la réduction de bruit et la distorsion. Ceci est réalisé par Doclo et al. [50], où l'ajout d'un facteur $\mu \in [0, \infty[$ ajuste le niveau de distorsion et/ou de bruit résiduel. Le critère (1.32) ainsi modifié s'écrit :

$$E\left[||\mathbf{G}^H(k,l)\mathbf{U}^b(k,l)||^2\right] + \frac{1}{\mu}E\left[||Y^s_{FBF}(k,l) - \mathbf{G}^H(k,l)\mathbf{U}^s(k,l)||^2\right] \tag{1.33}$$

où la pondération par μ aura l'effet suivant :

- $\mu = 1$: le critère du minimum d'erreur quadratique moyenne est atteint ;
- $\mu > 1$: l'accent est mis sur la réduction du bruit résiduel au prix de l'augmentation de la distorsion ;
- $\mu < 1$: l'inverse du cas précédent.

Ce principe, appliqué à l'annuleur de bruit multivoie du GSC, défini le *Speech Distortion Regularized-Generalized Sidelobe Canceller* (SDR-GSC). Une implémentation de ce critère selon une approche du gradient stochastique est donnée dans [172].

1.3 Performances du GSC

Tout d'abord, dans le cadre de l'aide aux malentendants, Spriet et al. ont comparé la robustesse du GSC à celle du filtrage de Wiener multicanal (FWM) [171]. De leur étude, il ressort que le FWM est plus robuste aux erreurs de modélisation que le GSC mais souffre néanmoins des erreurs d'estimation sur les statistiques d'ordre deux. Ainsi, le FWM perd son avantage dès lors que l'environnement est nettement non-stationnaire. Si l'on synthétise à présent les résultats spécifiques à l'évaluation

du GSC issus de [20, 21, 31, 150], nous obtenons en fonction des conditions de bruit les conclusions suivantes.

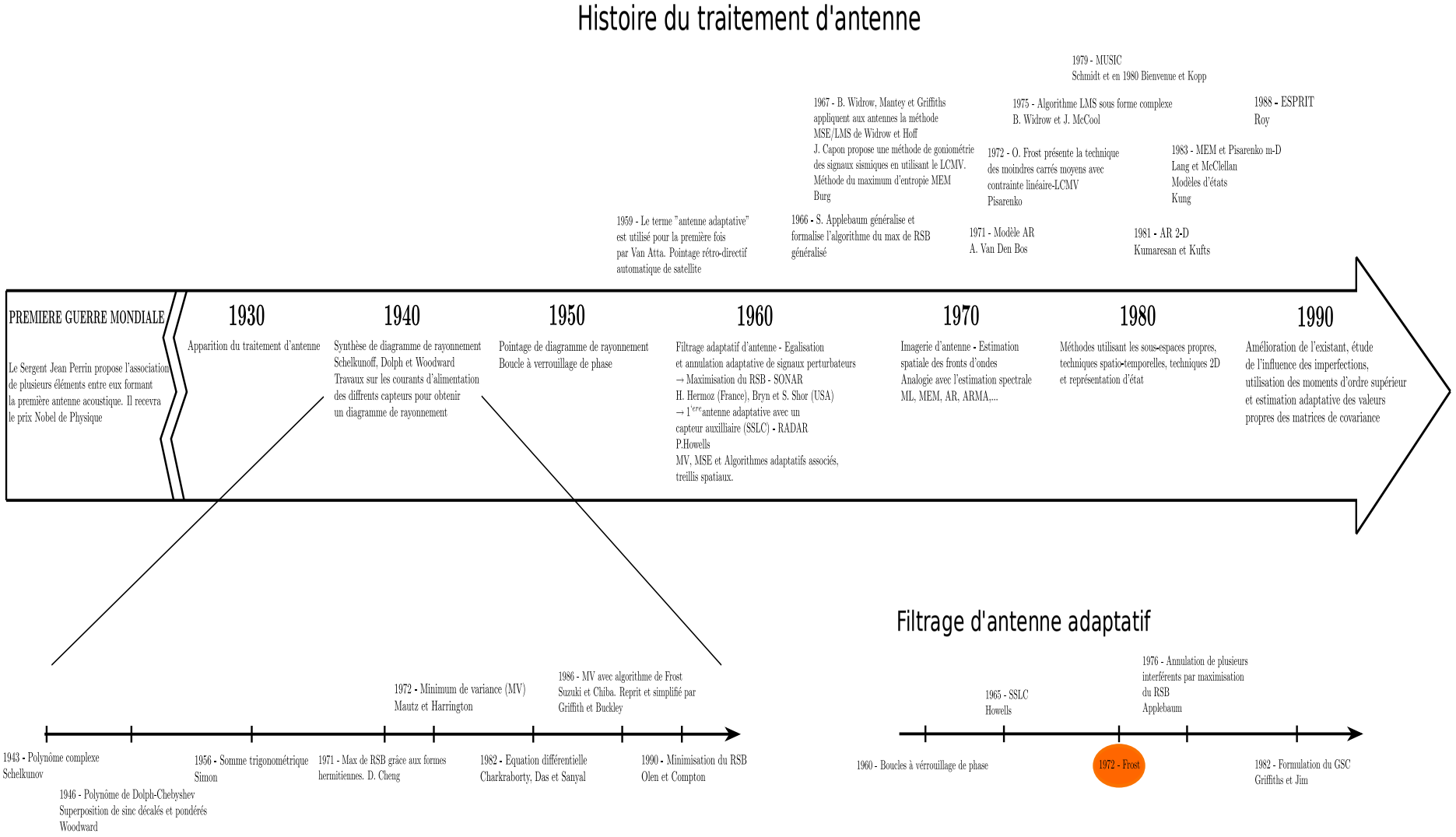
1. Lorsqu'une source interférente du même type que le signal utile (parole) est présente, les capacités du GSC à la supprimer sont indépendantes du type de champs de bruit. Les performances dépendent uniquement des contraintes imposées sur les blocs que sont le *Fixed Beamformer* (FBF) et l'ABM. Cependant, la précision des estimées des RTF est liée au type de champ de bruit, donc le champ de bruit influe quelque peu sur la suppression de l'interférent puisque l'estimation de la matrice de blocage sera imparfaite ;
2. Bruit cohérent : conditions de fonctionnement optimales où la réduction du bruit est en théorie infinie. En pratique, du fait de l'existence du bruit des microphones et de la sous-modélisation liée à des longueurs de filtres nécessairement plus courtes que les vraies FTA à estimer, ce n'est pas le cas ;
3. Bruit incohérent : la réduction de bruit est assurée exclusivement par le formateur de voies fixe (FBF). L'annuleur de bruit multivoie (*Adaptive Interference Canceller* (AIC)) est logiquement inutile pour ce type de bruit, nous sommes ramenés aux observations de Widrow sur les performances de l'annuleur de bruit adaptatif (*Adaptive Noise Canceller* (ANC)) ;
4. Bruit diffus : condition critique car la faible réduction de bruit que peut apporter l'AIC dans les basses fréquences est masquée par les performances médiocres du FBF. D'où l'utilisation d'une antenne superdirective pour pallier ce problème mais avec les défauts qui lui sont propres, cf. paragraphe 1.2.

La réduction de bruit est donc limitée par la précision de l'estimée des FTR, le champ de bruit et également par la séparation angulaire entre la source cible et l'interférent. De plus, en accord avec ce qui a été vu au paragraphe 1.2, la réverbération va diminuer les capacités de réduction de bruit du GSC de manière non négligeable à cause de la sensibilité aux erreurs de modélisation et de la troncature des FTR. Néanmoins, cette antenne reste très efficace dans le sens où elle est moins coûteuse que l'antenne de Frost et offre, grâce aux différentes contraintes que l'on peut lui ajouter, une bonne réduction du bruit et des interférents y compris lorsque l'environnement est non-stationnaire.

1.4 Synthèse du chapitre

Après avoir mis en avant les difficultés existantes dans les méthodes de traitement d'antenne, nous avons présenté au sein de ce chapitre un certain nombre de techniques développées dans le but de réduire la sensibilité des antennes aux différentes imperfections. Dans ce mémoire, les développements sont dédiés à la gestion

du défaut principal du GSC, à savoir la fuite du signal. Notre approche, décrite dans les prochains chapitres, vise à redéfinir la matrice de blocage et l'annuleur de bruit multivoie. Par conséquent, nous esquivons les problèmes de conception du frontal acoustique en utilisant une antenne à directivité contrôlée (cf. 1.1) sous l'hypothèse que la position des sources est connue et fixe. Par rapport à l'ensemble des techniques de la littérature dédiées à l'amélioration des blocs ABM et AIC, nous allons reprendre le concept de blocs complètement adaptatifs, d'une part, pour la capacité d'adaptation aux variations de l'environnement, et d'autre part, parce que la mise en œuvre adaptative de la matrice de blocage améliore la qualité du signal en sortie d'antenne [193]. Néanmoins, nous allons nous démarquer des méthodes de l'état de l'art en envisageant une implémentation récursive du GSC. Cette nouvelle architecture doit, selon nous, permettre une meilleure gestion du signal de fuite à condition d'avoir des algorithmes adaptatifs performants et adaptés à cette nouvelle structure.


Histoire du traitement d'antenne
1979 - MUSIC
Schmidt et en 1980 Bienvenue et Kopp
1967 - B. Widrow, Mantey et Griffiths
appliquent aux antennes la méthode
MSE/LMS de Widrow et Hoff
J. Capon propose une méthode de goniométrie
des signaux sismiques en utilisant le LCMV.
Méthode du maximum d'entropie MEM
Burg
1975 - Algorithme LMS sous forme complexe
B. Widrow et J. McCool
1988 - ESPRIT
Roy
1972 - O. Frost présente la technique
des moindres carrés moyens avec
contrainte linéaire-LCMV
Pisarenko
1983 - MEM et Pisarenko m-D
Lang et McClellan
Modèles d'états
Kung
1959 - Le terme "antenne adaptative"
est utilisé pour la première fois
par Van Atta. Pointage rétro-directif
automatique de satellite
1966 - S. Applebaum généralise et
formalise l'algorithme du max de RSB
généralisé
1971 - Modèle AR
A. Van Den Bos
1981 - AR 2-D
Kumaresan et Kufts
PREMIERE GUERRE MONDIALE
Le Sergent Jean Perrin propose l'association
de plusieurs éléments entre eux formant
la première antenne acoustique. Il recevra
le prix Nobel de Physique
1930
Apparition du traitement d'antenne
1940
Synthèse de diagramme de rayonnement
Schelkunoff, Dolph et Woodward
Travaux sur les courants d'alimentation
des diffrents capteurs pour obtenir
un diagramme de rayonnement
1950
Pointage de diagramme de rayonnement
Boucle à verrouillage de phase
1960
Filtrage adaptatif d'antenne - Egalisation
et annulation adaptative de signaux perturbateurs
→ Maximisation du RSB - SONAR
H. Hermoz (France), Bryn et S. Shor (USA)
→ 1ere antenne adaptative avec un
capteur auxilliaire (SSLC) - RADAR
P.Howells
MV, MSE et Algorithmes adaptatifs associés,
treillis spatiaux.
1970
Imagerie d'antenne - Estimation
spatiale des fronts d'ondes
Analogie avec l'estimation spectrale
ML, MEM, AR, ARMA,...
1980
Méthodes utilisant les sous-espaces propres,
techniques spatio-temporelles, techniques 2D
et représentation d'état
1990
Amélioration de l'existant, étude
de l'influence des imperfections,
utilisation des moments d'ordre supérieur
et estimation adaptative des valeurs
propres des matrices de covariance
1943 - Polynôme complexe
Schelkunov
1946 - Polynôme de Dolph-Chebyshev
Superposition de sinc décalés et pondérés
Woodward
1956 - Somme trigonométrique
Simon
1971 - Max de RSB grâce aux formes
hermitiennes. D. Cheng
1972 - Minimum de variance (MV)
Mautz et Harrington
1982 - Equation différentielle
Charkraborty, Das et Sanyal
1986 - MV avec algorithme de Frost
Suzuki et Chiba. Reprit et simplifié par
Griffith et Buckley
1990 - Minimisation du RSB
Olen et Compton
Filtrage d'antenne adaptatif
1960 - Boucles à vérrouillage de phase
1965 - SSLC
Howells
1972 - Frost
1976 - Annulation de plusieurs
interférents par maximisation
du RSB
Applebaum
1982 - Formulation du GSC
Griffiths et Jim

Chapitre 2

Vers la définition d'une cellule élémentaire

2.1 Constat d'une analogie 54
2.2 Annulation de bruit adaptative robuste à la diaphonie 56
2.3 Modèle de mélange, conditions de stabilité et positions des sources 60
2.3.1 Modélisation des canaux acoustiques 60
2.3.2 Simplification du modèle de mélange 63
2.3.3 Notions de stabilité 66
2.3.4 Influence de l'hypothèse de causalité sur la position des sources et les performances 68
2.4 Minimisation de l'EQM ou décorrélation ? 74
2.4.1 Minimisation de l'erreur quadratique moyenne 75
2.4.2 Illustration sur un cas simple 77
2.5 Synthèse du chapitre 81

Dans le chapitre précédent, nous avons présenté les différentes méthodes existantes pour pallier le défaut majeur du GSC, à savoir la fuite du signal désiré dans l'étage d'estimation du bruit et des interférents. Ce phénomène limitatif lié aux erreurs de pointage inévitables dans le contexte de la prise de son en environnement réel est la cause des plus fortes dégradations observées en sortie de cette antenne. Ce caractère inéluctable est pour nous un postulat qui s'explique par :

1. la difficulté d'estimer la position exacte des sources dans les milieux réverbérants pour lesquels la multiplicité des chemins de propagation source-capteur entre une source et un capteur donné rend difficile la décision sur le positionnement de celle-ci à partir des seules observations captées par les microphones,
2. la difficulté de suivre précisément la variabilité des paramètres statistiques due aux mouvements des sources et/ou à l'apparition soudaine de nouvelles sources,
3. les approximations et hypothèses faites pour le calcul des différents estimateurs mis en œuvre dans les algorithmes de localisation ajoutent également une incertitude et une erreur non négligeable sur la position estimée des sources.

Partant de ce postulat, nous choisissons de travailler sur une approche indirecte ayant pour but de soustraire ce signal de fuite à l'estimée du bruit faite par l'ANC. Il s'agit donc de traiter le pire cas, c'est-à-dire admettre qu'il est impossible d'obtenir une référence de bruit parfaite en sortie de la matrice de blocage.

2.1 Constat d'une analogie

Le schéma de la Figure 2.1 présente le GSC décliné dans sa version bicapteur et sur laquelle il est aisé de reconnaître la structure de l'ANC classique. Nous identifions une voie primaire sensée contenir principalement le signal désiré après pointage et sommation des signaux (Σ_1) et une voie de référence comportant le filtre $W(z)$, dont le rôle est de fournir une estimée du bruit perturbateur (bruit ambiant et interférents) présent sur la voie primaire et contenu dans le signal $X_p(z)$. Le signal $X_r(z)$ est obtenu par soustraction des signaux, ce qui correspond à l'hypothèse simple où l'on suppose une propagation des ondes selon un retard pur. En ce sens, dans le cas d'une propagation supposée en champ lointain, l'incertitude $\delta\varphi$ sur l'angle d'arrivée φ de la source d'intérêt influe sur le retard de pointage Δ, que l'on peut mettre sous la forme $\Delta = \frac{d \cdot sin(\varphi+\delta\varphi)}{c} \cdot f_s$. Si l'on suppose que cette erreur de pointage est relativement faible, alors nous pouvons faire les approximations suivantes : $sin(\delta\varphi) = \delta\varphi + o(\delta\varphi^3)$ et $cos(\delta\varphi) = 1 + o(\delta\varphi^2)$. Ainsi, le retard de pointage s'écrira $\Delta = \frac{d \cdot sin(\varphi)}{c} \cdot f_s + \frac{d \cdot \delta\varphi \cdot cos(\varphi)}{c} \cdot f_s = \delta + \Phi$. En conséquence, en prenant pour référence le point milieu entre les microphones et appartenant au plan médian de

cette antenne, les retards de pointage valent $D_1 = -\frac{\Delta}{2}$ et $D_2 = \frac{\Delta}{2}$. Compte-tenu des imperfections du pointage, les signaux en entrée de l'ANC s'écrivent :

$$X_p(z) = \frac{S(z)}{2}\left[z^{\frac{\Phi}{2}} + z^{-\frac{\Phi}{2}}\right] + \frac{1}{2}\left[B_1(z)z^{\frac{\Delta}{2}} + B_2(z)z^{-\frac{\Delta}{2}}\right] \quad (2.1)$$

$$X_r(z) = \frac{S(z)}{2}\left[z^{\frac{\Phi}{2}} - z^{-\frac{\Phi}{2}}\right] + \frac{1}{2}\left[B_1(z)z^{\frac{\Delta}{2}} - B_2(z)z^{-\frac{\Delta}{2}}\right] \quad (2.2)$$

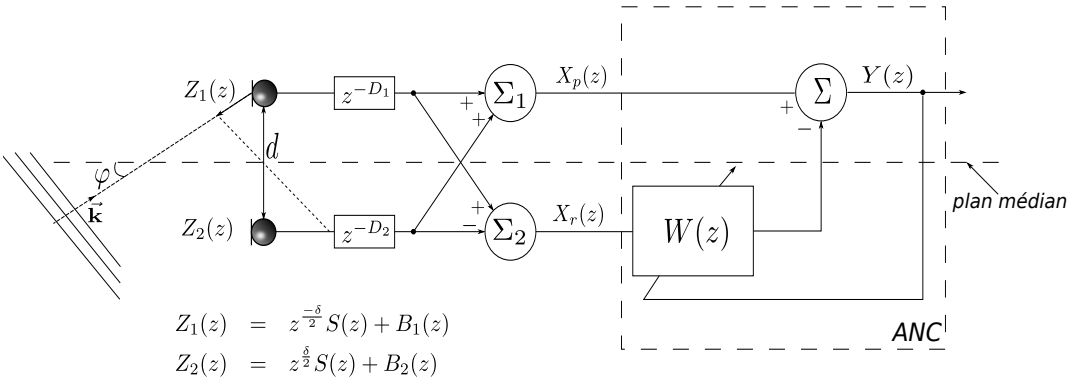


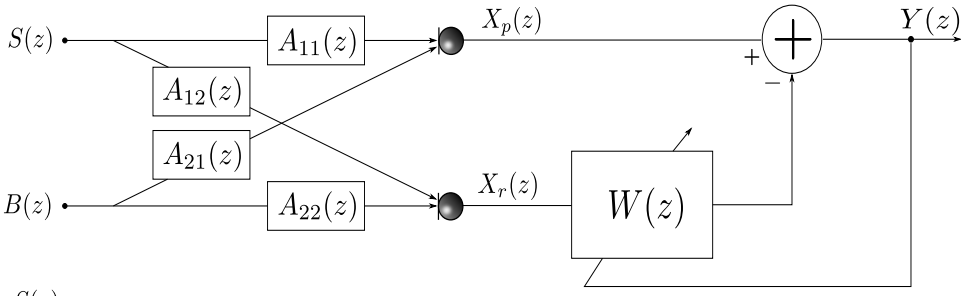
Figure 2.1 – GSC bicapteur. L'opérateur Σ_1 correspond à la sortie de l'antenne Retard-Somme, Σ_2 à la sortie de la matrice de blocage.

Ainsi, contrairement à ce qui est attendu en sortie de la matrice de blocage (Σ_2), des composantes du signal utile sont présentes sur la voie de référence dès lors que $\delta\varphi \neq 0$. Cette situation mise en exergue dans le cas d'un système bicapteur correspond exactement au phénomène dit de diaphonie ou *crosstalk* avec lequel nous établissons le parallèle et qui apparaît dès lors que la prise de son est réalisée par l'association de deux microphones sans barrière acoustique permettant d'isoler le microphone de référence du signal d'intérêt, Figure 2.2. Dans le cas contraire, on parle d'un système d'annulation de bruit avec référence de bruit seul et qui correspond au cas où $A_{12}(z) = 0$ ou bien d'un pointage idéal dans le cas du GSC. En effet, depuis le développement des méthodes adaptatives, l'enjeu majeur et déterminant pour l'efficacité des techniques bivoies en environnement réel réside dans notre capacité à fournir au système une référence de bruit fiable. Cette fiabilité s'évalue suivant le degré de corrélation qui existe entre le bruit perturbateur du signal de parole et celui que l'on prend pour image grâce au second microphone, voir Figure 2.2. Il est aisé de comprendre que la similitude entre le bruit de référence, ici $(\frac{1}{2}\left[B_1(z)z^{\frac{\Delta}{2}} - B_2(z)z^{-\frac{\Delta}{2}}\right])$ sur la Figure 2.1 et le bruit perturbateur $(\frac{1}{2}\left[B_1(z)z^{\frac{\Delta}{2}} + B_2(z)z^{-\frac{\Delta}{2}}\right])$ sera d'autant plus grande que les microphones seront placés suffisamment proche l'un de l'autre. De surcroît, cette proximité est aussi rendue indispensable pour limiter la longueur du filtre adaptatif $W(z)$ calculé au cours du temps, réduisant ainsi les désajustement qui produisent notamment le phénomène d'écho. Par conséquent, au même titre que

les erreurs de pointage engendrent la fuite du signal utile dans la voie de référence du GSC, la limite pratique sur la distance inter-microphone assurant d'une part une bonne corrélation des signaux de bruit sur le frontal acoustique et d'autre part le respect d'éventuelles contraintes imposées sur les dimensions du système, donne naissance à la diaphonie. C'est ce signal intrusif qui est à l'origine des principales dégradations sur le signal $Y(z)$ produit en sortie du système de réduction de bruit. L'analogie avec les systèmes de réduction de bruit bicapteur nous conduit donc à tenter de résoudre dans un premier temps le problème de la diaphonie pour ces derniers et ensuite étendre la méthode au GSC. De plus, nous utiliserons sans aucune ambiguité les termes signal de fuite et diaphonie puisqu'ils décrivent le même phénomène.

2.2 Annulation de bruit adaptative robuste à la diaphonie

L'influence de composante du signal utile dans la voie de référence est un problème connu et largement abordé selon différentes approches. Une première analyse sur les performances de l'ANC est donnée par Widrow dans [189] afin de justifier de l'intérêt de cette méthode dans ce type de situation. Il en ressort que les performances vont dépendre, logiquement, de la quantité de signal de fuite. En d'autres termes, les performances se dégraderont à mesure que le RSB sur la voie de référence augmente et ce, à RSB fixe sur la voie primaire. Ce phénomène est connu sous le nom "d'inversion de puissance". Pour mieux s'en convaincre, nous donnons une illustration du phénomène à partir de la Figure 2.2.



$S(z)$: *signal de parole*
$B(z)$: *signal perturbateur*
$X_p(z)$: *signal microphonique sur la voie primaire comprenant le signal utile plus le bruit*
$X_r(z)$: *signal microphonique sur la voie de référence comprenant le bruit plus le signal de fuite*
$A_{ij}(z)$: *canaux acoustiques de longueurs inconnues, variables dans le temps*
$W(z)$: *filtre d'identification de longueur* N
$Y(z)$: *signal de parole estimé*

Figure 2.2 – Schéma de l'annuleur de bruit adaptatif classique pour l'estimation d'un filtre RIF.

Le système de prise de son bicapteur enregistre deux signaux, $X_p(z)$ et $X_r(z)$ respectivement. Chacun de ces signaux étant le résultat de la somme des convolutions des signaux $S(z)$ et $B(z)$ avec le canal acoustique associé au chemin le reliant à chacun des microphones. Le filtre $W(z)$ est de longueur N, dont les coefficients $\{w(i)\}_{i=0,\ldots,N-1}$ sont adaptés dans le temps, et dont le but est de générer une estimée du bruit $A_{21}(z)B(z)$. Cette estimée nécessite, dans le cas où l'on cherche à minimiser la diaphonie résiduelle au prix d'une distorsion sur le signal utile, que le filtre $W(z)$ apprenne la réponse impulsionnelle $A_{21}(z)/A_{22}(z)$. Malheureusement, dans cette structure rien n'est mis en œuvre pour estimer le signal de diaphonie provenant de la convolution entre $S(z)$ et $A_{12}(z)$. Or la présence de ce signal fournit une solution optimale différente de celle du cas traité couramment [69], c'est-à-dire lorsque $A_{12}(z)$ est considérée comme nulle. En effet, en faisant l'hypothèse que $A_{11}(z) = A_{22}(z) = 1$ et en supposant que $s(n)$ et $b(n)$ ne sont pas corrélés, on montre que la solution optimale au sens de Wiener prenant en compte la fuite du signal de parole au travers $A_{12}(z)$ devient :

$$W^{opt}(z) = \frac{A_{21}(z)\Phi_{BB}(z) + A_{12}(z^{-1})\Phi_{SS}(z)}{\Phi_{BB}(z) + |A_{12}(z)|^2 \Phi_{SS}(z)} \tag{2.3}$$

où $\Phi_{BB}(z)$ et $\Phi_{SS}(z)$ sont les spectres complexes des signaux $B(z)$ et $S(z)$ dans le domaine de la transformée en Z. Cela prouve d'une part que l'annulation du bruit ne peut se faire totalement, puisque $W^{opt}(z) \neq A_{21}(z)$. D'autre part, en considérant que la diaphonie prédomine, alors l'équation (2.3) se résume à :

$$W^{opt}(z) \approx \frac{A_{12}(z^{-1})\Phi_{SS}(z)}{|A_{12}(z)|^2 \Phi_{SS}(z)} \approx \frac{1}{A_{12}(z)} \tag{2.4}$$

Dans ce cas, le signal en sortie du système, $Y(z)$, issu de la soustraction entre $X_p(z)$ et $W^{opt}(z)X_r(z)$ peut s'écrire :

$$\begin{aligned} Y(z) &= X_p(z) - W^{opt}(z)X_r(z) \\ &\approx S(z) + A_{21}(z)B(z) - \frac{1}{A_{12}(z)}B(z) - \frac{A_{12}(z)}{A_{12}(z)}S(z) \\ &\approx \left(A_{21}(z) - \frac{1}{A_{12}(z)}\right) B(z) \end{aligned} \tag{2.5}$$

La solution optimale conduit donc à l'annulation complète du signal. Un moyen simple de pallier ce problème est donc de stopper l'adaptation lorsque le signal de parole est présent. Une autre solution consiste à concevoir des structures prenant en compte de manière explicite la fuite du signal utile à travers l'utilisation d'un second filtre adaptatif. Le plus souvent, les solutions mises en œuvre emploient la combinaison de ces méthodes et se distinguent selon que les filtres adaptatifs soient disposés

dans une configuration directe ou récursive. Celles-ci sont décrites à la Figure 2.3 et extraites de [82].

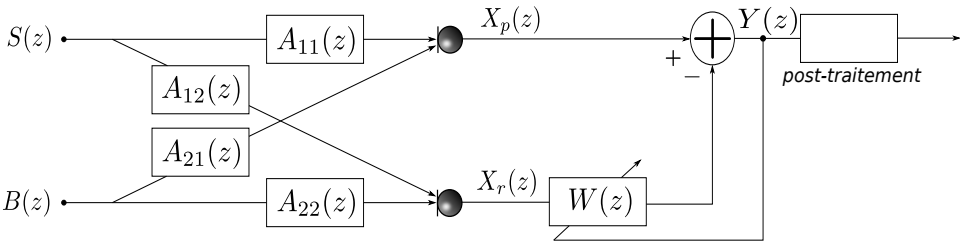
La première solution (a) est le pendant de l'utilisation du postfiltre dans le cas des antennes acoustiques, censé réduire le bruit diffus résiduel, réhausser le signal et améliorer la qualité du signal en termes de distorsion. Les solutions (b) et (c) sont très proches du point de vue conceptuel puisque l'idée est d'estimer, dans un premier temps, la diaphonie pour ensuite la soustraire au signal présent sur la voie de référence, et donc améliorer l'estimée du bruit. Cela est réalisé par l'ajout du filtre $W_{12}(z)$ qui prouve de manière explicite la capacité de ces méthodes à soustraire la diaphonie à la référence de bruit créée par le filtre $W_{21}(z)$ et permettant ainsi de réduire les distorsions sur le signal de parole estimé $Y(z)$. L'objectif de ce second filtre est donc d'estimer la réponse du canal acoustique $A_{12}(z)$ (en supposant $A_{11}(z) = A_{22}(z) = 1$) de manière adaptative au même titre que cela est fait pour le bruit au travers du filtre $W_{21}(z)$. Notons que la solution (c) est à la base du GSC robuste proposé par Hoshuyama et décrit au Chapitre 1. Enfin, la dernière solution donnée en (d) se distingue nettement des précédentes sur deux points.

1. Il s'agit de la seule structure capable de fournir deux signaux en sortie correspondant à l'estimée de la source utile et du bruit au facteur $A_{11}(z)$ et $A_{22}(z)$ près, ce qui la rend compatible avec les méthodes de séparation de sources,
2. Les filtres adaptatifs sont placés selon une configuration récursive. Par conséquent, il existe un couplage, une interdépendance forte entre les signaux de sortie qui n'existe pas dans sa version équivalente où les filtres sont placés de manière directe. En effet, chaque sortie du système de séparation est une combinaison linéaire d'une observation et de l'autre sortie. Ce point sera discuté en 2.4.

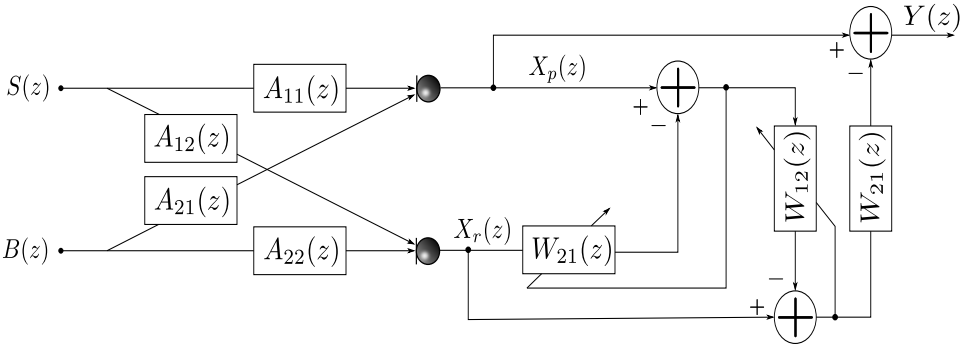
Pour répondre à l'objectif que nous nous sommes fixés, en l'occurence, définir une nouvelle architecture pour répondre au problème de la fuite du signal dans le GSC, notre choix s'est porté sur l'architecture (d) pour les raisons suivantes :

- l'absence de filtrage parasite supplémentaire pour obtenir les estimées,
- la réduction de l'ordre des filtres utilisés pour l'estimation des canaux acoustiques. Nous faisons allusion, ici, à la différence entre un filtre de type RIF et un filtre à Réponse Impulsionnelle Infinie (RII), ce qui sera discuté en 2.3.1,
- le potentiel et le caractère novateur que représente cette structure dans son utilisation pour le GSC. Ce point sera développé au Chapitre 6.
- la volonté d'appréhender et d'approfondir plus en avant les points forts et les points faibles de cette structure en regard de la littérature.

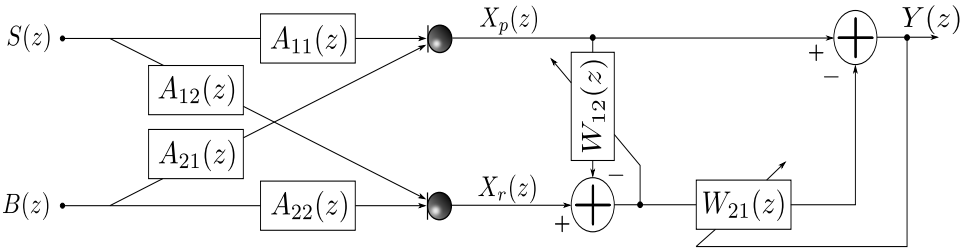
Par ailleurs, nous mènerons l'analyse et l'évaluation de cette architecture uniquement dans le cas d'un mélange de deux sources ponctuelles, comme cela est



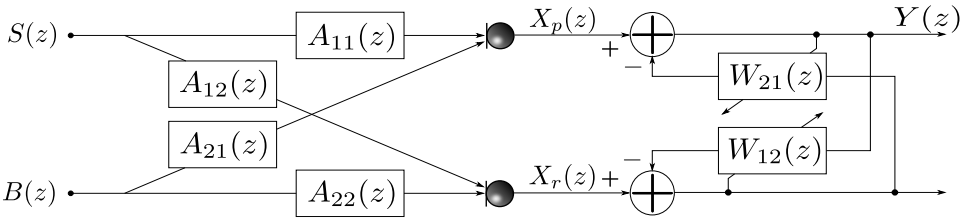
(a) Annuleur de bruit avec post-traitement.



(b) Annuleur de bruit avec pré-traitement.



(c) Annuleur de bruit fait d'une mise en cascade de l'annuleur de bruit classique.



(d) Filtres duaux.

Figure 2.3 – Les différentes architectures d'annuleur de bruit adaptatif robustes à la diaphonie.

sous-entendu sur les modèles de mélange donnés aux Figures 2.2 et 2.3 et discuté au paragraphe suivant. Ce choix est motivé par l'application visée. En effet, contrairement aux méthodes de séparation de source utilisant la même structure pour des mélanges comportant plus de deux sources, notre choix se justifie simplement par le fait que nous envisageons l'utilisation des filtres duaux derrière un frontal acoustique. Ainsi, là où la séparation de source travaille directement avec le mélange de N sources, ici, le pointage réalisé par le frontal acoustique ramènera l'étude à un ou N mélange(s) de sources prises deux à deux.

2.3 Modèle de mélange, conditions de stabilité et positions des sources

2.3.1 Modélisation des canaux acoustiques

Dans cette partie, nous discutons du choix possible entre deux modèles, à savoir tout-pôles ou tout-zéros pour modéliser les différents chemins acoustiques du modèle de mélange et les implications sur les résultats et les analyses qui seront faites dans la suite de ce document. Physiquement, une réponse impulsionnelle de salle est une fonction temporelle à énergie finie qui se caractérise par l'amplitude et la phase de sa fonction de transfert, $A(\omega) = |A(\omega)| \exp(j\Phi(\omega))$. Elle est par nature stable et causale et varie en fonction de la température, de la position des sources, des capteurs ainsi que des mouvements des locuteurs ou encore des nouvelles conditions dues à l'ouverture d'une porte, d'une fenêtre, etc. Sa fonction de transfert s'écrit de façon générale et selon la transformée en Z, sous la forme de la fonction rationnelle [134] suivante :

$$A_{zp}(z) = \frac{\prod_{i=1}^{K}(1 - g_i z^{-1}) \prod_{i=K+1}^{N}(1 - g_i z^{-1})}{\prod_{j=1}^{M}(1 - p_j z^{-1})} \tag{2.6}$$

ce qui correspond à un modèle zéros-pôles pour lequel rien ne garantit qu'il soit à phase minimale, propriété indispensable assurant que l'inverse de ce système soit stable. En raison de la stabilité intrinsèque de (2.6), tous les pôles sont à l'intérieur du cercle unité ($\mathcal{C}_1$), soit $|p_j| < 1, \quad j = 1, \ldots, M$. Par contre, les zéros ont une position quelconque dans le plan complexe. Le premier facteur du numérateur de $A_{zp}(z^{-1})$ concerne les K zéros situés dans $\mathcal{C}_1$, ($|g_i| < 1, \quad i = 1, \ldots, K$) tandis que le second facteur se rapporte à $N - K$ zéros placés en dehors de $\mathcal{C}_1$, ($|g_i| > 1, \quad i = K+1, \ldots, N$). Cette décomposition du numérateur nous fait voir un aspect important et délicat lié au problème d'inversion du canal acoustique pour l'annulation de l'effet de salle sur les signaux de parole enregistrés. En effet, le caractère à phase maximale du second facteur rend l'inversion de cette réponse instable et non-causale.

En pratique, seul le cas d'une prise de son dans un environnement fermé classique sans obstacles et avec une distance source-micro inférieure à 20 cm [136] vérifie, éventuellement, la propriété de chemins acoustiques à phase minimale. Quoi qu'il en soit des problèmes de stabilité de l'inverse du canal acoustique, nous devons faire le choix d'un modèle plausible décrivant au plus près celui-ci afin d'évaluer et développer nos algorithmes de débruitage dans l'optique d'une mise en œuvre en contexte réel. Divers modèles ont été développés et analysés [24, 77, 134, 136], et couvrent les trois classes de la modélisation que sont les modèles tout-pôles, tout-zéros et zéros-pôles. Les modèles tout-pôles et zéros-pôles conduisent à l'utilisation de filtres RII et présentent l'avantage de modéliser parfaitement les résonances de la salle qui varient peu avec la position de la source et du récepteur. En l'occurence, dans [77] les auteurs mettent à profit cette "invariablité" dans le calcul du modèle zéros-pôles proposé. D'autre part, la modélisation tout-pôles requiert un nombre réduit de coefficients, ce qui assure une bonne vitesse de convergence des filtres adaptatifs et un faible retard lié au traitement. A l'inverse, le modèle tout-zéros implémenté grâce à des filtres de type RIF, sera capable de modéliser parfaitement les zéros de la réponse impulsionnelle associés au temps de propagation de l'onde directe et les anti-résonnances. Néanmoins, ce modèle est logiquement très sensible à la position des sources et des microphones, et il faudra un nombre important de coefficients dès lors que le temps de réverbération est élevé. Typiquement, 4000 coefficients seront nécessaires pour modéliser une réponse impulsionnelle de salle ayant un temps de réverbération de 500 ms à une fréquence de 8 kHz. L'ordre élevé du filtre implique une faible convergence des filtres adaptatifs et une taille mémoire importante. Cependant, dans le cadre des applications qui nous intéressent, ce modèle reste celui le plus souvent utilisé du fait de sa simplicité : le filtre RIF est parfaitement maîtrisé, stable, moins sensible aux erreurs de quantification et donc préféré dans la pratique pour les techniques de réduction de bruit. D'autre part, il apporte une précision suffisante dans le contexte auquel nous nous restreignons pour les développements bicapteurs, cf. 2.3.2.

Modèle tout-zéros du canal acoustique

L'objectif est d'obtenir une représentation du canal acoustique sous la forme :

$$A_z(z) = \prod_{i=1}^{K}(1 - g_i z^{-1}) \prod_{i=K+1}^{N} (1 - g_i z^{-1}) \tag{2.7}$$

et dériver ensuite un "équivalent" à minimum de phase [134, 135] de $A_{zp}(z)$. L'obtention de ce modèle tout-zéros $A_z(z)$ résulte de la propriété de décomposition de tout système zéros-pôles à phase non minimale [136, 144] en un produit d'un filtre à

phase minimale par un filtre passe-tout.

$$A_{zp}(z) = A_{min}(z)A_{tout}(z) \tag{2.8}$$

Pour s'en convaincre il suffit de décomposer la phase du système initial (2.6) en la somme d'une composante à phase minimale et d'une composante représentant la déviation, $\Phi(\omega) = \Phi_{min}(\omega)+\Phi_d(\omega)$. Ce qui conduit à l'expression (2.8) avec $|A_{tout}(\omega)| = 1 \ \forall \ \omega$. Notons que $A_{zp}(z)$ sera à minimum de phase uniquement si $\Phi_d(\omega) = 0$, ce qui implique que $A_{tout}(\omega) = 1$. Une autre approche le justifiant s'obtient en écrivant $A_{zp}(z)$ sous la forme :

$$A_{zp}(z) = \frac{G_1(z)G_2(z)}{P(z)} \tag{2.9}$$

où $G_1(z)$ (respectivement $G_2(z)$) décrit le terme ayant toutes ses racines à l'intérieur (respectivement à l'extérieur) de $\mathcal{C}_1$. Alors $G_2(z^{-1})$ aura, au contraire, ses racines dans $\mathcal{C}_1$ et on pourra définir [144] :

$$A_{min}(z) = \frac{G_1(z)G_2(z^{-1})}{P(z)} \tag{2.10}$$

et

$$A_{tout}(z) = \frac{G_2(z)}{G_2(z^{-1})} \tag{2.11}$$

tel que $A_{zp}(z) = A_{min}(z)A_{tout}(z)$. Ce même raisonnement appliqué au modèle tout-zéros (2.7) conduit à :

$$A_z(z) = \underbrace{\prod_{i=1}^{K}(1 - g_i z^{-1}) \prod_{i=K+1}^{N} (1 - g_i z)}_{A_{min}(z)} \underbrace{\frac{\prod_{i=K+1}^{N}(1 - g_i z^{-1})}{\prod_{i=K+1}^{N}(1 - g_i z)}}_{A_{tout}(z)} \tag{2.12}$$

Le modèle tout-zéros ainsi obtenu en (2.12) est à la base des techniques de déconvolution appliquées pour la déréverbération, même si celles-ci ne sont pas efficaces en pratique [123]. En effet, il sera l'équivalent à minimum de phase de $A_{zp}(z)$ uniquement dans le cas où $A_{tout}(z) = 1$, comme cela a été vu plus haut pour la décomposition de $A_{zp}(z)$. Or, ce cas n'est rencontré potentiellement que dans des contextes de prises de son très particuliers et, de manière générale, la réverbération en partie concentrée dans la partie passe-tout mettra en défaut cette hypothèse et rendra de ce fait les techniques d'inversion caduques. Néanmoins, il reste pertinent et utile à l'étude de la stabilité de la structure récursive, voir 2.3.3, ainsi qu'à la restriction de l'ensemble des solutions de certaines méthodes [188]. Par conséquent, nous supposerons que la réponse impulsionnelle du canal acoustique peut être approximée par un filtre RIF pour tous les développements à venir et exposés dans ce mémoire.

2.3.2 Simplification du modèle de mélange

En général, pour des besoins de commodité et de simplicité des développements, le modèle de mélange des sources est, dans la littérature, très souvent simplifié. Néanmoins et comme nous allons l'exposer ci-après, cela n'enlève rien à la pertinence des résultats obtenus qui, à défaut d'être généralisables à tous les types de problème ou modèle de propagation acoustique, répondent à un contexte particulier. En effet, nous noterons que pour toutes les structures décrites à la Figure 2.3, l'intérêt est porté sur l'estimation des chemins croisés et non pas des chemins directs. La justification de cette observation tient dans le fait que pour des applications de débruitage de la parole et à condition que le milieu ne soit pas trop réverbérant ($T_{60} \leq 500ms$), il n'est en général pas gênant de conserver la réverbération introduite par la propagation. Si par contre l'impact de cette réverbération sur la qualité du signal en sortie est significative, il est alors possible d'utiliser la structure à filtres duaux agrémentée de filtres "directs", comme cela est présenté à la Figure 2.4. Dans ce cas, la relation

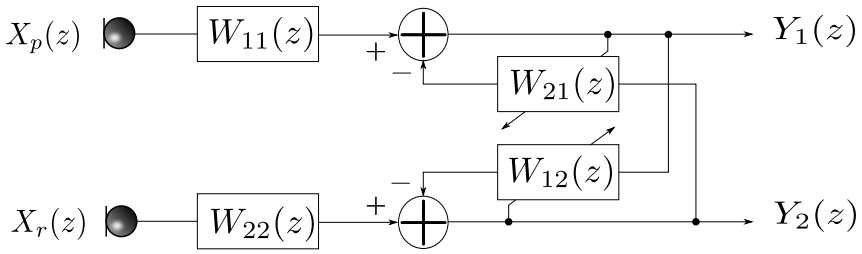


Figure 2.4 – Structure récursive à quatre filtres pour la séparation d'un mélange convolutif de deux sources.

entrée-sortie s'écrit :

$$\begin{aligned} Y_1(z) &= W_{11}(z)X_p(z) - W_{21}(z)Y_2(z) \\ Y_2(z) &= W_{22}(z)X_r(z) - W_{12}(z)Y_1(z) \end{aligned} \tag{2.13}$$

Ce système d'équations donne en fonction des observations :

$$\begin{cases} Y_1(z) = \frac{W_{11}(z)X_p(z) - W_{21}(z)W_{22}(z)X_r(z)}{1 - W_{21}(z)W_{12}(z)} \\ \\ Y_2(z) = \frac{W_{22}(z)X_r(z) - W_{12}(z)W_{11}(z)X_p(z)}{1 - W_{12}(z)W_{21}(z)} \end{cases} \tag{2.14}$$

En tenant compte du modèle de mélange des signaux sources, le système (2.14) devient :

$$\begin{aligned} Y_1(z) &= \frac{[W_{11}(z)A_{11}(z) - W_{21}(z)W_{22}(z)A_{12}(z)]S(z) + [W_{11}(z)A_{21}(z) - W_{21}(z)W_{22}(z)A_{22}(z)]B(z)}{1 - W_{21}(z)W_{12}(z)} \\ Y_2(z) &= \frac{[W_{22}(z)A_{12}(z) - W_{12}(z)W_{11}(z)A_{11}(z)]S(z) + [W_{22}(z)A_{22}(z) - W_{12}(z)W_{11}(z)A_{21}(z)]B(z)}{1 - W_{12}(z)W_{21}(z)} \end{aligned} \tag{2.15}$$

Hormis les éventuels problèmes de solutions "fantômes" [70], de permutations et de mise à l'échelle, la solution permettant d'obtenir $s(n)$ sur la sortie $y_1(n)$ et $b(n)$ sur la sortie $y_2(n)$ est donnée par :

$$\begin{aligned} W_{11}(z) &= \frac{1}{A_{11}(z)} \quad W_{12}(z) = \frac{A_{12}(z)}{A_{22}(z)} \\ W_{22}(z) &= \frac{1}{A_{22}(z)} \quad W_{21}(z) = \frac{A_{21}(z)}{A_{11}(z)} \end{aligned} \tag{2.16}$$

Ainsi, à travers les filtres $W_{11}(z)$ et $W_{22}(z)$, l'effet de salle contenu dans les chemins directs $A_{11}(z)$ et $A_{22}(z)$ peut, en théorie, être supprimé. Cependant, comme nous l'avons souligné auparavant, cette opération reste délicate du fait de la présence de la composante à phase maximale dans l'expression (2.12), ce qui peut conduire à des solutions instables pour les filtres $W_{11}(z)$ et $W_{22}(z)$. Néanmoins, il existe un contexte particulier pour lequel cette inversion n'est pas nécessaire. Cela concerne le cas où l'on modélise les chemins directs par une impulsion de Dirac, soit $A_{ii}(z) = 1$ impliquant par la même que $W_{ii}(z) = 1$. Par conséquent, le modèle de mélange se réduit [176] selon le schéma de la Figure 2.5. Cette simplification du modèle de mélange correspond au cas d'une propagation en champ proche. Cette hypothèse sera considérée comme vérifiée dès lors que la distance sources-antenne acoustique n'est pas d'un ordre de grandeur très significatif par rapport à la dimension du frontal acoustique. Pour être plus complet et surtout plus précis sur la possibilité de

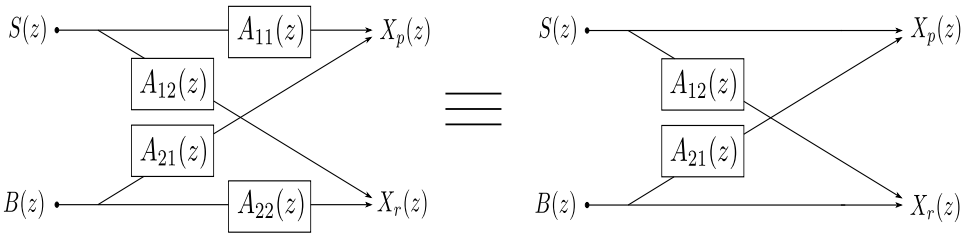


Figure 2.5 – Modèle de mélange simplifié compatible avec une propagation en champ proche.

cette simplification, il faut par ailleurs faire l'hypothèse d'une propagation homogène, isotrope et sans écho. Cette configuration des sources par rapport aux microphones correspond finalement à une situation courante de prise de son pour laquelle les grandeurs caractéristiques sont liées par la relation :

$$r < \frac{d_{tot}^2 f_s}{c} \tag{2.17}$$

Avec r la distance entre la source et le centre de l'antenne, $d_{tot} = d_{M_{mic}-1} - d_0$ la longueur totale de l'antenne, f_s la fréquence d'échantillonnage et $c = 340\ m.s^{-1}$ la célérité du son dans l'air. Pour mieux se convaincre du caractère commun de ce contexte de prise de son, le tableau 2.1 donne pour différentes valeurs de la distance

inter-microphone ($M_{mic} = 2$), la profondeur de champ correspondant à une propagation en champ proche pour deux fréquences d'échantillonnage usuelles du traitement de la parole. Un exemple typique où la courbure de l'onde acoustique ne peut être considérée comme plane est la prise de son en mode combiné ou en mains-libres avec un téléphone portable, où la distance maximale bouche-microphone ne dépasse pas 80 cm. Pour ce type de propagation, la modélisation des chemins croisés par des filtres RIF semble particulièrement adaptée [117, 153].
Dans le cas de la prise de son distante considérée dans ce travail de thèse, il est clair que le modèle simplifié décrit précédemment n'a plus de réalité physique. Il sert uniquement à faciliter les analyses. Sachant que nous considérons la structure de séparation à filtres duaux donnée en 2.3 (d), la réverbération n'est pas traitée mais supposée en partie supprimée par le frontal acoustique d'une part puis par l'ajout d'un postfiltre d'autre part. En effet, contrairement aux méthodes spectrales qui agissent sur le contenu fréquentiel et sont donc peu sensibles à la réverbération, il s'agit ici d'une méthode spatiale qui consiste à estimer les chemins acoustiques relatifs entre les sources et les micros. Les estimées obtenues en sortie de traitement pour les structures bicapteurs dans la suite du document sont donc celles obtenues à un filtrage près. Ceci rejoint en définitive le modèle de mélange simplifié qui n'a de sens physique que dans le cas où les sources sont situées suffisamment proches des microphones, mais pour lequel nous supposons :

$$S(z) = A_{11}^0(z)S^0(z) \quad B(z) = A_{22}^0(z)B^0(z)$$

$$A_{12}(z) = \frac{A_{12}^0(z)}{A_{22}^0(z)} \qquad A_{21}(z) = \frac{A_{21}^0(z)}{A_{11}^0(z)}$$

En ce sens, on parlera de modèle de mélange équivalent. Ainsi, la simplification du modèle de mélange est une manipulation avantageuse non réductrice puisqu'elle permet, quel que soit le type de propagation envisagée, d'approcher une scène sonore sous certaines hypothèses. Nous donnerons néanmoins au sein du Chapitre 6, section 6.3 une analyse de l'impact des chemins directs sur les estimées des filtres produites par la structure à filtres duaux.

Remarque 1 : *L'estimation des sources à un filtre près dans le cas des mélanges convolutifs peut influencer fortement l'intelligibilité du message. Cette remarque importante est donnée dans [79], où l'auteur dresse le parallèle avec l'estimation d'un processus AR. Accepter de retrouver ce processus à un filtre près est équivalent à dire que l'on restitue non pas le processus AR mais le processus d'innovation qui lui est associé. Par extension au cas de la parole et avec une certaine approximation, cela conduit à accepter l'excitation périodique de la glotte (pour les sons voisés) ou un bruit blanc (pour les sons non-voisés) comme sources représentatives de la parole,*

Distance inter-microphone	$f_s = 8000\ Hz$	$f_s = 16000\ Hz$
$5\ cm$	$5,9\ cm$	$11,8\ cm$
$10\ cm$	$23,5\ cm$	$47\ cm$
$15\ cm$	$52,9\ cm$	$1\ m$
$50\ cm$	$5,9\ m$	$11,8\ m$
$1\ m$	$23,5\ m$	$47\ m$

Table 2.1 – Calcul de la distance limite source-microphones pour l'approximation champ proche pour deux valeurs de fréquence caractéristiques. Le frontal est composé de deux microphones ($M = 2$).

sans considérer l'influence du conduit vocal.

2.3.3 Notions de stabilité

Lorsque des techniques adaptatives sont utilisées pour le traitement de la parole, un point d'étude important et même capital concerne la stabilité du système. En particulier pour les structures récursives comme celles à filtres duaux dont les problèmes d'instabilités sont liés à la causalité des filtres, la position des sources et la distance inter-capteur. Dans la plupart des configurations réelles telles que les télécommunications où la propagation de signaux acoustiques, le modèle de mélange à phase minimale n'est généralement pas raisonnable. Ainsi, les contraintes de causalité et de stabilité imposées par cette structure de séparation limitent fortement son champ d'application et par conséquent la structure directe lui est souvent préférée. Nous donnons ici les premiers éléments sur l'analyse de la stabilité ; l'étude de la convergence et des points d'équilibre de cette structure sera présentée au Chapitre 5. En effet, la stabilité du système global n'est pas garantie *a priori* à cause de la boucle qu'il contient, même en se restreignant au cas simplifié où tous les $A_{ij}(z)$ sont de type tout-zéros, impliquant des filtres de séparation également de type tout-zéros. La rétroaction réalise de manière implicite un filtrage comportant une partie tout-pôles qui apparaît dans les expressions données en (2.15) ou encore sur le schéma équivalent 2.6 extrait de [131], ce qui la différencie de la structure directe qui ne fait intervenir que des combinaisons linéaires de filtres tout-zéros dans sa version sans postfiltrage nécessaire pour recouvrer les sources sans effet de salle.

Les deux aspects de la stabilité

Pour discuter de la stabilité d'une méthode il faut tout d'abord distinguer de quelle notion de stabilité l'on parle. En effet, celle-ci se décompose en deux aspects

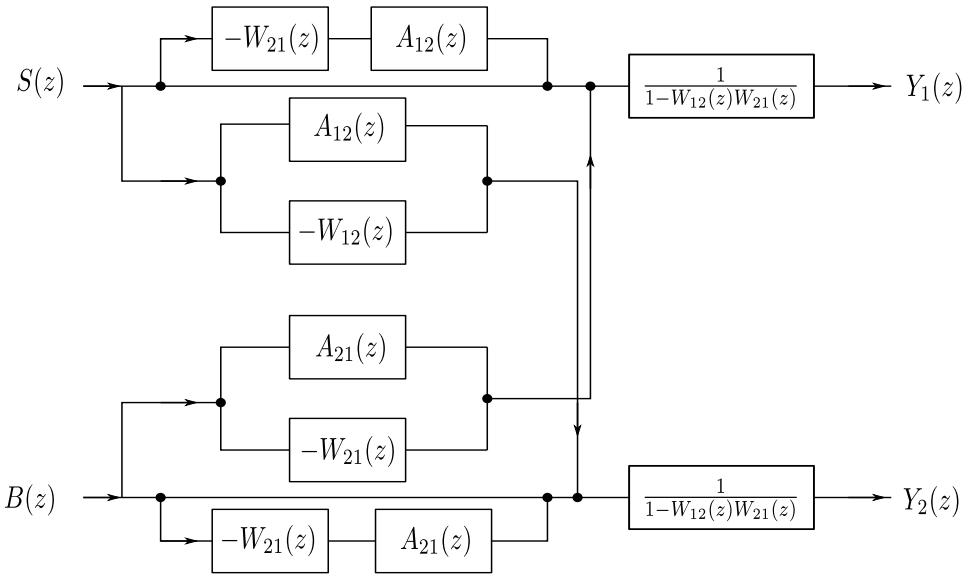


Figure 2.6 – Schéma bloc représentant les relations d'entrée-sortie associées à la structure récursive.

majeurs [69] selon qu'elle concerne l'architecture ou la règle d'adaptation choisie pour la mise à jour des coefficients des filtres adaptatifs. Nous distinguerons donc :

1. la stabilité structurelle liée à la stabilité de la boucle formée par les filtres $W_{12}(z)$ et $W_{21}(z)$. L'analyse de cette rétroaction révèle que ceci est équivalent à l'étude de la stabilité du filtre RII

$$F(z) = \frac{1}{1 - W_{12}(z)W_{21}(z)} \tag{2.18}$$

La stabilité de la structure requiert que les pôles de $F(z)$ soient à l'intérieur du cercle unité, $\mathcal{C}_1$. Ou bien, selon le critère de Nyquist, que $|W_{12}(e^{j\omega})W_{21}(e^{j\omega})| < 1$ avec $-\pi < \omega < \pi$. Cette condition est indépendante de l'algorithme utilisé pour l'adaptation des filtres. Or, l'objectif des algorithmes mis en œuvre est d'assurer à la convergence que les filtres adaptatifs $W_{12}(z)$ et $W_{21}(z)$ identifient $A_{12}(z)$ et $A_{21}(z)$. Par conséquent, ceux-ci doivent également remplir des conditions afin que la solution optimale soit stable. En définissant $\widetilde{F}(z)$ tel que :

$$\widetilde{F}(z) = \frac{1}{1 - A_{12}(z)A_{21}(z)} \tag{2.19}$$

la stabilité impose donc que $\widetilde{F}(z)$ soit stable, ce qui constitue une contrainte importante sur le mélange. Cette condition est également indépendante de l'algorithme et est appelée, lorsqu'elle est vérifiée, stabilité inhérente. Ainsi, la stabilité structurelle concerne le conditionnement des filtres adaptatifs et constitue un critère de viabilité de l'algorithme si ceux-ci vérifient que (2.18)

soit stable, y compris pendant la convergence. Sachant que les filtres sont choisis de type RIF, la stabilité peut être dérivée à partir de la valeur des coefficients, ce qui par la même, impose des contraintes sur la configuration des sources. Cet aspect sera étudié au paragraphe 2.3.4,

2. le deuxième aspect concerne la stabilité de l'algorithme adaptatif, que l'on décompose également en deux points :
   (a) le premier est appelé la stabilité asymptotique et étudie le comportement local de l'algorithme autour d'un point d'équilibre théorique et plus particulièrement autour de la solution optimale,
   (b) le second problème traite la stabilité d'un point de vue global en étudiant le comportement convergeant ou divergeant de l'algorithme.

   En comparaison avec le premier aspect de la stabilité, le second volet de l'analyse de la stabilité dépend clairement de l'algorithme utilisé. Ceci sera discuté au Chapitre 5.

2.3.4 Influence de l'hypothèse de causalité sur la position des sources et les performances

Nous discutons dans cette partie de la stabilité structurelle présentée précédemment et qui impose, comme cela est souvent fait dans la littérature, la causalité (et même stricte causalité) des filtres de mélange. Une étude très complète des différents critères de stabilité faite par Nguyen Thi et *al.* est donnée dans [175] et permet d'aboutir, grâce à une démarche successive, à une condition suffisante simple de stabilité. Nous résumons ci-après leur approche.
Supposons que l'ordre des filtres de mélange soit égal à N, que $W(z) = A_{12}(z)A_{21}(z) = \sum_{k=0}^{2N} w(k)z^{-k}$ et posons $P(z)$ tel que :

$$\begin{aligned} P(z) &= 1 - W(z) \\ &= 1 - \left(\sum_{i=0}^{N} a_{12}(i)z^{-i}\right)\left(\sum_{j=0}^{N} a_{21}(j)z^{-j}\right) \end{aligned} \tag{2.20}$$

D'après l'analyse précédente, la stabilité est assurée si tous les pôles de $F(z)$ sont dans $\mathcal{C}_1$, ce qui revient à dire que tous les zéros de $P(z)$ doivent être dans $\mathcal{C}_1$. En ce sens, l'étude de la stabilité est équivalente à l'étude de la position des racines du polynôme $P(z)$ d'ordre $2N$. Or, si nous portons une attention particulière au terme d'ordre zéro de ce polynôme, il est clair que si $p(0) = 1 - a_{12}(0)a_{21}(0) = 0$, alors $P(z)$ aura un zéro situé en $z = \infty$, ce qui est en contradiction avec la condition de stabilité donnée auparavant. Par conséquent, une condition nécessaire de stabilité est :

$$a_{12}(0)a_{21}(0) \neq 1 \tag{2.21}$$

et équivaut à dire qu'il n'existe pas de retour direct ou sans retard, comme cela est supposé par Mirchandani et *al.* dans [132] ou encore par Al-Kindi et Dunlop dans [10]. Pour satisfaire (2.21), une solution simple consiste à imposer un retard sur chacun des filtres adaptatifs afin de compenser le retard de la "boucle" récursive. Le problème est donc ramené à la recherche de la position des zéros du polynôme transformé suivant :

$$P(z)^{'} = 1 - \frac{\sum_{k=1}^{2N} w(k) z^{-k}}{1 - a_{12}(0)a_{21}(0)} \tag{2.22}$$

Les auteurs soulignent qu'en pratique il n'est pas nécessaire d'avoir tous les zéros de $P(z)$ dans $\mathcal{C}_1$ et qu'après développement, une condition adéquate mais très restrictive pour des filtres atténuateurs de type RIF dans un milieu isotrope s'écrit :

$$\sum_{i=0}^{N} |a_{12}(i)| < 1 \quad et \quad \sum_{j=0}^{N} |a_{21}(j)| < 1 \tag{2.23}$$

Le modèle de filtre atténuateur est en rapport avec les réponses impulsionnelles acoustiques qui sont par nature strictement causales, $a_{ij}(k) = 0$ avec $0 \leq k < L$ où $L << N$ [1] et vérifient dans la plupart des cas cette condition (2.23) qui peut encore conduire à :

$$\sum_{i=0}^{N} a_{12}^2(i) < 1 \quad et \quad \sum_{j=0}^{N} a_{21}^2(j) < 1 \tag{2.24}$$

D'après nos expérimentations, nous pourrions nous contenter de cette dernière expression (2.24) qui est systématiquement vérifiée. Malheureusement et comme souligné par les auteurs, celle-ci ne garantit pas la stabilité de $F(z)$.

Remarque 2 : *Il est aisé d'étendre cette analyse dans le cas général d'un mélange non simplifié, puisque les observations restent les mêmes pourvu que les retards introduits et associés aux chemins directs soient, par définition, inférieurs à ceux des chemins croisés.*

Relaxation de la contrainte

Il est très pratique et assez courant de faire l'hypothèse forte qui consiste à imposer que le système $P(z) = [1 - W_{12}(z)W_{21}(z)]$ soit à minimum de phase [117] de sorte à garantir l'inversibilité de $P(z)$ et la stabilité de $F(z)$. Cependant, en plus d'être une hypothèse peu réaliste, comme souligné en 2.3.1, elle est également stricte puisque équivalente à la condition de pôles tous situés dans $\mathcal{C}_1$. Or, dans [131] Mei et Yin montrent que cette contrainte n'est pas nécessaire pour des signaux dont le contenu spectral est à support limité, comme cela est le cas pour les signaux de

1. Notons que l'écart observé en pratique est lié au bruit de mesure introduit par les différents instruments de mesure utilisés.

parole. En effet, d'après le schéma 2.6, la relation liant les signaux sources aux sorties dans le cas du mélange simplifié donne :

$$\begin{aligned} Y_1(z) &= \frac{[1-W_{21}(z)A_{12}(z)]S(z)+[A_{21}(z)-W_{21}(z)]B(z)}{1-W_{21}(z)W_{12}(z)} \\ Y_2(z) &= \frac{[A_{12}(z)-W_{12}(z)]S(z)+[1-W_{12}(z)A_{21}(z)]B(z)}{1-W_{12}(z)W_{21}(z)} \end{aligned} \tag{2.25}$$

Or, si nous faisons l'hypothèse que les filtres $W_{12}(z)$, $W_{21}(z)$, $A_{12}(z)$ et $A_{21}(z)$ sont causaux, alors il est physiquement impossible que $W_{12}(z)A_{12}(z) = 1$, et $W_{21}(z)A_{21}(z) = 1 \ \forall \ z$. Par conséquent, les sorties vaudront respectivement $Y_1(z) = S(z)$ et $Y_2(z) = B(z)$ uniquement si :

$$\begin{aligned} [A_{12}(z) - W_{12}(z)]\, S(z) = 0 \\ [A_{21}(z) - W_{21}(z)]\, B(z) = 0 \end{aligned} \tag{2.26}$$

Ainsi, (2.26) est satisfaite dans deux cas. Ou bien, $W_{12}(z) = A_{12}(z)$, et $W_{21}(z) = A_{21}(z) \ \forall \ z$, ou uniquement lorsque $S(z) \neq 0$ et $B(z) \neq 0$. En d'autres termes, pour l'ensemble des composantes spectrales $z = e^{j\omega}$ telles que le spectre des signaux est nul, les filtres peuvent varier et les solutions ne sont donc pas uniques. Malheureusement, même si cette analyse est très intéressante, elle ne nous permet pas d'avancer plus en avant sur la manière de contraindre les filtres de façon plus lâche, même si en théorie cela vient d'être prouvé. Il ressort donc que les premiers coefficients des filtres de mélange jouent un rôle très important en ce qui concerne la stabilité des filtres duaux et nous allons voir que deux contraintes fortes en lien avec ces coefficients régissent le bon fonctionnement de cette architecture.

Contrainte structurelle conditionnant la position des sources

En observant ce qui se passe pour l'indice $i = j = 0$ dans l'expression des sorties à temps discret données ci-après :

$$y_1(n) = x_p(n) - \sum_{i=0}^{N} w_{21}(i) y_2(n-i) \tag{2.27}$$

$$y_2(n) = x_r(n) - \sum_{j=0}^{N} w_{12}(j) y_1(n-j) \tag{2.28}$$

nous sommes forcés de constater que l'estimation de filtres anti-causaux n'est, par construction, pas permise. En effet, le calcul à un instant n de la sortie d'un filtre comportant une partie non-causale nécessite de connaître à la fois le passé de son entrée pour la partie causale et l'avenir de cette entrée pour la partie non-causale. Or, l'entrée du filtre $\mathbf{w_{21}}$ (respectivement $\mathbf{w_{12}}$) est constituée de la sortie y_2 (respectivement y_1) du système, dont les valeurs futures ne sont pas encore disponibles

à l'instant n car elles dépendent également des valeurs futures et encore inconnues du signal de sortie y_1 (respectivement y_2) associé au filtre dual. A ce niveau d'analyse, nous pourrions penser que l'ajout d'un retard unitaire sur chacun des filtres adaptatifs afin de casser l'hyper-récursivité [10] permettrait également de résoudre le problème de la causalité, mais il n'en est rien. En effet, nous allons constater ci-après que le couplage entre les voies répercute le retard lié à la mémoire des filtres sur chacune d'elle et empêche la synchronisation des signaux. Pour cela, considérons la Figure 2.7 qui présente une configuration impliquant l'estimation d'un filtre anti-causal, b), et pour laquelle l'architecture sera inefficace. La nécessité d'avoir un filtre anti-causal pour w_{21} se justifie par le fait que l'on souhaite additionner ou soustraire des signaux synchrones vis-à-vis de la source d'intérêt et ceci ne peut se faire que dans le cas où l'on se fixe un micro de référence. De ce principe à la base des méthodes de formation de voies découle les notions de chemins directs associés par définition aux retards de propagation les plus courts, par opposition aux chemins croisés. Par conséquent, pour conserver la symétrie qui conserve ces notions de chemins directs et chemin croisés dans le cas décrit en b), il est nécessaire que w_{21} soit anti-causal de sorte que nous puissions supposer simplement $a_{11}(n) = \delta_{n,0}$, $a_{22}(n) = \delta_{n,0}$ et w_{12} causal. Observons ce qui se produit dans ce cas sur les estimées données en (2.27)

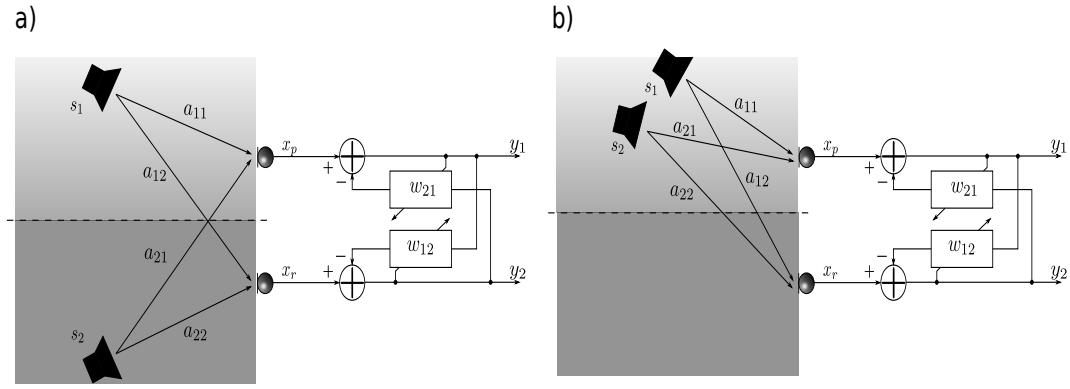


Figure 2.7 – Modèle de mélange de deux sources nécessitant a) uniquement des filtres de séparation causaux et b) un filtre causal pour w_{12} et un filtre anti-causal pour w_{21}.

et (2.28) et en tenant compte de l'anti-causalité de $\mathbf{w_{21}}$.

$$\begin{aligned} y_1(n) &= x_p(n) - \sum_{i=-N}^{0} w_{21}(i) y_2(n-i) \\ &= x_p(n) - [w_{21}(-N) y_2(n+N) + \cdots + w_{21}(0) y_2(n)] \end{aligned}$$

$$\begin{aligned} y_2(n) &= x_r(n) - \sum_{j=0}^{N} w_{12}(j) y_1(n-j) \\ &= x_r(n) - [w_{12}(0) y_1(n) + \cdots + w_{12}(N) y_1(n-N)] \end{aligned}$$

Or, pour produire l'estimée $y_1(n)$, nous devons disposer dans la mémoire du filtre w_{21} de l'échantillon $y_2(n+N)$, obtenu grâce au calcul suivant :

$$y_2(n+N) \quad = \quad x_r(n+N) - \sum_{j=0}^{N} w_{12}(j) \underbrace{y_1(n+N-j)}_{inconnu}$$

D'emblée, le problème de la causalité se pose puisque le calcul de l'estimée y_1 à l'instant n et nécessitant l'estimée y_2 à l'instant $n+N$ fait lui-même intervenir l'estimée de y_1 à l'instant $n+N$, alors que celle-ci est encore inconnue à l'instant n. Par conséquent, les filtres $\mathbf{w_{12}}$ et $\mathbf{w_{21}}$ sont donc nécessairement causaux et comme ceux-ci sont à l'optimum les estimées de a_{12} et a_{21}, cela implique que la structure à filtres duaux requiert une disposition des sources symétriques par rapport au plan médian, configuration a).

L'impact de la disposition des sources sur les performances est évalué et mis en évidence à la Figure 2.8 où les valeurs du RSB moyen, noté RSB_{moyen} et définit par Gannot et al. dans [66] sont tracées dans le cas d'un mélange anéchoique non simplifié (chemins directs ne sont pas réduits à un Dirac) comportant une voix d'homme (s_1) et un bruit de bureau (s_2). Ce RSB_{moyen} défini par l'expression donnée ci-après, compare l'énergie du signal utile à l'énergie du signal de bruit pour un signal $x(n)$ donné (variable muette).

$$RSB_{moyen} = \frac{\sum_{n \in T_{présence}} x^2(n) - \sum_{n \in T_{absence}} x^2(n)}{\sum_{n \in T_{absence}} x^2(n)}$$

Où $T_{présence}$ désigne les périodes où le signal utile est présent et $T_{absence}$ les périodes de bruit seul. Le RSB sur la voie primaire est de 5 dB et a été fixé grâce à l'outil normalisé sv56 (*speech voltmeter*) de l'*International Telecommunication Union* (ITU). A l'aide de l'outil *Roomsim* [26], nous générons l'ensemble des réponses impulsionnelles nécessaires pour simuler le déplacement de la source s_2 de son demi-plan au demi-plan de la source s_1. L'environnement sonore ainsi synthétisé est décrit à l'Annexe D. Les résultats obtenus montrent que les performances commencent à chuter dès lors que la source de bruit s_2 atteint le demi-plan situé sur la perpendiculaire au frontal acoustique et continuent de décroître à mesure que la source de bruit se déplace dans le plan de la source utile. Comme souligné dans [106], la contrainte de causalité intrinsèque aux filtres duaux fait que le signal supprimé sur une voie est celui qui correspond au plus long chemin source-microphone. En d'autres termes et de manière imagée, cette structure travaille de telle manière que les signaux restent dans le même demi-plan en entrée et en sortie du système. Ainsi, en respectant les dénominations de voie primaire et de voie de référence, toute source située dans le demi-plan inférieur, cf. Figure 2.7, sera considérée comme une source de bruit et vice versa. D'où la nécessité d'avoir un filtre anti-causal dans le cas b), afin que a_{22} corresponde effectivement au trajet direct et que la source s_2 se retrouve en y_2, c'est-à-dire

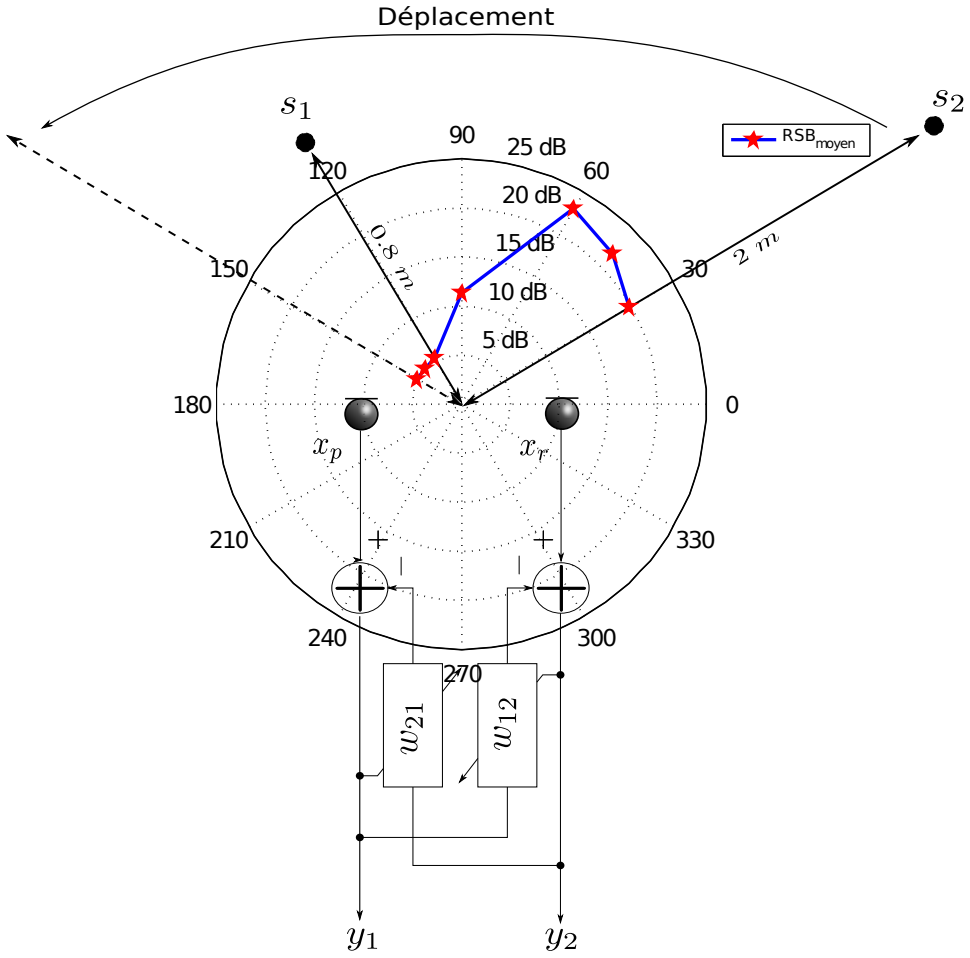


Figure 2.8 – Evolution du RSB en sortie de traitement en fonction de la position de la source s_2.

sans permutation. Or, la structure ne pouvant travailler qu'avec des filtres causaux, les auteurs introduisent sur la voie primaire un retard réalisant un pointage qui rompt la symétrie d'une part et ramène le problème à l'estimation de filtres causaux d'autre part. Cependant, il est à noter que cette méthode fonctionne uniquement pour un mélange de deux sources. Au délà, il est toujours possible d'avoir des sources situées dans le même demi-plan. Enfin, notons qu'il existe par ailleurs une zone d'ambiguïté en forme de cône dont l'axe principal est formé par le plan médian. Ainsi, une source appartenant à cette région sera difficile à extraire du fait d'une commutation incessante entre les deux voies conduisant à retrouver la source estimée alternativement sur la voie primaire et la voie de référence. Cette observation est en relation directe avec le critère de Rayleigh donnant le pouvoir de résolution des antennes [98] et dont les conséquences sont analysées par Cox dans [41].

Contrainte liée à l'implémentation

Nous avons vu précédemment que la structure à filtres duaux impose que les filtres $\mathbf{w_{21}}$ et $\mathbf{w_{12}}$ soient causaux. En fait, la contrainte est plus forte et exige qu'au moins un des filtres soit strictement causal. En effet, la mise en œuvre pratique des filtres duaux introduit de manière implicite un retard dans la structure qui n'est pas sans conséquence. La position de ce retard va dépendre de l'ordre dans lequel les signaux de sortie vont être calculés. Cette précédence dans la mise à jour des sorties, indispensable à la réalisation pratique de l'algorithme, se traduit donc simplement par l'ajout d'un retard sur la voie associée, par exemple y_1. En l'occurrence, (2.28) devient :

$$y_2(n) = x_r(n) - \sum_{i=0}^{N} w_{12}(i) y_1(n-1-i) \tag{2.29}$$

Ce choix est arbitraire et peut être inversé du fait de la symétrie de la structure. A partir de cette nouvelle expression, il est clair que le coefficient d'ordre zéro du filtre a_{12} ne pourra être identifié. L'hypothèse d'une initialisation à zéro de ce coefficient (filtre strictement causal) est un moyen simple et en adéquation avec la physique du problème. Par ce biais, la solution optimale théorique peut donc être atteinte. Cependant, ce ne sera jamais le cas en pratique, en raison du bruit de mesure qui fausse quelque peu cette hypothèse.

Remarque 3 : *L'analyse des contraintes imposées par la structure récursive montre que des précautions doivent être prises pour son utilisation. Nous remarquons en effet que certaines configurations pour les sources peuvent poser problème. Enfin, il faudra également en tenir compte pour la conception d'une antenne acoustique.*

2.4 Minimisation de l'EQM ou décorrélation ?

Les différentes utilisations des filtres duaux dans le domaine du traitement de la parole se distinguent selon que l'on considère le problème de la diaphonie, en l'occurrence celui qui nous concerne, ou bien la séparation de sources. Néanmoins, l'objectif est le même et l'on peut dire qu'historiquement c'est d'abord en tant que méthode de débruitage que cette méthode est née, puis ensuite reprise et généralisée pour la séparation de sources. Le lien est donc fort et la frontière infime si tant est qu'elle existe. La seule distinction tient dans l'utilisation des statistiques d'ordre supérieure qui a remplacé l'utilisation des statistiques d'ordre deux pour le mélange d'au moins deux sources sonores (paroles, parole et musique, parole et bruit ou l'ensemble). Alors que dans le problème de l'annulation de bruit, le mélange est plutôt restreint à un mélange de deux sources, dont l'une est une source de bruit ponctuelle

comme nous l'avons considéré jusqu'à présent. Nous traitons donc un sous-cas des problèmes de séparation de sources, mais en restreignant les solutions à celles basées sur les statistiques d'ordre deux. Pour garder en généralité, les sources seront donc notées dorénavant s_1 et s_2 et peuvent être indifféremment source utile ou bruit.
La littérature concernant l'étude et l'amélioration des algorithmes fondés sur les statistiques d'ordre deux pour cette architecture dans le cas de mélanges convolutifs se résume à quelques articles. En premier lieu, les travaux de Al-Kindi et Dunlop [10] qui formulent la première version de l'algorithme de décorrélation dans le domaine temporel. Par la suite, les travaux de Van Gerven et Van Compernolle [70, 71] ont conduit à l'algorithme nommé *Symmetric Adaptive Decorrelation* (SAD). En ce sens, cette approche est plus en phase avec l'idée de séparer les sources que l'approche classique qui consiste à minimiser l'erreur quadratique. La transposition dans le domaine fréquentiel de cette approche est donnée par Weinstein et al. [188] où l'algorithme vise à annuler l'interspectre des sorties. Cependant, comme nous allons le démontrer ci-après, le couplage entre les sorties des filtres duaux rend l'approche fondée sur la décorrélation caduque.

2.4.1 Minimisation de l'erreur quadratique moyenne

La construction d'un algorithme suivant un critère énergétique se base principalement sur le critère de minimisation de l'erreur quadratique moyenne définie par :

$$\xi = E\Big[e(n)^2\Big] = E\Big[(d(n) - \mathbf{w}^T(n)\mathbf{x}(n))^2\Big] \tag{2.30}$$

où $d(n)$ joue le rôle du signal primaire $x_p(n)$, $\mathbf{x}$(n) le vecteur contenant les $N+1$ dernières valeurs des signaux de référence $x_r(n)$, $\mathbf{w}$(n) le vecteur des coefficients du filtre (cf. Figure 2.2) et ^{T} l'opérateur de transposition. La minimisation de cette erreur va permettre d'obtenir la solution correspondant au filtre optimal $\mathbf{w}^*$ ou solution de Wiener, obtenue en cherchant la valeur qui annule le gradient de celle-ci par rapport à $\mathbf{w}$. Le gradient de l'équation (2.30) est donné à l'équation suivante :

$$\nabla_{\mathbf{w}}\xi = 2\mathbf{R}\mathbf{w} - \mathbf{2p} \tag{2.31}$$

avec $\mathbf{R} = E\Big[\mathbf{x}(n)\mathbf{x}^T(n)\Big]$ la matrice d'autocorrélation du signal $x(n)$ et $\mathbf{p} = E\left[d(n)\mathbf{x}(n)\right]$ le vecteur d'intercorrélation. L'annulation de ce gradient conduit à $\mathbf{w}^* = \mathbf{R}^{-1}\mathbf{p}$ ce qui correspond à la solution à minimum d'énergie. Ceci correspond à l'approche classique qui conduit à une mise en œuvre à l'aide de la méthode de descente du gradient. Mais ce n'est pas le seul moyen d'y parvenir, en écrivant l'intercorrélation notée $\mathbf{C}$ entre le signal de référence et le signal d'erreur, nous pouvons constater que :

$$\begin{aligned} \mathbf{C} &= E\left[e(n)\mathbf{x}(n)\right] = E\left[\Big(d(n) - \mathbf{w}^T(n)\mathbf{x}(n)\Big)\mathbf{x}(n)\right] \\ &= \mathbf{p} - \mathbf{R}\mathbf{w} = -\frac{\mathbf{1}}{\mathbf{2}}\nabla_{\mathbf{w}}\xi \end{aligned} \tag{2.32}$$

Par conséquent, l'annulation du gradient revient à annuler l'intercorrélation entre le signal d'erreur $e(n) = \Big(d(n) - \mathbf{w^T}(n)\mathbf{x}(n)\Big)$ et les données $\mathbf{x}(n)$. C'est de ce constat qu'est né l'algorithme de décorrélation, mis en œuvre numériquement par l'algorithme de Newton-Raphson. L'équivalence entre les deux approches est claire mais nous allons voir que les solutions optimales dépendent de l'architecture et que, contrairement à ce qui est avancé dans la littérature, le comportement de l'algorithme SAD avec la structure récursive n'est pas optimal car trop approximatif.
Appliquons le critère de l'EQM minimale à la structure récursive de la Figure 2.7 en supposant que $s_1(n)$ et $s_2(n)$ sont des processus aléatoires réels à temps discret. Les signaux d'erreurs à l'instant n sont donnés par :

$$y_1(n) = x_p(n) - \mathbf{w_{21}}^T(n)\mathbf{y_2}(n) \tag{2.33}$$

$$y_2(n) = x_r(n) - \mathbf{w_{12}}^T(n)\mathbf{y_1}(n) \tag{2.34}$$

où

$$\begin{aligned} \mathbf{w_{21}}^T(n) &= \Big[w_{21}^0(n)\ w_{21}^1(n)\ w_{21}^2(n) \dots w_{21}^N(n)\Big] \\ \mathbf{w_{12}}^T(n) &= \Big[w_{12}^0(n)\ w_{12}^1(n)\ w_{12}^2(n) \dots w_{12}^N(n)\Big] \end{aligned}$$

sont les coefficients variants dans le temps des filtres $\mathbf{w_{21}}$ et $\mathbf{w_{12}}$ d'ordre N, et

$$\begin{aligned} \mathbf{y_1}^T(n) &= [y_1(n)\ y_1(n-1) \dots y_1(n-N)] \\ \mathbf{y_2}^T(n) &= [y_2(n)\ y_2(n-1) \dots y_2(n-N)] \end{aligned}$$

Ainsi, l'erreur quadratique moyenne sur la voie primaire et associée à y_1 est :

$$\xi_1 = E\Big[y_1^2\Big] = E\Big[(x_p - \mathbf{w_{21}}^T\mathbf{y_2})^\mathbf{2}\Big]$$

où l'indice temporel n est omis pour faciliter la lecture et $\mathbf{y_2}$ désigne la mémoire du filtre w_{21}. En prenant le gradient de cette expression par rapport à $\mathbf{w_{21}}$ nous obtenons :

$$\nabla_{\mathbf{w_{21}}}\xi_1 = 2 \cdot E\Big[(x_p - \mathbf{w_{21}}^T\mathbf{y_2})(-\mathbf{y_2} - \mathbf{w_{21}}^\mathbf{T}\nabla_{\mathbf{w_{21}}}\mathbf{y_2})\Big] \tag{2.35}$$

Faisons maintenant l'hypothèse que $\nabla_{\mathbf{w_{21}}}\mathbf{y_2} = \mathbf{0}$, ce qui est théoriquement faux du fait du couplage existant dans cette structure et mis en exergue par les équations des sorties (2.33) et (2.34), nous écrivons :

$$\nabla_{\mathbf{w_{21}}}\xi_1 = 2 \cdot E\Big[(x_p - \mathbf{w_{21}}^T\mathbf{y_2})(-\mathbf{y_2})\Big] = -2 \cdot E[y_1\mathbf{y_2}] \tag{2.36}$$

Cette solution qui minimise la corrélation entre les sorties correspond à celle fournie par l'algorithme SAD proposé par Van Gerven dans [70]. Si au contraire nous ne négligeons pas le terme $\nabla_{\mathbf{w_{21}}}\mathbf{y_2}$, alors (2.35) devient :

$$\nabla_{\mathbf{w_{21}}}\xi_1 = -2 \cdot E\Big[y_1(\mathbf{y_2} + \mathbf{w_{21}}^\mathbf{T}\nabla_{\mathbf{w_{21}}}\mathbf{y_2})\Big] \tag{2.37}$$

ce qui correspond à l'approche de Zinser et Mirchandani [132, 202] où le couplage est donc clairement pris en compte et conduit donc aux solutions optimales exactes. Par opposition avec l'approche de Van Gerven ou de Al-Kindi et Dunlop où le fait de négliger le terme lié au couplage structurel rejoint une approche intuitive du gradient stochastique, exacte pour la structure directe, mais imprécise pour cette structure. En effet, que ce soit pour ces derniers ou Mei et Yin [131], ces auteurs vont, par abus, appliquer le principe d'orthogonalité sur des signaux très fortement corrélés de par la structure, ce qui conduit à la négligence de l'interdépendance lorsque ceux-ci définissent l'algorithme SAD. Néanmoins, il existe une subtilité que nous exposerons au Chapitre 4 et retrouverons au Chapitre 5 dans le sens où l'algorithme exact aura effectivement convergé vers la solution optimale une fois que l'erreur sera effectivement orthogonale au gradient de cette erreur. Mais nous considérons ici la solution après convergence, c'est-à-dire une fois que le mélange est séparé. Bien sûr, ceci ne peut pas être supposé d'emblée puisque c'est faux, car au moins à l'état initial, le gradient de cette erreur ne se résume pas aux données observées sur l'autre voie, par exemple ici $\mathbf{y_2}$. Un autre aspect en lien avec cette observation concerne l'unicité des solutions. Le problème de l'estimation par l'erreur quadratique moyenne admet une solution unique si les données $\mathbf{y_2}$ (respectivement $\mathbf{y_1}$) sont linéairement indépendantes, comme cela est traduit à la Figure 2.9. Or, d'après ce que nous venons de dire, cette condition n'est en rien garantie dans le cas de la structure à filtres duaux. Nous reviendrons plus en détail sur l'approche donnée par Zinser et Mirchandani au Chapitre 4.

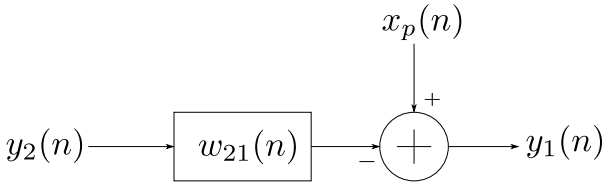
2.4.2 Illustration sur un cas simple

Pour appuyer notre analyse précédente montrant l'imprécision de l'algorithme SAD avec la structure récursive, nous illustrons le comportement de cet algorithme sur un cas simple. Tout d'abord, cette discussion sur la différence de performances des algorithmes que nous considérons montre que différentes approches peuvent être considérées pour la définition d'un algorithme adaptatif. En se restreignant aux algorithmes fondés sur la minimisation de l'énergie, nous pourrions ainsi envisager les trois approches suivantes :

$$\textit{Algorithme } A = \begin{cases} \mathbf{w_{21}}(n+1) = \mathbf{w_{21}}(n) - \mu_1 \nabla_{\mathbf{w_{21}}} \xi_1 \\ \mathbf{w_{12}}(n+1) = \mathbf{w_{12}}(n) - \mu_2 \nabla_{\mathbf{w_{12}}} \xi_2 \end{cases}$$

$$\textit{Algorithme } B = \begin{cases} \mathbf{w_{21}}(n+1) = \mathbf{w_{21}}(n) - \mu_1 \nabla_{\mathbf{w_{21}}} \xi_1 \\ \mathbf{w_{12}}(n+1) = \mathbf{w_{12}}(n) - \mu_2 \nabla_{\mathbf{w_{12}}} \xi_1 \end{cases}$$

$$\textit{Algorithme } C = \begin{cases} \mathbf{w_{21}}(n+1) = \mathbf{w_{21}}(n) - \mu_1 \nabla_{\mathbf{w_{21}}} \xi_2 \\ \mathbf{w_{12}}(n+1) = \mathbf{w_{12}}(n) - \mu_2 \nabla_{\mathbf{w_{12}}} \xi_2 \end{cases}$$



$$\widehat{x}_p(n) = \sum_{k=0}^{N} w_{21}(k) y_2(n-k)$$

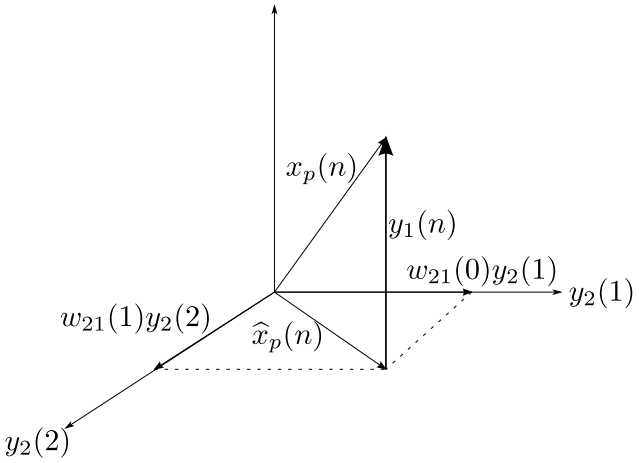


Figure 2.9 – Interprétation géométrique du problème d'estimation basé sur l'erreur quadratique moyenne. Cas du filtre à deux coefficients.

Pour l'*Algorithme A*, l'objectif est d'apprendre avec le même poids les chemins acoustiques croisés, ce qui correspond aux deux différentes approches illustrées par les équations (2.36) et (2.37) et menant aux relations de mise à jour des coefficients (a) et (b) du filtre $\mathbf{w_{21}}$.

$$\mathbf{w_{21}}(n+1) = \mathbf{w_{21}}(n) + 2\mu_{sad} E\left[y_1 \mathbf{y_2}\right] \quad \text{(a)}$$
$$\mathbf{w_{21}}(n+1) = \mathbf{w_{21}}(n) + 2\mu_{grad} E\left[y_1(\mathbf{y_2} + \mathbf{w_{21}}^{\mathbf{T}} \nabla_{\mathbf{w_{21}}} \mathbf{y_2})\right] \quad \text{(b)}$$

A l'inverse, les *Algorithme B* et *C* se focalisent préférentiellement sur l'un des chemins. Afin d'évaluer les performances de l'algorithme SAD, nous allons étudier la convergence des deux filtres adaptatifs $\mathbf{w_{21}}$ et $\mathbf{w_{12}}$ dans le cas très simple où les filtres de mélange et les solutions optimales sont réduits à un filtre à un seul coefficient. Les conditions de simulation sont les suivantes :

$$A_{12}(z) = 0,8 \quad A_{21}(z) = -0,8z^{-1}$$
$$\sigma^2_{s_1} = 0,2 \quad \sigma^2_{s_2} = 0,1$$
$$\mu_{sad} = 0,003 \quad \mu_{grad} = 0,005$$
$$s_1 \rightsquigarrow \mathcal{U}(0, \sigma^2_{s_1}) \quad s_2 \rightsquigarrow \mathcal{U}(0, \sigma^2_{s_2})$$

où $\mathcal{U}(0, \sigma_{s_i}^2)$ désigne un bruit blanc uniforme de moyenne nulle et de variance $\sigma_{s_i}^2$. Pour comparer les performances, les résultats obtenus avec un algorithme du gradient déterministe de type *Algorithme B* sont également donnés. Ceux associés à l'*Algorithme A* qui nous intéresse seront présentés au Chapitre 4, notamment parce qu'il est difficile d'obtenir des expressions simples pour les gradients et explicites des valeurs optimales des filtres, contrairement à l'*Algorithme B* pour lequel nous pouvons écrire :

$$\xi_1 = \frac{\sigma_p^2 + w_{21}^2\sigma_r^2 - 2w_{21}r_{pr}(0) - 2w_{21}^2w_{12}r_{pr}(-1)}{1-(w_{21}w_{12})^2} \tag{2.38}$$

Cette expression a été obtenue en faisant l'hypothèse de signaux sources blancs et indépendants. Dans ce cas et en tenant compte du retard lié à l'implémentation, les quantités suivantes apparaissent :

$$\begin{aligned} E\left[x_p(n-1)x_r(n)\right] &= a_{12}\cdot\sigma_{s_1}^2 = r_{pr}(1) = r_{pr}(-1) \\ E\left[x_p(n)x_r(n)\right] &= a_{21}\cdot\sigma_{s_2}^2 = r_{pr}(0) \\ r_{pp}(0) &= \sigma_{s_1}^2 + a_{21}\cdot\sigma_{s_2}^2 \\ r_{rr}(0) &= \sigma_{s_2}^2 + a_{12}\cdot\sigma_{s_1}^2 \end{aligned}$$

Les gradients de ξ_1 par rapport à chacun des filtres et donnés ci-après sont déduits de (2.38).

$$\frac{\partial\xi_1}{\partial w_{21}} = \frac{2w_{21}\sigma_r^2 - 2r_{pr}(0) - 4w_{21}w_{12}r_{pr}(-1) - 2w_{21}^2w_{12}^2r_{pr}(0) + 2w_{21}w_{12}^2\sigma_p^2}{(1-(w_{21}w_{12})^2)^2}$$

$$\frac{\partial\xi_1}{\partial w_{12}} = \frac{-2w_{21}^2r_{pr}(-1) - 2w_{21}^4w_{12}^2r_{pr}(-1) + 2w_{21}^4w_{12}\sigma_r^2 + 2w_{12}w_{21}^2\sigma_p^2 - 4w_{21}^3w_{12}r_{pr}(0)}{(1-(w_{21}w_{12})^2)^2}$$

Ces quantités sont celles utilisées dans la mise en œuvre du gradient déterministe selon l'*Algorithme B*. La Figure 2.10 présente les résultats de simulation obtenus pour les deux algorithmes considérés. Malgré un mélange de sources très simple, nous constatons que l'algorithme SAD est peu précis. L'erreur commise sur les estimées au regard de la simplicité du mélange montre l'inefficacité de cet algorithme avec la structure à filtres duaux. Enfin sur la Figure 2.11, nous traçons les courbes de niveau de la fonction de coût ξ_1 (2.38) de l'*Algorithme B* et la trajectoire du gradient déterministe. D'emblée, la structure modale de l'EQM alors que les signaux sont blancs met en avant le couplage structurel qui caractérise la structure de séparation et qui influence le conditionnement des matrices de corrélation mises en jeu. Par ailleurs, en observant la solution optimale de Wiener obtenue en fixant $\mathbf{w_{12}} = \mathbf{0}$, ce qui revient à supprimer le filtre $\mathbf{w_{12}}$ et retrouver l'annuleur de bruit classique, nous constatons que celle-ci se trouve à "mi-chemin" et à proximité de la trajectoire du

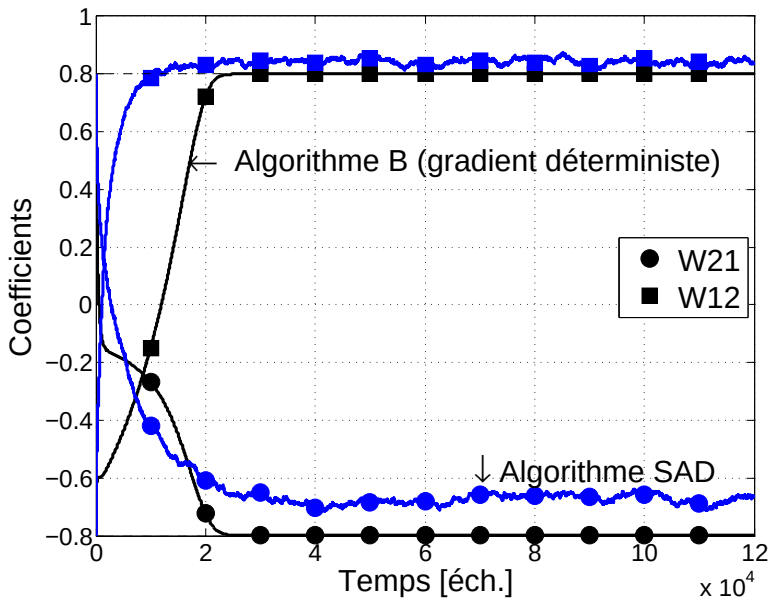


Figure 2.10 – Evolution des coefficients des filtres $\mathbf{w_{21}}$ et $\mathbf{w_{12}}$ pour un bruit blanc uniforme de moyenne nulle. En noir, les résultats obtenus à partir de la version déterministe de l'*Algorithme B*, en bleu ceux obtenus avec l'algorithme SAD.

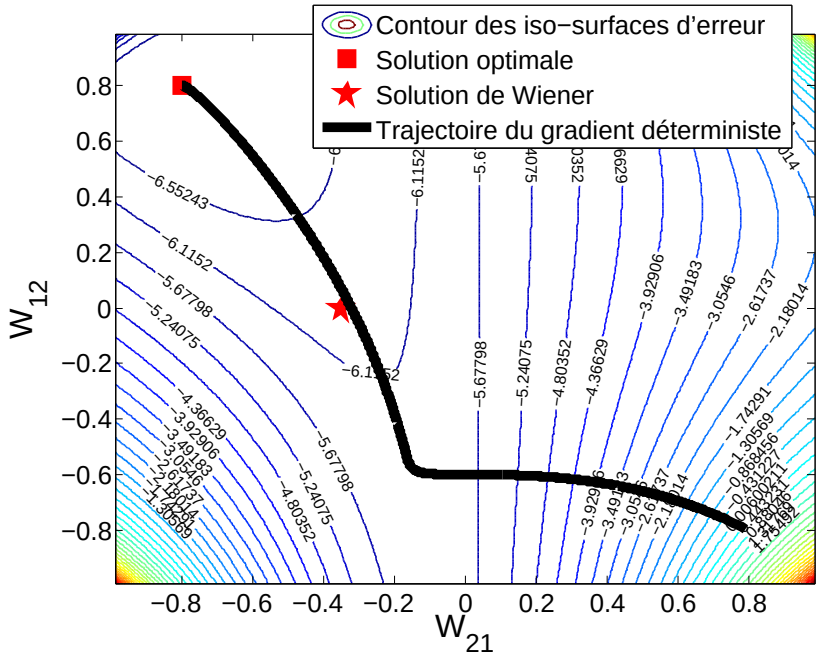


Figure 2.11 – Courbes de niveau de la fonction de coût ξ_1 et trajectoire du gradient déterministe selon l'*Algorithme B* pour un mélange de deux bruits blancs uniformes.

gradient vers la solution optimale. Ceci montre par conséquent la nécessité du second filtre pour la minimisation de l'EQM ξ_1. Les observations faites à partir de ces deux figures soulignent donc l'importance de l'interdépendance qui existe entre les sorties du système à filtres duaux ainsi que l'apport de l'utilisation d'un second filtre pour l'estimation du chemin de diaphonie.

2.5 Synthèse du chapitre

En définissant et démontrant les intérêts, les avantages et les limites de la structure récursive, ce chapitre établit la transition entre les méthodes de formation de voies de l'état de l'art et celles que nous allons proposer. En effet, en établissant un parallèle avec l'antenne GSC bicapteur, nous avons ouvert la voie vers une étude et la recherche d'une cellule de base, dans un premier temps bicapteur, afin de corriger l'un de ses défauts majeurs : la fuite du signal. Ainsi, après avoir discuté de la stabilité structurelle et des contraintes qui sont imposées sur le mélange et la position des sources, nous avons montré que, dans l'optique d'un traitement destiné à l'amélioration de la qualité vocale basé sur les statistiques d'ordre deux, la méthode reposant sur la décorrélation des sorties du système ne semble pas la plus efficace pour définir un algorithme adaptatif. En ce sens, les deux chapitres qui suivent vont se concentrer sur l'obtention d'algorithmes efficaces en adéquation avec les propriétés intrinsèques de la structure récursive qui ont été exposées dans ce chapitre.

Chapitre 3

Solution architecturale : le DCTRANC

3.1 Présentation du DCTRANC 84
3.1.1 Description générale 84
3.1.2 Découplage simple - Filtre adaptatif $\mathbf{w_1}$ 85
3.1.3 Version hybride - Filtre adaptatif $\mathbf{w_2}$ 86
3.1.4 Gestion du bruit du gradient - Contrainte algorithmique 87
3.2 Approches dédiées à l'amélioration de la gestion des pas 89
3.2.1 Règles algorithmiques pour la gestion des pas 90
3.2.2 Bilan des méthodes existantes 93
3.2.3 Méthodes proposées 93
3.3 Etude des solutions optimales 104
3.3.1 Solutions analytiques théoriques 104
3.3.2 Réflexion sur le DCTRANC 108
3.4 Implémentation fréquentielle et performances en environnement réel 108
3.4.1 Influence de la sous-modélisation 110
3.4.2 Prise de son binaurale 111
3.5 Synthèse du chapitre 112

Des techniques de réduction de bruit résistantes à la diaphonie dans le contexte de la prise de son bicapteur ont d'ores et déjà été proposées. Elles ont été décrites à la section 2.2 du chapitre précédent. Ces architectures de base se déclinent sous deux classes principales selon qu'elles correspondent à une implémentation directe (*feedforward*) ou récursive (*feedback*). La mise en œuvre de ces algorithmes conduit généralement en l'ajout d'un filtre supplémentaire permettant l'apprentissage du signal contaminant la référence de bruit. Elles peuvent consister également à introduire un bloc permettant le calcul du pas d'adaptation des différents filtres afin de régler au mieux le taux d'apprentissage, ce qui revient plus ou moins à l'élaboration d'une détection d'activité vocale. D'ailleurs, nous trouverons dans la littérature des techniques qui utilisent de manière distincte une DAV couplée à un étage de détermination du pas. En effet, un moyen simple de pallier la diaphonie consiste à stopper l'adaptation lorsque le signal du locuteur est présent. Nous pourrions donc imaginer de supprimer le second filtre en reportant toute la complexité sur l'élaboration d'une DAV qui devra être sans faille, ce qui en pratique est irréalisable surtout à faible RSB (< 5 dB) et dès lors que la scène sonore contient deux locuteurs. Ainsi, dans ce chapitre, on se propose de résoudre deux problématiques. D'une part, résoudre le problème du couplage existant entre les filtres et caractéristique de l'implémentation récursive qui nous intéresse. D'autre part, fournir une solution efficace pour la gestion des pas des filtres adaptatifs mis en œuvre. Les travaux décrits dans ce chapitre ont été en partie publiés à *International Workshop on Acoustic Echo and Noise Control* (IWAENC) 2008, voir [112], et ont également fait l'objet d'un brevet [113].

3.1 Présentation du DCTRANC

3.1.1 Description générale

Le DCTRANC, pour *Decoupled Crosstalk Resistant Adaptive Noise Canceller*, est une solution qui reporte la complexité algorithmique sur le matériel. En effet, nous avons vu au chapitre précédent qu'il existait un couplage, une interdépendance, entre les signaux de sortie non pris en compte par l'algorithme basé sur la décorrélation (SAD), le rendant par conséquent imprécis. Par ailleurs, la complexité de l'approche basée sur le critère de minimisation de l'EQM [202], n'a pas contribué à une utilisation massive de cette approche. Cependant, nous verrons dans le chapitre suivant que des optimisations sont possibles afin de tirer le meilleur parti de celle-ci. Au sein du présent chapitre, nous cherchons plutôt à adapter la structure à filtres duaux de sorte qu'elle puisse permettre l'utilisation de l'algorithme du gradient stochastique classique (*Least Mean Square* (LMS) ou *Normalized Least Mean Square* (NLMS)), c'est-à-dire en faisant abstraction du couplage structurel. Le schéma donné à la Fi-

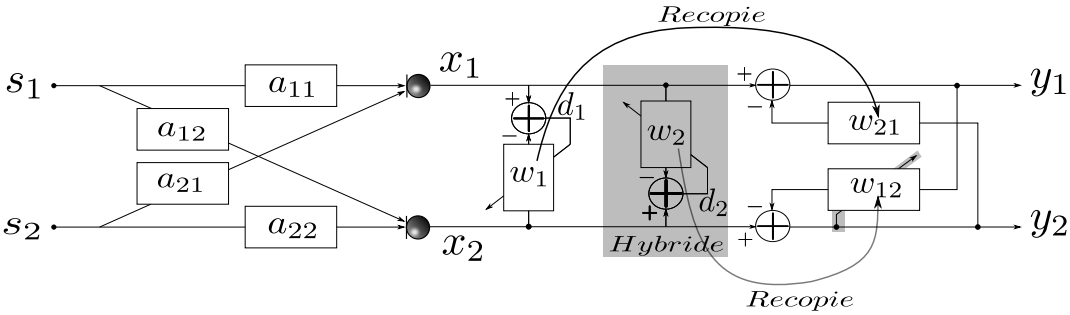


Figure 3.1 – Schéma descriptif du DCTRANC. Il est issu de la juxtaposition du CTRANC avec un étage de découplage formé par les filtres $\mathbf{w_1}$ et $\mathbf{w_2}$.

gure 3.1 décrit l'architecture du DCTRANC qui repose sur la mise en œuvre des fonctionnalités suivantes :

1. filtrage adaptatif $\mathbf{w_1}$: un premier filtre adaptatif permet de générer une estimée du signal s_2 (bruit) ;
2. [*Version hybride*] filtrage adaptatif $\mathbf{w_2}$: un second filtre adaptatif dédié à l'apprentissage du signal s_1 (diaphonie). Notons qu'il n'est pas nécessaire d'utiliser ce second filtre pour découpler les sorties, nous le montrerons en 3.1.2. Cette version du DCTRANC correspond à une méthode hybride puisqu'elle réalise l'adaptation des deux filtres selon une structure directe et les applique effectivement au sein de la structure récursive. L'intérêt d'une telle structure fait l'objet des prochains paragraphes ;
3. recopie de ce(s) filtre(s) adaptatif(s) dans son (leurs) homologue(s) fixe(s) placé(s) selon une configuration récursive propre aux filtres duaux,
4. contrainte algorithmique : lorsque la décision d'adapter le filtre destiné à l'apprentissage du signal de diaphonie est prise alors l'adaptation de $\mathbf{w_1}$ est gelée.

Au delà de notre choix de se restreindre à l'algorithme du gradient stochastique, il faut bien voir que grâce au DCTRANC, l'utilisation intuitive des algorithmes fondés sur les statistiques d'ordre deux et vérifiant le principe d'orthogonalité va être rendue possible.

3.1.2 Découplage simple - Filtre adaptatif $\mathbf{w_1}$

Ce bloc réalise une opération de filtrage adaptatif dont le but est d'identifier le canal acoustique a_{21} et de suivre ses variations dans le temps afin de fournir une estimée du bruit. Une fois le filtre calculé, celui-ci est ensuite recopié à chaque itération dans son homologue fixe. Ainsi, à chaque instant, nous avons :

$$\mathbf{w_{21}} = \mathbf{w_1} \tag{3.1}$$

En réalisant l'adaptation des filtres de cette manière, le problème du couplage entre les sorties disparaît, rendant légitime l'utilisation de l'algorithme du gradient classiquement mise en œuvre dans l'ANC. En effet, reconsidérons l'expression du signal de sortie y_1 pour la structure récursive :

$$y_1(n) = x_1(n) - \sum_{i=0}^{N} w_{21}(i) \cdot y_2(n-i) \tag{3.2}$$

Avec $y_2(n) = x_2(n) - \sum_{i=0}^{N} w_{12}(i) \cdot y_1(n-i)$. Le découplage simple implique que la relation (3.2) est aussi égale à :

$$y_1(n) = x_1(n) - \sum_{i=0}^{N} w_1(i) \cdot y_2(n-i)$$

Or, à chaque itération, $\mathbf{w_1}$ est calculé à partir du signal $d_1(n)$ donné par :

$$d_1(n) = x_1(n) - \sum_{i=0}^{N} w_1(i) \cdot x_2(n-i) \tag{3.3}$$

De toute évidence, l'interdépendance existante dans l'adaptation des filtres est supprimée puisqu'il n'existe pas de relation liant le signal x_2 au signal x_1 au travers $\mathbf{w_2}$ comme précédemment. Le découplage de l'adaptation est donc réalisé. Nous insistons sur le fait qu'il s'agit effectivement du découplage du processus d'adaptation et non pas de la structure qui, si cela était le cas, ne serait plus récursive.

3.1.3 Version hybride - Filtre adaptatif $\mathbf{w_2}$

De par la symétrie de la structure, le même raisonnement peut être tenu pour l'obtention d'une estimée du signal de diaphonie. Ceci conduit à la structure hybride évoquée plus haut, qui nous en conviendrons affiche une complexité matérielle importante. Pour être acceptable, elle doit se justifier par un gain en performance significatif par rapport à l'utilisation de la structure à filtres duaux classique, ou bien même par rapport au DCTRANC avec un seul filtre de découplage.
Avant même de considérer plus finement les résultats de simulation, nous pouvons mettre en avant un point important qui peut d'ores et déjà justifier la version hybride. Supposons que seul le filtre $\mathbf{w_1}$ soit utilisé, le filtre $\mathbf{w_2}$ n'est donc plus présent sur la Figure 3.1, l'adaptation est faite par $\mathbf{w_{12}}$. Ce cas correspond à ce que nous appelons un découplage simple. L'équation (3.3) montre que l'estimée du bruit va pouvoir être calculée classiquement sans que l'estimée du signal de diaphonie fournie par $\mathbf{w_{12}}$ ne vienne perturber l'adaptation. Or, par construction et selon notre hypothèse, ce n'est pas le cas pour le filtre $\mathbf{w_{12}}$ qui reste dépendant de y_1, donc de l'estimée bonne ou mauvaise du bruit. Or, l'objectif étant d'utiliser l'algorithme du

gradient stochastique classique y compris pour le filtre $\mathbf{w_{12}}$, l'adaptation est entachée d'erreur. Par contre, si nous utilisons les deux filtres $\mathbf{w_1}$ et $\mathbf{w_2}$, l'adaptation est totalement découplée, les éventuels désajustements de l'un des filtres n'influent plus sur la convergence de son dual.

Remarque 1 : *En supposant que l'on dispose d'une DAV idéale, la convergence de* $\mathbf{w_1}$ *vers la solution optimale de Wiener peut être supposée acquise*[1]*. Par contre, il sera toujours très difficile de fournir une estimée optimale du signal de diaphonie simplement parce qu'il n'existe pas d'instant pendant lequel le bruit peut être considéré comme absent, empêchant ainsi la convergence vers l'optimal de* $\mathbf{w_2}$*. Par conséquent, si, en raison de contraintes de complexité matérielle, l'utilisation de la version hybride du DCTRANC est remise en cause, la solution du découplage simple reste un bon compromis.*

3.1.4 Gestion du bruit du gradient - Contrainte algorithmique

La nécessité de cette contrainte est tirée de l'observation faite par Al-Kindi et Dunlop [10] et de l'analyse comportementale de l'agorithme LMS [191] en ce qui concerne l'impact de l'approximation stochastique sur la convergence. Par suite à ce qui a été dit au chapitre précédent où l'on donne l'expression de l'EQM (2.30) et la manière dont on obtient la solution optimale $\mathbf{w}^*$ qui correspond au minimum de l'EQM notée ξ_{min}, l'expression de l'EQM en fonction de la déviation des coefficients par rapport à l'optimum peut être réécrite comme suit :

$$\xi = \xi_{min} + (\mathbf{w} - \mathbf{w}^*)^T \mathbf{R} (\mathbf{w} - \mathbf{w}^*) \tag{3.4}$$

où $\mathbf{R}$ est la matrice d'autocorrélation du signal observé sur les micros. La déviation des coefficients traduite par $(\mathbf{w} - \mathbf{w}^*)$ est à l'image de la distance existant entre la solution fournie par le processus d'adaptation et la solution de Wiener. En théorie, si l'algorithme converge, cette distance est nulle et $\xi = \xi_{min}$. Cependant, du fait de l'approximation stochastique qui remplace les moyennes d'ensemble intervenant dans le calcul du gradient de l'EQM par des estimateurs statistiques, un bruit appelé bruit du gradient ou *gradient noise* est introduit. Par conséquent, la solution atteinte variera aléatoirement autour de la solution optimale et on aura, pour chaque itération k, $\xi_k > \xi_{min}$. Cet écart, observé à l'état stable de l'algorithme se quantifie par l'erreur quadratique moyenne en excès, obtenue par le moyennage des EQM instantanées ξ_k :

$$EQM_{excès} = E[\xi_k - \xi_{min}] = E\left[(\mathbf{w}_k - \mathbf{w}^*)^T \mathbf{R} (\mathbf{w}_k - \mathbf{w}^*)\right] \tag{3.5}$$

1. Nous modulerons ce propos avec l'analyse comportementale de l'algorithme LMS qui met en évidence l'influence du bruit du gradient sur la convergence.

Jusqu'ici, nous avons supposé de manière sous-jacente que toutes les quantités étaient des processus stationnaires. Or, un environnement acoustique se caractérise par des variations qui engendrent un changement fréquent des chemins acoustiques et donc de la solution optimale. Dans ce cas, Widrow et al. [190] montrent que la déviation des coefficients se décompose en deux termes (3.6), l'un lié au bruit du gradient, l'autre associé à l'erreur de suivi de la solution optimale changeante. De ce fait, il en va de même pour l'$EQM_{excès}$.

$$(\mathbf{w}_k - \mathbf{w}^*{}_k) = \underbrace{(\mathbf{w}_k - E[\mathbf{w}_k])}_{Bruit\ du\ gradient} + \underbrace{(E[\mathbf{w}_k] - \mathbf{w}_k^*)}_{erreur\ de\ suivi} \tag{3.6}$$

Le plus souvent, l'hypothèse de faibles variations rend le second terme négligeable et revient à dire que la solution optimale est invariante dans le temps, soit donc $\mathbf{w}^*{}_k \equiv \mathbf{w}^*$. En pratique, cela signifie qu'en cas de changement rapide, l'objectif est davantage de maintenir la stabilité de l'algorithme que de garantir la qualité du débruitage. Une étude des capacités de suivi des non-stationarités engendrées par le déplacement du locuteur est donnée dans [181], où, pour un contexte de simulation particulier, il est montré que la supériorité du filtrage adaptatif sur les méthodes fixes est remise en cause. Néanmoins depuis ces travaux datant de 1992, différentes déclinaisons de l'algorithme LMS ont été proposées afin de pallier ce problème. Un travail particulier sur le pas d'adaptation a permis, notamment, d'améliorer la vitesse de convergence et la poursuite des variations. Cet aspect sera décrit ultérieurement au sein de ce chapitre.

Pour quantifier les performances de l'algorithme LMS (ou NLMS), plusieurs critères sont utilisés notamment celui de la déviation quadratique moyenne ou *Mean Square Deviation* (MSD). Dans le cas de signaux blancs stationnaires, Park et al. [140] ont explicité de manière élégante et plus précise que dans la littérature initiale traitant de ce sujet, la valeur de celle-ci pour un pas d'adaptation compris dans l'intervalle $[0, 1]$. En définissant le vecteur déviation comme suit $\widetilde{\mathbf{w}}_k = (\mathbf{w}^* - \mathbf{w}_k)$, cette relation s'exprime en fonction du RSB suivant :

$$MSD = E\left[\widetilde{\mathbf{w}}_k^T \widetilde{\mathbf{w}}_k\right] = \mu \xi_{min} \mathbf{I} = \frac{\mu}{2-\mu} \frac{\sigma_n^2}{\sigma_s^2} = \frac{\mu}{2-\mu} \frac{1}{RSB} \tag{3.7}$$

L'expression de la MSD ainsi obtenue, (3.7), aide à la compréhension de l'observation faite par Al-Kindi et Dunlop qui consiste à dire que de fortes composantes de signal aux points de sommation de la structure nommée *Crosstalk Resistant Adaptive Noise Canceller* (CTRANC) entraînent un bruit du gradient important. Au vu de la relation, nous proposons donc de gérer le seul paramètre maîtrisable, c'est-à-dire le pas d'adaptation μ. Pour ce faire, nous proposons simplement de mettre en œuvre un test permettant de décider si l'adaptation du filtre produisant l'estimée du bruit,

en l'occurrence $\mathbf{w_1}$, doit être gelée ou non. Ce choix est entre autre justifié par la remarque 1. De nombreuses méthodes sur la gestion des pas d'adaptation existent dans la littérature et, après avoir cité les plus courantes, nous décrivons notre méthode dans la partie ci-après.

3.2 Approches dédiées à l'amélioration de la gestion des pas

La gestion du pas d'adaptation pour les méthodes adaptatives utilisées dans le traitement de la parole est une étape cruciale qui conditionne les performances des algorithmes. Nous avons vu précédemment que le pas influait sur le bruit du gradient, mais nous savons aussi que sa valeur rythme la rapidité de convergence jusqu'à une valeur limite, $\mu < \frac{2}{\lambda_{max}}$ avec λ_{max} la valeur propre maximale de la matrice d'autocorrélation du signal d'entrée du filtre. Cette valeur limite peut également s'exprimer en fonction de la densité spectrale de puissance du signal reçu ou de la trace de la matrice d'auto-corrélation, mais le point important réside dans la dépendance de λ_{max} vis-à-vis de l'énergie du signal reçu sur les micros. En conséquence, la technique qui consiste à fixer hors ligne un pas maximal unique est adaptée au cas de signaux stationnaires pour lesquels on connaît et maîtrise la dynamique maximale. Dans le cas de signaux de parole, par définition non-stationnaires et dont la dynamique évolue rapidement, une normalisation est indispensable (NLMS) afin de conditionner les éventuels pics d'énergie et ainsi garantir la stabilité et la convergence. Le pas associé au filtre $\mathbf{w_i}$ sera donc toujours de la forme $\mu = \frac{cste}{Energie_i}$ avec des règles et paramètres supplémentaires qui permettront de le piloter en fonction de son dual, $\mathbf{w_j}$ avec $j \neq i$. Basiquement, le terme d'énergie dénoté $Energie_i$ est calculé à partir du produit scalaire du signal mémorisé par bloc de taille L correspondant à la longueur du filtre $\mathbf{w_i}$ sur la voie i. En ce sens, pour satisfaire aux conditions de stabilité des filtres adaptatifs constituant le CTRANC, Zinser et Mirchandani proposent d'utiliser la puissance reçue sur chacune des voies afin de déterminer une valeur normalisée du pas. Ainsi, une expression de μ_i qui satisfasse au comportement souhaité est définie en (3.8) (cf. [202]), où (a_i,b_i) sont des constantes. On note $\sigma_i^2(n)$ l'estimation de la puissance du signal considéré à l'instant n sur la voie i.

$$\mu_i(n) = \frac{a_i}{\sigma_i^2(n) + b_i} \tag{3.8}$$

où $\sigma_i^2(n) = (1-\alpha)\sigma_i^2(n-1) + \alpha y_j^2(n)$ correspond à une estimée de la puissance par lissage exponentiel. L'indice j évolue de manière symétrique à i, c'est-à-dire que le calcul de la puissance sur la voie 1 associée au signal $x_1(n)$ fait intervenir l'erreur (l'estimée) y_j calculée au même instant sur la voie 2 et associée au signal $x_2(n)$, et inversement.

3.2.1 Règles algorithmiques pour la gestion des pas

Nous donnons dans cette partie deux exemples de méthodes adaptées aux filtres duaux pour la définition de règles dictant l'évolution des pas de chacun des filtres. D'autres méthodes issues de la littérature et traitant de ce sujet, notamment pour l'ANC, pourraient être transposées au CTRANC. D'ailleurs, pour n'en citer que quelques unes, nous renvoyons le lecteur aux références suivantes :

- Park et al. [140] : l'interprétation géométrique de la MSD leur fournit une courbe objectif qui devra être suivie à chaque instant en ajustant les pas d'adaptation. Ces pas d'adaptation sont pré-définis et stockés dans une mémoire. La taille de la mémoire est à fixer en fonction de la granularité voulue tout en tenant compte du coût et du contexte ;
- Benesty et al. [17] ($VSS\ NLMS$) : obtention d'un pas variable non paramétrique. Le nouvel algorithme est dérivé de manière élégante, sans hypothèse particulière et est en quelque sorte une première généralisation de la méthode pour la famille des algorithmes à pas variable. Ces travaux s'appuient sur ceux de Kwong et Johnston [107] ainsi que sur ceux de Aboulnasr et Mayyas [2] ;
- Mathews et Xie [125] : pilotage du pas par un algorithme du gradient qui vise à minimiser l'erreur d'estimation au carré ;
- Cui et al. [43] ($PNLMS$, $IPNLMS$) : l'évolution de chacun des coefficients du filtre adaptatif est guidée par un pas qui lui est propre. Ce contrôle fin du filtre d'abord proposé par Duttweiler dans [51] a été amélioré par Benesty et Gay [15, 16].

Il existe donc deux classes de méthodes. Celles qui, à l'aide d'une fonction de coût à minimiser, vont guider l'évolution des pas de manière adaptative. Et celles qui s'appuient sur des fonctions objectifs afin d'utiliser les valeurs optimales des pas stockées dans une mémoire. Comme nous allons le voir ci-après, ces deux types d'approches sont transposées aux structures bicapteurs.

- *Méthode développée par Ikeda* [88, 89]

Cette technique propose de faire évoluer μ dans un intervalle $[\mu_{min}, \mu_{max}]$ en fonction des niveaux énergétiques présents sur chacun des microphones et du filtre auquel ils sont associés, dans le but notamment de réduire la réverbération. En effet, ces pas d'adaptation doivent agir de manière inverse puisque pour l'un il s'agit d'apprendre le bruit et pour l'autre le signal de diaphonie. Pour le pas associé à $\mathbf{w_{21}}$, le comportement est le suivant :

- $\mu(n) \rightarrow \mu_{max}$ en période de bruit seul afin de faire converger $\mathbf{w_{21}}$,
- $\mu(n) \rightarrow \mu_{min}$ en phase de parole utile seule afin de stabiliser l'estimée de $\mathbf{a_{21}}$

faite par $\mathbf{w_{21}}$ et donc ne pas perturber l'apprentissage par le signal de parole fuyant dans la référence de bruit,

– $\mu \in [\mu_{min}, \mu_{max}]$ en phase de parole bruitée. Il faut que $\mu(n) \rightarrow \mu_{min}$ si la parole utile est prépondérante devant le bruit et que $\mu(n) \rightarrow \mu_{max}$ dans le cas contraire.

Le raisonnement inverse vaut pour le pas associé à $\mathbf{w_{12}}$. Les valeurs seuils de ces pas sont prises dans l'intervalle [0, 0.1]. Elles sont attribuées en fonction des valeurs des rapports signaux à bruit sur chacune des voies, eux-mêmes seuillés. Finalement, un moyen simple choisi par les auteurs pour se conformer à ce comportement est d'utiliser une fonction affine qui va décrire l'évolution du pas (linéaire dans ce cas), dans cet intervalle à l'aide des valeurs des RSB calculés au cours du temps ainsi que des seuils qui leurs sont associés. Ceci est schématisé à la Figure 3.2 alors que la Figure 3.3 décrit la struture dans son ensemble. On identifie deux étages de traitement synchronisés par l'ajout d'un retard, bloc z^{-L}. Ceux-ci sont mis en oeuvre pour permettre à l'étage inférieur de calculer une estimation de deux RSB (RSB_1 et RSB_2) dont dépendent les pas ($\mu_{w_{21}}$ et $\mu_{w_{12}}$) ainsi rendu adaptatifs. Le premier RSB estimé relié à $f(.)$ est associé à un filtrage adaptatif traditionnel, c'est-à-dire que l'on cherche à obtenir une estimée du bruit (e_2) pour le soustraire au signal de parole (e_1). Le second RSB estimé relié à $g(.)$ est une structure analogue mais placée de manière symétrique afin d'obtenir une estimation du signal de diaphonie pour la soustraire au bruit. Ainsi, les puissances calculées pour la parole et le bruit sont *a priori* plus précises que dans le cas monocapteur et permettent d'affiner au maximum les valeurs des RSB permettant de piloter les pas des deux filtres adaptatifs de l'étage supérieur. Finalement, l'objectif est ainsi de guider l'adaptation afin que le filtrage des signaux de parole et de bruit soit effectué aux instants adéquats. De plus, cette structure permet de réduire en partie la distorsion introduite lorsque le filtre $\mathbf{w_{21}}$ converge. Dans [156], Sato complètera la méthode en rendant les pas des filtres SAF_1 et SAF_2 eux-mêmes adaptatifs au prix d'une complexité accrue puisqu'il faudra fixer des seuils supplémentaires pour ces deux filtres, avec toute la difficulté que cela représente.

- *Méthode développée par Zhang* [201]

Il s'agit de réduire la complexité de la structure mise en œuvre par Ikeda pour le calcul du pas. Pour cela, il propose d'utiliser deux DAV. L'une fondée sur le calcul de l'énergie, l'autre sur la corrélation des signaux reçus et en sortie des filtres afin de piloter un système de décision permettant de sélectionner, via un compteur, la valeur optimale des pas prédéfinis et fonction du contexte.
Ainsi, il n'y a plus de continuité dans l'évolution des pas en dehors des zones seuils.

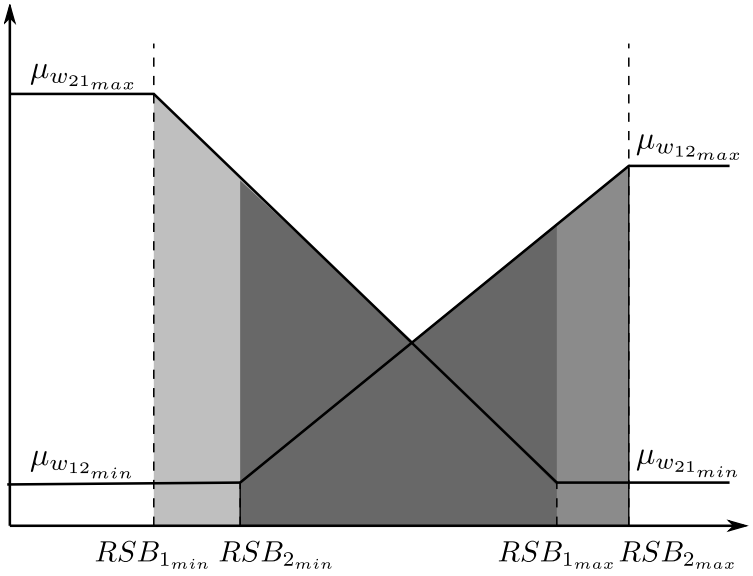


Figure 3.2 – Illustration de l'évolution des pas par rapport aux RSB en fonction de leurs seuils maximum et minimum sur chacune des voies.

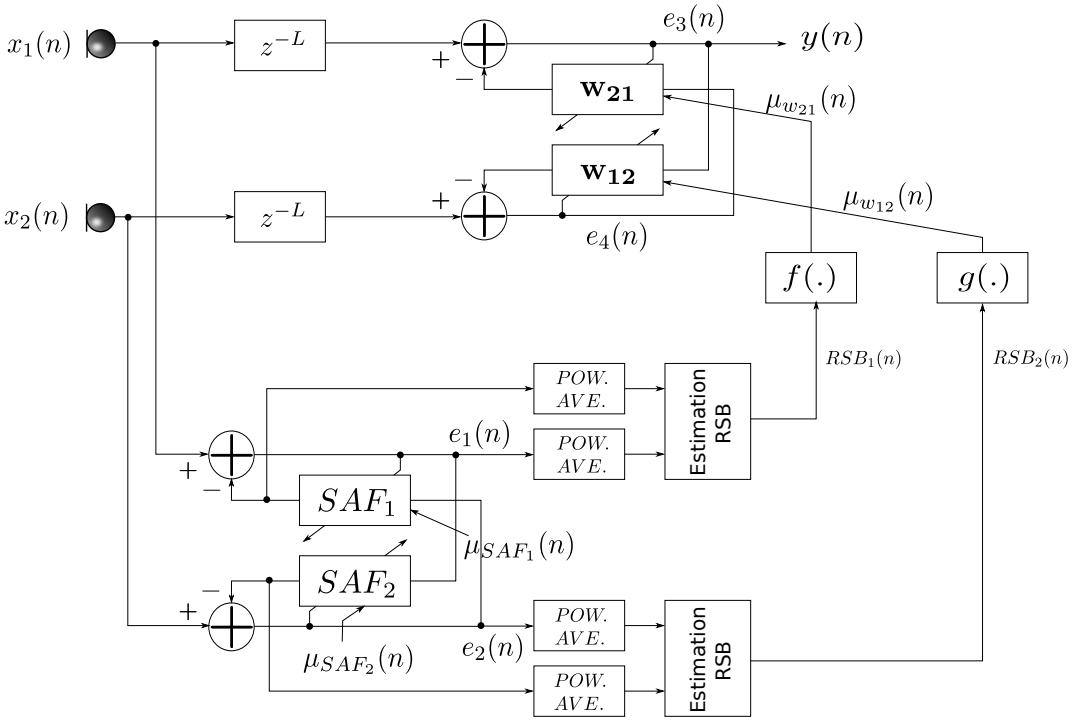


Figure 3.3 – Structure proposée par Ikeda et Sugiyama. Les blocs SAF_x (*Sub Adaptive Filter*) sont des filtres adaptatifs. Les blocs $POW.\ AVE.$ (*Power average*) correspondent au calcul de la puissance et les deux blocs $f(.)$ et $g(.)$ traduisent l'utilisation de fonction permettant le calcul des pas en fonction des RSB estimés.

Six valeurs de pas différentes sont utilisées, celles-ci étant indexées par les deux types de DAV mises en œuvre. Les filtres adaptatifs ($SAF_{1/2}$) nécessaires aux calculs des RSB de la méthode précédente sont donc supprimés.

3.2.2 Bilan des méthodes existantes

Le bruit introduit par l'algorithme du gradient, amplifié par l'interdépendance des sorties de la structure à filtres duaux, empêche une bonne convergence des filtres $\mathbf{w_{21}}$ et $\mathbf{w_{12}}$ (voir [10]) déjà pénalisée par l'utilisation d'un algorithme peu précis. Lorsque cela est possible, l'introduction d'une connaissance *a priori* sur l'un des filtres et consistant à supposer que l'un d'eux est connu par avance permet de pallier ce problème. Le cas échéant, l'utilisation d'une DAV pour la mise à jour aux instants adéquats de chacun des filtres est une solution inévitable pour permettre un fonctionnement efficace, en particulier dans le traitement de la parole. Les méthodes exposées précédemment sont donc un moyen de résoudre ce problème mais souffrent, en plus du problème du bruit du gradient, d'une part de la complexité qu'elles introduisent et d'autre part d'un libre choix difficile sur la valeur des seuils pour la gestion des pas. Or, ceux-ci ne peuvent être choisis au hasard. Si la méthode proposée par Zhang permet de réduire la complexité en termes de coût de calcul par rapport à la méthode de Ikeda, il reste cependant de nombreux seuils à fixer et ceci pour chacun des blocs utilisés : à savoir quatre blocs de DAV plus un système de décision fixant les valeurs des pas à attribuer.
Enfin, dans un contexte très adverse, la mise en œuvre d'une DAV est souvent un handicap très pénalisant car les fausses détections entraînent des distorsions très marquées sur le signal de sortie.

3.2.3 Méthodes proposées

La première solution que nous allons décrire ci-après est basée sur la méthode de Ikeda. L'objectif est néanmoins de montrer qu'il est possible de l'améliorer en y ajoutant une simple contrainte dans le but de réduire le bruit du gradient. La seconde méthode quant à elle a été conçue pour réduire de façon drastique la complexité du bloc de gestion des pas ainsi que le nombre de paramètres fixés empiriquement.

Solution 1

Dans nos travaux publiés dans [112], nous avons utilisé et amélioré le mécanisme de contrôle des pas proposé par Ikeda en procédant comme suit. Si la valeur calculée du pas $\mu_{12}(n)$ associé au filtre dédié à la gestion de la diaphonie est différente de sa valeur minimale alors nous forçons le pas $\mu_{21}(n)$ associé au filtre dédié à l'apprentissage du bruit à zéro. Ainsi, en garantissant que les filtres ne s'adaptent pas

conjointement dans la phase critique où l'énergie du signal de diaphonie est significative, nous réduisons le bruit du gradient.

Remarque 2 : *Cette condition supplémentaire sous-entend que le bruit perturbateur ne correspond pas structurellement à un signal de parole de sorte que la distinction entre le signal de diaphonie et le bruit perturbateur soit suffisamment aisée.*

Par ailleurs, les seuils fixés pour les RSB ont été choisis pour couvrir une plage de RSB en entrée en accord avec une utilisation de la vie courante, soit pour des RSB supérieurs à 5 dB. Ces seuils sont les suivants :

$$RSB_{1_{min}} = 5 \text{ dB et } RSB_{1_{max}} = 15 \text{ dB}$$
$$RSB_{2_{min}} = -8 \text{ dB et } RSB_{2_{max}} = 14 \text{ dB}$$

Ces réglages associés à la condition de non-adaptation conjointe garantissent que :

1. les filtres suivent les non-stationnarités des signaux,
2. le pas associé au filtre dédié à l'apprentissage du bruit se comporte quasiment comme un détecteur de bruit, soit l'inverse d'une DAV.

Nous donnons ci-après la description algorithmique de la méthode, **Algorithme 1**, accompagnée du diagramme d'activité, Figure 3.4. Les indices 1 et 2 font référence aux différentes voies identifiées sur les Figures 3.1 et 3.3. Les fonctions f et g sont des fonctions (non restreintes à des fonctions affines) qui décrivent l'évolution des pas d'adaptation en fonction des valeurs extrémales de ces derniers (fixées par avance) et des seuils des RSB décrits ci-dessus. Les lignes 16 et 17 reflètent la condition qui gèle l'adaptation du filtre $\mathbf{w_{21}}$ dès lors que l'énergie du signal de diaphonie est significative et peut perturber l'adaptation de ce même filtre.
Les Figures 3.5 à 3.7 donnent les résultats de simulation obtenus dans le cas d'un mélange anéchoïque simplifié correspondant à une scène sonore créée à l'aide de l'outil *Roomsim*. Elle correspond au cas où les deux sources sont placées de manière symétrique à 40 *cm* des microphones. La distance inter-capteur est de 15 *cm* ce qui correspond, peu ou prou, à la distance interaurale. Cette scène est entièrement décrite en Annexe D. Le mélange est fait de deux doubles phrases, l'une prononcée par un homme, l'autre par une femme. Le signal perturbateur est un bruit de bureau réglé à l'aide de l'outil SV56 pour un RSB en entrée de 10 dB. L'influence de la condition de non-adaptation conjointe sur les instants propices à la diaphonie est visible à la Figure 3.5. Pour cette configuration, les résultats estampillés DCTRANC sont en fait obtenus sans modification de la structure CTRANC, seule la condition

Algorithme 1 Algorithme de gestion des pas basé sur la méthode de Ikeda

Début
si $RSB_1 > RSB_{1_{max}}$ **alors**
  $\mu_{w_{21}} \leftarrow \mu_{w_{21_{min}}}$
sinon si $RSB_1 < RSB_{1_{min}}$ **alors**
  $\mu_{w_{21}} \leftarrow \mu_{w_{21_{max}}}$
sinon
  $\mu_{w_{21}} \leftarrow f\left(\mu_{w_{21_{min}}}, \mu_{w_{21_{max}}}, RSB_{1_{min}}, RSB_{1_{max}}\right)$
fin si
si $RSB_2 < RSB_{2_{min}}$ **alors**
  $\mu_{w_{12}} \leftarrow \mu_{w_{12_{min}}}$
sinon si $RSB_2 > RSB_{2_{max}}$ **alors**
  $\mu_{w_{12}} \leftarrow \mu_{w_{12_{max}}}$
sinon
  $\mu_{w_{12}} \leftarrow g\left(\mu_{w_{12_{min}}}, \mu_{w_{12_{max}}}, RSB_{2_{min}}, RSB_{2_{max}}\right)$
fin si
si $\mu_{w_{12}} <> \mu_{w_{12_{min}}}$ **alors**
  $\mu_{w_{21}} \leftarrow 0$
fin si
Fin

décrite à l'**Algorithme 1** est ajoutée. L'impact de cette condition est visible sur le comportement du pas $\mu_{w_{21}}$ associé au DCTRANC et qui est à comparer à celui de Ikeda. Nous observons très nettement le comportement "binaire" du pas et non plus principalement linéaire. Ce comportement est au bénéfice de la convergence du filtre $\mathbf{w_{21}}$ qui comme nous pouvons le voir à partir de la déviation normalisée de ce même filtre, n'est plus soumis aux fluctuations dues aux composantes de parole. Nous résolvons donc le problème soulevé par Al-Kindi et Dunlop en ce qui concerne l'impact de forte composante de signal au point de sommation de la structure et qui contribue au bruit du gradient. La Figure 3.6 démontre l'apport du premier filtre de découplage (cf. Figure 3.1) sur la suppression de bruit et la précision de l'estimation du filtre $\mathbf{w_{12}}$. En effet, par comparaison avec la déviation normalisée du filtre $\mathbf{w_{12}}$ obtenue à la Figure 3.5, nous observons que la tendance s'améliore même si la convergence n'est pas encore atteinte. Ceci s'explique par le phénomène identifié en 3.1.3 vis-à-vis de l'adaptation du filtre $\mathbf{w_{12}}$ à partir de l'algorithme NLMS "classique", équivalent dans cette configuration à l'algorithme de décorrélation. L'interdépendance des sorties dans le processus d'adaptation n'est pas prise en compte. Par conséquent, l'estimation du filtre est imparfaite et est de plus tributaire des

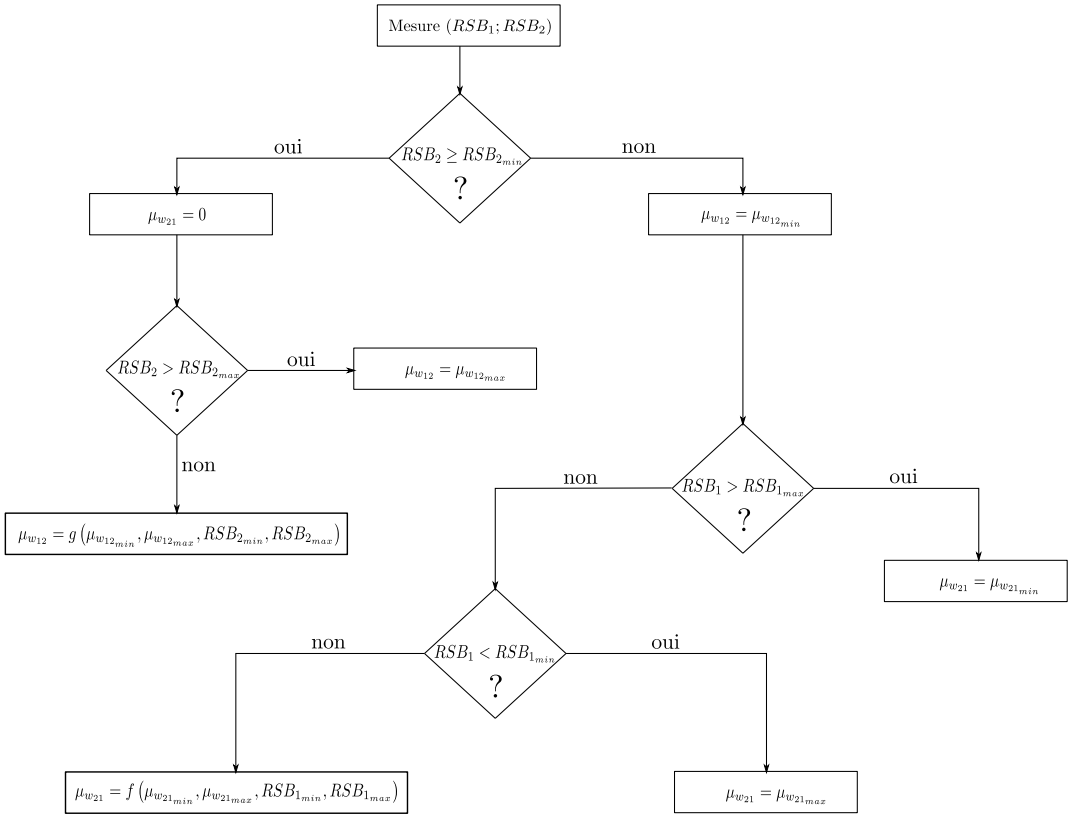


Figure 3.4 – Diagramme d'activités de la gestion des pas fondé sur la méthode de Ikeda. La première condition testée est la condition de non-adaptation conjointe ajoutée à cette méthode afin de réduire le bruit du gradient.

erreurs d'estimation faites sur le filtre $\mathbf{w_{21}}$ qui viennent s'ajouter aux erreurs déjà commises par l'algorithme. L'impact de ces erreurs est visible sur l'estimée produite par le DCTRANC où les instants de divergence et d'oscillations correspondent à ceux observés sur la déviation normalisée du filtre $\mathbf{w_{12}}$ associé au DCTRANC. Ces défauts sont complètement corrigés par la version hybride, voir Figure 3.7, ce qui prouve également que l'algorithme de décorrélation ou du gradient "classique" est en adéquation avec une implémentation des filtres selon une structure directe et non pas récursive. Le signal estimé est sans artefacts, clairement débruité et les filtres présentent une bonne convergence.

Néanmoins et bien que la solution proposée soit efficace, celle-ci est biaisée. En effet, dans la méthode de Ikeda l'étage de gestion des pas est construit à partir d'une recopie de la structure à filtres duaux. Or, dans leurs travaux, les auteurs utilisent l'algorithme du gradient classique, soit donc de décorrélation, pour la structure récursive et y compris pour la gestion des pas. Ainsi, pour être réellement optimale, la solution devrait également remplacer les filtres duaux de l'étage de gestion des pas par le DCTRANC de sorte à obtenir de meilleures estimées pour les RSB. Ici,

nous nous sommes contentés de corriger uniquement l'étage produisant les estimées des signaux utiles. Nous pourrions donc encore améliorer la méthode mais cela n'est pas dans notre intérêt et ne correspond pas à nos objectifs d'une solution optimale répondant à des contraintes de coût minimal et s'affranchissant au maximum de paramètres réglés de manière empirique.

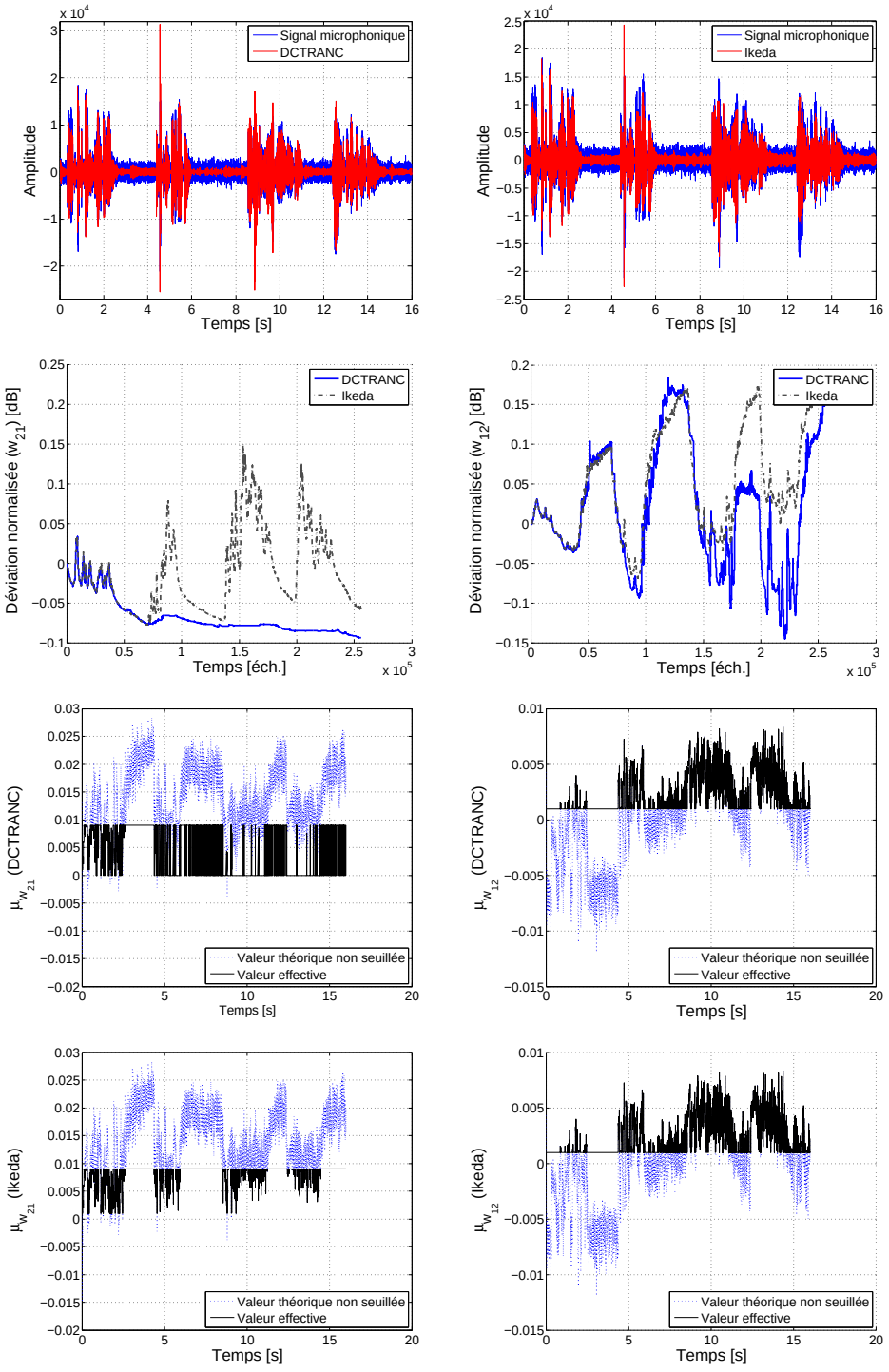


Figure 3.5 – Influence de la condition de non adaptation conjointe pendant les périodes de présence de parole sur les performances de la méthode de Ikeda.

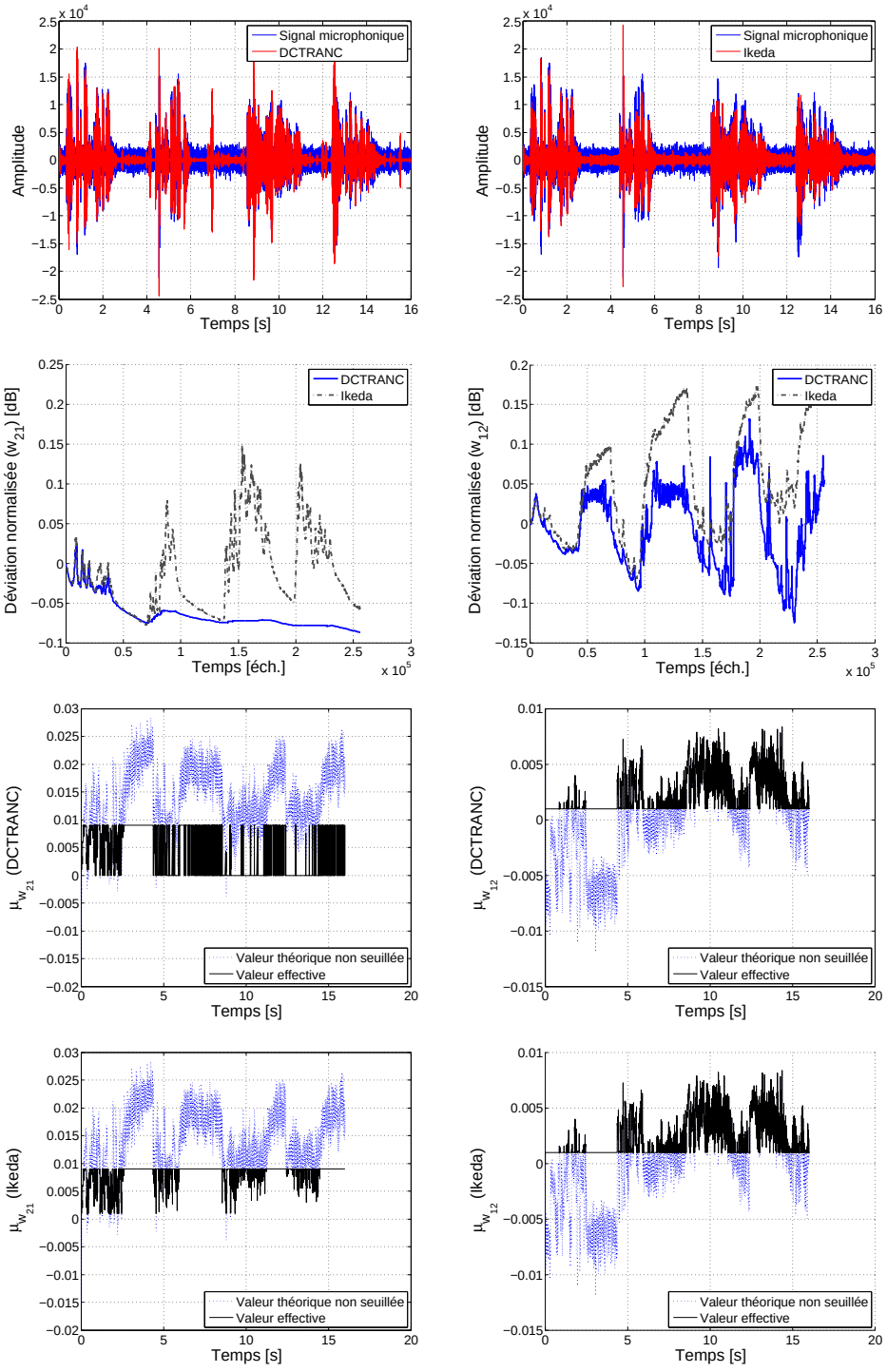


Figure 3.6 – Influence du découplage simple réalisé par le DCTRANC. Cependant, l'inadéquation de l'algorithme avec un processus d'adaptation récursif pour le filtre $\mathbf{w_{12}}$ cause des instabilités et des artefacts sur le signal estimé.

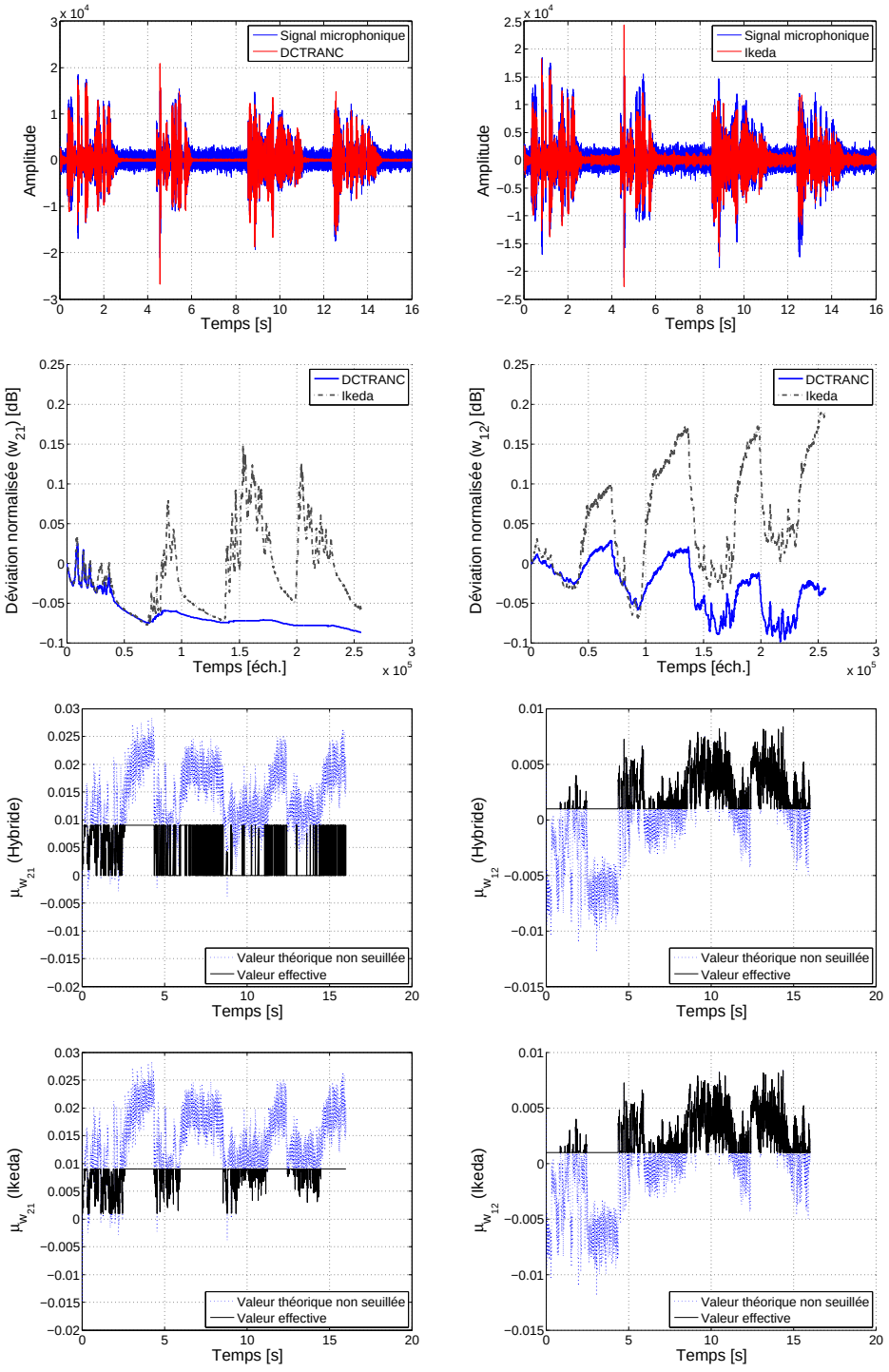


Figure 3.7 – Importance du découplage complet de l'adaptation réalisé par la version hybride du DCTRANC. L'algorithme NLMS "classique" est dans ce cas en parfaite adéquation avec le schéma d'adaptation.

Solution 2

Nous venons de voir qu'il était important de geler l'adaptation des filtres aux moments opportuns. Nous allons conserver ce comportement dans la méthode proposée ci-après et la transposer dans le domaine fréquentiel. Notre but : piloter l'adaptation sur chaque composante fréquentielle de manière plus précise. En pratique, la dynamique des pas d'adaptation pour le traitement de la parole est comprise entre 0 (gel de l'adaptation) et 0,1 (apprentissage maximal). Il nous faut donc trouver une méthode de variation des pas dans le domaine fréquentiel qui soit à l'image du comportement souhaité et étudié précédemment, et ce, pour chacun des filtres.
Pour ce faire, nous allons nous appuyer sur les méthodes monocapteurs de soustraction spectrale, cf. Annexe B, et plus particulièrement sur la technique en deux passes appelée *Two-Step Noise Reduction* (TSNR) décrite dans [141, 142]. Comme toute méthode de ce type, son efficacité est conditionnée par l'estimation de la DSP du bruit permettant le calcul des RSB *a priori* et *a posteriori*. Dans nos travaux, l'estimation de la DSP du bruit a été envisagée de deux manières. La première utilise une DAV idéale et calcule la DSP du bruit par lissage exponentiel pendant les périodes d'inactivité vocale. La dernière valeur calculée est réutilisée pendant les instants de parole. La seconde méthode est fondée sur une technique d'estimation continue décrite dans [75] et connue sous le nom d'automate de Cucchi. La DSP du bruit est obtenue par la pondération de son spectre par un coefficient non linéaire α calculé en fonction du RSB *a posteriori* :

$$\alpha = L + (1 - L) \cdot g \cdot \frac{RSB_{post}}{1 + g \cdot RSB_{post}} \cdot \left(1 + \frac{1}{1 + b \cdot g \cdot RSB_{post}}\right) \tag{3.9}$$

où

$$RSB_{post} = \frac{|X_2(k,l)|^2}{DSP_{bruit}}$$

avec b le gain maximal à appliquer d'une trame à l'autre, $g = \frac{1}{1-b}$ et L la variable qui ajuste la décroissance exponentielle du spectre. Dans la description algorithmique donnée à l'**Algorithme 2**, cette étape d'estimation de la DSP du bruit, nécessaire aux calculs des RSB *a posteriori* (RSB_{post}) et RSB *a priori* (RSB_{prio}), est réalisée par la fonction CalculDSPbruit.

Remarque 3 : $\alpha \approx L$ *durant les périodes de bruit seul.*

L'estimateur du RSB *a priori* de la ligne 8, nommé *Decision-Directed* (DD) [57], est obtenu à partir du calcul de la DSP estimée du signal de parole DSP_W. Cette estimée est calculée à l'aide du $Gain_{W_2}$ correspondant au filtre de Wiener estimé à la

Algorithme 2 Algorithme pour la gestion des pas basé sur la méthode TSNR

Initialisations et définitions des matrices

1: $DSP_{bruit} \leftarrow |X_2|^2 \qquad DSP_W \leftarrow [0,\ldots,0]^T$

2: $Gain_{W_1} \leftarrow [0,\ldots,0]^T \quad Gain_{W_2} \leftarrow [0,\ldots,0]^T \quad Gain_{W_{2_{seuillé}}} \leftarrow [0,\ldots,0]^T$

3: Début

4: **pour chaque nouveau bloc de M échantillons faire**

5: CalculDSPbruit(X_2)

6: $RSB_{post} \leftarrow \frac{|X_2|^2}{DSP_{bruit}}$ //Calcul du RSB a posteriori

7: $DSP_W \leftarrow Gain^2_{W_2}|X_2|^2$

8: //Première passe

$RSB_{prio} \leftarrow \lambda\frac{DSP_W}{DSP_{bruit}} + (1-\lambda)RSB_{post}$ //Calcul du RSB a priori

9: $Gain_{W_2} \leftarrow \frac{RSB_{prio}}{1+RSB_{prio}}$ //Calcul du gain

10: //Seconde passe

$RSB_{prio} \leftarrow Gain^2_{W_2}RSB_{post}$ //Calcul du RSB a priori

11: $Gain_{W_2} \leftarrow \frac{RSB_{prio}}{1+RSB_{prio}}$ //Calcul du gain

12: //Mise en forme des gains et condition de non adaptation conjointe

$Gain_{W_1} \leftarrow \frac{1}{1+RSB_{prio}}$

13: $Gain_{W_{2_{seuillé}}} \leftarrow Gain_{W_2} \underbrace{(Gain_{W_2} > seuil)}_{valeur\ binaire\ (0\ ou\ 1)}$ //Masque

14: **fin pour**

15: Fin

trame précédente. Ensuite, nous reconnaissons les deux passes caractéristiques de la méthode TSNR qui permettent le calcul du RSB *a priori* à la trame $l+1$ dans le but de corriger le biais de l'estimateur DD. En effet, l'estimée du RSB *a priori* produite à la ligne 10 est obtenue de manière similaire à celle de la ligne 8, avec pour unique différence le fait que nous fixons $\lambda = 1$ afin de ne pas ajouter un retard algorithmique supplémentaire [141]. Pour s'en convaincre, observons l'estimée théorique du RSB *a priori* de la ligne 10 tel que nous l'utilisons :

$$\begin{aligned} RSB_{prio}(k,l+1) &= \lambda\frac{DSP_W(k,l)}{DSP_{bruit}(k,l)} + (1-\lambda)RSB_{post}(k,l+1) \\ &= \lambda\, Gain^2_{W_2}RSB_{post}(k,l) + (1-\lambda)RSB_{post}(k,l+1) \end{aligned} \tag{3.10}$$

En fixant $\lambda = 1$, il n'est plus nécessaire de disposer de la trame future $X_2(k,l+1)$ pour le calcul de $RSB_{post}(k,l+1)$ et donc l'expression du RSB *a priori* se résume à

la relation (3.11) suivante.

$$RSB_{prio}(k,l) = Gain_{W_2}^2 RSB_{post}(k,l) \tag{3.11}$$

Au final, nous obtenons grâce à cet algorithme les deux gains souhaités, c'est-à-dire que ceux-ci évoluent effectivement de manière symétrique et permettent d'obtenir deux nouveaux pas d'adaptation non paramétriques appropriés au pilotage des filtres $\mathbf{w_1}$ et $\mathbf{w_2}$. La Figure 3.8 illustre le comportement des gains à la fréquence discrète $f = 1000\ Hz$. Sur cet exemple, nous choisissons de geler l'adaptation du filtre $\mathbf{w_2}$ dès lors que l'amplitude de $Gain_{W_1}$ est supérieure à 0,5. Les pas variables

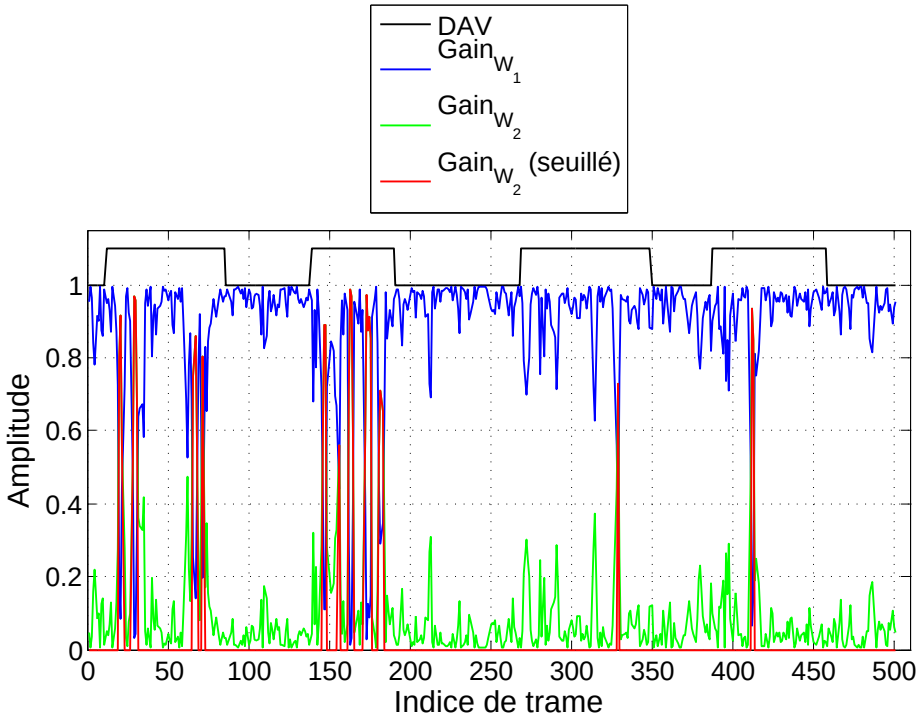


Figure 3.8 – Exemple de comportement des gains de l'**Algorithme 2** pour un bruit de bureau à 10 dB.

qui en découlent sont définis aux équations (3.12) et (3.13), où les constante $cste_{W_1}$ et $cste_{W_2}$ permettent d'ajuster la dynamique de ces derniers afin qu'ils soient compatibles avec le traitement de la parole.

$$\mu_{W_1}(k,l) = \frac{Gain_{W_1}(k,l)}{cste_1} \tag{3.12}$$

$$\mu_{W_2}(k,l) = \frac{Gain_{W_2}(k,l)}{cste_2} \tag{3.13}$$

Typiquement, nous fixerons $cste_{W_2}$ et $cste_{W_1}$ de sorte à garantir un bon apprentissage du bruit et éviter les divergences du filtre $\mathbf{w_2}$ avec un pas trop fort. Notons que la définition du pas μ_{W_2} donnée à l'équation (3.13) ne fait volontairement pas intervenir le gain $Gain_{W_{2_{seuillé}}}$, ce qui peut paraître surprenant compte tenu des résultats précédents (Solution 1). Cependant, ce choix se justifie par les raisons suivantes :

1. les pas d'adaptation sont calculés dans le domaine fréquentiel, par conséquent ils agissent par bande de fréquence et sont donc plus fins que ceux fixés de manière globale dans le domaine temporel. En d'autres termes, nous bénéficions d'une granularité plus fine simplement en travaillant dans le domaine fréquentiel. De plus, la complémentarité observée sur chaque bande de fréquence permet déjà un pilotage fin des filtres adaptatifs,
2. en regard de l'évolution indépendante de chacune des composantes fréquentielles des pas $\mu_{W_i}(k,l)$ $i = 1, 2$; fixer un seuil identique et arbitraire est une solution pauvre qui en quelque sorte dégrade le potentiel de la solution proposée. En effet, plutôt que d'inclure un paramètre réglé de manière empirique, nous pourrions utiliser une modélisation paramétrique de la source utile pour calculer les seuils optimaux pour chaque bande de fréquence,
3. les gains ne seront obtenus sans erreurs que si une DAV idéale est utilisée pour garantir une bonne estimation de la DSP du bruit. Or, dans la pratique, la définition d'une DAV robuste relève d'une tâche difficile, en particulier lorsque le bruit est non-stationnaire. En imaginant simplement que le mélange soit fait de deux sources de parole dont une est à extraire, nous percevons la difficulté à discerner la source utile de la source interférente, et par conséquent à définir une DAV efficace. Une définition du seuil qui a un sens pratique doit tenir compte de cet aspect.

Selon nous, une étude supplémentaire est donc nécessaire, c'est pourquoi nous utiliserons les définitions données aux équations (3.12) et (3.13) dans nos simulations. Mais il est certain que les résultats qui seront présentés sont perfectibles à l'aide de l'opération de seuillage.

3.3 Etude des solutions optimales

Dans cette partie, nous allons décrire les solutions optimales des filtres $W_1(z)$ et $W_2(z)$ au sens du minimum de l'erreur quadratique moyenne.

3.3.1 Solutions analytiques théoriques

Supposons que l'on dispose des deux observations $x_1(n)$ et $x_2(n)$ pour $n = 0, \ldots, N-1$ et observons tout d'abord ce qui se passe pour le filtre $W_1(z)$. No-

tons $W_1^{opt}(z)$ cette solution optimale, celle-ci s'obtient en calculant :

$$W_1^{opt}(z) = \frac{\Phi_{X_1X_2}(z)}{\Phi_{X_2X_2}(z)} \tag{3.14}$$

Or, en notant que la transformée de Fourier du processus tronqué $x_i(n)$ sur $(0, N-1)$ s'écrit :

$$X_{i_N}(f) = \sum_{n=0}^{N-1} x_i(n)e^{-j2\pi nfT}$$

nous pouvons en déduire l'expression de l'interspectre donnée ci-après :

$$\Phi_{X_1X_2}(f) = \lim_{N\to\infty} \frac{1}{N} E\Big[X_{1_N}(f)X^*_{2_N}(f)\Big] \tag{3.15}$$

En évaluant cette quantité uniquement sur le cercle unité $\mathcal{C}_1$, celle-ci peut être transposée dans le domaine de la transformée en Z comme suit :

$$\begin{array}{rcl} \Phi_{X_1X_2}(z) & = & \lim\limits_{N\to\infty} \frac{1}{N} E\Big[X_{1_N}(z)X^*_{2_N}(\frac{1}{z^*})\Big] \\ & \overset{(Signaux\ réels)}{=} & \lim\limits_{N\to\infty} \frac{1}{N} E\Big[X_{1_N}(z)X^*_{2_N}(z^{-1})\Big] \end{array}$$

Pour alléger l'écriture, nous ometterons dans ce qui suit l'indice N signifiant que le processus est observé sur une fenêtre de longueur N. En faisant intervenir les grandeurs constituant le mélange, nous obtenons :

$$\begin{aligned} &E\left[X_1(z)X_2^*(z^{-1})\right] = \\ &\qquad E\left[\left(A_{11}(z)S_1(z) + A_{21}(z)S_2(z)\right)\left(A_{22}(z^{-1})S_2(z^{-1}) + A_{12}(z^{-1})S_1(z^{-1})\right)\right] \end{aligned}$$

En supposant que s_1 et s_2 sont décorrélés, nous obtenons :

$$\Phi_{X_1X_2}(z) = A_{11}(z)A_{12}(z^{-1})\Phi_{S_1S_1}(z) + A_{21}(z)A_{22}(z^{-1})\Phi_{S_2S_2}(z) \tag{3.16}$$

$$\Phi_{X_2X_2}(z) = A_{22}(z)A_{22}(z^{-1})\Phi_{S_2S_2}(z) + A_{12}(z)A_{12}(z^{-1})\Phi_{S_1S_1}(z) \tag{3.17}$$

En restreignant le domaine d'analyse au cercle unité $\mathcal{C}_1$, (3.17) est équivalente à :

$$\Phi_{X_2X_2}(z) = |A_{22}(z)|^2\Phi_{S_2S_2}(z) + |A_{12}(z)|^2\Phi_{S_1S_1}(z) \tag{3.18}$$

Ainsi, on en déduit la solution optimale suivante :

$$W_1^{opt}(z) = \frac{A_{11}(z)A_{12}(z^{-1})\Phi_{S_1S_1}(z) + A_{21}(z)A_{22}(z^{-1})\Phi_{S_2S_2}(z)}{|A_{22}(z)|^2\Phi_{S_2S_2}(z) + |A_{12}(z)|^2\Phi_{S_1S_1}(z)} \tag{3.19}$$

Dans le cas du DCTRANC, cette solution est ré-injectée dans $W_{21}(z)$ et il en sera de même pour le couple ($W_2(z)$,$W_{12}(z)$). Nous étudions maintenant les signaux d'erreurs $d_1(n)$ et $d_2(n)$ qui servent uniquement à la mise à jour des filtres $\mathbf{w_1}$ et $\mathbf{w_2}$. En

l'occurrence, c'est la principale différence entre l'étage de découplage du DCTRANC et la structure directe pour qui ces signaux d'erreurs sont également les sorties du système. Par conséquent, nous allons au travers le DCTRANC établir un parallèle entre la structure directe et la structure récursive.

$$\begin{aligned} D_1(z) &= X_1(z) - W_1^{opt}(z)X_2(z) \\ &= \left[A_{11}(z) - \frac{A_{11}(z)|A_{12}(z)|^2\Phi_{S_1S_1}(z) + A_{21}(z)A_{22}(z^{-1})A_{12}(z)\Phi_{S_2S_2}(z)}{|A_{22}(z)|^2\Phi_{S_2S_2}(z) + |A_{12}(z)|^2\Phi_{S_1S_1}(z)}\right] S_1(z) \\ &+ \left[A_{21}(z) - \frac{A_{11}(z)A_{12}(z^{-1})A_{22}(z)\Phi_{S_1S_1}(z) + A_{21}(z)|A_{22}(z)|^2\Phi_{S_2S_2}(z)}{|A_{22}(z)|^2\Phi_{S_2S_2}(z) + |A_{12}(z)|^2\Phi_{S_1S_1}(z)}\right] S_2(z) \end{aligned}$$

Après simplification, le signal d'erreur $D_1(z)$ s'écrit :

$$\begin{aligned} D_1(z) &= \left[\frac{A_{11}(z)|A_{22}(z)|^2 - A_{21}(z)A_{22}(z^{-1})A_{12}(z)}{|A_{12}(z)|^2\Phi_{S_1S_1}(z) + |A_{22}(z)|^2\Phi_{S_2S_2}(z)}\Phi_{S_2S_2}(z)\right] S_1(z) \\ &+ \left[\frac{A_{21}(z)|A_{12}(z)|^2 - A_{11}(z)A_{12}(z^{-1})A_{22}(z)}{|A_{12}(z)|^2\Phi_{S_1S_1}(z) + |A_{22}(z)|^2\Phi_{S_1S_1}(z)}\Phi_{S_2S_2}(z)\right] S_2(z) \end{aligned} \tag{3.20}$$

A partir des équations (3.19) et (3.20), nous sommes en mesure de qualifier les solutions selon les trois cas décrits ci-après.

1. Supposons qu'une barrière acoustique résolve le problème de la diaphonie, soit donc $A_{12}(z) = 0$.

$$\begin{aligned} W_1^{opt}(z) &\approx \frac{A_{21}(z)A_{22}(z^{-1})\Phi_{S_2S_2}(z)}{|A_{22}(z)|^2\Phi_{S_2S_2}(z)} \\ D_1(z) &\approx A_{11}(z)S_1(z) \end{aligned} \tag{3.21}$$

   La solution optimale pour le filtre $W_1(z)$ correspond à celle de l'ANC, le signal d'erreur est une version filtrée du signal source.

2. Supposons à présent que nous disposions d'une DAV idéale qui assure la mise à jour des coefficients uniquement pendant les périodes de silence.

$$\begin{aligned} W_1^{opt}(z) &\approx \frac{A_{21}(z)A_{22}(z^{-1})\Phi_{S_2S_2}(z)}{|A_{22}|^2\Phi_{S_2S_2}(z)} \overset{\mathcal{C}_1}{\approx} \frac{A_{21}(z)}{A_{22}(z)} \\ D_1(z) &\approx \left[A_{11}(z) - \frac{A_{21}(z)}{A_{22}(z)}A_{12}(z)\right] S_1(z) \end{aligned} \tag{3.22}$$

   Dans le cas d'un mélange simplifié, il apparaît le terme $[1 - A_{21}(z)A_{12}(z)]$ en facteur de $S_1(z)$ caractéristique de la structure directe. Pour retrouver exactement la source, il faudra ajouter un bloc de postfiltrage. Ce bloc correspond au filtre RII $\widetilde{F}(z)$, déjà étudié au Chapitre 2 et réalisé intrinsèquement par la connexion récursive des filtres dans le cas de la structure à filtres duaux.

3. Supposons enfin que le signal de diaphonie soit prédominant dans le mélange.

$$\begin{aligned} W_1^{opt}(z) &\overset{\mathcal{C}_1}{\approx} \frac{A_{11}(z)}{A_{12}(z)} \\ D_1(z) &\approx \left[A_{21}(z) - \frac{A_{11}(z)}{A_{12}(z)} A_{22}(z)\right] S_2(z) \end{aligned} \tag{3.23}$$

Sachant que, implicitement, nous supposons que s_1 correspond au signal désiré, nous constatons en (3.23) que celui-ci est annulé. Le signal d'erreur d_1 ne contient plus que le signal perturbateur. Par ailleurs, ceci ne correspond pas à un problème de permutation entre les voies puisque en élaborant la même analyse pour le filtre $W_2(z)$ on montre que la solution optimale pour le filtre $W_2(z)$ s'écrit :

$$W_2^{opt}(z) = \frac{A_{12}(z)A_{11}(z^{-1})\Phi_{S_1S_1}(z) + A_{22}(z)A_{21}(z^{-1})\Phi_{S_2S_2}(z)}{|A_{11}(z)|^2\Phi_{S_2S_2}(z) + |A_{21}(z)|^2\Phi_{S_1S_1}(z)} \tag{3.24}$$

Or, dans le cas particulier considéré ici, celle-ci devient :

$$\begin{aligned} W_2^{opt}(z) &\overset{\mathcal{C}_1}{\approx} \frac{A_{12}(z)}{A_{11}(z)} = \frac{1}{W_1^{opt}(z)} \\ D_2(z) &\approx \left[A_{22}(z) - \frac{A_{12}(z)}{A_{11}(z)} A_{21}(z)\right] S_2(z) \end{aligned} \tag{3.25}$$

Les solutions optimales au sens du minimum de l'EQM montrent donc que le signal d'intérêt est complètement supprimé.

Remarque 4 : *Dans l'analyse du comportement de $W_2(z)$, faire l'hypothèse d'un signal de diaphonie prédominant équivaut à supposer que nous disposons d'une DAV idéale qui garantit la mise à jour du filtre uniquement pendant les périodes de présence du signal utile.*

Les trois cas qui viennent d'être considérés décrivent les solutions fournies par la structure directe avant l'opération de postfiltrage. Or, observons ce qui se produit pour les estimées y_1 et y_2 en sortie du DCTRANC et dont les expressions sont données à l'équation (2.15) du Chapitre 2, lorsque l'on ré-injecte les solutions optimales obtenues sous l'hypothèse d'une DAV idéale dans leurs homologues fixes.

$$\begin{aligned} Y_1(z) &= \frac{[1-W_{21}(z)A_{12}(z)]S_1(z)+[A_{21}(z)-W_{21}(z)]S_2(z)}{1-W_{21}(z)W_{12}(z)} \\ Y_2(z) &= \frac{[A_{12}(z)-W_{12}(z)]S_1(z)+[1-W_{12}(z)A_{21}(z)]S_2(z)}{1-W_{12}(z)W_{21}(z)} \end{aligned} \tag{3.26}$$

Avec

$$W_{21}(z) = W_1^{opt}(z) \overset{DAV_idéale}{=} \frac{A_{21}(z)}{A_{22}(z)} \overset{Mélange_simplifié}{=} A_{21}(z)$$

$$W_{12}(z) = W_2^{opt}(z) \overset{DAV_idéale}{=} \frac{A_{12}(z)}{A_{11}(z)} \overset{Mélange_simplifié}{=} A_{12}(z)$$

les estimées deviennent :

$$\begin{aligned} Y_1(z) &= \frac{[1-A_{21}(z)A_{12}(z)]S_1(z)}{1-A_{21}(z)A_{12}(z)} = S_1(z) \\ Y_2(z) &= \frac{[1-A_{12}(z)A_{21}(z)]S_2(z)}{1-A_{12}(z)A_{21}(z)} = S_2(z) \end{aligned} \tag{3.27}$$

Par conséquent, nous venons de démontrer que le DCTRANC correspond à un moyen de réaliser le postfiltrage indispensable à la structure directe. Le DCTRANC est donc équivalent à la structure directe intégrant le postfiltrage. Des travaux concernant la mise en œuvre algorithmique efficace de cette opération de postfiltrage sont donnés dans [45].

3.3.2 Réflexion sur le DCTRANC

Nous venons de démontrer que la solution proposée est en fait une solution intermédiaire, un pont entre les structures récursive et directe. En effet, en cherchant à rendre l'algorithme du gradient "classique" compatible avec la structure récursive nous avons été amené à ajouter un filtre de découplage. Pour plus de stabilité, nous avons vu que l'ajout d'un deuxième filtre de découplage amenant à une adaptation des deux filtres selon une structure directe améliorait les performances, notamment parce que l'algorithme du gradient usuellement utilisé est compatible uniquement dans ce cas de figure. D'un autre point de vue, nous pouvons considérer le DCTRANC comme une nouvelle solution (matérielle) pour la réalisation du bloc de postfiltrage de la structure directe.
En résumé, l'évaluation des performances du DCTRANC donnée ci-après sera l'occasion de comparer l'efficacité des algorithmes et des structures directe et récursive analysées dans ce mémoire.

3.4 Implémentation fréquentielle et performances en environnement réel

L'implémentation fréquentielle du DCTRANC, donnée à l'**Algorithme 3**, s'appuie sur l'algorithme LMS fréquentiel (FLMS) basé sur la technique de l'*Overlap-Save* (OLS) décrit par Shynk dans [165]. Dans nos simulations, la quantité 2μ faisant référence à un pas fixe, sera remplacée par les pas variables non paramétriques proposés au paragraphe 3.2.3. Par conséquent, l'**Algorithme 2** est placé au début de la boucle de l'**Algorithme 3** comme première instruction.

Algorithme 3 Algorithme DCTRANC dans le domaine fréquentiel - Méthode OLS

Initialisations et définitions des matrices

$\mathbf{W_1}(0) \leftarrow [0,\ldots,0]^T \qquad \mathbf{W_2}(0) \leftarrow [0,\ldots,0]^T$

$P_{r_m}(0) = P_{p_m}(0) \leftarrow \epsilon_m \quad m = 0,\ldots,2M-1$

$$\mathbf{g} \leftarrow \begin{bmatrix} \mathbf{I_M} & \mathbf{0_M} \\ \mathbf{0_M} & \mathbf{0_M} \end{bmatrix}$$

ă $\mathbf{k} \leftarrow [\mathbf{0_M}, \mathbf{I_M}]$: fenêtrage $\mathbf{F}$ = matrice de la TFD $2M \times 2M$

Début

pour chaque nouveau bloc de M échantillons faire

//Signal sur la voie de référence et associé à s_2 (Partie Découplage)

$\mathbf{X_r}(k) \leftarrow diag\{\mathbf{F}[x_2(kM-M),\ldots,x_2(kM),\ldots,x_2(kM+M-1)]^T\}$

$\mathbf{E_1}(k) \leftarrow \mathbf{X_r}(k)\mathbf{W_1}(k)$ //Convolution

$\mathbf{e_1}(k) \leftarrow \mathbf{kF}^{-1}\mathbf{E_1}(k)$ //OLS

$\mathbf{d_1}(k) \leftarrow \mathbf{x_1}(k) - \mathbf{e_1}(k)$ //Calcul de l'erreur

$\mathbf{D_1}(k) \leftarrow \mathbf{Fk}^T\mathbf{d_1}(k)$

$P_{r_m}(k) \leftarrow \lambda P_{r_m}(k-1) + (1-\lambda)\left|\mathbf{X_r}_m(k)\right|^2 \quad m = 0,\ldots,2N-1$

$\mu_{\mathbf{r}}(k) \leftarrow diag\{P_{r_0}^{-1}(k),\ldots,P_{r_{2N-1}}^{-1}(k)\}$ //Pas

//Gradient stochastique

$\mathbf{W_1}(k+1) \leftarrow \mathbf{W_1}(k) + 2\mu\mathbf{FgF}^{-1}\mu_{\mathbf{r}}(k)\mathbf{D_1}(k){\mathbf{X_r}}^H(k)$

//Signal sur la voie primaire et associé à s_1 (Partie Découplage)

$\mathbf{X_p}(k) \leftarrow diag\{\mathbf{F}[x_1(kM-M),\ldots,x_1(kM),\ldots,x_1(kM+M-1)]^T\}$

$\mathbf{E_2}(k) \leftarrow \mathbf{X_p}(k)\mathbf{W_2}(k)$ //Convolution

$\mathbf{e_2}(k) \leftarrow \mathbf{kF}^{-1}\mathbf{E_2}(k)$ //OLS

$\mathbf{d_2}(k) \leftarrow \mathbf{x_2}(k) - \mathbf{e_2}(k)$ //Calcul de l'erreur

$\mathbf{D_2}(k) \leftarrow \mathbf{Fk}^T\mathbf{d_2}(k)$

$P_{p_m}(k) \leftarrow \lambda P_{p_m}(k-1) + (1-\lambda)\left|\mathbf{X_p}_m(k)\right|^2 \quad m = 0,\ldots,2N-1$

$\mu_{\mathbf{r}}(k) \leftarrow diag\{P_{r_0}^{-1}(k),\ldots,P_{r_{2N-1}}^{-1}(k)\}$ //Pas

//Gradient stochastique

$\mathbf{W_2}(k+1) \leftarrow \mathbf{W_2}(k) + 2\mu\mathbf{FgF}^{-1}\mu_{\mathbf{r}}(k)\mathbf{D_2}(k){\mathbf{X_p}}^H(k)$

//Signal sur la voie primaire et associé à s_1 (Partie CTRANC)

$\mathbf{Y_1}(k) \leftarrow diag\{\mathbf{F}[y_1(kM-M),\ldots,y_1(kM),\ldots,y_1(kM+M-1)]^T\}$

$\mathbf{U}(k) \leftarrow \mathbf{Y_1}(k)\mathbf{W_2}(k)$ //Convolution

$\mathbf{u}(k) \leftarrow \mathbf{kF}^{-1}\mathbf{U}(k)$ //OLS

$\mathbf{y_2}(k) \leftarrow \mathbf{x_2}(k) - \mathbf{u}(k)$ //Calcul de l'erreur

//Signal sur la voie de référence et associé à s_2 (Partie CTRANC)

$\mathbf{Y_2}(k) \leftarrow diag\{\mathbf{F}[y_2(kM-M),\ldots,y_2(kM),\ldots,y_2(kM+M-1)]^T\}$

$\mathbf{V}(k) \leftarrow \mathbf{Y_2}(k)\mathbf{W_1}(k)$ //Convolution

$\mathbf{v}(k) \leftarrow \mathbf{kF}^{-1}\mathbf{V}(k)$ //OLS

$\mathbf{y_1}(k) \leftarrow \mathbf{x_1}(k) - \mathbf{v}(k)$ //Calcul de l'erreur

fin pour

Fin

3.4.1 Influence de la sous-modélisation

Une deuxième scène sonore, composée de deux sources placées de manière symétrique, décrit le cas où la source de parole est placée à 80 *cm* et la source de bruit à 2 *m* des microphones. Les réponses impulsionnelles sont générées pour un environnement de bureau classique à partir de l'outil *Roomsim*. L'objectif est d'évaluer l'influence de la sous-modélisation sur les performances du DCTRANC pour un mélange simplifié. Les critères utilisés à présent sont la réduction de bruit (**NR**), le gain en RSB (**G**) et la distance cepstrale (**DC**). Ceux-ci sont décrits en Annexe C. Le signal utile est fait de la concaténation de deux doubles phrases, l'une prononcée par un homme, l'autre par une femme. Le mélange résulte de l'ajout d'un bruit de bureau au signal utile. Le RSB en entrée est de 5 *dB*. La longueur de chacun des deux filtres est prise égale à 512 coefficients pour le DCTRANC et 256 coefficients pour l'algorithme SAD. Ce choix se justifie simplement par les conditions d'équivalence sur l'ordre entre le filtrage par un filtre FIR et un filtre RII. La longueur des réponses impulsionnelles notée L_{RI} varie de 256 à 2000 coefficients. Enfin, pour ne pas conditionner les résultats aux performances de la DAV, nous considérons une DAV idéale pour le calcul de la DSP du bruit. Les résultats obtenus au tableau 3.1

	DCTRANC			SAD		
L_{RI}	**NR** [dB]	**G** [dB]	**DC**	**NR** [dB]	**G** [dB]	**DC**
256	9,7	9,8	0,07	**15,9**	**15,5**	0,08
512	**9,6**	**9,8**	0,16	8,3	9	0,14
1024	**4,4**	**5**	0,2	3,8	4,8	0,2
2000	**3,8**	**4,6**	0,2	3,6	4,8	0,3

Table 3.1 – Réduction de bruit (**NR**), gain en RSB (**G**) et distance cepstrale (**DC**) mesurés pour un bruit de bureau de RSB égal à 5 *dB*.

montrent la robustesse de l'algorithme proposé vis-à-vis de la sous-modélisation. En effet, à distance cepstrale quasiment équivalente, les performances de l'algorithme DCTRANC en termes de gain en RSB et de réduction de bruit sont systématiquement supérieures à celles de l'algorithme SAD. Excepté pour le cas $L_{RI} = 256$, où le DCTRANC est victime du bruit lié à la sur-modélisation. Ces résultats prouvent également l'erreur commise lorsque l'on applique le principe d'orthogonalité sans précaution et amenant à l'utilisation de l'algorithme du gradient "classique" (algorithme de décorrélation), incompatible avec la structure à filtre duaux.

3.4.2 Prise de son binaurale

Cette dernière partie compare les performances des deux algorithmes DCTRANC et SAD pour un environnement réel de prise son selon la technique binaurale. Le contexte est décrit en Annexe E. Le mélange se compose de quatre réponses impulsionnelles mesurées en conditions réelles. Or, nous avons considéré jusqu'à présent des structures sans filtrage des trajets directs. Par conséquent, les performances seront sous-optimales et pourraient être améliorées en ajoutant deux filtres derrière chacun des micros. Par exemple, nous pourrions réutiliser les filtres de type Wiener calculés pour la gestion des pas adaptatifs afin de réduire la réverbération et une partie du bruit pour un coût supplémentaire faible. Cependant, notre but ultime n'est pas d'utiliser ces techniques directement en tant que frontal acoustique mais au sein d'une antenne. Nous conservons donc les structures telles qu'elles ont été étudiées jusqu'ici afin d'en apprécier les performances dans les conditions où elles seront utilisées. La même double phrase que précédemment est utilisée pour générer deux mélanges, l'un avec un bruit de bureau, l'autre à partir d'un bruit de restaurant pour deux RSB différents, soit 5 et 10 dB. L'estimation de la DSP du bruit est faite grâce à une DAV idéale. Enfin, dans ce cas particulier d'un mélange réel et par rapport au fait que les techniques ne soient pas en mesure de supprimer la réverbération, le signal de référence utilisé pour la mesure de la distance cepstrale est le signal de parole non bruité convolué par le chemin direct $\mathbf{a_{11}}$.

	DCTRANC			SAD		
RSB [dB]	**NR** [dB]	**G** [dB]	**DC**	**NR** [dB]	**G** [dB]	**DC**
5	7,2	7	0,04	**7,5**	**7,6**	0,05
10	6,2	6	0,04	**6,9**	**6,9**	0,04

Table 3.2 – Réduction de bruit (**NR**), gain en RSB (**G**) et distance cepstrale (**DC**) mesurés pour un bruit de bureau.

	DCTRANC			SAD		
RSB [dB]	**NR** [dB]	**G** [dB]	**DC**	**NR** [dB]	**G** [dB]	**DC**
5	**5,3**	4,8	0,04	5,2	**5**	0,05
10	4,7	4,2	0,04	**5**	**4,8**	0,05

Table 3.3 – Réduction de bruit (**NR**), gain en RSB (**G**) et distance cepstrale (**DC**) mesurés pour un bruit de restaurant.

Les résultats donnés aux tableaux 3.2 et 3.3 montrent, à distance cepstrale équivalente, la supériorité et la robustesse de la structure à filtre duaux sur la structure directe, même lorsque nous utilisons un algorithme sous-optimal. En effet, que ce soit

pour la réduction de bruit ou le gain en RSB, les performances de l'algorithme SAD sont entre 0,2 et 0,9 dB supérieures. Dans ces conditions difficiles et avec des réglages strictement identiques pour les estimateurs statistiques et les pas des filtres adaptatifs, la structure récursive se montre donc plus efficace. Cette dernière remarque est cependant surprenante compte tenu de l'étude précédente et surtout remet en cause l'intérêt du DCTRANC. Pour le moins, nous avons vu précédemment en 4.4.1 que le DCTRANC était plus performant que l'algorithme SAD dans le cas d'un mélange simplifié. Par conséquent, l'infériorité de la structure directe dans le cas d'un mélange réel semble être liée à l'estimation des postfiltres (recopie), éventuellement moins performante que celle que l'on peut trouver dans la littérature [45]. Ceci sous-entend donc que la structure directe est plus sensible à l'influence des chemins directs que la structure récursive. Le cas de mélange réel révèle que cette sensibilité est forte, au point qu'elle semble prendre le pas sur la précision de l'algorithme utilisé.

3.5 Synthèse du chapitre

Dans ce chapitre, une première solution tenant compte du couplage existant dans la structure à filtre duaux a été proposée. Une analyse approfondie a permis de mettre en avant les avantages et les inconvénients de la méthode, en particulier pour le cas d'un mélange réel de signaux. En effet, en explicitant tout d'abord l'équivalence entre le DCTRANC et la structure directe, paragraphe 3.3, nous avons pu mettre en avant deux points :

1. dans un cadre théorique maîtrisé avec un mélange simplifié, le découplage à l'aide des filtres supplémentaires $\mathbf{w_1}$ et $\mathbf{w_2}$ est profitable aux performances du système comparé à l'utilisation de l'algorithme LMS classique ou de décorrélation avec la structure à filtres duaux,

2. l'influence des chemins directs dans le cas d'un mélange réel de signaux est particulièrement significative lorsque les filtres sont adaptés selon une configuration directe. Au paragraphe 3.4.2, nous avons montré que les performances du DCTRANC sont inférieures à celles de l'algorithme SAD avec la structure à filtres duaux. Or, si le DCTRANC se révèle être une solution pour effectuer l'estimation des postfiltres éventuellement moins efficace que celles de la littérature, nous montrons au travers ces expérimentations que l'implémentation récursive est plus robuste et finalement plus simple. En effet, par rapport aux techniques de l'état de l'art qui traitent du calcul des postfiltres, la mise en œuvre de la structure à filtre duaux avec l'algorithme SAD s'avère être une solution simple et efficace qui mérite d'être considérée.

Après avoir établi ce lien entre les structures directe et récursive et dressé les conclusions sur les performances de ces différentes approches, nous allons montrer dans le chapitre suivant qu'une solution algorithmique performante existe, résolvant les problèmes liés à la fois au couplage et à l'estimation des postfiltres.

Chapitre 4

Solution algorithmique optimisée : le LCCTRANC

4.1 Algorithme optimal de Zinser, Mirchandani et Evans 116
4.1.1 Analyse de l'erreur quadratique moyenne 116
4.1.2 Filtres optimaux 120
4.1.3 Algorithme des moindres carrés transverses 122
4.1.4 Evaluation de la complexité 124
4.2 Réduction de la complexité de l'algorithme optimal 126
4.2.1 Concept à la base de nos simplifications algorithmiques 126
4.2.2 Obtention du nouvel algorithme 127
4.2.3 Evaluation de la complexité du LCCTRANC 129
4.3 Comparaison des performances 131
4.3.1 Mélange de signaux synthétiques 131
4.3.2 Mélange de signaux réels 133
4.3.3 Gestion des instabilités - Réduction de l'ordre 135
4.4 Implémentation fréquentielle et performances en environnement réel 139
4.4.1 Influence de la sous-modélisation 139
4.4.2 Prise de son binaurale 141
4.5 Synthèse du chapitre 142

Après avoir mis en avant les spécificités du CTRANC et l'imprécision de l'algorithme de décorrélation pour cette structure au Chapitre 2, nous avons donné, au Chapitre 3, une solution qui a permis de garder un algorithme simple et maîtrisé au dépend d'une architecture plus complexe. Il s'agissait donc d'une première solution dont nous avons montré l'efficacité. Dans ce chapitre, nous revenons sur l'algorithme basé sur les statistiques d'ordre deux et conçu pour cette structure. L'objectif est d'appréhender les performances de cette méthode proposée par Zinser et al. [132, 202] dans le cas du traitement de la parole et également réduire sa complexité pour permettre son utilisation dans notre antenne. Afin d'être en mesure d'apprécier la précision et la complexité de l'algorithme, nous décrivons dans un premier temps les différentes étapes qui conduisent à cet algorithme optimal. Ensuite, nous présentons nos optimisations qui aboutissent à une version efficace, moins coûteuse, transposable dans le domaine fréquentiel et plus stable que l'algorithme initial dans le contexte du traitement de la parole. Les travaux décrits dans ce chapitre ont été en partie publiés à *European Signal Processing Conference* (EUSIPCO) 2009 [114].

4.1 Algorithme optimal de Zinser, Mirchandani et Evans

Nous donnons dans cette partie, l'analyse complète tirée de [132] et menant à l'algorithme stochastique approprié à la structure récursive.

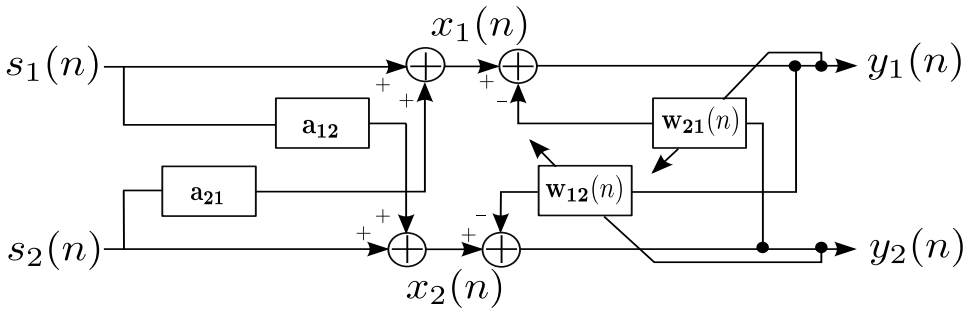


Figure 4.1 – Schéma descriptif du CTRANC proposé par Zinser et al. avec son modèle de mélange.

4.1.1 Analyse de l'erreur quadratique moyenne

On rappelle que les signaux de sorties, également appelés signaux d'erreurs dans la convention habituelle du filtrage adaptatif, sont donnés par :

$$\begin{aligned} y_1(n) &= x_1(n) - \mathbf{w_{21}}^T(n)\mathbf{y_2}(n) \\ y_2(n) &= x_2(n) - \mathbf{w_{12}}^T(n)\mathbf{y_1}(n) \end{aligned} \tag{4.1}$$

En supposant que les filtres $\mathbf{w_{21}}$ et $\mathbf{w_{12}}$ soient fixes, l'erreur quadratique moyenne s'exprime donc de la manière suivante :

$$\begin{aligned}
\xi_1 &= E\left[\left(x_1(n) - \sum_{k=0}^{N} w_{21}^k \left(x_2(n-k) - \sum_{s=0}^{N} w_{12}^s\ y_1(n-k-s)\right)\right)^2\right] \\
&= E\left[x_1^2(n)\right] - 2E\left[\sum_{m=0}^{N} x_1(n)x_2(n-m)w_{21}^m\right] \\
&\quad +E\left[\sum_{k=0}^{N}\sum_{m=0}^{N} x_2(n-k)x_2(n-m)w_{21}^m w_{21}^k\right] \\
&\quad +2E\left[\sum_{t=0}^{N}\sum_{m=0}^{N} x_1(n)y_1(n-t-m)w_{21}^m w_{12}^t\right] \\
&\quad -2E\left[\sum_{k=0}^{N}\sum_{t=0}^{N}\sum_{m=0}^{N} x_2(n-k)y_1(n-m-t)w_{21}^k w_{21}^m w_{12}^t\right] \\
&\quad +E\left[\sum_{k=0}^{N}\sum_{s=0}^{N}\sum_{t=0}^{N}\sum_{m=0}^{N} y_1(n-k-s)y_1(n-m-t)w_{21}^k w_{21}^s w_{12}^m w_{12}^t\right]
\end{aligned} \tag{4.2}$$

En faisant entrer l'espérance statistique dans les sommations et en supposant que x_1, x_2, y_1 et y_2 sont des processus stationnaires au sens large, l'expression précédente devient :

$$\begin{aligned}
\xi_1 &= r_{x_1x_1}(0) - 2\sum_{m=0}^{N} w_{21}^m r_{x_1x_2}(m) + 2\sum_{t=0}^{N}\sum_{m=0}^{N} w_{21}^t w_{12}^m r_{x_1y_1}(m+t) \\
&\quad +\sum_{k=0}^{N}\sum_{m=0}^{N} w_{21}^m w_{21}^k r_{x_2x_2}(k-m) \\
&\quad -2\sum_{k=0}^{N}\sum_{m=0}^{N} w_{21}^k w_{21}^m \left(\sum_{t=0}^{N} w_{12}^t r_{x_2y_1}(t+m-k)\right) \\
&\quad +\sum_{k=0}^{N}\sum_{m=0}^{N} w_{21}^k w_{21}^m \left(\sum_{s=0}^{N}\sum_{t=0}^{N} w_{12}^s w_{12}^t r_{y_1y_1}(m+t-k-s)\right)
\end{aligned} \tag{4.3}$$

avec

$$\begin{aligned}
r_{x_ix_i}(n) &= E\left[x_i(m)x_i(m-n)\right], \quad i = 1,2 \\
r_{y_iy_i}(n) &= E\left[y_i(m)y_i(m-n)\right], \quad i = 1,2 \\
r_{x_iy_i}(n) &= E\left[x_i(m)y_i(m-n)\right], \quad i = 1,2
\end{aligned}$$

Pour parvenir à une écriture matricielle de (4.3), il nous faut porter une attention particulière aux termes dans la triple somme ainsi qu'au dernier terme. Mais avant cela, nous rappelons quelques propriétés sur l'intercorrélation, la relation liant l'intercorrélation avec le produit de convolution et enfin la formule des interférents qui vont

nous être utiles. Posons $u(n)$ et $v(n)$ deux séquences réelles et d'énergie finie, alors nous pouvons écrire, sous l'hypothèse d'ergodicité, les deux relations qui suivent :

$$\boxed{\begin{array}{rcl} r_{uv}(n) & = & r_{vu}(-n) \\ r_{uv}(n) & = & u(n) * v(-n) \end{array}} \tag{4.4}$$

Supposons maintenant que la séquence $v(n)$ soit issue du processus de filtrage de $u(n)$ par un système linéaire invariant de réponse impulsionnelle $h(n)$, alors il existe une relation entre la corrélation du signal de sortie $v(n)$ et la corrélation du signal d'entrée $u(n)$ donnée par la formule des interférents telle que décrit ci-après.

$$\boxed{r_{vv}(n) = r_{uu}(n) * h(n) * h(-n)} \tag{4.5}$$

Ainsi, à partir de (4.4) et (4.5), nous pouvons réécrire les termes de la triple somme comme suit :

$$\begin{array}{rcl} \sum_{t=0}^{N} w_{12}^{t} r_{x_2 y_1}(t+m-k) & = & \sum_{t=0}^{N} w_{12}^{t} r_{y_1 x_2}((k-m)-t) \\ & = & w_{12}^{k-m} * r_{y_1 x_2}(k-m) \\ & = & w_{12}^{k-m} * r_{x_2 y_1}(-(k-m)) \\ & = & r_{w_{12} r_{x_2 y_1}}(k-m) \end{array} \tag{4.6}$$

La triple somme fait donc apparaître la convolution entre l'intercorrélation $r_{y_1 x_2}$ et les coefficients du filtre $\mathbf{w_{12}}$. Observons maintenant le dernier terme de l'expression (4.3) et faisant intervenir une quadruple somme.

$$\begin{aligned} \sum_{k=0}^{N}\sum_{m=0}^{N} & w_{21}^{k} w_{21}^{m} \left(\sum_{s=0}^{N}\sum_{t=0}^{N} w_{12}^{s} w_{12}^{t} r_{y_1 y_1}(m+t-k-s) \right) \\ &= \sum_{k=0}^{N}\sum_{m=0}^{N} w_{21}^{k} w_{21}^{m} \sum_{t=0}^{N} w_{12}^{t} \sum_{s=0}^{N} w_{12}^{s} r_{y_1 y_1}((m-k+t)-s) \\ &= \sum_{k=0}^{N}\sum_{m=0}^{N} w_{21}^{k} w_{21}^{m} \sum_{t=0}^{N} w_{12}^{t} \left[w_{12}^{m-k+t} * r_{y_1 y_1}(m-k+t) \right] \end{aligned}$$

En posant $t' = -t$ et $z(m-k+t) = w_{12}^{m-k+t} * r_{y_1 y_1}(m-k+t)$

$$\begin{aligned} \sum_{k=0}^{N}\sum_{m=0}^{N} & w_{21}^{k} w_{21}^{m} \sum_{t=0}^{N} w_{12}^{t} \left[w_{12}^{m-k+t} * r_{y_1 y_1}(m-k+t) \right] \\ &= \sum_{k=0}^{N}\sum_{m=0}^{N} w_{21}^{k} w_{21}^{m} \sum_{t'=-N}^{0} w_{12}^{-t'} z((m-k)-t') \\ &= \sum_{k=0}^{N}\sum_{m=0}^{N} w_{21}^{k} w_{21}^{m} \sum_{t'=-N}^{0} \widetilde{w}_{12}^{t'} z((m-k)-t') \end{aligned}$$

avec $\widetilde{w}_{12}^{x} = w_{12}^{-x}$. En identifiant la convolution et en re-développant, nous obtenons :

$$\begin{aligned}
\sum_{k=0}^{N}\sum_{m=0}^{N} & w_{21}^{k} w_{21}^{m} \left(\widetilde{w}_{12}^{m-k} * z(m-k)\right) \\
&= \sum_{k=0}^{N}\sum_{m=0}^{N} w_{21}^{k} w_{21}^{m} \left(\widetilde{w}_{12}^{m-k} * w_{12}^{m-k} * r_{y_1 y_1}(m-k)\right) \\
&= \sum_{k=0}^{N}\sum_{m=0}^{N} w_{21}^{k} w_{21}^{m} \left(w_{12}^{k-m} * w_{12}^{m-k} * r_{y_1 y_1}(m-k)\right) \\
&= \sum_{k=0}^{N}\sum_{m=0}^{N} w_{21}^{k} w_{21}^{m} r_{d_2 d_2}(m-k) \qquad (4.7)
\end{aligned}$$

où le signal $d_2 = \sum_{k=0}^{N} w_{12}(k) y_1(n-k)$ désigne le signal en sortie du filtre $\mathbf{w_{12}}$. Ces longs développements permettent d'aboutir à l'expression finale de ξ_1 sous forme matricielle.

$$\begin{aligned}
\xi_1 = & \mathbf{r_{x_1 x_1}}(0) - 2\mathbf{r_{x_1 x_2}}^{T}\mathbf{w_{21}} + \mathbf{2w_{21}}^{\mathbf{T}}\mathbf{R_{x_1 y_1}}\mathbf{w_{12}} \\
& + \mathbf{w_{21}}^{T}\mathbf{R_{x_2 x_2}}\mathbf{w_{21}} - \mathbf{2w_{21}}^{\mathbf{T}}\mathbf{R_{w_{12}}}\mathbf{w_{21}} + \mathbf{w_{21}}^{\mathbf{T}}\mathbf{R_{w_{12} w_{12}}}\mathbf{w_{21}} \qquad (4.8)
\end{aligned}$$

où

$$\mathbf{R_{x_1 y_1}} = \begin{bmatrix} r_{x_1 y_1}(0) & r_{x_1 y_1}(1) & \cdots & r_{x_1 y_1}(N) \\ r_{x_1 y_1}(1) & r_{x_1 y_1}(2) & \cdots & r_{x_1 y_1}(N+1) \\ \vdots & \vdots & \cdots & \vdots \\ r_{x_1 y_1}(N) & r_{x_1 y_1}(N+1) & \cdots & r_{x_1 y_1}(2N) \end{bmatrix}$$

$$\mathbf{r_{x_1 x_2}}^{T} = [r_{x_1 x_2}(0)\ r_{x_1 x_2}(1)\ r_{x_1 x_2}(2) \ldots r_{x_1 x_2}(N)]$$

$$\mathbf{R_{x_2 x_2}} = [r_{x_2 x_2}(i-j)] \quad 0 \leq i, j \leq N$$

$$\mathbf{R_{w_{12}}} = \left[r_{w_{12} r_{x_2 y_1}}(i-j)\right] \quad 0 \leq i, j \leq N$$

$$\mathbf{R_{w_{12} w_{12}}} = [r_{d_2 d_2}(i-j)] \quad 0 \leq i, j \leq N$$

De la même manière, en remplaçant respectivement les quantités x_1, y_1, d_1 et w_{21} par leurs duales x_2, y_2, d_2 et w_{12}, l'erreur quadratique moyenne sur l'autre voie s'écrit :

$$\begin{aligned}
\xi_2 = & \mathbf{r_{x_2 x_2}}(0) - 2\mathbf{r_{x_2 x_1}}^{T}\mathbf{w_{12}} + \mathbf{2w_{12}}^{\mathbf{T}}\mathbf{R_{x_2 y_2}}\mathbf{w_{12}} \\
& + \mathbf{w_{12}}^{T}\mathbf{R_{x_1 x_1}}\mathbf{w_{12}} - \mathbf{2w_{12}}^{\mathbf{T}}\mathbf{R_{w_{12}}}\mathbf{w_{12}} + \mathbf{w_{12}}^{\mathbf{T}}\mathbf{R_{w_{21} w_{21}}}\mathbf{w_{12}} \qquad (4.9)
\end{aligned}$$

Les équations (4.8) et (4.9) mettent en exergue la complexité de cette structure, où ξ_1 et ξ_2 apparaissent respectivement comme des formes quadratiques des vecteurs $\mathbf{w_{21}}$ et $\mathbf{w_{12}}$. Notons que chacune de ces EQM fait intervenir les coefficients des deux filtres. Par conséquent, la démarche classique amenant à une expression explicite de

la solution théorique optimale se révèle inefficace, puisqu'il est impossible d'aboutir à une relation pour l'un des filtres indépendante de l'autre filtre. Ceci est démontré ci-après par le calcul des gradients de chacune de ces quantités.

4.1.2 Filtres optimaux

Selon l'approche des moindres carrés, les filtres optimaux sont obtenus en minimisant les EQM ξ_1 et ξ_2. En accord avec l'approche décrite par l'*Algorithme A* donné en 2.4.2, nous cherchons donc à annuler les gradients de ξ_1 et ξ_2 par rapport à $\mathbf{w_{21}}$ et $\mathbf{w_{12}}$, respectivement.

$$\nabla_{\mathbf{w_{21}}}\xi_1 = \left[\frac{\partial\xi_1}{\partial_{w_{21}^0}}\ \frac{\partial\xi_1}{\partial_{w_{21}^1}}\ \frac{\partial\xi_1}{\partial_{w_{21}^2}}\cdots\frac{\partial\xi_1}{\partial_{w_{21}^N}}\right]^T = \mathbf{0}$$

Cependant et comme souligné par les auteurs, les deux dernières quantités de (4.8) n'étant pas des termes linéaires ou quadratiques à cause de la convolution qui introduit des termes d'ordre supérieur à 2, il est plus facile d'exprimer le gradient à partir des quantités scalaires, soit en calculant :

$$\frac{\partial\xi_1}{\partial_{w_{21}^q}} = 0 \quad 0 \leq q \leq N$$

D'après l'expression (4.3) cette quantité équivaut à :

$$\begin{aligned}
\frac{\partial\xi_1}{\partial_{w_{21}^q}} = & -2r_{x_1x_2}(q) + 2\sum_{k=0}^{N} w_{12}^k r_{x_1y_1}(q+k) \\
& +2\sum_{k=0}^{N} w_{21}^k r_{x_2x_2}(k-q) \\
& -2\sum_{k=0}^{N}\sum_{m=0}^{N} w_{21}^k w_{21}^m r_{x_2y_1}(m+q-k) \\
& -2\sum_{k=0}^{N}\sum_{m=0}^{N} w_{21}^k w_{21}^m r_{x_2y_1}(m+k-q) \\
& +\sum_{k=0}^{N}\sum_{m=0}^{N}\sum_{p=0}^{N} w_{21}^k w_{12}^m w_{12}^p r_{y_1y_1}(m+k-q-p) \\
& +\sum_{k=0}^{N}\sum_{m=0}^{N}\sum_{p=0}^{N} w_{21}^k w_{12}^m w_{12}^p r_{y_1y_1}(m+q-p-k) \\
& = 0 \qquad\qquad \text{pour} \quad 0 \leq q \leq N
\end{aligned}$$

Soit encore

$$\begin{aligned}\frac{\partial \xi_1}{\partial_{w_{21}^q}} =& -2r_{x_1x_2}(q) + 2\left[r_{x_1y_1}(q)\ r_{x_1y_1}(1+q) \dots\ r_{x_1y_1}(n+q)\right]\mathbf{w_{12}} \\ &+2\left[r_{x_2x_2}(-q)\ r_{x_2x_2}(1-q) \dots\ r_{x_2x_2}(N-q)\right]\mathbf{w_{21}} \\ &-2\ \mathbf{w_{21}}^T\mathbf{R_{x_2y_1}}(q)\mathbf{w_{12}} - 2\ \mathbf{w_{21}}^T\widetilde{\mathbf{R}}_{\mathbf{x_2y_1}}(q)\mathbf{w_{12}} \\ &+\left[\mathbf{w_{21}}^T\mathbf{A_0}(q)\mathbf{w_{12}} \quad \mathbf{w_{21}}^T\mathbf{A_1}(q)\mathbf{w_{12}} \dots \mathbf{w_{21}}^T\mathbf{A_N}(q)\mathbf{w_{12}}\right]\mathbf{w_{12}} \\ &+\left[\mathbf{w_{21}}^T\mathbf{B_0}(q)\mathbf{w_{12}} \quad \mathbf{w_{21}}^T\mathbf{B_1}(q)\mathbf{w_{12}} \dots \mathbf{w_{21}}^T\mathbf{B_N}(q)\mathbf{w_{12}}\right]\mathbf{w_{12}} \\ &= 0 \qquad\qquad \text{pour} \quad 0 \leq q \leq N\end{aligned} \tag{4.10}$$

avec

$$\mathbf{R_{x_2y_1}}(q) = \begin{bmatrix} r_{x_2y_1}(q) & r_{x_2y_1}(q+1) & \cdots & r_{x_2y_1}(q+N) \\ r_{x_2y_1}(q-1) & r_{x_2y_1}(q) & \cdots & r_{x_2y_1}(q+N+1) \\ \vdots & \vdots & \ddots & \vdots \\ r_{x_2y_1}(q-N) & r_{x_2y_1}(q-N+1) & \cdots & r_{x_2y_1}(q) \end{bmatrix}$$

$$\widetilde{\mathbf{R}}_{\mathbf{x_2y_1}}(q) = \begin{bmatrix} r_{x_2y_1}(-q) & r_{x_2y_1}(1-q) & \cdots & r_{x_2y_1}(N-q) \\ r_{x_2y_1}(1-q) & r_{x_2y_1}(2-q) & \cdots & r_{x_2y_1}(N+1-q) \\ \vdots & \vdots & \vdots & \vdots \\ r_{x_2y_1}(N-q) & r_{x_2y_1}(N+1-q) & \cdots & r_{x_2y_1}(2N-q) \end{bmatrix}$$

$$\mathbf{A_i}(q) = \begin{bmatrix} r_{y_1y_1}(-q+i) & r_{y_1y_1}(-q+i-1) & \cdots & r_{y_1y_1}(-q+i-N) \\ r_{y_1y_1}(-q+i+1) & r_{y_1y_1}(-q+i) & \cdots & r_{y_1y_1}(-q+i+1-N) \\ \vdots & \vdots & \ddots & \vdots \\ r_{y_1y_1}(-q+i+N) & r_{y_1y_1}(-q+i+N-1) & \cdots & r_{y_1y_1}(-q+i) \end{bmatrix}$$

$$\mathbf{B_i}(q) = \begin{bmatrix} r_{y_1y_1}(q+i) & r_{y_1y_1}(q+i-1) & \cdots & r_{y_1y_1}(q+i-N) \\ r_{y_1y_1}(q+i-1) & r_{y_1y_1}(q+i) & \cdots & r_{y_1y_1}(q+i-1+N) \\ \vdots & \vdots & \ddots & \vdots \\ r_{y_1y_1}(q+i-N) & r_{y_1y_1}(q+i-1+N) & \cdots & r_{y_1y_1}(q+i) \end{bmatrix}$$

Par extension sur tous les coefficients de chacun des filtres, la nullité du gradient $\nabla_{\mathbf{w_{21}}}\xi_1 = \mathbf{0}$ conduit à la relation suivante :

$$
\begin{aligned}
&2\ \mathbf{R_{x_1y_1}w_{12}} + 2\ \mathbf{R_{x_2x_2}w_{21}} - 2\ \mathbf{r_{x_1x_2}}(q) = \\
&\begin{bmatrix}
2\ \mathbf{w_{21}}^T\Big(\mathbf{R_{x_2y_1}}(0) + \widetilde{\mathbf{R}}_{\mathbf{x_2y_1}}(0)\Big)\mathbf{w_{12}} & -\Big[\mathbf{w_{21}}^T\mathbf{A_0}(0)\mathbf{w_{12}} \ldots \mathbf{w_{21}}^T\mathbf{A_N}(0)\mathbf{w_{12}}\Big]\mathbf{w_{12}} \\
& -\Big[\mathbf{w_{21}}^T\mathbf{B_0}(0)\mathbf{w_{12}} \ldots \mathbf{w_{21}}^T\mathbf{B_N}(0)\mathbf{w_{12}}\Big]\mathbf{w_{12}} \\
& \vdots \\
2\ \mathbf{w_{21}}^T\Big(\mathbf{R_{x_2y_1}}(q) + \widetilde{\mathbf{R}}_{\mathbf{x_2y_1}}(q)\Big)\mathbf{w_{12}} & -\Big[\mathbf{w_{21}}^T\mathbf{A_0}(q)\mathbf{w_{12}} \ldots \mathbf{w_{21}}^T\mathbf{A_N}(q)\mathbf{w_{12}}\Big]\mathbf{w_{12}} \\
& -\Big[\mathbf{w_{21}}^T\mathbf{B_0}(q)\mathbf{w_{12}} \ldots \mathbf{w_{21}}^T\mathbf{B_N}(q)\mathbf{w_{12}}\Big]\mathbf{w_{12}} \\
& \vdots \\
2\ \mathbf{w_{21}}^T\Big(\mathbf{R_{x_2y_1}}(N) + \widetilde{\mathbf{R}}_{\mathbf{x_2y_1}}(N)\Big)\mathbf{w_{12}} & -\Big[\mathbf{w_{21}}^T\mathbf{A_0}(N)\mathbf{w_{12}} \ldots \mathbf{w_{21}}^T\mathbf{A_N}(N)\mathbf{w_{12}}\Big]\mathbf{w_{12}} \\
& -\Big[\mathbf{w_{21}}^T\mathbf{B_0}(N)\mathbf{w_{12}} \ldots \mathbf{w_{21}}^T\mathbf{B_N}(N)\mathbf{w_{12}}\Big]\mathbf{w_{12}}
\end{bmatrix}
\end{aligned}
\tag{4.11}
$$

En interchangeant les indices des signaux, les matrices $\mathbf{A}$ et $\mathbf{B}$ ainsi que les filtres $\mathbf{w_{21}}$ et $\mathbf{w_{12}}$, une relation similaire est obtenue pour $\nabla_{\mathbf{w_{12}}}\xi_2 = \mathbf{0}$. De part la relation (4.11), nous sommes forcés de constater que tenter d'obtenir une expression du filtre $\mathbf{w_{21}}$ explicite et simple relève d'une tâche difficile, voir même impossible, par comparaison avec le cas simple de la dérivation du filtre de Wiener.

4.1.3 Algorithme des moindres carrés transverses

L'approximation stochastique consistant à remplacer le gradient de l'EQM, $\nabla_{\mathbf{w_i}}\xi_i = \nabla_{\mathbf{w_i}}E\left[y_i(n)^2\right] = 2E\left[y_i(n)\nabla_{\mathbf{w_i}}y_i(n)\right]$, par sa valeur instantanée, $y_i(n)\nabla_{\mathbf{w_i}}y_i(n)$, permet d'aboutir aux équations récursives de mise à jour des coefficients des filtres adaptatifs. Nous définissons les deux vecteurs gradients suivants :

$$
\nabla^T_{\mathbf{w_{21}}}\mathbf{y_1}(n) = \left[\frac{\partial y_1(n)}{\partial w^0_{21}(n)}\ \frac{\partial y_1(n)}{\partial w^1_{21}(n)} \cdots \frac{\partial y_1(n)}{\partial w^N_{21}(n)}\right] \tag{4.12}
$$

$$
\nabla^T_{\mathbf{w_{12}}}\mathbf{y_2}(n) = \left[\frac{\partial y_2(n)}{\partial w^0_{12}(n)}\ \frac{\partial y_2(n)}{\partial w^1_{12}(n)} \cdots \frac{\partial y_2(n)}{\partial w^N_{12}(n)}\right] \tag{4.13}
$$

Par ailleurs, en remarquant que chacun des signaux estimés peut être mis sous la forme

$$
y_i(n) = x_i(n) - \sum_{k=0}^{N} w^k_{ji}(n)\left[x_j(n-k) - \sum_{s=0}^{N} w^s_{ij}(n-k)y_i(n-k-s)\right]
\quad \text{avec } i = 1,2 \quad j = 1,2 \quad \text{et} \quad i \neq j \tag{4.14}
$$

le calcul des gradients de chacune des sorties pourra être factorisé en regroupant les termes par rapport au retard des dérivées partielles, voir (4.27). Pour s'en convaincre, il suffit comme précédemment de calculer les quantités scalaires des vecteurs gradients définis auparavant. Ainsi, en introduisant les quantités $C_i(n)$ et $D_i(n)$ tel que :

$$C_i(n) = \begin{cases} \sum_{j=0}^{i} w_{21}^{j}(n) w_{12}^{i-j}(n-j) & 1 \leq i \leq N \\ \sum_{j=i-N}^{N} w_{21}^{j}(n) w_{12}^{i-j}(n-j) & N+1 \leq i \leq 2N \end{cases}$$

$$D_i(n) = \begin{cases} \sum_{j=0}^{i} w_{12}^{j}(n) w_{21}^{i-j}(n-j) & 1 \leq i \leq N \\ \sum_{j=i-N}^{N} w_{12}^{j}(n) w_{21}^{i-j}(n-j) & N+1 \leq i \leq 2N \end{cases}$$

nous obtenons :

$$\nabla_{\mathbf{w_{21}}}\mathbf{y_1}(n) - \left[w_{21}^{0}(n) w_{12}^{0}(n)\right] \nabla_{\mathbf{w_{21}}}\mathbf{y_1}(n) = -\mathbf{y_2}(n) + \sum_{k=1}^{2N} C_k(n) \nabla_{\mathbf{w_{21}}}\mathbf{y_1}(n-k) \tag{4.15}$$

$$\nabla_{\mathbf{w_{12}}}\mathbf{y_2}(n) - \left[w_{21}^{0}(n) w_{12}^{0}(n)\right] \nabla_{\mathbf{w_{12}}}\mathbf{y_2}(n) = -\mathbf{y_1}(n) + \sum_{k=1}^{2N} D_k(n) \nabla_{\mathbf{w_{12}}}\mathbf{y_2}(n-k) \tag{4.16}$$

Or, ces quantités interviennent dans les équations du gradient stochastique données ci-après :

$$\mathbf{w_{21}}(n+1) = \mathbf{w_{21}}(n) - 2\mu_1(n) y_1(n) \nabla_{\mathbf{w_{21}}} y_1(n) \tag{4.17}$$

$$\mathbf{w_{12}}(n+1) = \mathbf{w_{12}}(n) - 2\mu_2(n) y_2(n) \nabla_{\mathbf{w_{12}}} y_2(n) \tag{4.18}$$

et bien que l'obtention des formules (4.15) et (4.16) ait déjà nécessité des développements relativement longs, celles-ci ne sont pas encore directement exploitables pour l'implémentation. En effet, elles ne forment pas une récurrence explicite du fait de la dépendance des termes $\nabla_{\mathbf{w_{21}}}\mathbf{y_1}(n-k)$ et $\nabla_{\mathbf{w_{12}}}\mathbf{y_2}(n-k)$ par rapport aux coefficients $w_{21}(n)$ et $w_{12}(n)$ calculés à l'instant n et non pas $n-k$. Pour trouver cette relation de récurrence, les auteurs utilisent la décomposition suivante :

$$\boxed{\frac{\partial y_i(n-k)}{\partial w_{ji}^{m}(n)} = \frac{\partial y_i(n-k)}{\partial w_{ji}^{m}(n-k)} \frac{\partial w_{ji}^{m}(n-k)}{\partial w_{ji}^{m}(n)} \quad \begin{matrix} i=1,2 & j=1,2 & i \neq j & m=0,\ldots,N \end{matrix}} \tag{4.19}$$

En effet, en itérant k fois l'une des relations (4.17) ou (4.18) par rapport au $m^{ème}$ coefficient, nous obtenons la relation (4.20). Puis, en calculant la dérivée partielle de

cette relation (4.20) par rapport à $w_{ji}^m(n)$ nous faisons apparaître le terme $\partial w_{ji}^m(n-k)/\partial w_{ji}^m(n)$ de la relation (4.19) que nous explicitons à l'équation (4.21).

$$w_{ji}^m(n-k) = w_{ji}^m(n) + 2\sum_{l=1}^{k} \mu_i(n-l) y_i(n-l) \frac{\partial y_i(n-l)}{\partial w_{ji}^m(n-l)} \tag{4.20}$$

$$\frac{\partial w_{ji}^m(n-k)}{\partial w_{ji}^m(n)} = 1 + 2\frac{\partial}{\partial w_{ji}^m(n)} \left[\sum_{l=1}^{k} \mu_i(n-l) y_i(n-l) \frac{\partial y(n-l)}{\partial w_{ji}^m(n-l)} \right] \tag{4.21}$$

La somme présente dans (4.20) est une moyenne pondérée par μ_i de l'erreur et de la dérivée de cette erreur. En supposant de faibles variations du signal et donc du pas d'adaptation, le pas peut être mis en facteur de cette somme. Or, la condition traduisant la bonne convergence des filtres décrits en (4.17) et (4.18) est, qu'à l'optimum, l'erreur (soit donc le signal estimé y_i) et la dérivée de cette erreur soient orthogonales. Ainsi, la somme est nulle et le terme $\partial w_{ji}^m(n-k)/\partial w_{ji}^m(n)$ peut donc être négligé puisqu'il vaut 1. Ainsi, $\frac{\partial y_i(n-k)}{\partial w_{ji}^m(n)} \approx \frac{\partial y_i(n-k)}{\partial w_{ji}^m(n-k)}$ et les récursions (4.15) et (4.16) deviennent :

$$\nabla_{\mathbf{w_{21}}}\mathbf{y_1}(n) = C_0(n) \left[\sum_{k=1}^{2N} C_k(n) \nabla_{\mathbf{w_{21}}}^k \mathbf{y_1}(n-k) - \mathbf{y_2}(n) \right] \tag{4.22}$$

$$\nabla_{\mathbf{w_{12}}}\mathbf{y_2}(n) = D_0(n) \left[\sum_{k=1}^{2N} D_k(n) \nabla_{\mathbf{w_{12}}}^k \mathbf{y_2}(n-k) - \mathbf{y_1}(n) \right] \tag{4.23}$$

avec

$$C_0(n) = D_0(n) = \frac{1}{1 - w_{21}^0(n) w_{12}^0(n)}$$

et où la notation ∇^k désigne le calcul du gradient à partir des coefficients des filtres estimés aux instants $n-k$.

4.1.4 Evaluation de la complexité

Afin d'être en mesure d'établir des comparaisons et apprécier les optimisations algorithmiques que nous proposons, nous évaluerons la complexité des deux approches. Pour plus de simplicité, nous ne parlerons pas de la complexité globale de l'algorithme au sens habituel. Pour nos besoins, nous définissons simplement celle-ci en termes du nombre de points mémoires, de multiplications et d'additions nécessaires pour calculer les gradients propres à chacun des algorithmes. Ce critère est suffisant pour rendre compte des améliorations qui seront apportées d'autant plus que celles-ci sont principalement basées sur une optimisation du calcul de ces gradients. L'analyse faite à partir de notre définition de la complexité vaudra également pour la complexité telle que nous l'entendons habituellement.

Représentation schématique de l'algorithme

La Figure 4.2 illustre la mise en œuvre du calcul du gradient $\nabla_{\mathbf{w_{21}}}\mathbf{y_1}(n)$ tel qu'il est dicté dans l'algorithme de Zinser et al. Il s'agit plus précisément d'une mise en image de la formule décrite par $C_i(n)$, où l'on comprend que pour calculer les termes allant de 0 à N, la première colonne de la matrice des coefficients de $\mathbf{w_{21}}$ est multipliée par les coefficients de la matrice stockant les coefficients de $\mathbf{w_{12}}$ sur un horizon de N valeurs et situé sur les diagonales montantes. Chaque terme issu de la sommation de ces produits pris ensuite ligne par ligne, correspond au terme en facteur de chacune des colonnes de la matrice qui stocke les composantes des gradients sur un horizon de $2N$. Ainsi, dans la fenêtre $0 \leq i \leq N$, le calcul de $\nabla_{\mathbf{w_{21}}}\mathbf{y_1}(n)$ nécessite la matrice triangulaire supérieure des coefficients de $\mathbf{w_{12}}$, diagonale incluse. Au delà, pour $N+1 \leq i \leq 2N$, ce sont les termes de la matrice triangulaire inférieure qui sont utilisés. Par conséquent, le calcul de $\nabla_{\mathbf{w_{12}}}\mathbf{y_2}(n)$ qui est le symétrique de $\nabla_{\mathbf{w_{21}}}\mathbf{y_1}(n)$

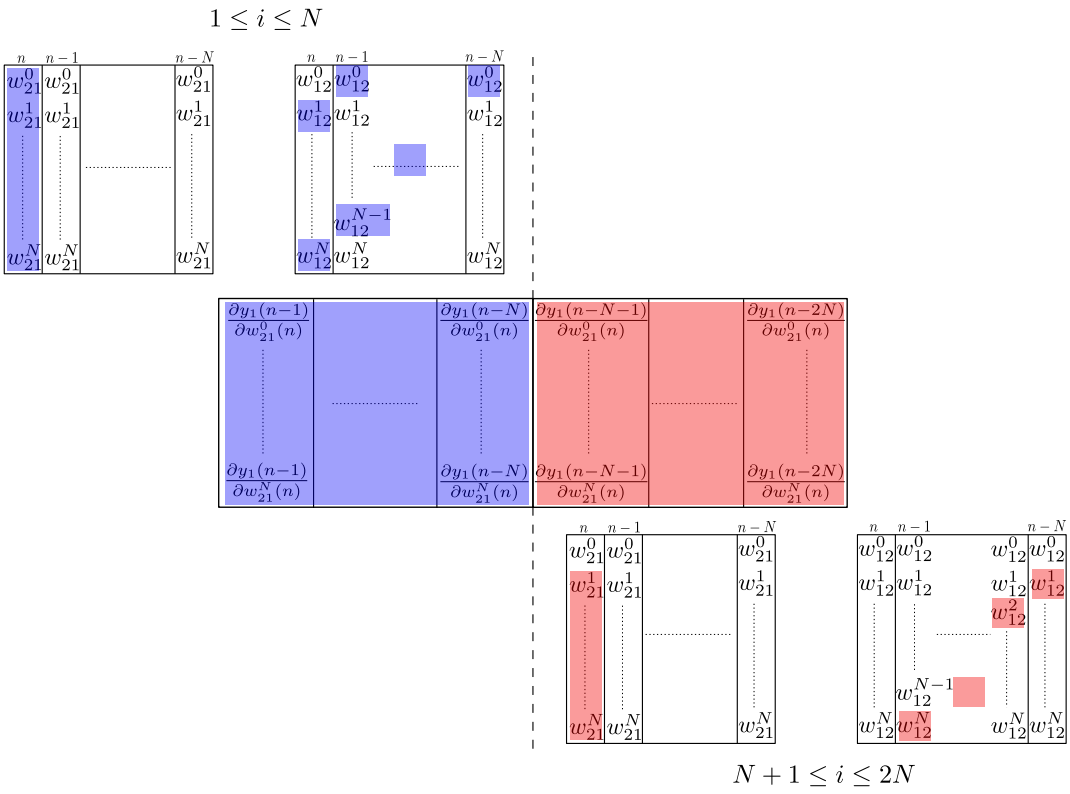


Figure 4.2 – Schéma descriptif du calcul du gradient $\nabla_{\mathbf{w_{21}}}\mathbf{y_1}(n)$ tel qu'il est dicté dans l'algorithme de Zinser et al.

se fera de la même manière en inversant les rôles des filtres, d'où la nécessité de stocker des matrices de coefficients et non pas uniquement la première colonne (vecteur des estimées courantes des filtres), comme cela est finalement traduit par la formule de $D_i(n)$.

Coût mémoire

La représentation schématique précédente montre que chaque gradient est une fonction des coefficients des filtres adaptatifs w_{21}^k et w_{12}^k et qui plus est dépendante de l'indice temporel n. Ceci implique donc qu'il faille stocker $2N(N+1)$ valeurs de gradient ainsi que $2(N+1)^2$ coefficients.

Complexité algorithmique

Le tableau 4.1 donne la complexité en termes du nombre de multiplications et d'additions nécessaires au calcul d'une composante du gradient à l'instant n.

Dérivée partielle	Multiplications	Additions
$\frac{\partial y_k(n)}{\partial w_k^i(n)} \quad k=1,2$	$2N^2+4N+2$	N^2+2N+1

Table 4.1 – Complexité algorithmique du CTRANC.

La force de l'algorithme proposé par Zinser et al. réside dans sa précision, son adéquation avec le couplage introduit par la structure et qui est clairement pris en compte par l'algorithme tel qu'il a été décliné dans les parties précédentes. Cependant, nous verrons qu'il existe un effet pernicieux lié à l'utilisation d'une mémoire des coefficients trop importante, en particulier dans le cas du traitement du signal de parole. De plus, l'évaluation de la complexité montre que le coût de cet algorithme par rapport à ceux couramment utilisés en pratique est très important. D'ailleurs, cette complexité pourrait expliquer en partie le fait que cet algorithme n'ait jamais été réutilisé par la suite dans la littérature, à l'inverse de l'algorithme de décorrélation.

4.2 Réduction de la complexité de l'algorithme optimal

4.2.1 Concept à la base de nos simplifications algorithmiques

Pour réduire la complexité d'une part et faciliter l'implémentation dans le domaine fréquentiel d'autre part, nous proposons d'appliquer une hypothèse courante du filtrage adaptatif, à savoir :

– $(\mathcal{H}_1)$: les coefficients des filtres adaptatifs varient lentement au cours du temps, ce qui permet l'approximation suivante : $w_{12,21}^i(n) \approx w_{12,21}^i(n-k)$ avec $k=1,\ldots,N$.

Grâce à cette hypothèse, nous allons être en mesure de réécrire chacune des expressions des gradients (4.22) et (4.23) comme une convolution permettant à la fois de réduire la complexité et simplifier la mise en œuvre fréquentielle.

4.2.2 Obtention du nouvel algorithme

Considérons les équations des signaux estimés (4.1). En développant celle associée à y_1, nous obtenons

$$\begin{aligned} y_1(n) &= x_1(n) - \sum_{k=0}^{N} w_{21}^k(n) x_2(n-k) \\ &+ \sum_{s=0}^{N} w_{21}^s(n) \sum_{t=0}^{N} w_{12}^t(n-s) y_1(n-s-t) \end{aligned} \tag{4.24}$$

Le gradient de cette dernière équation (4.24) pour chaque composante du vecteur qui en résulte $\nabla_{\mathbf{w_{21}}}\mathbf{y_1}(n)$ donne

$$\begin{aligned} \frac{\partial y_1(n)}{\partial w_{21}^l(n)} &= -x_2(n-l) + \sum_{t=0}^{N} w_{12}^t(n-l) y_1(n-l-t) \\ &+ \sum_{s=0}^{N} w_{21}^s(n) \sum_{t=0}^{N} w_{12}^t(n-s) \frac{\partial y_1(n-s-t)}{\partial w_{21}^l(n)} \end{aligned} \tag{4.25}$$

Les deux premiers termes de (4.25) correspondent au signal de sortie $-y_2(n-l)$. Par ailleurs, en développant le terme de la double somme de sorte à faire apparaître la dérivée partielle de y_1 aux instants courants n, nous obtenons :

$$\begin{aligned} \frac{\partial y_1(n)}{\partial w_{21}^l(n)} &= -y_2(n-l) + w_{21}^0(n) w_{12}^0(n) \frac{\partial y_1(n)}{\partial w_{21}^l(n)} \\ &+ w_{21}^0(n) \sum_{k=1}^{N} w_{12}^k(n) \frac{\partial y_1(n-k)}{\partial w_{21}^l(n)} \\ &+ \sum_{s=1}^{N} w_{21}^s(n) \sum_{t=0}^{N} w_{12}^t(n-s) \frac{\partial y_1(n-s-t)}{\partial w_{21}^l(n)} \end{aligned} \tag{4.26}$$

Après factorisation des deux derniers termes de (4.26) par rapport aux retards des dérivées partielles $\frac{\partial y_1(n-i)}{\partial w_{21}^l(n)}$, ceux-ci peuvent être réécrits tel que :

$$\begin{aligned} &\dots \quad w_{21}^0(n) \sum_{k=1}^{N} w_{12}^k(n) \frac{\partial y_1(n-k)}{\partial w_{21}^l(n)} \\ &+ \sum_{s=1}^{N} w_{21}^s(n) \sum_{t=0}^{N} w_{12}^t(n-s) \frac{\partial y_1(n-s-t)}{\partial w_{21}^l(n)} \\ &= \begin{cases} \sum_{j=0}^{i} w_{21}^j(n) w_{12}^{i-j}(n-j) \frac{\partial y_1(n-i)}{\partial w_{21}^l(n)} & 1 \le i \le N \\ \sum_{j=i-N}^{N} w_{21}^j(n) w_{12}^{i-j}(n-j) \frac{\partial y_1(n-i)}{\partial w_{21}^l(n)} & N+1 \le i \le 2N \end{cases} \end{aligned} \tag{4.27}$$

où l'on reconnaît le terme $C_i(n) \times \frac{\partial y_1(n-i)}{\partial w_1^l(n)}$. Appliquons maintenant l'hypothèse $(\mathcal{H}_1)$ à $C_i(n)$ sans omettre le premier terme pour $i = 0$, il vient immédiatement :

$$\begin{aligned} \sum_{i=0}^{N}\sum_{j=0}^{N} w_{21}^i(n) w_{12}^j(n-i) &\overset{(\mathcal{H}_1)}{\equiv} \sum_{k=0}^{i} w_{21}^k(n) w_{12}^{i-k}(n) \qquad 0 \leq i \leq 2N \\ &= \mathbf{w_{21}}(n) * \mathbf{w_{12}}(n) = G(n) \end{aligned} \tag{4.28}$$

La somme des produits des coefficients des filtres $\mathbf{w_{12}}$ et $\mathbf{w_{21}}$ sur un horizon de N est transformée en une convolution à partir des coefficients calculés à l'instant n. La Figure 4.3 met en image l'identification des termes de la convolution à partir de l'approximation $(\mathcal{H}_1)$ sur le développement du terme $C_k(n)$ complété par les termes d'indice $i = 0$. Cette approximation, ré-injectée dans les expressions des gradients,

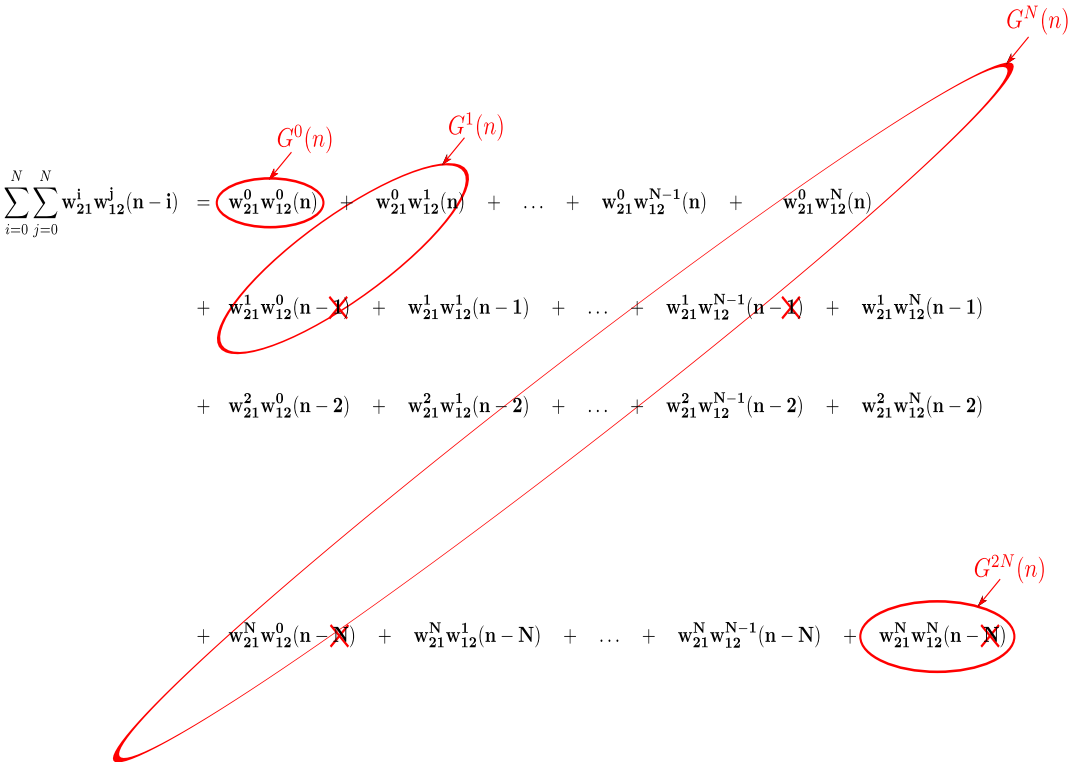


Figure 4.3 – Identification des $2N+1$ termes de la convolution par application de l'hypothèse $(\mathcal{H}_1)$ sur la somme des produits des coefficients des filtres $\mathbf{w_{12}}$ et $\mathbf{w_{21}}$.

implique que leurs nouvelles estimées deviennent :

$$\nabla_{\mathbf{w_{21}}}\mathbf{y_1}(n) = \frac{1}{1-G^0(n)}\left[\sum_{i=1}^{2N} G^i(n) \cdot \nabla_{\mathbf{w_{21}}}\mathbf{y_1}(n-i) - \mathbf{y_2}(n)\right] \tag{4.29}$$

$$\nabla_{\mathbf{w_{12}}}\mathbf{y_2}(n) = \frac{1}{1-G^0(n)}\left[\sum_{i=1}^{2N} G^i(n) \cdot \nabla_{\mathbf{w_{12}}}\mathbf{y_2}(n-i) - \mathbf{y_1}(n)\right] \tag{4.30}$$

où $\frac{1}{1-G^0(n)} = C_0(n) = D_0(n)$. Le stockage des matrices de coefficients des filtres n'est donc plus nécessaire et est réduit à la mémorisation d'un vecteur de coefficients qui correspond à la première colonne de chacune de ces matrices. Cette première optimisation apportée par notre hypothèse ($\mathcal{H}_1$) n'est pas la seule. En effet, le stockage des vecteurs gradients (4.29) et (4.30) va pouvoir également être réduit à la mémorisation d'un vecteur. Plus précisément, l'analyse de ces vecteurs montre une propriété de quasi-circularité induite par notre hypothèse de faible variation des coefficients. La Figure 4.4 explique le calcul à l'instant n de chacune des dérivées partielles qui constitue le vecteur $\nabla_{\mathbf{w_{21}}}\mathbf{y_1}(n)$. Or, les entrées de chacun de ces sous-blocs de filtrage sont issues de la mémorisation des signaux estimés sur l'autre voie (y_2) et mis dans une mémoire (*buffer*) de taille $N+1$ correspondant à la longueur des filtres. Ainsi, à chaque itération un nouvel échantillon est mémorisé et utilisé pour le calcul des gradients. Du fait de ce mécanisme de mémoire (*buffer*) circulaire, l'échantillon $y_2(n-1)$ qui va servir au calcul de $\frac{\partial y_1(n)}{\partial w_{21}^1(n)}$ est le même que celui qui a servi à l'instant précédent pour le calcul de $\frac{\partial y_1(n-1)}{\partial w_{21}^0(n)}$. Ainsi, en admettant que l'hypothèse ($\mathcal{H}_1$) est effectivement vérifiée, le filtrage par $\widetilde{G}(n)$ qui est une version tronquée du premier terme de $G(n)$ n'influe que faiblement et donc les quantités $\frac{\partial y_1(n)}{\partial w_{21}^1(n)}$ et $\frac{\partial y_1(n-1)}{\partial w_{21}^0(n)}$ sont sensiblement égales. En d'autres termes et par extension de ce raisonnement sur un horizon de $2N$, les dérivées partielles constitutives de la matrice de mémorisation des gradients, voir Figure 4.2, ont des composantes quasiment égales sur toutes les diagonales montantes. En exploitant cette propriété de quasi-circularité, il est alors possible de réduire le coût de calcul du vecteur $\nabla_{\mathbf{w_{21}}}\mathbf{y_1}(n)$ en ne calculant que le premier coefficient, à chaque itération, du vecteur constitué par la première ligne de cette matrice. Un *buffer* circulaire de taille $2N$ permettra ensuite d'obtenir toutes les composantes de celui-ci. Ce même raisonnement vaut pour les gradients du second signal estimé y_2 sur la voie duale.

L'ensemble de ces modifications décrivent le nouvel algorithme proposé et nommé *Low Complexity Crosstalk Resistant Adaptive Noise Canceller* (LCCTRANC).

4.2.3 Evaluation de la complexité du LCCTRANC

Coût mémoire

Grâce à notre hypothèse, chaque gradient est une fonction dépendante des dernières estimées des filtres $\mathbf{w_{21}}(n)$ et $\mathbf{w_{12}}(n)$ et des vecteurs gradients $\nabla_{\mathbf{w_{21}}}\mathbf{y_1}(n)$ et $\nabla_{\mathbf{w_{12}}}\mathbf{y_2}(n)$. Par conséquent, il faudra conserver les $2N$ composantes du vecteur gradient ainsi que les $2(N+1)$ coefficients des deux filtres adaptatifs. Ceci réduit de manière importante le coût mémoire par rapport à l'algorithme initial.

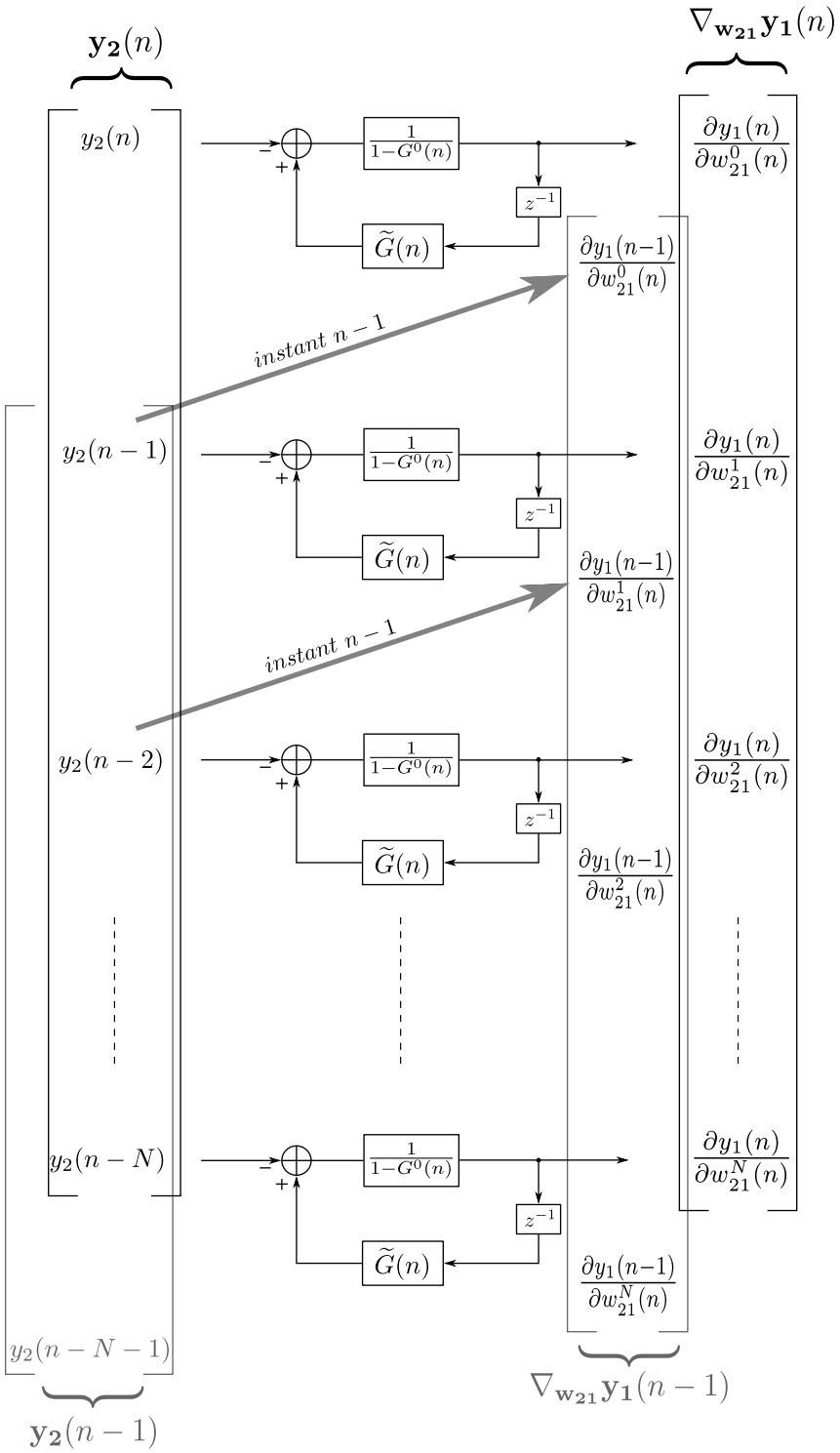


Figure 4.4 – Schéma de principe descriptif du calcul de chacune des composantes du vecteur gradient $\nabla_{\mathbf{w_{21}}}\mathbf{y_1}(n)$, à l'instant n. $\widetilde{G}(n)$ correspond au vecteur des coefficients issus de la convolution $\mathbf{w_{21}}(n) * \mathbf{w_{12}}(n)$ privé du premier terme $G^0(n) = w_{12}^0(n)w_{21}^0(n)$.

Complexité algorithmique

Comme nous l'avons fait pour le CTRANC, le tableau 4.2 fournit la complexité du LCCTRANC en termes de coût de calcul pour une composante du gradient. Par comparaison avec les résultats associés au CTRANC, nous constatons que le

Dérivée partielle	Multiplications	Additions
$\frac{\partial y_k(n)}{\partial w_k^i(n)} \quad k=1,2$	N^2+3N+2	$\frac{1}{2}N^2+\frac{3}{2}N+1$

Table 4.2 – Complexité algorithmique du LCCTRANC.

LCCTRANC fait l'économie de N^2+N multiplications et $\frac{1}{2}N^2+\frac{1}{2}N$ additions par composante de gradient calculée à chaque instant n.

Résumé

Afin de mieux comparer le gain en complexité, nous calculons le coût mémoire total (C.M.) obtenu en additionnant toutes les composantes nécessaires aux calculs d'une composante de gradient dans les algorithmes étudiés. La même procédure est mise en œuvre pour le calcul de la complexité algorithmique (C.A.) en sommant respectivement le nombre de multiplications et d'additions que requiert chacun des algorithmes et donnés aux Tableaux 4.1 et 4.2. Les résultats sont tracés à la Figure 4.5.

4.3 Comparaison des performances

Pour la comparaison des performances, nous créons un contexte synthétique à partir de l'outil *Roomsim*. Nous simulons une première scène sonore composée de deux sources placées de manière symétrique, à 40 *cm* des microphones et dans un environnement anéchoïque. La distance inter-microphone est de 15 *cm* ce qui correspond, peu ou prou, à la distance interaurale. Le but est de disposer dans un premier temps d'un contexte simple avec des réponses impulsionnelles courtes. Ce contexte est décrit plus précisément en Annexe D.

4.3.1 Mélange de signaux synthétiques

Les signaux sources sont deux bruits blancs gaussiens de même puissance $\sigma_{s_1}^2 = \sigma_{s_2}^2 = 1$. Le mélange est fait en utilisant uniquement les chemins croisés générés par l'outil *Roomsim*, ceci en accord avec le modèle de mélange simplifié discuté au Chapitre 2. Les performances des algorithmes vont être comparées en observant le comportement de la déviation normalisée, Figure 4.6, et en analysant l'erreur

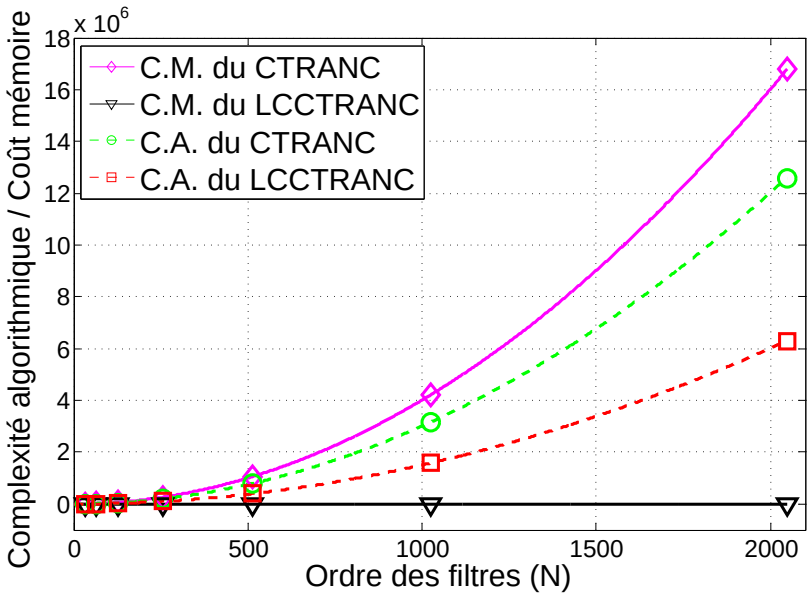


Figure 4.5 – Comportement du coût mémoire (C. M.) et de la complexité algorithmique (C. A.) pour le calcul d'une composante de gradient. C.M. est obtenu en sommant toutes les mémoires utilisées pour le stockage des coefficients des filtres ainsi que les valeurs passées des composantes du gradient. C.A. est la somme des multiplications et des additions nécessaires.

quadratique moyenne résiduelle après que les algorithmes aient convergé, Figure 4.7. Les résultats de la déviation normalisée sont issus d'une moyenne sur 30 tirages.

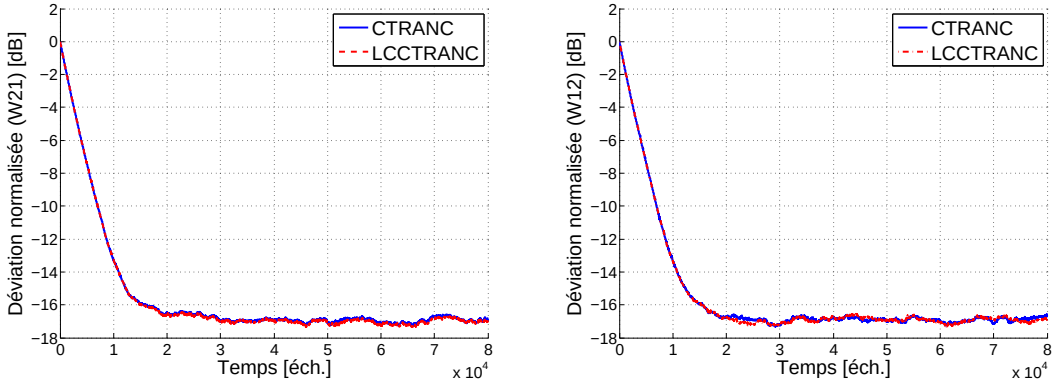


Figure 4.6 – Comportement de la déviation normalisée de chaque filtre en fonction de l'algorithme.

D'après les résultats obtenus, nous constatons que l'hypothèse ($\mathcal{H}_1$) qui nous a permis de décliner notre version à faible complexité de l'algorithme est valide et justifiée puisque les performances sont très proches. En effet, dans les deux cas, l'EQM résiduelle est très proche et atteint environ -20 dB. Par ailleurs, les estimées des filtres sont entachées de la même erreur au vue du comportement de la déviation

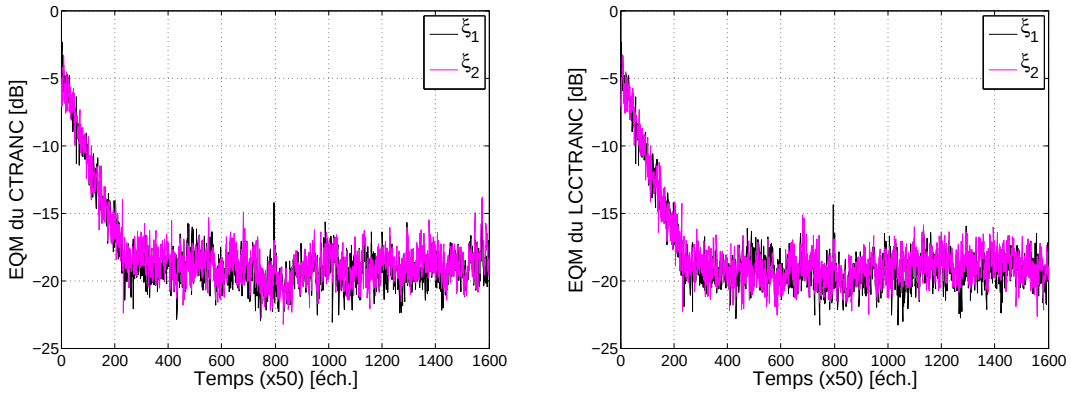


Figure 4.7 – Comportement de l'EQM sur chaque voie en fonction de l'algorithme.

normalisée de chaque filtre qui se superposent de manière quasi parfaite. Nous venons donc d'illustrer que les performances entre l'algorithme optimal de Zinser et al. celui que nous proposons sont identiques dans le cas d'un mélange simplifié de signaux blancs gaussiens. Nous allons à présent appréhender le comportement des deux algorithmes pour un mélange de sources non-stationnaires.

4.3.2 Mélange de signaux réels

Le mélange est maintenant constitué d'une source de parole et d'un bruit de bureau pour un RSB de 10 dB. Ce choix de RSB est arbitraire et les observations que nous allons faire ici sont valables quelle que soit sa valeur. Le signal de parole est issu de la concaténation de deux doubles phrases alternant une voix d'homme et une voix de femme. Sur la Figure 4.8, l'allure des signaux estimés montre que les performances globales vis-à-vis de la réduction de bruit semblent préservées. La moyenne et l'écart type de l'erreur entre ces deux estimées ($y_{1_{CTRANC}} - y_{1_{LCCTRANC}}$) sont respectivement de -2,4.10^{-3} et 13,2. Au regard de ces valeurs, nous constatons que l'amplitude de l'erreur est très faible et le fait que la moyenne soit négative tend à montrer qu'en moyenne l'amplitude de $y_{1_{LCCTRANC}}$ est supérieure à celle de $y_{1_{CTRANC}}$, ce qui sur des instants de bruit seul est équivalent à dire que l'algorithme LCCTRANC est à peine moins efficace que l'algorithme CTRANC. Néanmoins, la Figure 4.9 montre que la distance cepstrale obtenue au cours du temps pour l'algorithme LCCTRANC se superpose à celle obtenue avec l'algorithme CTRANC. Les simplifications amenant au LCCTRANC ne semblent donc pas altérer significativement les performances de l'algorithme initial, y compris pour des signaux non-stationnaires. Bien que le LCCTRANC puisse être quelque peu moins efficace en termes de réduction de bruit que le CTRANC, cette légère perte de performance est sans commune mesure avec le gain en complexité algorithmique et la réduction du coût mémoire auxquels nous

sommes parvenus grâce à nos optimisations.

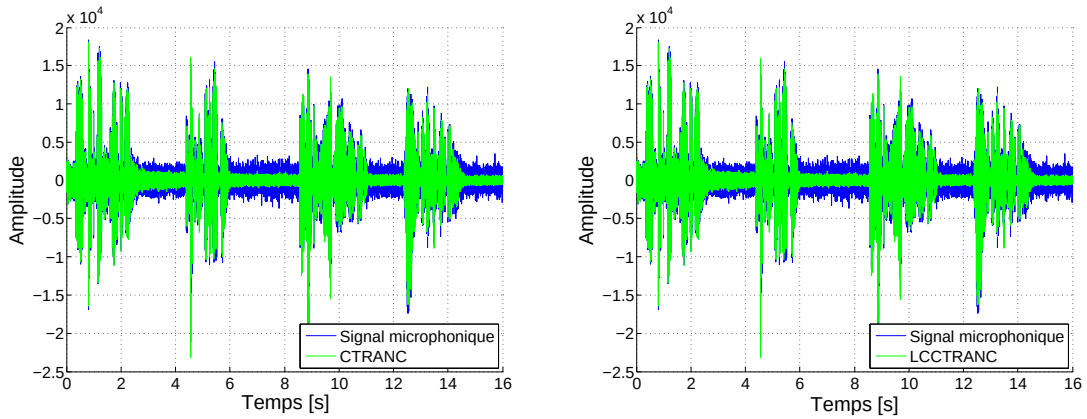


Figure 4.8 – Résultats de simulation pour un mélange fait d'une source de parole et d'un bruit de bureau.

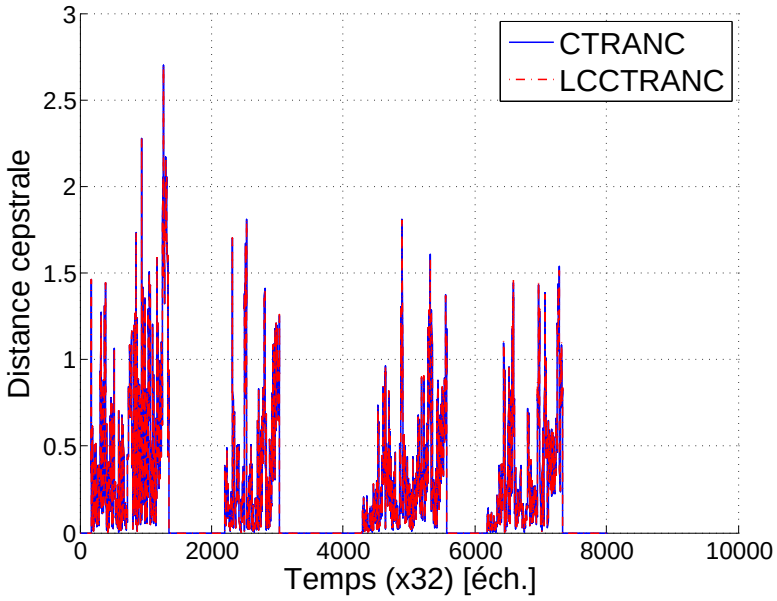


Figure 4.9 – Evolution de la distance cepstrale au cours du temps pour les deux algorithmes considérés.

En ce sens et pour justifier de la pertinence de l'observation faite sur la propriété de quasi-circularité des matrices de gradients, nous donnons à la Figure 4.10 les valeurs de celles-ci codées sur un jeu de couleur à l'image de la valeur relative de chacun des coefficients qui les constituent. Le contexte de simulation est le même que précédemment avec toutefois des filtres plus courts et comptant 25 coefficients de sorte à améliorer la lisibilité. Cependant, cela implique que les résultats sont

influencés par la sous-modélisation, sans conséquence ici puisque nous identifions aisément les diagonales montantes à l'image de la circularité des matrices de gradient.

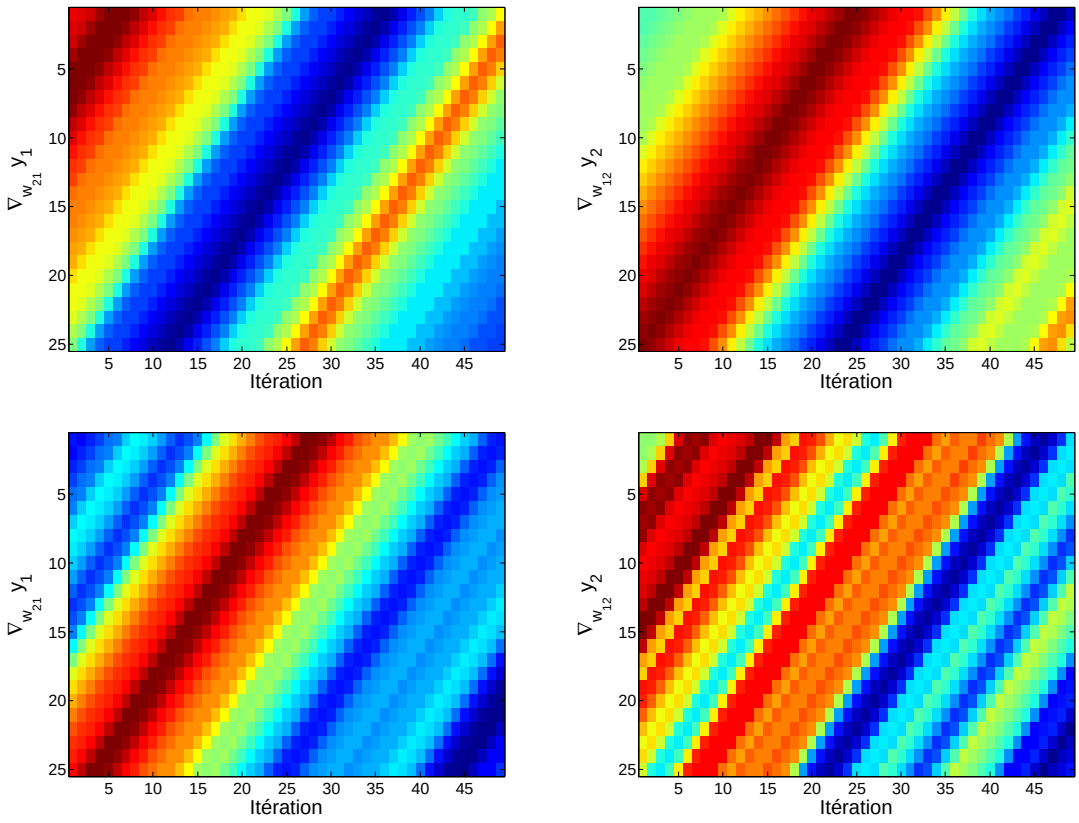


Figure 4.10 – Observation de la quasi-circularité des matrices des gradients aux instants $t = 1$ (haut) et $t = 16$ secondes (bas). Pour plus de lisibilité, les filtres sont réduits à 25 coefficients.

4.3.3 Gestion des instabilités - Réduction de l'ordre

Dans cette partie, nous souhaitons attirer l'attention sur une source potentielle d'instabilité présente dans les deux algorithmes étudiés dans ce chapitre et directement liée à la physique du problème de l'estimation de réponse impulsionnelle acoustique. En effet, les expressions des gradients données en (4.22), (4.23), (4.29) et (4.30) sont à la fois à l'origine de la précision des estimées comparée aux algorithmes de décorrélation mais également la source d'instabilités dues au mauvais conditionnement des réponses impulsionnelles que l'on cherche à estimer.
A l'aide de la modélisation par des schémas blocs du calcul des gradients de l'algorithme proposé (Figure 4.4), nous pouvons extraire, grâce à la transformée en Z, la fonction de transfert du système relatif au calcul d'une composante de gradient.

Celle-ci est donnée ci-après.

$$\frac{TZ\left[\frac{\partial y_1(n)}{\partial w_{21}^l(n)}\right]}{TZ\left[y_2(n)\right]} = -\frac{-1}{1 - G^0(z) - \widetilde{G}(z)z^{-1}} \tag{4.31}$$

Cette fonction de transfert renvoie à l'étude de la stabilité du filtre RII d'expression $F(z) = \frac{1}{1-W_{12}(z)W_{21}(z)}$, vu au Chapitre 2 et pour lequel nous avons mis en avant les instabilités causées par la partie des réponses impulsionnelles (queue) correspondant à la réverbération. Par conséquent, plus la longueur des filtres est importante, plus la convolution comportera des termes associés à la réverbération et donc pourra potentiellement engendrer des instabilités. Cela sous-entend donc qu'il faille trouver un compromis sur la longueur des filtres. Ou de façon plus judicieuse, réduire la longueur de la mémoire utilisée pour stocker les gradients intervenant dans le calcul de leurs estimées courantes. En agissant sur la longueur de la récursion et non pas sur la longueur des filtres, nous cherchons finalement à trouver un compromis entre les méthodes dites de décorrélation et celles que nous considérons. En effet, si l'on conserve uniquement le terme $G^0(n)$, alors l'algorithme LCCTRANC et l'algorithme SAD sont équivalents. Ceci se produit lorsque les filtres de mélange forment un mélange instantané et n'ont donc qu'un seul coefficient. Dans les autres cas, nous proposons de ne considérer qu'un nombre limité de termes pris dans l'intervalle $[0\ 2N]$, afin d'améliorer les estimées mais sans introduire d'instabilités. Cette troncature sur l'ordre de la récursion, en plus d'un gain mémoire, va permettre une dérivation simple de l'algorithme dans le domaine fréquentiel.
En ce sens, les Figures 4.11 et 4.12 présentent l'évolution de la déviation normalisée en fonction de la longueur de la récursion pour les deux types de mélange de signaux donnés précédemment : synthétiques et réels. Plusieurs choses sont à remarquer à partir des résultats obtenus :

1. l'ordre de la récursion doit au minimum être égal à la longueur des filtres ($N = 64$ dans le cas considéré ici). En deçà de cette valeur, les performances du CTRANC et du LCCTRANC sont équivalentes à celles de l'algorithme SAD. A partir de cet ordre, la valeur de la déviation normalisée pour chaque filtre est systématiquement meilleure que celle du SAD, ce qui signifie que les estimées sont plus précises même si en moyenne le gain est inférieur à 1 dB,
2. la troncature aux N premiers coefficients stabilise les algorithmes CTRANC et LCCTRANC, en les rendant plus robustes aux non-stationnarités et aux bruits de calcul. En effet, en considérant les résultats pour le mélange synthétique, nous noterons que le CTRANC peut pâtir de la prise en compte de tous les termes de la récursion. Ceci s'observe nettement sur la déviation normalisée associée au filtre $\mathbf{w_{12}}$ où les moins bonnes estimées sont celles fournies par l'algorithme CTRANC, y compris sans troncature. Ce phénomène s'explique par

une mise à jour des gradients tenant compte de valeurs passées trop lointaines qui vont perturber l'adaptation plutôt que l'affiner. Ceci est aussi en lien avec le pas d'adaptation faible, en adéquation avec un comportement asymptotique qui ne tient donc plus compte de la mémoire des valeurs passées. Enfin, la robustesse vis-à-vis des non-stationnarités est visible pour le mélange comprenant un signal de parole, où la troncature permet à l'algorithme LCCTRANC de fournir l'estimée la plus fine,

3. la supériorité de l'algorithme optimal et du LCCTRANC pour un mélange de signaux réels est rendue visible par la déviation normalisée du filtre $\mathbf{w_{21}}$ qui passe au dessous de -3 dB, soit environ 2 dB de mieux que l'algorithme SAD. L'estimation du filtre associé au signal stationnaire sera plus précise et plus robuste avec les algorithmes CTRANC et LCCTRANC non tronqués et toujours supérieure à l'algorithme SAD pourvu que l'on conserve au moins N valeurs de la récursion.

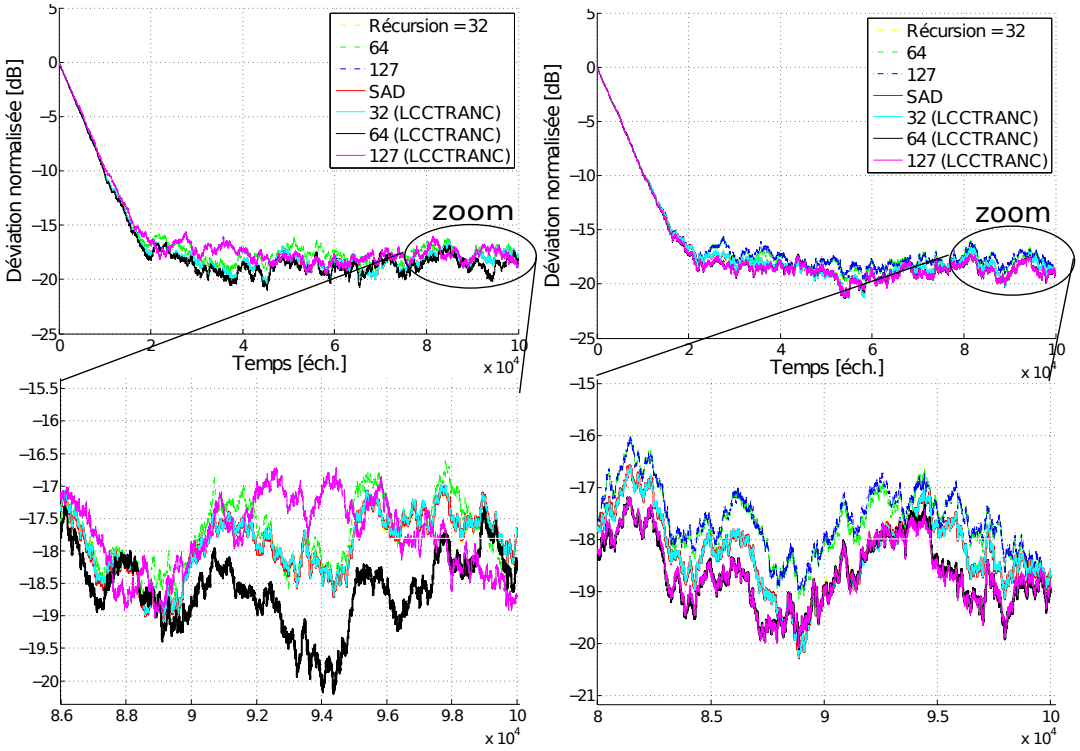


Figure 4.11 – Evolution de la déviation normalisée des filtres $\mathbf{w_{21}}$ (gauche) et $\mathbf{w_{12}}$ (droit) pour un mélange de deux sources blanches gaussiennes de même puissance $\sigma^2_{s_1} = \sigma^2_{s_2} = 1$. Les pas d'adaptation sont fixés à 0,01.

Ainsi, il est nécessaire de restreindre la longueur de la mémoire des gradients, y compris pour des mélanges anéchoïques. Dans le cas réaliste de mélange échoïques cela sera d'autant plus nécessaire que la sous-modélisation est inévitable, et donc

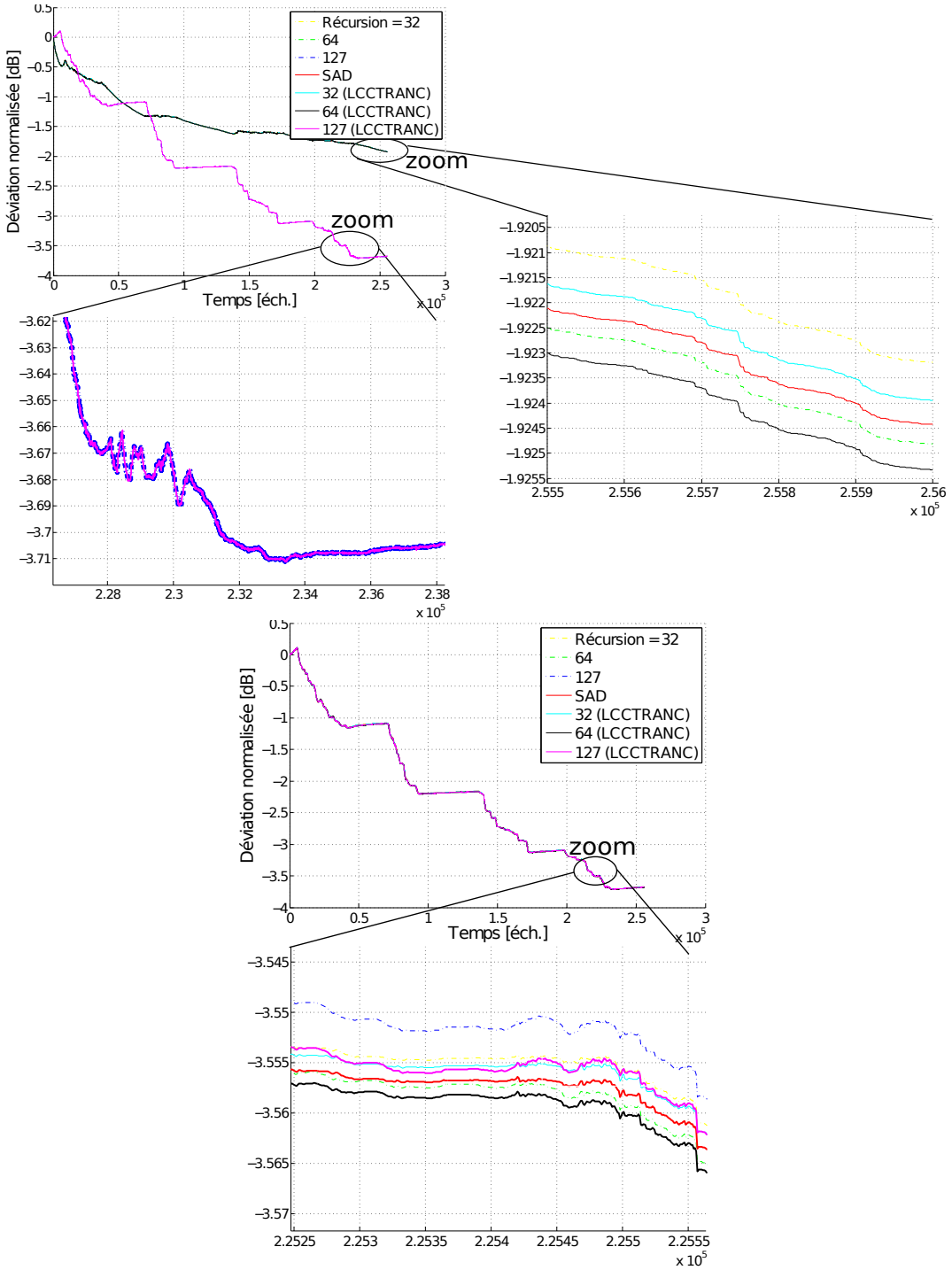


Figure 4.12 – Evolution de la déviation normalisée des filtres $\mathbf{w_{21}}$ (haut) et $\mathbf{w_{12}}$ (bas) pour un mélange composé d'une source de parole et d'un bruit de bureau pour un RSB de 10 dB. Les pas d'adaptation sont fixés à 0,001.

accentuera l'impact du bruit de calcul. A partir des observations précédentes, la mémoire des gradients sera donc tronquée aux N premières valeurs, ce qui revient à

considérer que l'ordre de la récursion est égal à l'ordre des filtres adaptatifs utilisés. De ce fait, le support du vecteur gradient est le même que celui des filtres adaptatifs, ce qui permet de traiter le vecteur gradient de la même manière que les filtres dans l'implémentation fréquentielle qui fait l'objet du paragraphe 4.4.

4.4 Implémentation fréquentielle et performances en environnement réel

Le nouvel algorithme que nous proposons rend la déclinaison dans le domaine fréquentiel du CTRANC plus aisée. En effet, l'apparition de la convolution dans l'expression des gradients nous permet de mettre à profit l'algorithme de transformée de Fourier rapide (*Fast Fourier Transform* (FFT)) et donc de bénéficier de traitements plus fins en comparaison avec les traitements temporels, rapides et peu coûteux grâce notamment aux versions optimisées de la FFT. L'implémentation fréquentielle du LCCTRANC est donnée à l'**Algorithme 4** et s'appuie sur l'algorithme LMS fréquentiel (FLMS) basé sur la technique de l'OLS décrit par Shynk dans [165]. La différence majeure avec le FLMS est visible aux lignes 14, 15 et 23, faisant apparaître la convolution et la corrélation de trois vecteurs dans le calcul des gradients. De ce fait, afin d'éviter les problèmes de repliement temporel, nous doublons la taille de la TFD qui sera prise égale $M = 2N$, avec N la longueur des filtres, et non pas $M = N$ comme cela est fait classiquement. Enfin, dans nos simulations, la quantité 2μ faisant intervenir un pas fixe sera remplacée par les pas variables non paramétriques proposés au paragraphe 3.2.3. Par conséquent, l'**Algorithme 2** est placé au début de la boucle de l'**Algorithme 4** comme première instruction.

4.4.1 Influence de la sous-modélisation

Une deuxième scène sonore, toujours composée de deux sources placées de manière symétrique, décrit le cas où la source de parole est placée à 80 *cm* et la source de bruit à 2 *m* des microphones. Les réponses impulsionnelles sont générées pour un environnement de bureau classique toujours à partir de l'outil *Roomsim*. L'objectif est d'évaluer l'influence de la sous-modélisation sur les performances du LCCTRANC pour un mélange simplifié. Les critères utilisés à présent sont la réduction de bruit (**NR**), le gain en RSB (**G**) et la distance cepstrale (**DC**). Ceux-ci sont décrits en Annexe C. Le mélange résulte de la concaténation de deux doubles phrases, l'une prononcée par un homme, l'autre par une femme et d'un bruit de bureau. Le RSB en entrée est de 5 *dB* et le pas du filtre adaptatif $\mathbf{w_{21}}$ est ajusté pour converger de manière optimale dans chacun des algorithmes. Le pas du filtre adaptatif $\mathbf{w_{12}}$ est quant à lui identique pour les deux algorithmes. La longueur de chacun des deux

Algorithme 4 Algorithme LCCTRANC dans le domaine fréquentiel - Méthode OLS

Initialisations et définitions des matrices

$\mathbf{W_{21}}(0) \leftarrow [0,\ldots,0]^T \qquad \mathbf{GradE_{1old}}(0) \leftarrow [0,\ldots,0]^T$

$\mathbf{W_{12}}(0) \leftarrow [0,\ldots,0]^T \qquad \mathbf{GradE_{2old}}(0) \leftarrow [0,\ldots,0]^T$

$\mathbf{G}(0) \leftarrow [0,\ldots,0]^T \qquad P_{r_m}(0) = P_{p_m}(0) \leftarrow \epsilon_m \quad m = 0,\ldots,2M-1$

$$\mathbf{g} \leftarrow \begin{bmatrix} \mathbf{I_M} & \mathbf{0_M} \\ \mathbf{0_M} & \mathbf{0_M} \end{bmatrix}$$

ă $\quad \mathbf{k} \leftarrow [\mathbf{0_M}, \mathbf{I_M}]$: fenêtrage $\qquad \mathbf{F}$ = matrice de la TFD $2M \times 2M$

Début

pour chaque nouveau bloc de M échantillons faire

//Signal sur la voie de référence et associé à s_2

$\mathbf{Y_1}(k) \leftarrow diag\{\mathbf{F}[y_1(kM-M),\ldots,y_1(kM),\ldots,y_1(kM+M-1)]^T\}$

$\mathbf{D_2}(k) \leftarrow \mathbf{Y_1}(k)\mathbf{W_{12}}(k)$ //Convolution

$\mathbf{d_2}(k) \leftarrow \mathbf{kF}^{-1}\mathbf{D_2}(k)$ //OLS

$\mathbf{y_2}(k) \leftarrow \mathbf{x_2}(k) - \mathbf{d_2}(k)$ //Calcul de l'erreur

$\mathbf{Y_2}(k) \leftarrow \mathbf{Fk}^T\mathbf{y_2}(k)$

$P_{r_m}(k) \leftarrow \lambda P_{r_m}(k-1) + (1-\lambda)\left|\mathbf{Y_1}_m(k)\right|^2 \quad m = 0,\ldots,2N-1$

$\mu_{\mathbf{r}}(k) \leftarrow diag\{P_{r_0}^{-1}(k),\ldots,P_{r_{2N-1}}^{-1}(k)\}$ //Pas

$\mathbf{G}(k) \leftarrow \mathbf{W_{21}}(k)\mathbf{W_{12}}(k)$ //Gradient stochastique

$\mathbf{W_{12}}(k+1) \leftarrow \mathbf{W_{12}}(k) + 2\mu\mathbf{FgF}^{-1}\mu_{\mathbf{r}}(k)\mathbf{Y_2}(k)\left[\mathbf{Y_1}^H(k) - (\mathbf{G}(k)\mathbf{GradE_{2old}}(k))^H\right]$

$\mathbf{GradE_{2old}}(k) \leftarrow \mathbf{FgF}^{-1}\mu_{\mathbf{r}}(k)\mathbf{Y_2}(k)\left[\mathbf{Y_1}^H(k) - (\mathbf{G}(k)\mathbf{GradE_{2old}}(k))^H\right]$

//Signal sur la voie primaire et associé à s_1

$\mathbf{Y_2}(k) \leftarrow diag\{\mathbf{F}[y_2(kM-M),\ldots,y_2(kM),\ldots,y_2(kM+M-1)]^T\}$

$\mathbf{D_1}(k) \leftarrow \mathbf{Y_2}(k)\mathbf{W_{21}}(k)$ //Convolution

$\mathbf{d_1}(k) \leftarrow \mathbf{kF}^{-1}\mathbf{D_1}(k)$ //OLS

$\mathbf{y_1}(k) \leftarrow \mathbf{x_1}(k) - \mathbf{d_1}(k)$ //Calcul de l'erreur

$\mathbf{Y_1}(k) \leftarrow \mathbf{Fk}^T\mathbf{y_1}(k)$

$P_{p_m}(k) \leftarrow \lambda P_{p_m}(k-1) + (1-\lambda)\left|\mathbf{Y_2}_m(k)\right|^2 \quad m = 0,\ldots,2N-1$

$\mu_{\mathbf{p}}(k) \leftarrow diag\{P_{p_0}^{-1}(k),\ldots,P_{p_{2N-1}}^{-1}(k)\}$ //Pas

//Gradient stochastique

$\mathbf{W_{21}}(k+1) \leftarrow \mathbf{W_{21}}(k) + 2\mu\mathbf{FgF}^{-1}\mu_{\mathbf{p}}(k)\mathbf{Y_1}(k)\left[\mathbf{Y_2}^H(k) - (\mathbf{G}(k)\mathbf{GradE_{1old}}(k))^H\right]$

$\mathbf{GradE_{1old}}(k) \leftarrow \mathbf{FgF}^{-1}\mu_{\mathbf{p}}(k)\mathbf{Y_1}(k)\left[\mathbf{Y_2}^H(k) - (\mathbf{G}(k)\mathbf{GradE_{1old}}(k))^H\right]$

fin pour

Fin

filtres est fixée à 256 coefficients, celle des réponses impulsionnelles notée L_{RI} varie de 256 à 2000 coefficients. Enfin, pour ne pas conditionner les résultats aux performances de la DAV, nous considérons une DAV idéale pour le calcul de la DSP du bruit. Les résultats obtenus au tableau 4.3 montrent la robustesse de l'algorithme

	LCCTRANC			SAD		
L_{RI}	**NR** [dB]	**G** [dB]	**DC**	**NR** [dB]	**G** [dB]	**DC**
256	13,7	14	0,09	**15,9**	**15,5**	0,08
512	**13,3**	**14**	0,2	8,3	9	0,14
1024	**5,1**	**6,1**	0,3	3,8	4,8	0,2
2000	**4,4**	**5,6**	0,3	3,6	4,8	0,3

Table 4.3 – Réduction de bruit (**NR**), gain en RSB (**G**) et distance cepstrale (**DC**) mesurés pour un bruit de bureau de RSB égal à 5 dB.

proposé vis-à-vis de la sous-modélisation. En effet, à distance cepstrale quasiment équivalente, les performances de l'algorithme LCCTRANC en termes de gain en RSB et de réduction de bruit sont systématiquement supérieures à celles de l'algorithme SAD.

4.4.2 Prise de son binaurale

Cette dernière partie compare les performances des deux algorithmes LCCTRANC et SAD pour un environnement réel de prise son selon la technique binaurale. Le contexte est décrit en Annexe E. Le mélange se compose donc de quatre réponses impulsionnelles mesurées en conditions réelles. Notons que nous avons considéré jusqu'à présent des structures sans filtrage des trajets directs. Par conséquent, les performances seront sous-optimales et pourraient être améliorées en ajoutant deux filtres derrière chacun des micros. Par exemple, nous pourrions réutiliser les filtres de type Wiener calculés pour la gestion des pas adaptatifs afin de réduire la réverbération et une partie du bruit pour un coût supplémentaire faible. Cependant, notre but n'est pas d'utiliser ces techniques directement en tant que frontal acoustique mais au sein d'une antenne. Nous conservons donc les structures telles qu'elles ont été étudiées jusqu'ici afin d'en apprécier les performances dans les conditions où elles seront utilisées. La même double phrase que précédemment est utilisée pour générer deux mélanges, l'un fait avec un bruit de bureau, l'autre fait à partir d'un bruit de restaurant pour deux RSB différents, soit 5 et 10 dB. L'estimation de la DSP du bruit est faite grâce à une DAV idéale. Enfin, dans ce cas particulier d'un mélange réel et par rapport au fait que les techniques ne soient pas en mesure de supprimer la réverbération, le signal de référence utilisé pour la mesure de la distance cepstrale est le signal de parole non bruité convolué par le chemin direct $\mathbf{a_{11}}$.

	LCCTRANC			SAD		
RSB [dB]	**NR** [dB]	**G** [dB]	**DC**	**NR** [dB]	**G** [dB]	**DC**
5	**9**	**9**	0,05	7,5	7,6	0,05
10	**8,3**	**8**	0,04	6,9	6,9	0,04

Table 4.4 – Réduction de bruit (**NR**), gain en RSB (**G**) et distance cepstrale (**DC**) mesurés pour un bruit de bureau.

	LCCTRANC			SAD		
RSB [dB]	**NR** [dB]	**G** [dB]	**DC**	**NR** [dB]	**G** [dB]	**DC**
5	**6,4**	**5,9**	0,05	5,2	5	0,05
10	**6**	**5,4**	0,04	5	4,8	0,04

Table 4.5 – Réduction de bruit (**NR**), gain en RSB (**G**) et distance cepstrale (**DC**) mesurés pour un bruit de restaurant.

Les résultats donnés aux tableaux 4.4 et 4.5 montrent, à distance cepstrale équivalente, une amélioration quasi-systématique de 1 dB des grandeurs évaluées pour l'algorithme LCCTRANC par rapport à l'algorithme SAD. Ces derniers résultats valident donc l'ensemble de l'analyse et des hypothèses formulées dans ce chapitre qui ont permis de définir le nouvel algorithme et justifier sa supériorité par rapport à l'algorithme de décorrélation.

4.5 Synthèse du chapitre

Ce chapitre clos l'étude des méthodes bicapteurs présentées dans ce mémoire et qui ont pour objectif de définir une cellule élémentaire optimale constitutive de l'antenne que nous souhaitons concevoir. Après avoir extrait de l'état de l'art l'algorithme le plus précis du point de vue théorique, ce chapitre a proposé un nouvel algorithme plus robuste, moins coûteux et facilement transposable dans le domaine fréquentiel. Celui-ci, construit à partir de l'algorithme optimal selon le critère du minimum de l'erreur quadratique moyenne, affiche logiquement des performances supérieures à celles obtenues à partir de l'algorithme basé sur la décorrélation des sorties.

Chapitre 5

Analyse de la convergence et de la stabilité de la structure récursive

5.1 Données et points bloquants du problème 144
5.1.1 Description du problème 144
5.1.2 Absence de solution analytique facilement exploitable 146
5.2 Recherche de l'état d'équilibre 149
5.2.1 Analyse de l'erreur résiduelle sur les signaux estimés 149
5.2.2 Conservation de l'énergie de l'erreur résiduelle 150
5.2.3 Analyse des Rapports Signaux à Artefacts (RSA) 152
5.2.4 Analyse de l'équilibre à partir du principe d'égalité 152
5.3 Résultats expérimentaux avec des signaux synthétiques 155
5.3.1 Convergence initiale et asymptotique 155
5.3.2 RSA sur chaque voie 156
5.4 Résultats expérimentaux avec des signaux de parole 160
5.4.1 Conservation de l'énergie de l'erreur résiduelle - Principe d'égalité . . . 160
5.4.2 RSA sur chaque voie - Relation d'équilibre 161
5.5 Etude des résultats expérimentaux dans le cas dissymétrique . . 166
5.5.1 Cas des sources de même puissance 168
5.5.2 Cas des sources de puissance différente 168
5.6 Synthèse du chapitre 173

Le chapitre 2 a été l'occasion de faire le point sur les conditions de stabilité du CTRANC, où nous avons discuté en détail de la stabilité structurelle. Mais, comme nous l'avons souligné, cela n'est qu'un aspect de la stabilité, il faut également évaluer le comportement asymptotique lié à l'algorithme mis en œuvre pour l'estimation des filtres. Plus particulièrement, ce chapitre se veut être une étude des propriétés de convergence et de l'existence de points d'équilibres de la structure dans le but de statuer non seulement sur la stabilité globale mais aussi sur les performances et la précision des estimées produites. Outre le fait qu'il faille pour toute méthode adaptative être en mesure d'évaluer ses capacités de convergence, sa stabilité et quantifier l'erreur résiduelle subsistant à l'optimum, l'analyse menée dans ce chapitre est d'autant plus importante et justifiée qu'il existe, au final, peu de travaux qui traitent ces différents aspects pour la structure récursive. Lorsque le problème est considéré, il est traité le plus souvent dans (i) des cas simples [71] (filtres de mélange ayant entre 1 et 4 coefficients, signaux stationnaires) ou (ii) autour de la solution optimale [29, 30] grâce à des outils mathématiques appropriés. Ceci s'explique principalement en raison de la complexité des équations décrivant le fonctionnement du CTRANC. Or, l'approche (i) ne permet pas de statuer précisément sur le comportement du CTRANC dans le cas d'un mélange complexe réaliste composé de réponses impulsionnelles acoustiques. Quant à l'approche (ii), elle impose une solution optimale théorique sans être en mesure de dire si celle-ci est unique ou non et surtout si elle est réellement atteinte en pratique. D'autre part, bien que cette solution découle logiquement du type d'algorithme considéré, toutes les analyses données sur les propriétés de convergence et de stabilité selon cette approche sont, par conséquent, dépendantes de l'algorithme mis en œuvre. A contrario, nous tentons ici de répondre aux questions de stabilité et d'existence de point d'équilibre en analysant la convergence initiale et asymptotique sans pré-supposer de la solution optimale que peut nous fournir un algorithme basé sur la minimisation de l'énergie (EQM ou décorrélation). Grâce à cette analyse, nous espérons être en mesure de fournir les clés du problème de la convergence des filtres afin de mieux piloter, contrôler cette structure pour ensuite être capable d'améliorer les performances de l'antenne dans laquelle elle sera introduite. Les travaux décrits dans ce chapitre ont été en partie publiés à SSPD 2010, voir [158].

5.1 Données et points bloquants du problème

5.1.1 Description du problème

En accord avec ce qui a été observé sur la contrainte liée à l'implémentation du CTRANC, cf. Chapitre 2, nous considérons dans l'étude qui suit la structure de la Figure 5.1, où l'on fait apparaître clairement ce retard placé de manière arbitraire

sur l'une des voies. Rappelons qu'il fait également écho aux contraintes structurelles données également dans ce même chapitre.

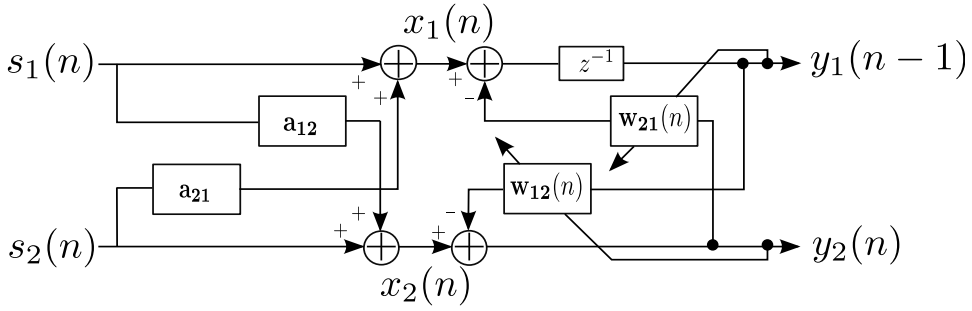


Figure 5.1 – Schéma descriptif du CTRANC faisant apparaître la contrainte liée à l'implémentation ainsi que le modèle du mélange.

En s'appropriant la notation matricielle propre à la séparation de sources, les signaux d'observation, dans le domaine de la transformée en Z, s'écrivent :

$$\begin{bmatrix} X_1(z) \\ X_2(z) \end{bmatrix} = \boldsymbol{M}(z) \begin{bmatrix} S_1(z) \\ S_2(z) \end{bmatrix}, \quad \boldsymbol{M}(z) = \begin{bmatrix} 1 & A_{21}(z) \\ A_{12}(z) & 1 \end{bmatrix}$$

avec $\boldsymbol{M}(z)$ la matrice de mélange. En pratique, les filtres de mélange sont inconnus et doivent être estimés. Ainsi, pour annuler le couplage entre les observations et recouvrer les signaux sources, l'utilisation du CTRANC présente un avantage majeur par rapport à la structure directe. En effet, théoriquement il est clair que si l'on est capable de calculer, d'estimer la matrice inverse de $\boldsymbol{M}(z)$ notée $\boldsymbol{G}(z)$ et de l'appliquer sur les signaux d'observation, alors la séparation sera effectuée et l'on retrouvera en sortie de traitement $y_1 = s_1$ et $y_2 = s_2$.

$$\boldsymbol{G}(z) = \frac{1}{1 - A_{12}(z)A_{21}(z)} \begin{bmatrix} 1 & -A_{21}(z) \\ -A_{12}(z) & 1 \end{bmatrix}$$

L'estimation de la matrice inverse $\boldsymbol{G}(z)$ et plus particulièrement du filtre RII qui la compose est une tâche délicate, déjà discutée précédemment mais néanmoins nécessaire lorsque l'on choisit une structure adaptative directe bicapteur pour l'estimation des filtres croisés. A contrario, cette opération algorithmique n'est plus nécessaire lorsque l'on choisit la structure récursive pour laquelle les filtres adaptatifs se contentent d'estimer les filtres des chemins croisés, l'inversion étant réalisée par la connection des filtres en récursif. C'est donc l'ensemble du système (algorithme + architecture) qui réalise l'estimation de $\boldsymbol{G}(z)$ et non pas seulement l'algorithme. En posant $\boldsymbol{W}(z)$ la matrice associée à l'algorithme d'estimation et en tenant compte du retard implicite, la relation d'entrée-sortie "pratique" s'écrit :

$$\begin{bmatrix} Y_1(z) \\ Y_2(z) \end{bmatrix} = \frac{1}{1 - W_{12}(z)W_{21}(z)z^{-1}} \begin{bmatrix} 1 & -W_{21}(z) \\ -W_{12}(z)z^{-1} & 1 \end{bmatrix} \begin{bmatrix} X_1(z) \\ X_2(z) \end{bmatrix}$$

Les problèmes de stabilité liés au filtre RII, $[1 - W_{12}(z)W_{21}(z)z^{-1}]^{-1}$, restent identiques à ceux étudiés au Chapitre 2. Par conséquent, dans tous les développements qui vont suivre, nous garantirons la stabilité en imposant, en accord avec la Figure 5.1, une hypothèse de causalité stricte pour le filtre $\mathbf{a_{12}}$. Cette hypothèse sera notée $(\mathcal{H}_2)$.

5.1.2 Absence de solution analytique facilement exploitable

Nous avons vu au chapitre précédent que l'approche classique pour la dérivation des filtres optimaux aboutit à des expressions complexes mais surtout dépendantes l'une de l'autre, voir (4.11). De ce fait, l'évaluation, l'étude et les améliorations potentielles pouvant être apportées à l'algorithme optimal dans un contexte particulier, comme par exemple celui du traitement de la parole, deviennent délicates. Or, ne pas pouvoir caractériser simplement les solutions optimales stables d'un algorithme adaptatif amène d'une part un sentiment de frustration et d'inachevé et d'autre part conduit surtout à ne pas l'employer en pratique. Notons par ailleurs que l'étude de la matrice Hessienne statuant habituellement sur la convergence et les conditions de stabilité, échoue également avec la structure récursive, simplement parce que même avec des sources stationnaires la matrice Hessienne associée à celle-ci est dépendante du temps, voir [164].
Une solution palliative pourrait être une approche numérique pas à pas qui permettrait d'analyser la convergence au fil de l'eau, en calculant à tour de rôle les coefficients des filtres après que ceux-ci auraient atteint un niveau de convergence fixé, puis en les ré-injectant successivement dans leur dual. Par ce biais et bien que les expressions soient dépendantes, nous pourrions vérifier si la valeur finale atteinte par chacun des filtres est bien celle attendue et, par la même, statuer sur les performances de l'algorithme utilisé. La littérature offre quelques pistes sur le comportement de la structure récursive, mais toujours sous l'oeil de l'algorithme mis en œuvre pour la mise à jour des filtres. Ces algorithmes sont basés ou bien sur les statistiques d'ordres supérieures ou bien sur des statistiques d'ordres deux et sont du type Robbins-Monro (adaptation stochastique de l'algorithme de Newton-Raphson). Van Gerven et al. analysent le comportement de l'algorithme SAD, qui est comme nous l'avons vu au Chapitre 2 une version simplifiée de l'algorithme de Newton-Raphson où l'on suppose des sources décorrélées, dans le cas de mélanges simples de sources stationnaires [70]. De manière plus complète et plus générale, les propriétés de convergence et de stabilité de la structure récursive ont été étudiées par Charkani et al. dans [29] et [30] autour du point d'équilibre supposé atteint. En effet, à l'aide de l'hypothèse de l'indépendance des sources plus forte que la simple décorrélation, les auteurs affirment que pour un algorithme adaptatif construit sur

cette statistique, le point d'équilibre est obtenu lorsque :

$$E\left[f_i(y_i(n))g_i(y_j(n-k))\right] = 0, \quad k \in [0, Q], \quad i \neq j \in 1, 2 \tag{5.1}$$

où Q est l'ordre des filtres de mélange et f_i et g_i les fonctions séparatrices qui peuvent être du type $f_i(x) = g_i(x) = x$ (ordre deux), ou encore $f_i(x) = x^3$ et $g_i(x) = x$, etc. Partant de ce postulat, ils déterminent ensuite à quelles conditions ce point est un équilibre stable de la règle d'adaptation, ce qui permet d'introduire des restrictions sur les fonctions f_i et g_i afin de garantir la stabilité et la minimisation de la variance de l'erreur asymptotique. Ce travail, qui définit les classes de fonctions séparatrices optimales, s'appuie sur les équations différentielles ordinaires (*Ordinary Differential Equation* (ODE)) qui se révèlent être une technique très efficace pour l'analyse de la convergence. Néanmoins, son défaut majeur est qu'elle ne permet pas de prouver que l'algorithme ainsi étudié converge vers un minimum global, sauf dans le cas particulier où les surfaces d'erreurs sont convexes. Ce qui n'est pas le cas d'après les expressions des EQM (4.8) et (4.9) que nous avons explicitées au Chapitre 4. Des hypothèses supplémentaires sont donc nécessaires. Nous prenons donc conscience des difficultés apportées par la connexion récursive des filtres qui, à la différence de la structure directe, ne permettent pas une étude aisée de leurs comportements. Par ailleurs, le fait que notre système soit dédié aux traitements de signaux de parole nous oblige à considérer l'algorithme des moindres carrés normalisé (NLMS). Or, si les propriétés statistiques de l'algorithme LMS ont été largement traitées dans la littérature, ce n'est pas le cas pour le NLMS pour qui la normalisation effectuée complique l'évaluation des moyennes d'ensemble. Néanmoins, des travaux récents menés par Eweda dans [60] statuent sur le comportement du NLMS pour un pas d'adaptation compris entre 0 et 2 en levant les hypothèses usuelles (i) d'indépendance entre les vecteurs régresseur successifs, (ii) d'indépendance entre les éléments du vecteur régresseur, (iii) sur la longueur du filtre adaptatif ou encore (iv) sur le type de distribution de l'entrée du filtre et du bruit. En ce sens, cette étude est d'un grand intérêt en pratique puisque, pour des contextes réalistes tels que l'annulation d'écho, le débruitage de la parole ou la formation de voies dans les antennes acoustiques, les hypothèses (i) et (ii) ne tiennent pas, les vecteurs d'entrées successifs sont fortement dépendants. A la lumière de ces travaux, nous déclinons le raisonnement pour la structure récursive en cherchant à analyser le comportement du filtre $\mathbf{w_{21}}$ avec l'approche associée à un problème d'identification de filtre. D'après la Figure 5.1, le schéma bloc associé au filtre $\mathbf{w_{21}}$ est comme donné à la Figure 5.2 ci-après. En notant Δy_1 et Δy_2 les erreurs sur les signaux estimés, nous pouvons écrire ceux-ci sous la forme

$$\begin{aligned} y_1(n) &= s_1(n) + \Delta y_1(n) \\ y_2(n) &= s_2(n) + \Delta y_2(n) \end{aligned}$$

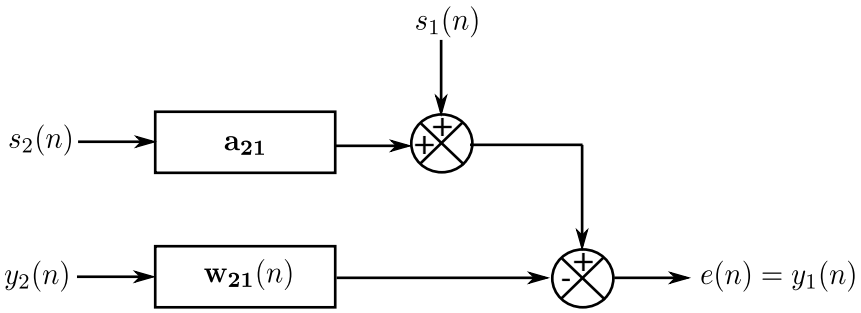


Figure 5.2 – Schéma bloc descriptif de l'environnement pour l'adaptation du filtre $\mathbf{w_{21}}$ dans la structure récursive.

A partir de ces écritures et après quelques manipulations, nous transformons le schéma bloc de la Figure 5.2 en celui de la Figure 5.3, propre à un problème d'identification. Or d'emblée et à la différence du cas traité par Eweda avec un nombre

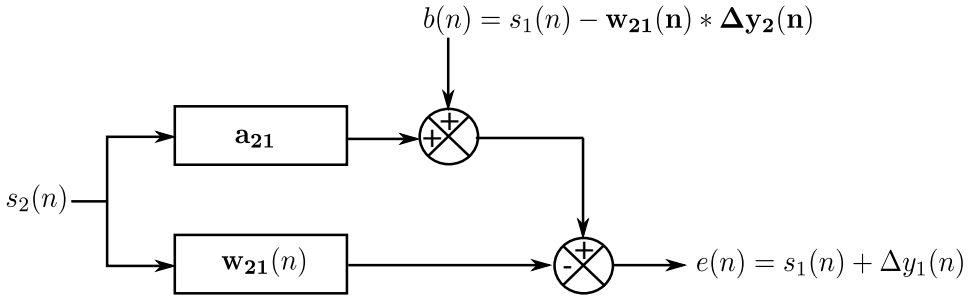


Figure 5.3 – Identification adaptative du filtre $\mathbf{a_{21}}$.

d'hypothèses très réduit, lorsque l'on calque le raisonnement pour la structure récursive comme nous venons de le faire, nous constatons que l'une des hypothèses nécessaires à l'étude du NLMS n'est pas vérifiée. En effet, dans notre cas d'analyse, le bruit $b(n)$ est tel que :

- $s_2(n)$ et $b(n)$ ne sont pas mutuellement indépendants car Δy_2 est corrélé à s_2,
- $b(n)$ est non-stationnaire du fait de la présence de $\mathbf{w_{21}}$,
- $b(n)$ n'est pas indépendant de $\mathbf{w_{21}}$.

Ainsi, les bornes fournies pour la déviation et l'erreur quadratique moyenne en excès ne s'appliquent pas. Des expressions analytiques théoriques sur les statistiques de l'algorithme NLMS appliqué à la structure récursive sont à notre connaissance inexistantes dans la littérature et restent un sujet de recherche ouvert. En regard de ces différentes approches extraites de la littérature, nous proposons dans ce chapitre d'analyser les propriétés de convergence de la structure récursive mais indépendamment de la règle d'adaptation choisie. De même, nous ne nous restreignons pas à un mélange de sources particuliers, si ce n'est qu'il faille que celui-ci vérifie les conditions de stabilité et de bon fonctionnement de la structure récursive.

5.2 Recherche de l'état d'équilibre

Notre approche est fondée sur l'analyse des relations liant les signaux sources, les estimées et leurs artefacts au sein de la structure CTRANC. Elle prend en considération l'influence de chacun de ces signaux non pas par rapport à un algorithme adaptatif particulier mais uniquement d'un point de vue général. En effet, les estimées étant imparfaites, les artefacts seront toujours présents mais de façon plus ou moins importante. L'importance de ces artefacts sera, quant à elle, déterminée par le choix des algorithmes utilisés.

5.2.1 Analyse de l'erreur résiduelle sur les signaux estimés

D'après le modèle de mélange et de séparation de la Figure 5.1, l'erreur résiduelle sur le signal estimé $y_1(n)$ définie par $\Delta y_1(n) = y_1(n) - s_1(n)$ s'écrit :

$$\begin{aligned}\Delta y_1(n) &= x_1(n) - \mathbf{w_{21}}^T(n)\mathbf{y_2}(n) - s_1(n) \\ &= \mathbf{a_{21}}^T\mathbf{s_2}(n) - \mathbf{w_{21}}^T(n)\Big[\mathbf{x_2}(n) - diag\left\{\mathbf{W_{12}}^T(n)\widehat{\mathbf{M}}_\mathbf{1}(n-1)\right\}\Big] \end{aligned} \quad (5.2)$$

avec

$$\mathbf{a_{21}} = [a_{21}^0\ a_{21}^1 \dots a_{21}^{N-1}]^T$$
$$\mathbf{w_{21}}(n) = [w_{21}^0(n)\ w_{21}^1(n) \dots w_{21}^{N-1}(n)]^T$$
$$\mathbf{s_2}(n) = [s_2(n)\ s_2(n-1) \dots s_2(n-N+1)]^T$$
$$\mathbf{y_2}(n) = [y_2(n)\ y_2(n-1) \dots y_2(n-N+1)]^T$$
$$\mathbf{x_2}(n) = [x_2(n)\ x_2(n-1) \dots x_2(n-N+1)]^T$$
$$\widehat{\mathbf{M}}_\mathbf{1}(n-1) = [\mathbf{y_1}(n-1)\ \mathbf{y_1}(n-2) \dots \mathbf{y_1}(n-N)]$$
$$\mathbf{W_{12}}(n) = [\mathbf{w_{12}}(n)\ \mathbf{w_{12}}(n-1) \dots \mathbf{w_{12}}(n-N+1)]$$

Appliquons l'hypothèse $(\mathcal{H}_1)$ du chapitre précédent qui consiste à supposer que les coefficients des filtres varient lentement au cours du temps, soit donc que $\mathbf{w_{12}}(n-j) = \mathbf{w_{12}}(n)\ \forall\ j = 0, \dots, N-1$. Du point de vue du filtrage adaptatif, cela est équivalent à fixer un pas d'adaptation faible (comportement asymptotique). En notant que $\mathbf{x_2}(n) = \mathbf{s_2}(n) + \mathbf{M_1}^T(n)\mathbf{a_{12}}$, la relation (5.2) devient alors :

$$\Delta y_1(n) = \Delta\mathbf{a_{21}}^T(n)\mathbf{s_2}(n) - \mathbf{w_{21}}^T(n)\left[\mathbf{M_1^T}(n)\mathbf{a_{12}} - \widehat{\mathbf{M}}_\mathbf{1}^T(n-1)\mathbf{w_{12}}(n)\right] \quad (5.3)$$

où $\Delta\mathbf{a_{21}}(n) = \mathbf{a_{21}} - \mathbf{w_{21}}(n)$ est le vecteur de déviation et $\mathbf{M_1}(n) = [\mathbf{s_1}(n)\ \mathbf{s_1}(n-1) \dots \mathbf{s_1}(n-N+1)]$ la matrice des signaux sources. Introduisons maintenant la matrice $\Delta\mathbf{Y_1}(n) = \mathbf{M_1}(n) - \widehat{\mathbf{M}}_\mathbf{1}(n)$, l'expression entre parenthèses s'exprime alors comme suit :

$$\mathbf{M_1}^T(n)\mathbf{a_{12}} - \widehat{\mathbf{M}}_\mathbf{1}^T(n-1)\mathbf{w_{12}}(n) = \mathbf{M_1}^T(n)\mathbf{a_{12}} - \Big[\mathbf{M_1}^T(n-1) - \Delta\mathbf{Y_1}^T(n-1)\Big]\mathbf{w_{12}}(n)$$

avec

$$\mathbf{M_1}^T(n)\mathbf{a_{12}} - \mathbf{M_1}^T(n-1)\mathbf{w_{12}}(n) = \sum_{i=0}^{N-1} \mathbf{s_1}(n-i)a_{12}(i) - \sum_{j=0}^{N-1} \mathbf{s_1}(n-1-j)w_{12}(j) \tag{5.4}$$

Imposons la contrainte de stricte causalité ($\mathcal{H}_2$) sur $\mathbf{a_{12}}$, c'est-à-dire $a_{12}(0) = 0$ qui implique en toute logique que $w_{12}(N-1) = 0$, l'équation (5.4) conduit à

$$\begin{aligned}
\mathbf{M_1}^T(n)\mathbf{a_{12}} - \mathbf{M_1}^T(n-1)\mathbf{w_{12}}(n) &= \sum_{i=1}^{N-1} \mathbf{s_1}(n-i)a_{12}(i) - \sum_{j=0}^{N-2} \mathbf{s_1}(n-1-j)w_{12}(j) \\
&= \sum_{k=0}^{N-2} \mathbf{s_1}(n-1-k)\left[a_{12}(k+1) - w_{12}(k)\right] \\
&= \widetilde{\mathbf{M}}_1(n-1)\left(\tilde{\mathbf{a}}_{12} - \tilde{\mathbf{w}}_{12}(n)\right)
\end{aligned}$$

où $\widetilde{\mathbf{M}}_1(n-1)$ est une version tronquée de $\mathbf{M_1}(n-1)$ et de taille $N \times (N-1)$. De la même manière, $\tilde{\mathbf{a}}_{12} = \mathbf{a_{12}}(i) \mid_{i=1,\ldots,N-1}$ et $\tilde{\mathbf{w}}_{12} = \mathbf{w_{12}}(i) \mid_{i=0,\ldots,N-2}$. Ces développements conduisent à l'expression finale suivante de l'erreur résiduelle sur le signal $y_1(n)$.

$$\begin{aligned}
\Delta y_1(n) = \Delta\mathbf{a_{21}}^T(n)\mathbf{s_2}(n) - \\
\mathbf{w_{21}}^T(n)\left[\widetilde{\mathbf{M}}_1(n-1)\Delta\tilde{\mathbf{a}}_{12}(n) + \Delta\mathbf{Y_1}^T(n-1)\mathbf{w_{12}}(n)\right]
\end{aligned} \tag{5.5}$$

Les mêmes manipulations peuvent être appliquées pour le second signal estimé $y_2(n)$ et donnent

$$\begin{aligned}
\Delta y_2(n) = \Delta\tilde{\mathbf{a}}_{12}^T(n)\tilde{\mathbf{s}}_1(n-1) - \mathbf{w_{12}}^T(n)\mathbf{M_2}^T(n-1)\Delta\mathbf{a_{21}}(n-1) - \\
\mathbf{w_{12}}^T(n)\Delta\mathbf{Y_2}^T(n-1)\mathbf{w_{21}}(n-1)
\end{aligned} \tag{5.6}$$

Les équations (5.5) et (5.6) définissent les erreurs résiduelles commises sur l'estimation des sources et associés à la structure récursive de la Figure 5.1. Le réseau global descriptif correspondant à ces deux relations est donné à la Figure 5.4. A partir de ce réseau, nous discutons de l'existence d'un point d'équilibre dans le paragraphe suivant.

5.2.2 Conservation de l'énergie de l'erreur résiduelle

Nous allons faire l'hypothèse ($\mathcal{H}_3$), à vérifier, que l'état d'équilibre du CTRANC est obtenu lorsque l'énergie des signaux d'erreurs résiduelles sur chaque voie est égale, soit $E\left[\Delta y_1(n)^2\right] = E\left[\Delta y_2(n)^2\right]$. En considérant toujours de faibles variations des filtres dans le temps, i.e. $\mathbf{w_{21}}(n) * \mathbf{w_{12}}(n) \approx \mathbf{w_{12}}(n) * \mathbf{w_{21}}(n-1)$, l'égalité des énergies des erreurs résiduelles implique que les corrélations des signaux en A et B

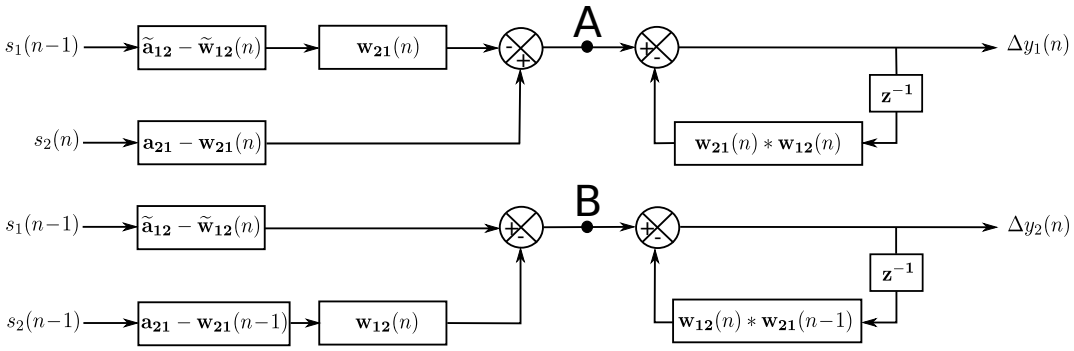


Figure 5.4 – Schéma du réseau permettant d'obtenir l'expression des signaux d'erreurs résiduelles sur chacune des voies.

(voir Figure 5.4) soient identiques, soit encore que les densités spectrales de puissance soient égales. En notant $\Phi_{AA}(e^{j\omega})$ et $\Phi_{BB}(e^{j\omega})$ ces densités spectrales définies comme ci-après :

$$\begin{aligned}\Phi_{AA}(e^{j\omega}) &= \sum_{k=-\infty}^{\infty} r_{AA}(k)e^{-jk\omega} \\ \Phi_{BB}(e^{j\omega}) &= \sum_{k=-\infty}^{\infty} r_{BB}(k)e^{-jk\omega}\end{aligned}$$

où $r_{AA}(k) = E\left[A(n)A(n+k)\right]$ est la matrice de corrélation du signal au point A et $r_{BB}(k) = E\left[B(n)B(n+k)\right]$ celle au point B, l'hypothèse $(\mathcal{H}_3)$ se traduit par la relation suivante :

$$\Phi_{AA}(e^{j\omega}) = \Phi_{BB}(e^{j\omega}) \tag{5.7}$$

Supposons $(\mathcal{H}_4)$ que $s_1(n)$ et $s_2(n)$ sont de moyennes nulles et indépendants, nous obtenons :

$$\begin{aligned}\Phi_{AA}(e^{j\omega}) &= \Phi_{S_1S_1}(e^{j\omega})\left|\widetilde{A}_{12}(e^{j\omega}) - \widetilde{W}_{12}(e^{j\omega})\right|^2\left|W_{21}(e^{j\omega})\right|^2 \\ &\quad + \Phi_{S_2S_2}(e^{j\omega})\left|A_{21}(e^{j\omega}) - W_{21}(e^{j\omega})\right|^2\end{aligned} \tag{5.8}$$

Le même raisonnement vaut également en B avec $r_{BB}(k)$. En faisant apparaître le vecteur de déviation, les densités spectrales de puissance peuvent être réécrites comme suit :

$$\begin{aligned}\Phi_{AA}(e^{j\omega}) &= \Phi_{S_1S_1}(e^{j\omega})\left|\Delta\widetilde{A}_{12}(e^{j\omega})\right|^2\left|W_{21}(e^{j\omega})\right|^2 \\ &\quad + \Phi_{S_2S_2}(e^{j\omega})\left|\Delta A_{21}(e^{j\omega})\right|^2\end{aligned} \tag{5.9}$$

$$\begin{aligned}\Phi_{BB}(e^{j\omega}) &= \Phi_{S_2S_2}(e^{j\omega})\left|\Delta A_{21}(e^{j\omega})\right|^2\left|W_{12}(e^{j\omega})\right|^2 \\ &\quad + \Phi_{S_1S_1}(e^{j\omega})\left|\Delta\widetilde{A}_{12}(e^{j\omega})\right|^2\end{aligned} \tag{5.10}$$

En remplaçant (5.9) et (5.10) dans (5.7), nous obtenons la relation d'équilibre supposée (5.11) et associée à la structure.

$$\frac{\left|\Delta\widetilde{A}_{12}(e^{j\omega})\right|^2}{|\Delta A_{21}(e^{j\omega})|^2} = \frac{\Phi_{S_2S_2}(e^{j\omega})}{\Phi_{S_1S_1}(e^{j\omega})} \cdot \frac{|W_{12}(e^{j\omega})|^2 - 1}{|W_{21}(e^{j\omega})|^2 - 1} \tag{5.11}$$

Notons que cette relation est valide quelle que soit la valeur des coefficients des filtres adaptatifs. Seules des hypothèses de stationnarité ont été nécessaires.

5.2.3 Analyse des Rapports Signaux à Artefacts (RSA)

Grâce à l'étude précédente, le RSA sur chaque voie peut être facilement explicité :

$$\begin{aligned} RSA_1(e^{j\omega}) &= \frac{\Phi_{S_1S_1}(e^{j\omega})}{\Phi_{\Delta S_1\Delta S_1}(e^{j\omega})} \\ &= \Phi_{S_1S_1}(e^{j\omega})\frac{\left|1 - W_{12}(e^{j\omega})W_{21}(e^{j\omega})e^{-j\omega}\right|^2}{\Phi_{AA}(e^{j\omega})} \end{aligned} \tag{5.12}$$

$$\begin{aligned} RSA_2(e^{j\omega}) &= \frac{\Phi_{S_2S_2}(e^{j\omega})}{\Phi_{\Delta S_2\Delta S_2}(e^{j\omega})} \\ &= \Phi_{S_2S_2}(e^{j\omega})\frac{\left|1 - W_{12}(e^{j\omega})W_{21}(e^{j\omega})e^{-j\omega}\right|^2}{\Phi_{BB}(e^{j\omega})} \end{aligned} \tag{5.13}$$

Ensuite, à partir du principe d'égalité énoncé précédement en (5.7) nous avons $\Phi_{AA}(e^{j\omega}) = \Phi_{BB}(e^{j\omega})$. Ainsi, la relation d'équilibre peut également se mettre sous la forme :

$$\frac{RSA_1(e^{j\omega})}{RSA_2(e^{j\omega})} = \frac{\Phi_{S_1S_1}(e^{j\omega})}{\Phi_{S_2S_2}(e^{j\omega})} \tag{5.14}$$

Cette relation décrit le comportement de la structure récursive et est valide quelles que soient la nature des signaux sources et la structure des réponses impulsionnelles des filtres de mélange $\mathbf{a_{12}}(n)$ et $\mathbf{a_{21}}(n)$.

Remarque 1 : *Notons que si les deux sources $s_1(n)$ et $s_2(n)$ correspondent à des bruits blancs, de puissances respectives $\sigma^2_{s_1}$ et $\sigma^2_{s_2}$, le rapport donné par (5.14) est égal à celui des puissances moyennes $\sigma^2_{s_1}/\sigma^2_{s_2}$ des deux sources.*

5.2.4 Analyse de l'équilibre à partir du principe d'égalité

Du point de vue du principe d'équilibre énoncé précédemment, si l'on analyse les relations existantes entre les signaux sources et leurs artefacts (cf. Figure 5.4), alors nous sommes ramené à l'étude d'un problème d'optimisation dont les inconnues sont les filtres $\mathbf{w_{12}}(n)$ et $\mathbf{w_{21}}(n)$. En effet, un algorithme sera considéré efficace s'il

minimise l'énergie des signaux d'erreurs résiduelles $\Delta y_1(n)$ et $\Delta y_2(n)$, soit idéalement $E\left[\Delta y_1(n)^2\right] = E\left[\Delta y_2(n)^2\right] = 0$. Or, du point de vue énergétique, la contribution de la rétroaction faite par la convolution des filtres $\mathbf{w_{12}}$ et $\mathbf{w_{21}}$ est identique sur chacun des sous-réseaux. Nous pouvons donc faire abstraction de cette boucle et conclure que l'algorithme aura atteint son point d'équilibre optimal lorsque l'énergie des signaux en A et B sera nulle. Ceci sera assuré en A (respectivement B) si l'énergie de la voie associée au signal de diaphonie $s_2(n)$ (respectivement $s_1(n-1)$) est compensée par l'énergie de la voie associée au signal de distorsion $s_1(n-1)$ (respectivement $s_2(n-1)$). Notons que pour les deux sous-réseaux, l'énergie du chemin de diaphonie est comptée positivement. Ce qui veut dire que d'un point de vue comportemental et d'après notre principe d'égalité, les filtres adaptatifs seront calculés de sorte que l'énergie de $\Delta\widetilde{\mathbf{a}}_{\mathbf{12}}(n) * \mathbf{w_{21}}(n)$ compense l'énergie de $\Delta\mathbf{a_{21}}(n)$ et dans le même temps, l'énergie de $\Delta\mathbf{a_{21}}(n-1) * \mathbf{w_{12}}(n)$ devra compenser celle de $\Delta\widetilde{\mathbf{a}}_{\mathbf{12}}(n)$. Notons que ces observations sont valides, stricto sensu, uniquement pour des signaux blancs. Par ailleurs, on peut s'attendre à ce que, théoriquement, cette solution conduise asymptotiquement vers les réponses impulsionnelles de couplage, soit $\lim_{n\to\infty} \mathbf{w_{21}}(n) = \mathbf{a_{21}}$ et $\lim_{n\to\infty} \mathbf{w_{12}}(n) = \mathbf{a_{12}}$. D'où l'écriture du double problème d'optimisation suivant :

$$\min_{W_{21}(e^{j\omega})} \quad \left|\Delta A_{21}(e^{j\omega})\right|^2 - \left|\Delta \widetilde{A}_{12}(e^{j\omega})\right|^2 \left|W_{21}(e^{j\omega})\right|^2$$

$$\text{s.c.q} \qquad \left|\Delta A_{21}(e^{j\omega})\right|^2 \leq \epsilon_1$$

Et

$$\min_{W_{12}(e^{j\omega})} \quad \left|\Delta \widetilde{A}_{12}(e^{j\omega})\right|^2 - \left|\Delta A_{21}(e^{j\omega})\right|^2 \left|W_{12}(e^{j\omega})\right|^2$$

$$\text{s.c.q} \qquad \left|\Delta \widetilde{A}_{12}(e^{j\omega})\right|^2 \leq \epsilon_2$$

La décomposition du problème global de la convergence des deux filtres en deux sous-problèmes interdépendants est volontaire, car comme nous allons le constater, la convergence des filtres sera effectivement dictée par le principe d'égalité mais différemment selon trois cas de figure. Ces différents cas sont répertoriés à la Figure 5.5 et résumés dans le tableau 5.1. Le premier cas, estampillé Cas 1, correspond à une disposition symétrique des sources et dont la trajectoire se situe sur les cercles et dans le plan des microphones. De plus, en supposant une propagation homogène, isotrope et sans obstacles, le Cas 1 couvre l'ensemble des configurations pour lesquelles l'énergie des chemins croisés est identique et inférieure à celle des chemins directs, ce qui implique que les déplacements des deux sources se font également de manière symétrique tout en s'assurant que chacune d'elle reste dans son propre demi-plan.

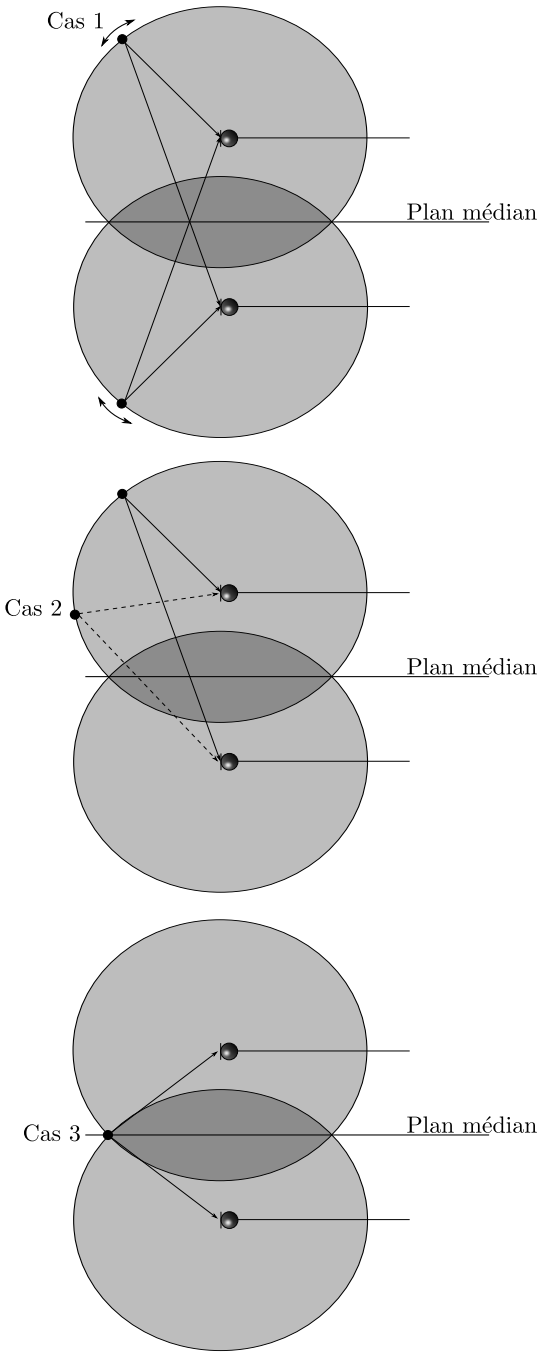


Figure 5.5 – Configuration des sources amenant à la résolution du double problème d'optimisation donnant les solutions optimales selon trois cas.

Cas 1	$\| a_{12} \|^2 = \| a_{21} \|^2 \ll 1$	Disposition symétrique
Cas 2	$\| a_{12} \|^2 \ll 1$, $\| a_{21} \|^2 \ll 1$ et $\| a_{12} \|^2 > \| a_{21} \|^2$	Cas dissymétrique
Cas 3	$\| a_{12} \|^2 = \| a_{21} \|^2 = 1$	Sources non séparables

Table 5.1 – Lien entre la configuration des sources et l'énergie des chemins croisés pour une propagation homogène, isotrope et sans obstacles. Les chemins directs sont choisis tel que $\| a_{11} \|^2 = \| a_{22} \|^2 = 1$.

Le second cas, estampillé Cas 2, couvre l'ensemble des configurations pour lesquelles l'énergie des chemins croisés est différente. Par conséquent, en fixant la position de l'une des sources, le Cas 2 correspond pour l'autre source à l'ensemble des positions où l'énergie de son chemin croisé est strictement supérieure ou inférieure à celle du chemin croisé de la source fixe. Cela revient donc à rapprocher ou éloigner dans son demi-plan l'une des sources. Il correspond également à des déplacements selon la même direction et sur les cercles des deux sources. Enfin, le dernier cas correspond à la position critique où les sources se superposent ou sont très proches, de sorte que l'énergie des chemins croisés est égale à celle des chemins directs. Dans ce cas, nous avons montré au Chapitre 2 que la structure CTRANC est inefficace. De ce fait, nous savons que dès lors que les sources se trouvent dans cette région appelée cône de confusion, les filtres ne seront jamais correctement estimés. Pour résumer, nous avons donc discerné trois cas dont seuls deux sont à analyser.

Les sections 5.3 et 5.4 présentent les résultats expérimentaux dans le premier cas, c'est-à-dire pour une prise de son particulière où l'on suppose que les sources sont placées de part et d'autre de l'axe médian au plan des microphones. La section 5.5 sera consacrée à l'analyse des résultats dans le cas dissymétrique, simulant ainsi le cas où les deux sources sont tournées préférentiellement vers un microphone.

5.3 Résultats expérimentaux avec des signaux synthétiques

5.3.1 Convergence initiale et asymptotique

Pour illustrer la validité de notre approche et de nos hypothèses, nous donnons dans cette partie les résultats de simulation obtenus dans un premier temps avec des signaux synthétiques. Le signal $s_2(n)$ correspond à un bruit blanc gaussien de densité spectrale de puissance $\Phi_{S_2S_2}(e^{j\omega}) = \sigma_{s_2}^2$. Le signal d'entrée $s_1(n)$ est issu d'un processus autorégressif d'ordre 1, AR(1), tel que :

$$s_1(n) = \alpha s_1(n-1) + v(n)$$

avec le pôle $\alpha \in [0, 1[$. Le processus d'innovation provient de la réalisation d'un bruit blanc gaussien de variance σ_v^2 reliée à la variance du signal d'entrée par la relation suivante :

$$\sigma_v^2 = (1 - \alpha^2)\sigma_{s_1}^2$$

Pour effectuer l'estimation des filtres $\mathbf{w_{12}}(n)$ et $\mathbf{w_{21}}(n)$, nous choisissons l'algorithme de décorrélation, [70] :

$$\mathbf{w_{ji}}(n+1) = \mathbf{w_{ji}}(n) + \mu\ y_i(n)\mathbf{y_j}(n) \qquad \forall\ i \neq j \in \{1, 2\} \tag{5.15}$$

où $\mathbf{y_j}(n) = [y_j(n) \dots y_j(n - N + 1)]^T$ désigne le vecteur des N derniers échantillons du signal estimé y_j. L'intérêt de ce choix est double. D'une part, cet algorithme est simple, d'autre part cela va dans le sens de montrer l'indépendance de l'analyse vis-à-vis de l'algorithme employé, y compris lorsque celui-ci est approximatif, cf. Chapitre 2. Afin de ne pas introduire d'autres effets perturbateurs dans l'analyse, nous assurons une modélisation parfaite en utilisant des filtres de séparation de la même longueur que ceux des filtres de mélange [1], soit $N = 20$. Le pas d'adaptation est pris tel que $\mu = 10^{-4}$.
La Figure 5.6 montre pour un RSB $(\sigma_{s_1}^2/\sigma_{s_2}^2)$ de 10 dB avec un signal coloré $s_1(n)$ $(\alpha = 0.4)$ et un bruit blanc gaussien $s_2(n)$, la convergence au cours du temps des coefficients de $\mathbf{w_{12}}(n)$ et $\mathbf{w_{21}}(n)$. Nous pouvons constater que la phase transitoire est différente entre les deux voies et que le filtre associé au signal de plus grande énergie converge plus facilement et plus rapidement. Ceci est rendu plus visible sur la Figure 5.7 où l'on trace la déviation normalisée des coefficients des filtres qui révèle l'effet de la corrélation des signaux sources uniquement sur la convergence initiale. Les performances asymptotiques restant similaires.

5.3.2 RSA sur chaque voie

Nous allons maintenant considérer la relation précédemment établie et exprimant le RSA sur chaque voie. En observant les valeurs prises par chacune des grandeurs mises en jeu à chaque instant n, il est possible de calculer les DSP $\Phi_{AA}(e^{j\omega})$ et $\Phi_{BB}(e^{j\omega})$ données en (5.9) et (5.10). Une fois ces quantités connues, nous pouvons exprimer à partir des expressions (5.12) et (5.13), la valeur des RSA par composante spectrale sur chacune des voies. Ceux-ci sont tracés sur la Figure 5.8 (haut) pour un signal $s_1(n)$ de type AR (1), de pôle $\alpha = 0,8$ et un RSB tel que $\sigma_{s_1}^2/\sigma_{s_2}^2 = 10$.

1. $\mathbf{a_{12}} = [0, -0.6060, -0.0532, 0.0131, -0.0914, -0.0415, -0.2946, -0.2086, -0.1436, 0.0614, -0.0934, 0.2331, 0.0202, -0.2556, -0.0593, -0.0810, 0.0122, -0.0248, -0.0298, -0.0746]^T$
et
$\mathbf{a_{21}} = [-0.4266, 0.1721, -0.3538, -0.1682, -0.0353, 0.0571, -0.0168, 0.3619, 0.3212, 0.1180, 0.0165, -0.1377, -0.0748, -0.2168, -0.0549, -0.0657, 0.2325, 0.0951, -0.0078, -0,0334]^T$

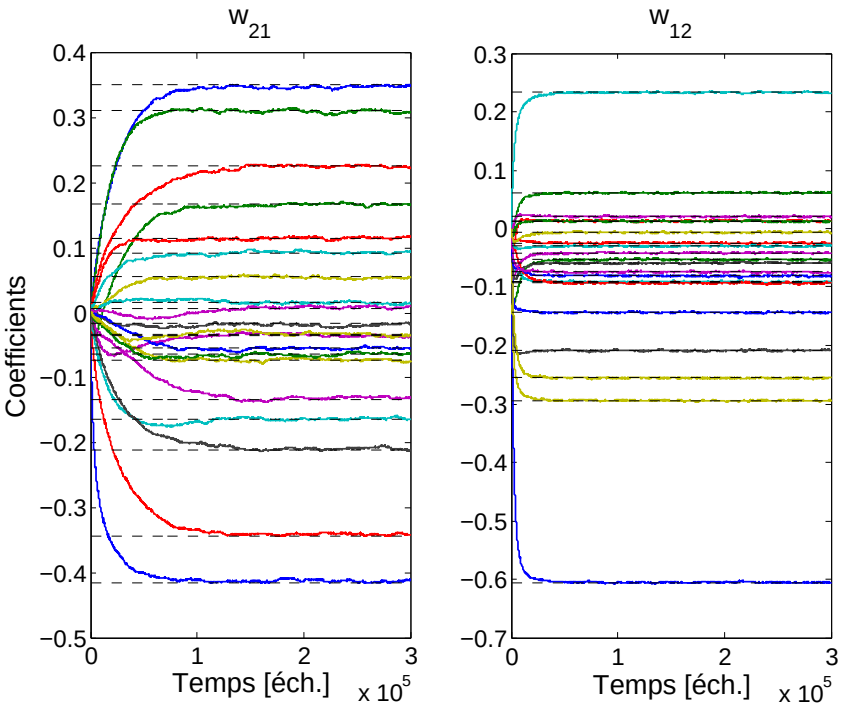


Figure 5.6 – Convergence de $\mathbf{w_{12}}(n)$ et $\mathbf{w_{21}}(n)$: (voie 1) signal AR(1) avec $\alpha = 0,4$ et $\sigma^2_{s_1} = 10$ (voie 2) bruit blanc gaussien avec $\sigma^2_{s_2} = 1$.

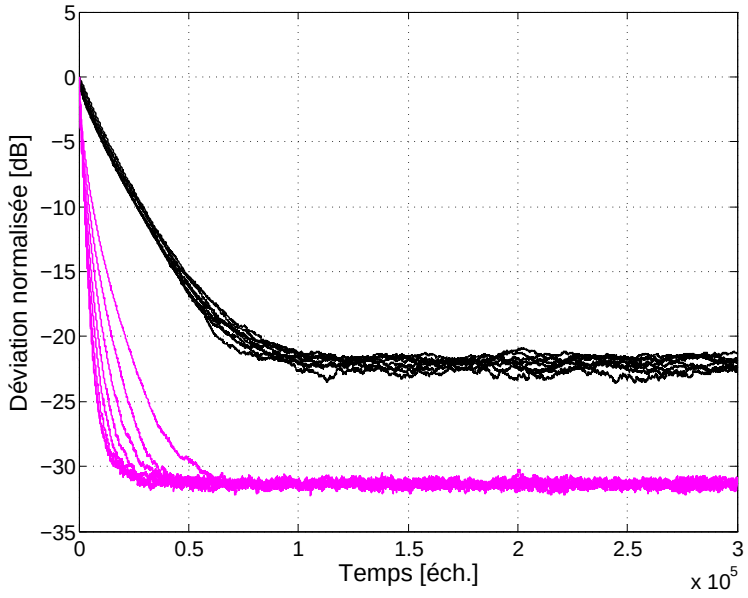


Figure 5.7 – Evolution temporelle de $\|\Delta\mathbf{a_{21}}(n)\|^2 / \|\mathbf{a_{21}}(n)\|^2$ (noir) et $\|\Delta\mathbf{a_{12}}(n)\|^2 / \|\mathbf{a_{12}}(n)\|^2$ (magenta) pour différentes valeurs du pôle $\alpha = 0, 0.1, 0.2, \cdots, 0.9$.

Par ailleurs, nous avons considéré dans nos calculs les valeurs moyennes asymptotiques des filtres, $\overline{\mathbf{w}}_{\mathbf{12}} = \lim_{n\to\infty} E\left[\mathbf{w_{12}}(n)\right]$ et $\overline{\mathbf{w}}_{\mathbf{21}} = \lim_{n\to\infty} E\left[\mathbf{w_{21}}(n)\right]$, ce qui revient à l'évaluation du comportement asymptotique des filtres adaptatifs. De la

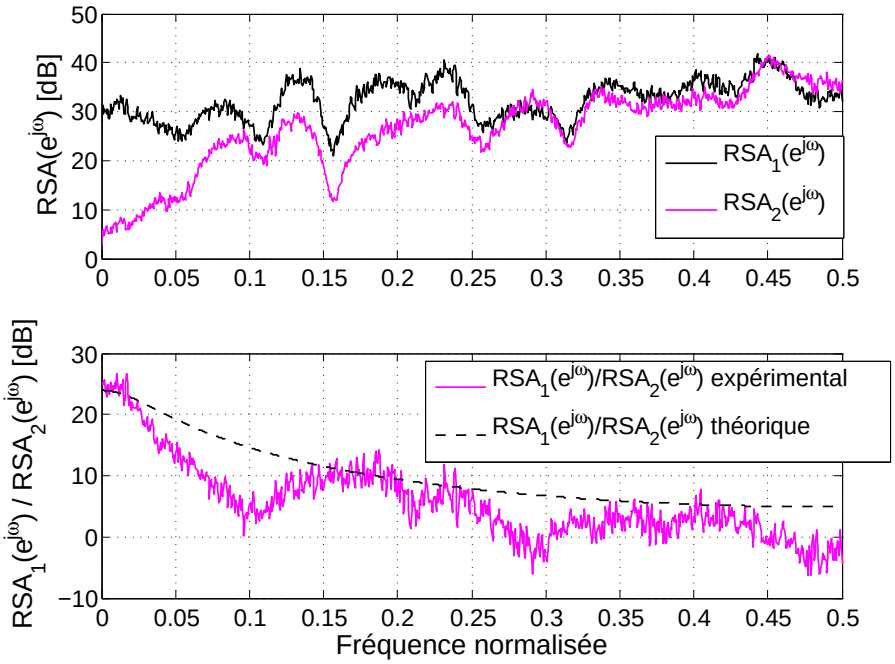


Figure 5.8 – Comportement fréquentiel du $RSA_1(e^{j\omega})$ et $RSA_2(e^{j\omega})$ (haut) et du rapport $RSA_1(e^{j\omega})/RSA_2(e^{j\omega})$ (bas).

même manière, la Figure 5.8 (bas) donne l'évolution du rapport des RSA obtenus en (5.14). Ce rapport est à l'image de la relation d'équilibre évoquée en 5.2.2 et 5.2.3. En remarquant que la DSP théorique d'un processus AR(1) s'écrit :

$$\Phi_{S_1S_1}(e^{j\omega}) = \sum_k r_{S_1S_1}(k)e^{-jk\omega} = \frac{(1-\alpha^2)\sigma_{s_1}^2}{1+\alpha^2-2\alpha cos(\omega)} \tag{5.16}$$

il est alors possible de représenter la valeur théorique de ce rapport sur cette même figure (trait discontinu) et donnée par :

$$\frac{RSA_1(e^{j\omega})}{RSA_2(e^{j\omega})} = \frac{\sigma_{s_1}^2}{\sigma_{s_2}^2} \times \frac{(1-\alpha^2)}{1+\alpha^2-2\alpha cos(\omega)} \tag{5.17}$$

L'allure des courbes obtenues par l'expérience montre la capacité de cette relation (5.17) à prédire le comportement expérimental.

Remarque 2 : *Nous rappellons que la DSP donnée par (5.16) et associée à un processus AR ne modélise que les maximas de l'enveloppe spectrale.*

En ce sens, la précision de la prédiction théorique tend à valider l'hypothèse ($\mathcal{H}_3$) faite lors de l'analyse théorique et stipulant qu'à chaque instant n, l'état stable de la structure récursive est obtenu lorsque la condition d'égalité des puissances moyennes des signaux d'erreurs résiduelles, $E\left[\Delta y_1(n)^2\right] = E\left[\Delta y_2(n)^2\right]$, est vérifiée.
Nous allons maintenant analyser les résultats en pleine bande. Pour ce faire, intégrons sur tout le domaine spectral le numérateur et le dénominateur de (5.14), nous obtenons l'expression (5.18) dont on trace la valeur expérimentale sur la Figure 5.9 (bas).

$$\frac{RSA_1^{fb}}{RSA_2^{fb}} = \frac{\int_{-\pi}^{\pi} RSA_1(e^{j\omega})d\omega}{\int_{-\pi}^{\pi} RSA_2(e^{j\omega})d\omega} = \frac{\sigma_{s_1}^2}{\sigma_{s_2}^2} \tag{5.18}$$

Le résultat obtenu est en adéquation avec l'analyse théorique. En fait, grâce aux

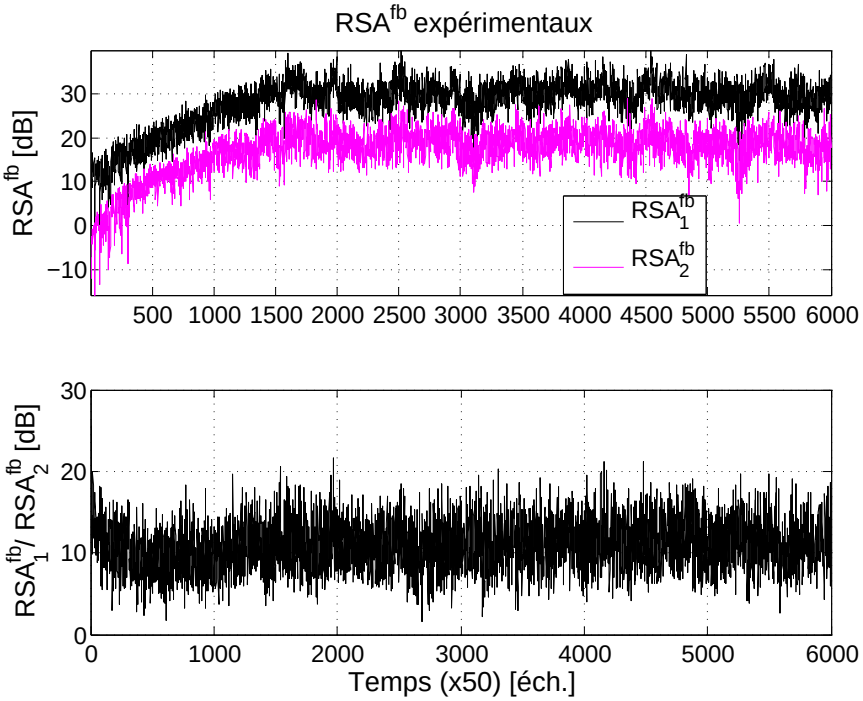


Figure 5.9 – Evolution temporelle des RSA mesurés en pleine bande. RSA_1^{fb} et RSA_2^{fb} (haut), et RSA_1^{fb}/RSA_2^{fb} (bas).

courbes données à la Figure 5.9 (haut), nous constatons que la différence (en dB) entre les RSA^{fb} de chacune des voies est égal à 10 dB, ce qui est en accord avec le RSB fixé, $\sigma_{s_1}^2/\sigma_{s_2}^2 = 10$, et y compris pendant la convergence initiale. Pour conforter notre analyse nous donnons à la Figure 5.10, pour différentes valeurs du RSB en entrée et deux bruits blancs gaussiens, l'évolution des RSA^{fb} avec en correspondance le rapport des deux. Comme nous l'avions supposé, celui-ci se comporte comme le dicte l'expression théorique (5.18) et ce, quels que soient les RSB envisagés.

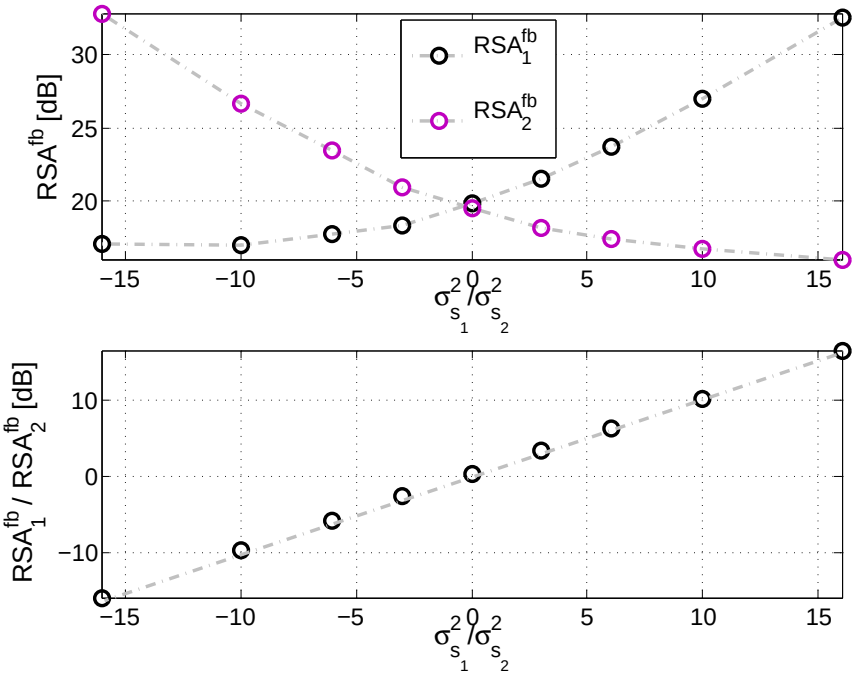


Figure 5.10 – RSA mesurés en pleine bande pour différentes valeurs de $\sigma^2_{s_1}/\sigma^2_{s_2}$.

5.4 Résultats expérimentaux avec des signaux de parole

Pour garantir de la généralité de notre approche, les signaux synthétiques s_1 et s_2 sont maintenant remplacés par des signaux de parole. Les réponses impulsionnelles restent inchangées tandis que les RSB sont fixés grâce à l'outil normalisé SV56. Le mélange est fait d'une voix d'homme et d'une voix de femme actifs aux mêmes instants, voir Figure 5.11.

5.4.1 Conservation de l'énergie de l'erreur résiduelle - Principe d'égalité

Nous allons une nouvelle fois évaluer la véracité de l'hypothèse ($\mathcal{H}_3$) dans ce contexte particulier. Les DSP expérimentales sont calculées à partir des signaux expérimentaux d'erreurs résiduelles restreints à la fenêtre temporelle qui correspond à la convergence des filtres adaptatifs. Les DSP théoriques (5.19) et (5.20), quant à elles, sont calculées à partir des transformées de Fourier des valeurs moyennes asymptotiques des filtres $\overline{\mathbf{w}}_{\mathbf{12}}$ et $\overline{\mathbf{w}}_{\mathbf{12}}$.

$$\Phi_{\Delta Y_1 \Delta Y_1}(e^{j\omega}) = \frac{\Phi_{AA}(e^{j\omega})}{|1 - W_{12}(e^{j\omega})W_{21}(e^{j\omega})e^{-j\omega}|^2} \tag{5.19}$$

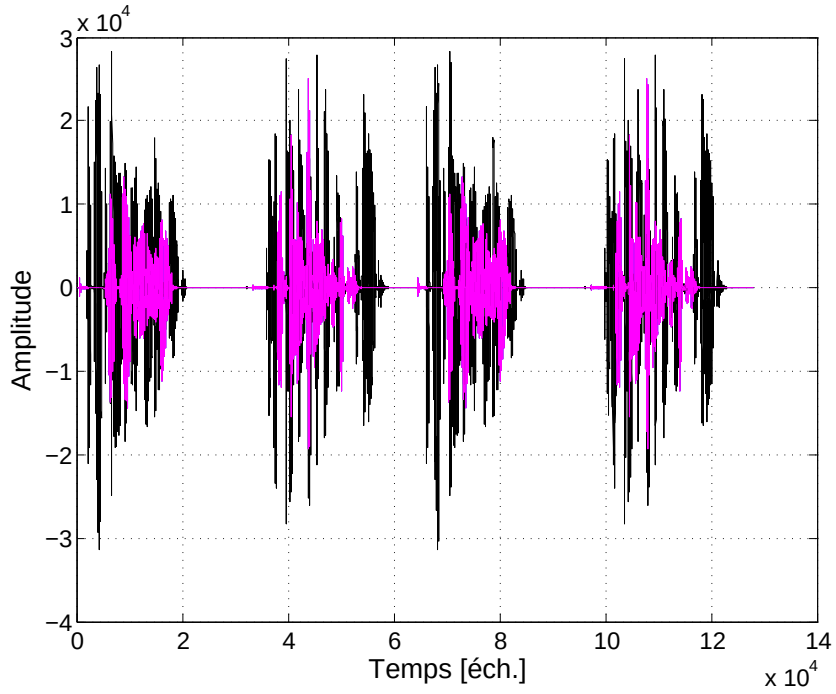


Figure 5.11 – Sources de parole actives aux mêmes instants. En noir la voix d'homme, en magenta la voix de femme.

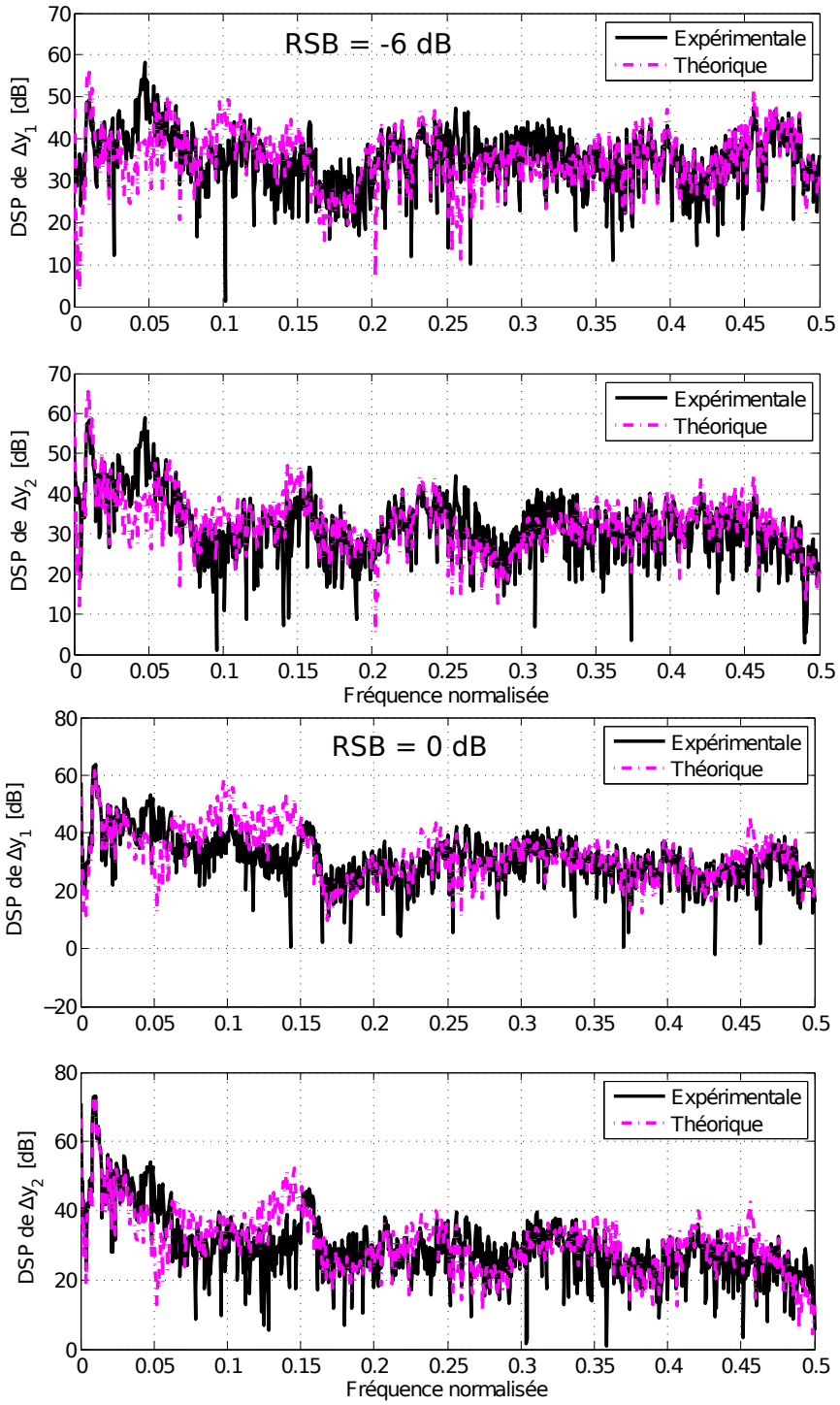
$$\Phi_{\Delta Y_2 \Delta Y_2}(e^{j\omega}) = \frac{\Phi_{BB}(e^{j\omega})}{|1 - W_{12}(e^{j\omega})W_{21}(e^{j\omega})e^{-j\omega}|^2} \tag{5.20}$$

Les résultats sont représentés à la Figure 5.12 où l'on trace les DSP théoriques et expérimentales. Comme nous pouvons le constater et pour toutes les conditions de RSB choisies, les DSP théoriques identifient suffisamment précisément les DSP expérimentales pour que nous soyons en mesure de valider le principe d'égalité des énergies des erreurs résiduelles sur chacune des voies.

5.4.2 RSA sur chaque voie - Relation d'équilibre

Nous venons de voir dans le paragraphe précédent que le principe d'égalité est vérifié y compris pour des signaux de parole. Par conséquent, la relation d'équilibre devrait, *a fortiori*, également être vérifiée. Conformément aux relations (5.12) et (5.13), les RSA sur chaque voie peuvent être réécrits comme suit :

$$\begin{aligned} RSA_1(e^{j\omega}) &= \frac{\Phi_{S_1S_1}(e^{j\omega})}{\Phi_{\Delta Y_1 \Delta Y_1}(e^{j\omega})} \\ &= \frac{\left|1 - W_{12}(e^{j\omega})W_{21}(e^{j\omega})e^{-j\omega}\right|^2}{\left|\Delta\widetilde{A}_{12}(e^{j\omega})\right|^2 |W_{21}(e^{j\omega})|^2 + \frac{\Phi_{S_2S_2}(e^{j\omega})}{\Phi_{S_1S_1}(e^{j\omega})} |\Delta A_{21}(e^{j\omega})|^2} \end{aligned} \tag{5.21}$$



$$
\begin{aligned}
RSA_2(e^{j\omega}) &= \frac{\Phi_{S_2S_2}(e^{j\omega})}{\Phi_{\Delta Y_2 \Delta Y_2}(e^{j\omega})} \\
&= \frac{\left|1 - W_{12}(e^{j\omega}) W_{21}(e^{j\omega}) e^{-j\omega}\right|^2}{\left|\Delta A_{21}(e^{j\omega})\right|^2 \left|W_{12}(e^{j\omega})\right|^2 + \frac{\Phi_{S_1S_1}(e^{j\omega})}{\Phi_{S_2S_2}(e^{j\omega})} \left|\Delta \widetilde{A}_{12}(e^{j\omega})\right|^2}
\end{aligned}
\tag{5.22}
$$

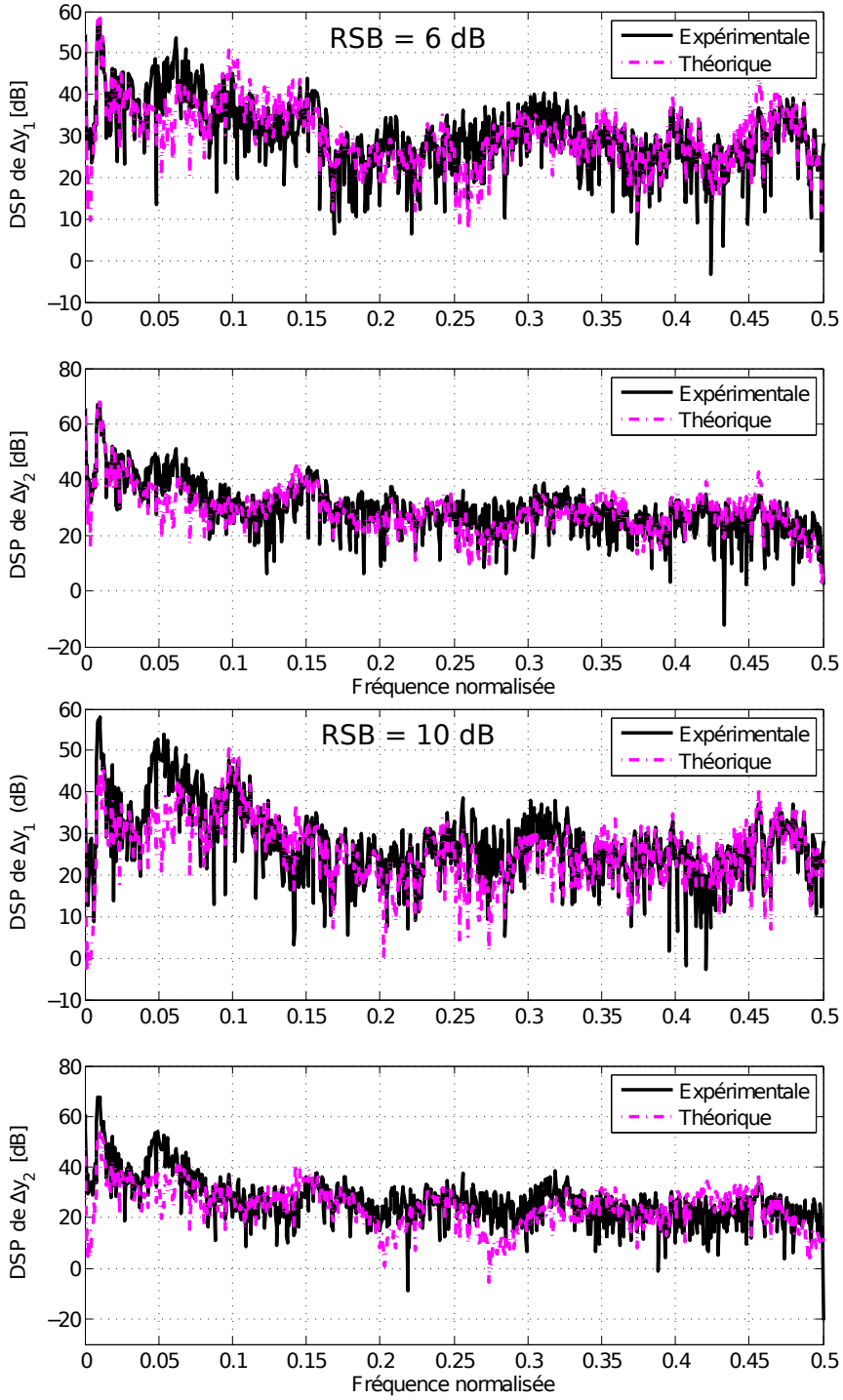
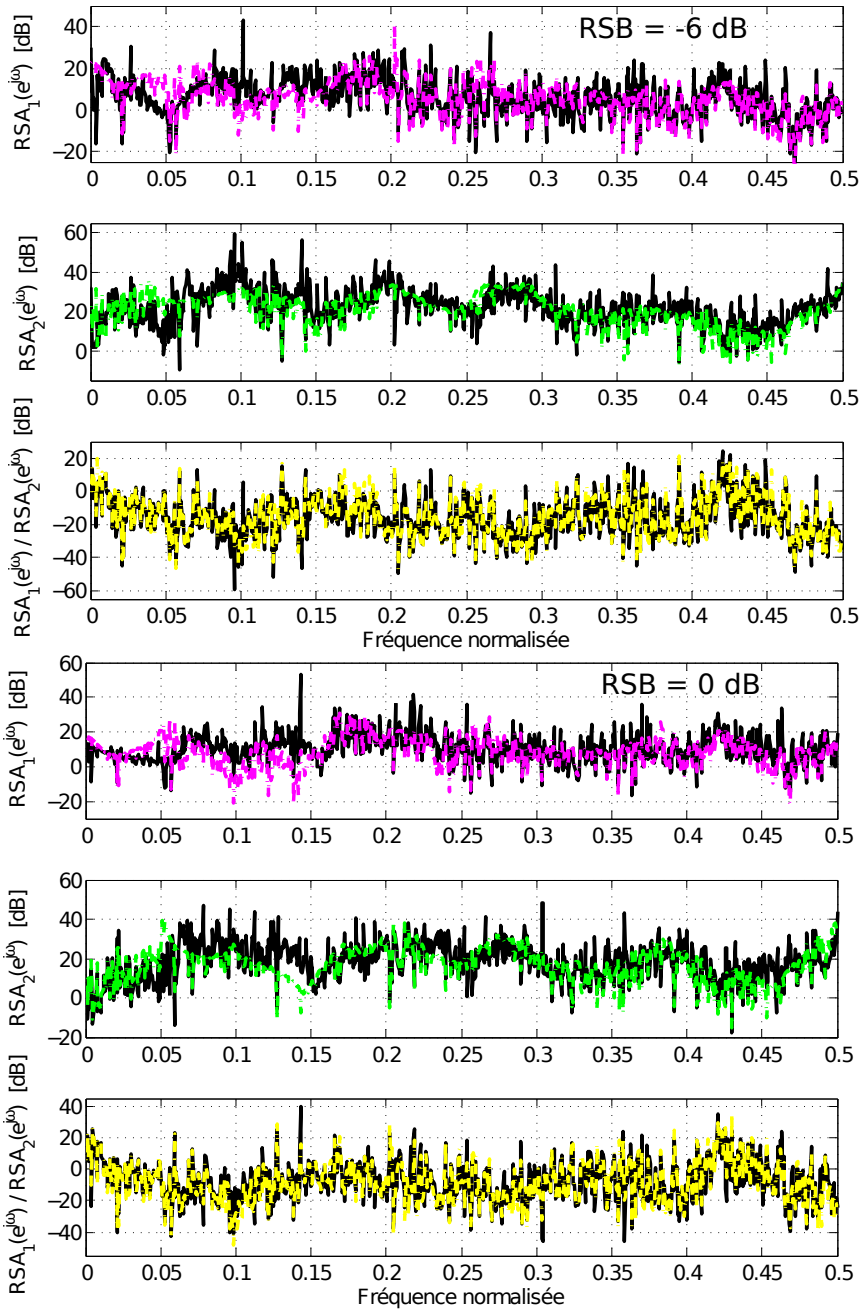


Figure 5.12 – Densités spectrales de puissance des signaux d'erreurs résiduelles pour des RSB en entrée de $\{-6, 0, 6, 10\}(dB)$.

En procédant comme précédemment, nous traçons ces expressions théoriques (pointillés) ainsi que les valeurs expérimentales (traits continus) obtenues sur la Figure 5.13.



Une fois de plus, les courbes expérimentales corroborent les courbes théoriques. Pour tous les RSB considérés, nous constatons que les courbes expérimentales se superposent aux courbes obtenues théoriquement. Enfin, la Figure 5.14 atteste de la robustesse de notre approche en montrant l'évolution du rapport des RSA^{fb} qui correspond pour chaque cas considéré, au rapport des puissances moyennes des sources de parole utilisées pour le mélange. L'ensemble des résultats expérimentaux montrent donc que, dans le cas où les sources sont disposées symétriquement par rapport au plan médian des microphones, l'hypothèse de l'existence d'un point d'équilibre

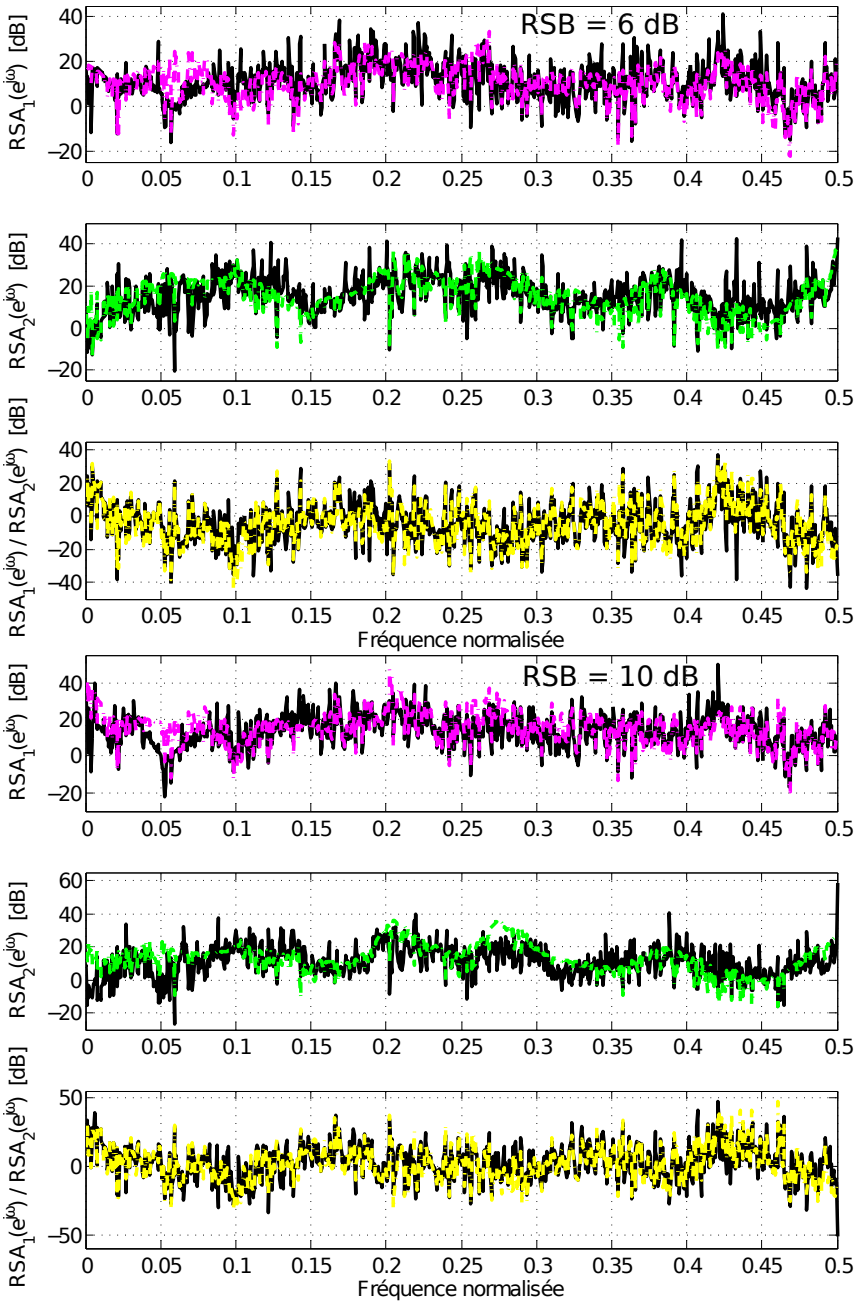


Figure 5.13 – Evolution du $RSA_1(e^{j\omega})$, $RSA_2(e^{j\omega})$ et $RSA_1(e^{j\omega})/RSA_2(e^{j\omega})$ pour un mélange de signaux de parole. Les lignes discontinues dénotent les courbes théoriques et les traits pleins, les courbes expérimentales. Les RSB en entrée valent $\{-6, 0, 6, 10\}(dB)$.

stable atteint lorsque l'énergie des signaux d'erreurs résiduelles s'égalise est valide. Cette hypothèse est de plus vérifiée quelle que soit la nature des sources. L'étude de l'influence de la position des sources et de leurs puissances sur nos analyses théoriques concernant la stabilité et la position du point d'équilibre est donnée dans la section 5.5 ci-après.

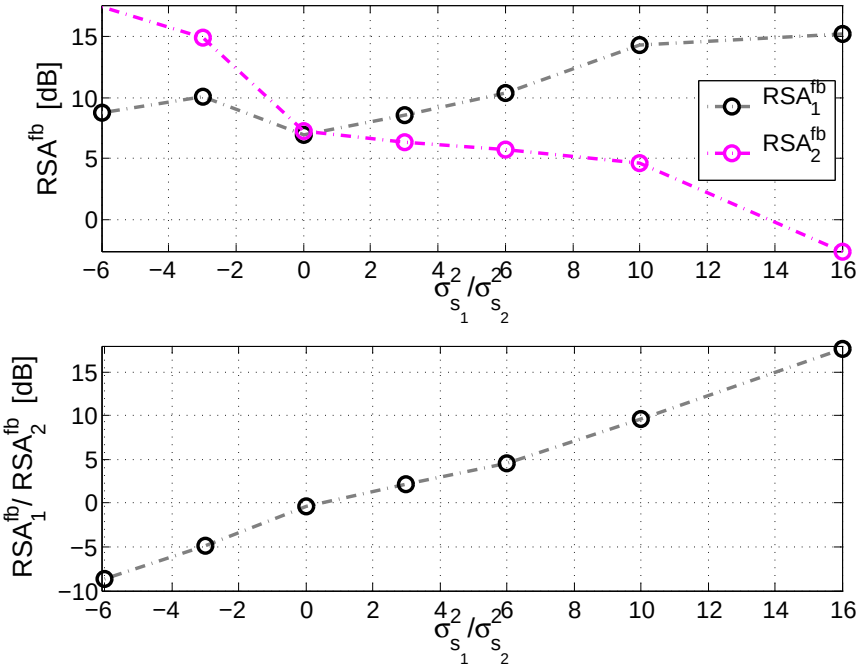


Figure 5.14 – RSA mesurés en pleine bande pour un mélange de signaux de parole et pour différentes valeurs de $\sigma^2_{s_1}/\sigma^2_{s_2}$.

5.5 Etude des résultats expérimentaux dans le cas dissymétrique

Jusqu'à présent, nous avons montré que notre analyse théorique est valide dans le cas où les sources sont placées symétriquement par rapport au plan médian des microphones avec des chemins croisés ayant la même énergie. Pour nos besoins, le modèle de mélange de la Figure 5.1 peut être représenté de manière équivalente par celui donné à la Figure 5.15. A l'aide de ce modèle nous allons être en mesure de simuler les conditions du Cas 2, en s'appuyant sur le Cas 1 déjà validé. En effet, d'après ce modèle, en fixant une valeur de Δ quelconque (cf. Figure 5.15), on impose une différence d'énergie entre les chemins qui physiquement correspond à deux cas de figure :

- $\Delta > 0$: les sources s_1 et s_2 se déplacent vers le haut dans le demi-plan du microphone 1 ;
- $\Delta < 0$: les sources s_1 et s_2 se déplacent vers le bas dans le demi-plan du microphone 2.

Pour simuler le déplacement d'une seule source, l'autre restant fixe, il suffira de placer ce Δ d'énergie sur l'un des chemins croisés uniquement, plutôt que de le répartir équitablement sur chacune des voies.

Remarque 3 : *le Cas 2 et le Cas 1 sont identiques lorsque le gain* Δ *est pris égal à* 0. *En toute rigueur et pour être en adéquation avec le schéma de la Figure 5.5, nous devrions inverser les gains* $-\frac{\Delta}{2}$ *et* $\frac{\Delta}{2}$ *sur la Figure 5.15 de sorte à simuler effectivement un rapprochement de la source* s_2. *Mais cela n'a pas d'importance sur le raisonnement, les résultats sont transposables quelle que soit la source choisie.*

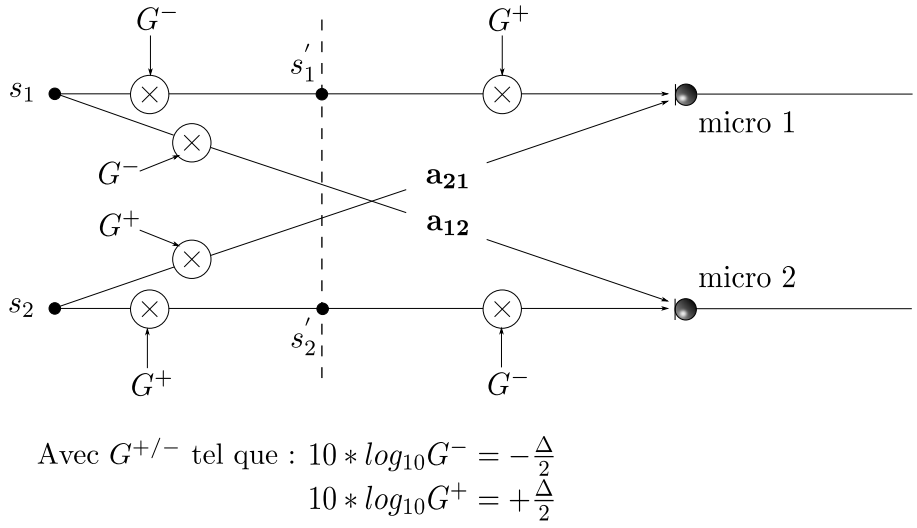


Figure 5.15 – Modèle de mélange équivalent permettant de simuler les différents cas associés aux conditions décrites au tableau 5.1.

Enfin, pour mieux comprendre les différents phénomènes qui régissent le fonctionnement de la structure récursive dans le cas d'une disposition dissymétrique des sources, nous affinons l'analyse en calculant expérimentalement les énergies de tous les signaux présents sur le réseau de la Figure 5.4, à savoir la distorsion et la diaphonie sur chacune des voies. Pour ce faire, nous appliquons, en parallèle et à chaque itération, le filtre calculé sur chacune des voies aux signaux bruités, au signal utile seul ainsi qu'au signal de bruit seul. Cette procédure est décrite en Annexe C. Le premier de ces signaux permettra donc de calculer le nouveau filtre pour la prochaine itération, le second permettra d'extraire le signal de distorsion et le dernier, le signal de diaphonie. En d'autres termes, nous décomposons donc l'erreur d'estimation Δy_i $i = 1, 2$ en deux composantes comme suit :

$$\Delta y_i(n) = c_i(n) + d_i(n) \quad i = 1, 2 \tag{5.23}$$

où c_i est le signal de diaphonie et d_i le signal de distorsion sur la voie i.

5.5.1 Cas des sources de même puissance

Dans un premier temps, nous prenons des signaux source de même puissance. Le comportement sera évalué pour deux configurations : soit les deux sources se déplacent vers le demi-plan supérieur et selon une trajectoire circulaire, soit seule une source se rapproche du plan médian, l'autre restant fixe. Ces deux configurations couvrent le Cas 2 décrit à la Figure 5.5 et que l'on peut modéliser grâce au modèle de mélange de la Figure 5.15. Par ailleurs, la nature des signaux (processus AR ou blanc) ayant peu d'influence, nous utiliserons pour nos simulations deux bruits blancs gaussiens qui suffiront à mettre en évidence le comportement de l'algorithme. Les quantitésE_{c1}, E_{c2}, E_{d1} et E_{d2} désignent respectivement les énergies des signaux de diaphonie et de distorsion sur chacune des voies 1 (micro 1) et 2 (micro 2).
D'après les résultats obtenus aux Figures 5.16 (deux sources mobiles) et 5.17 (une source mobile et l'autre fixe), nous constatons que le comportement est identique quelle que soit la configuration des sources. En effet, dans les deux cas, l'écart d'énergie sur les chemins croisés se répercute sur la précision des filtres estimés. En effet, nous observons que la convergence la plus fine, mise en image par la déviation normalisée la plus faible ($M_{W_{21}}$) est associée au chemin de plus forte énergie. La déviation normalisée $M_{W_{12}}$ se situe 5 dB au dessus, ce qui signifie que l'erreur sur le filtre $\mathbf{w_{12}}$ est plus grande. Ce comportement est en fait celui attendu puisque, d'après le principe d'égalité $\Phi_{AA}(e^{j\omega}) = \Phi_{BB}(e^{j\omega})$ (cf. 5.7) et l'équilibre sera obtenu lorsque $E\left[\Delta y_1(n)^2\right] = E\left[\Delta y_2(n)^2\right]$. Or, tout algorithme basé sur la minimisation de l'EQM aura convergé vers la solution optimale si $E\left[\Delta y_1(n)^2\right] = E\left[\Delta y_2(n)^2\right] = 0$. Ainsi, pour que l'énergie de la distorsion E_{d1} compense effectivement l'énergie de la diaphonie E_{c1}, il faut que l'énergie de $\Delta\tilde{\mathbf{a}}_{\mathbf{12}}$ soit suffisamment importante, ce qui ne peut être obtenu que si l'erreur d'estimation est à la hauteur de cette différence d'énergie. Et comme nous pouvons le constater, cet équilibre est atteint car, sur chacune des voies, l'énergie des signaux de distorsion E_{d1} et E_{d2} s'égalise de même que celle des signaux de diaphonie, $E_{c1} = E_{c2}$. La convergence des filtres s'arrête donc dès lors que les relations d'équilibre sont vérifiées. Nous mettons donc en exergue le fait que les artefacts sur le signal estimé dirigent l'adaptation en dictant le niveau maximal de convergence que peut atteindre chacun des filtres.

Remarque 4 : *En prenant* $\Delta^{'} = -\Delta$ *les propriétés de convergence s'inversent.*

5.5.2 Cas des sources de puissance différente

Nous allons maintenant étudier l'influence de la puissance des sources sur la convergence. Nous considérons uniquement le cas où une seule source est mobile

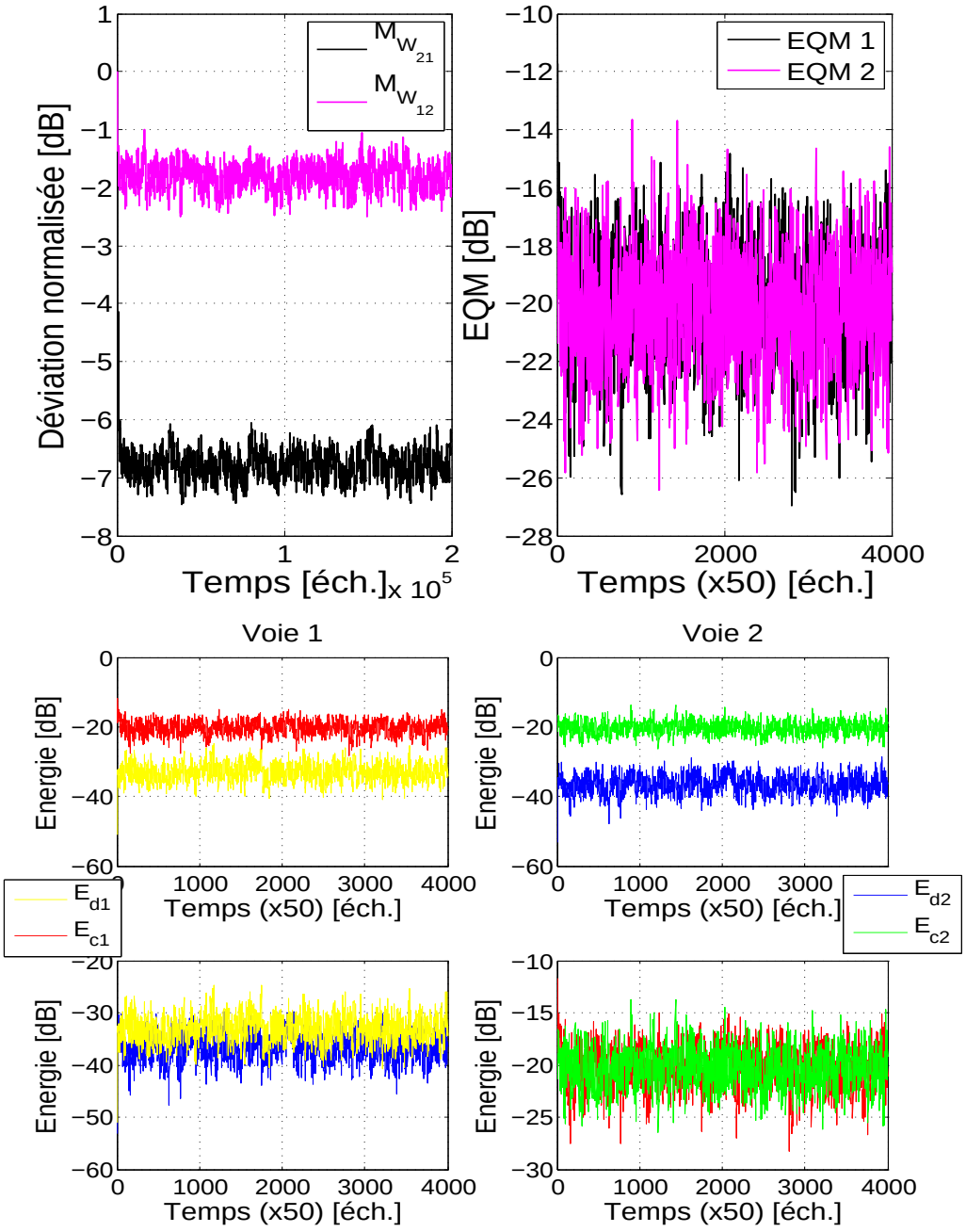


Figure 5.16 – Résultats de simulation dans le cas de deux sources mobiles. La différence d'énergie entre les chemins croisés est de $\Delta = 5\ dB$.

puisque nous avons vu précédemment que le comportement est identique dans les deux configurations. Cependant, deux cas seront appréhendés. Dans un cas, la source de plus forte puissance est la source fixe, ce qui dans notre modèle correspond au chemin croisé de plus faible énergie, Figure 5.18. Dans l'autre cas, la source rapprochée

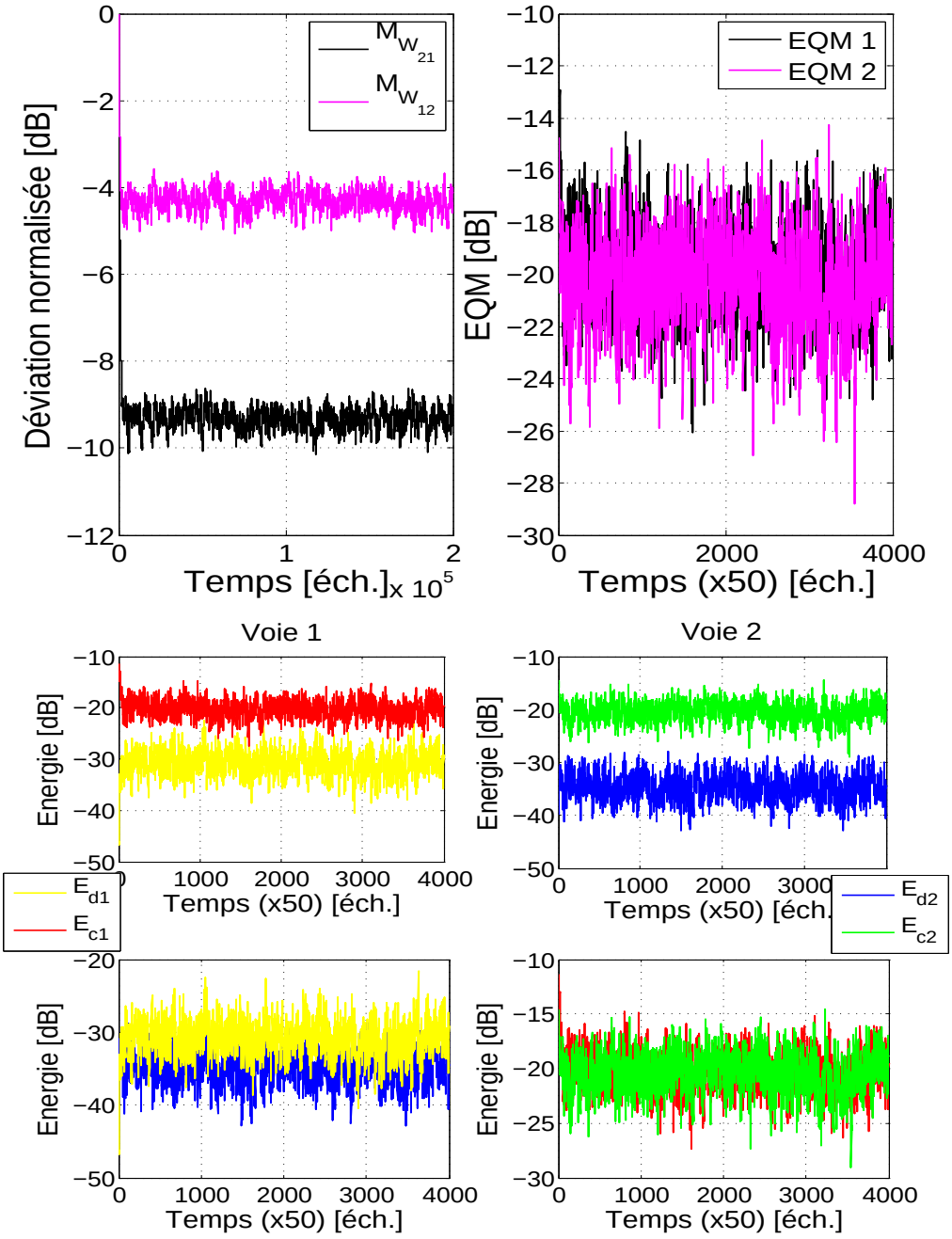


Figure 5.17 – Résultats de simulation dans le cas où une seule source se déplace. La différence d'énergie entre les chemins croisés est de $\Delta = 5\ dB$.

du plan médian sera de puissance supérieure à celle de la source fixe, Figure 5.19. Le RSB (relatif) en entrée sera de 10 dB.

D'emblée, il ressort que la puissance des sources peut contrebalancer la différence d'énergie des chemins. En effet, sur la Figure 5.18 où sont représentées les évolutions

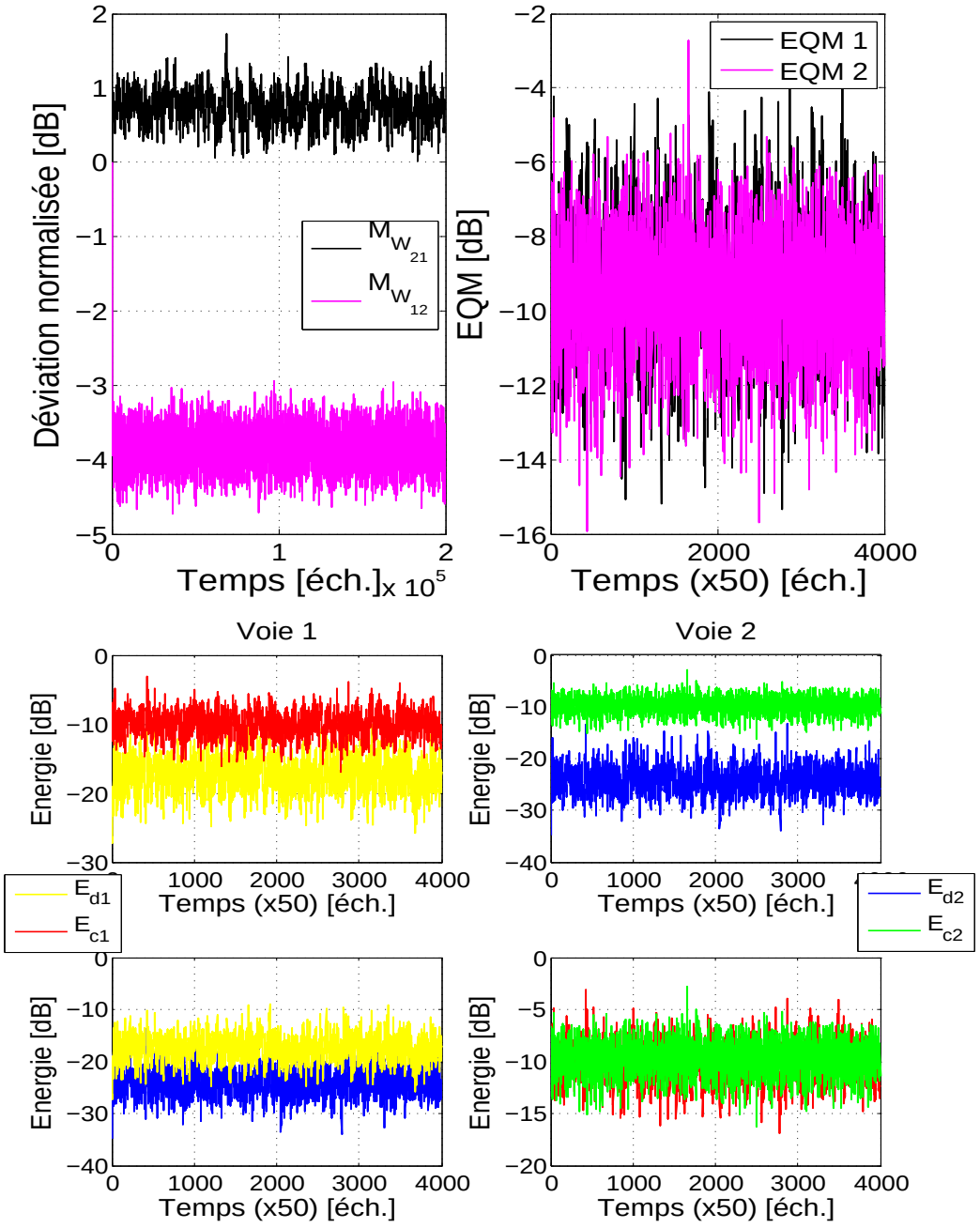


Figure 5.18 – Résultats de simulation dans le cas où une seule source se déplace. La différence d'énergie entre les chemins croisés est de $\Delta = 5\ dB$, la puissance de s_1 est de 10 dB, celle de s_2 de 1 dB.

de $M_{W_{12}}$ et $M_{W_{21}}$, nous constatons, en comparaison avec le cas traité précédemment avec des sources de même puissance, que $M_{W_{12}} < M_{W_{21}}$. Ce qui signifie que cette fois-ci le filtre $\mathbf{a_{12}}$ est le mieux estimé. A l'inverse, la puissance des sources peut

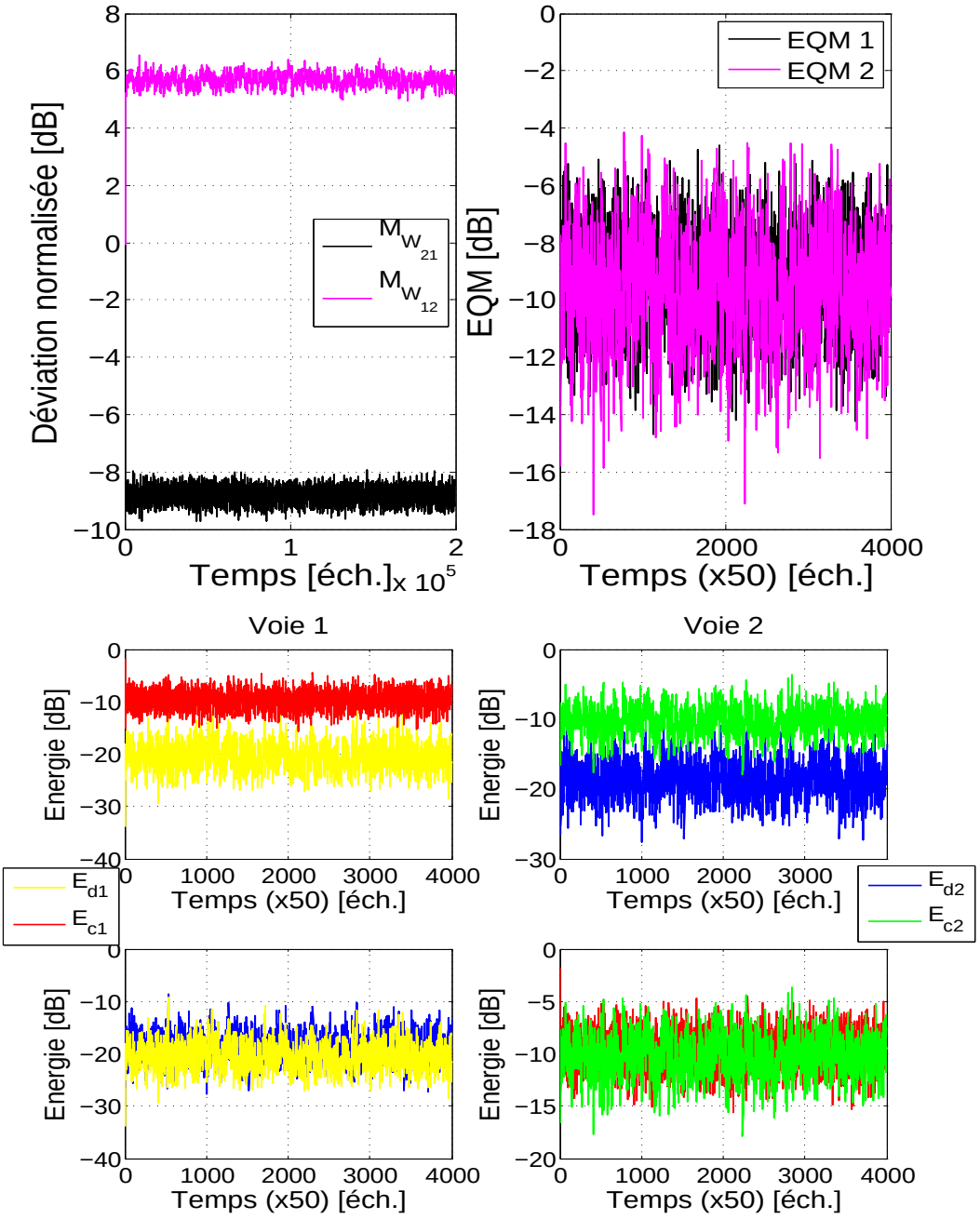


Figure 5.19 – Résultats de simulation dans le cas où une seule source se déplace. La différence d'énergie entre les chemins croisés est de $\Delta = 5\ dB$, la puissance de s_1 est de 1 dB, celle de s_2 de 10 dB.

également accentuer la différence et l'erreur nécessaire à l'équilibre sur le filtre $\mathbf{w_{12}}$. Ce cas s'observe à la Figure 5.19, où 15 dB sépare $M_{W_{12}}$ et $M_{W_{21}}$, écart dû à l'accumulation de la différence d'énergie de 5 dB entre les chemins croisés et 10 dB entre

les sources. Ainsi, ce sont deux facteurs supplémentaires qui viennent perturber la convergence des filtres et leurs possibilités d'atteindre la solution optimale, ou pour le moins tendre vers celle-ci avec une erreur la plus faible possible. Sachant que nous avons écarté le problème de la sous-modélisation, nous concevons à partir des résultats obtenus l'importance de la puissance des sources et du milieu de propagation sur l'analyse des performances des différents algorithmes qui peuvent être utilisés avec la structure récursive. Enfin, pour conforter nos résultats théoriques vis-à-vis de l'état d'équilibre et y compris dans le cas d'une disposition dissymétrique des sources, nous traçons à la Figure 5.20 les RSA expérimentaux. Si l'on s'intéresse plus particulièrement au rapport des RSA, nous pouvons observer que celui-ci correspond effectivement au rapport des puissances des sources. En effet, lorsque les sources sont de même puissance, Figure 5.20 (haut), nous avons $\Delta RSA = 0\ dB$. Et dans le cas où le $RSB = -10\ dB$ (la source s_2 étant considérée comme la source de bruit, le RSB est négatif), $\Delta RSA = -10\ dB$. Par conséquent, le principe d'égalité qui conduit à la relation d'équilibre supposée est une fois de plus vérifié. L'ensemble des résultats obtenus aux sections 5.3, 5.4 et 5.5 confirme donc l'étude théorique et les hypothèses sous-jacentes formulées au début du chapitre, après avoir confronté celles-ci à des expérimentations couvrant l'ensemble des configurations et types de sources. Par ailleurs, nos résultats sont également plus consistants et vont au-delà de la simple observation qui consiste à dire que la structure récursive converge uniquement dans le cas où le RSB sur l'un des microphones est favorable, voir [28] et [75]. De même, et contrairement à ce qui est dit dans [188], le principe d'inversion de puissance ne suffit pas à décrire complètement le comportement. En effet, en traçant à la Figure 5.21 les Rapports Signaux à Interférents (RSI) tels qu'ils sont définis par Weinstein et al., nous constatons que contrairement à la théorie exposée dans cet article, il existe des cas où les RSI calculés sur chaque sortie de la structure ne s'égalisent pas. En l'occurrence, cela se produit pour des sources de puissance différente comme le montre notre simulation.

5.6 Synthèse du chapitre

Au cours de ce chapitre, nous sommes donc parvenu à comprendre le comportement de la structure récursive en explicitant son état d'équilibre qui conditionne la convergence des filtres et en discutant de sa stabilité. L'originalité de ce travail réside dans le fait que l'analyse a été conduite sans tenir compte de la règle d'adaptation choisie. En mettant en évidence les relations existantes entre les signaux sources et leurs estimées, nous sommes allé plus en profondeur sur le comportement de cette structure que l'ensemble des travaux trouvés dans la littérature, dans le sens où nous n'avons pas eu besoin de restreindre le cadre de notre étude à un contexte particu-

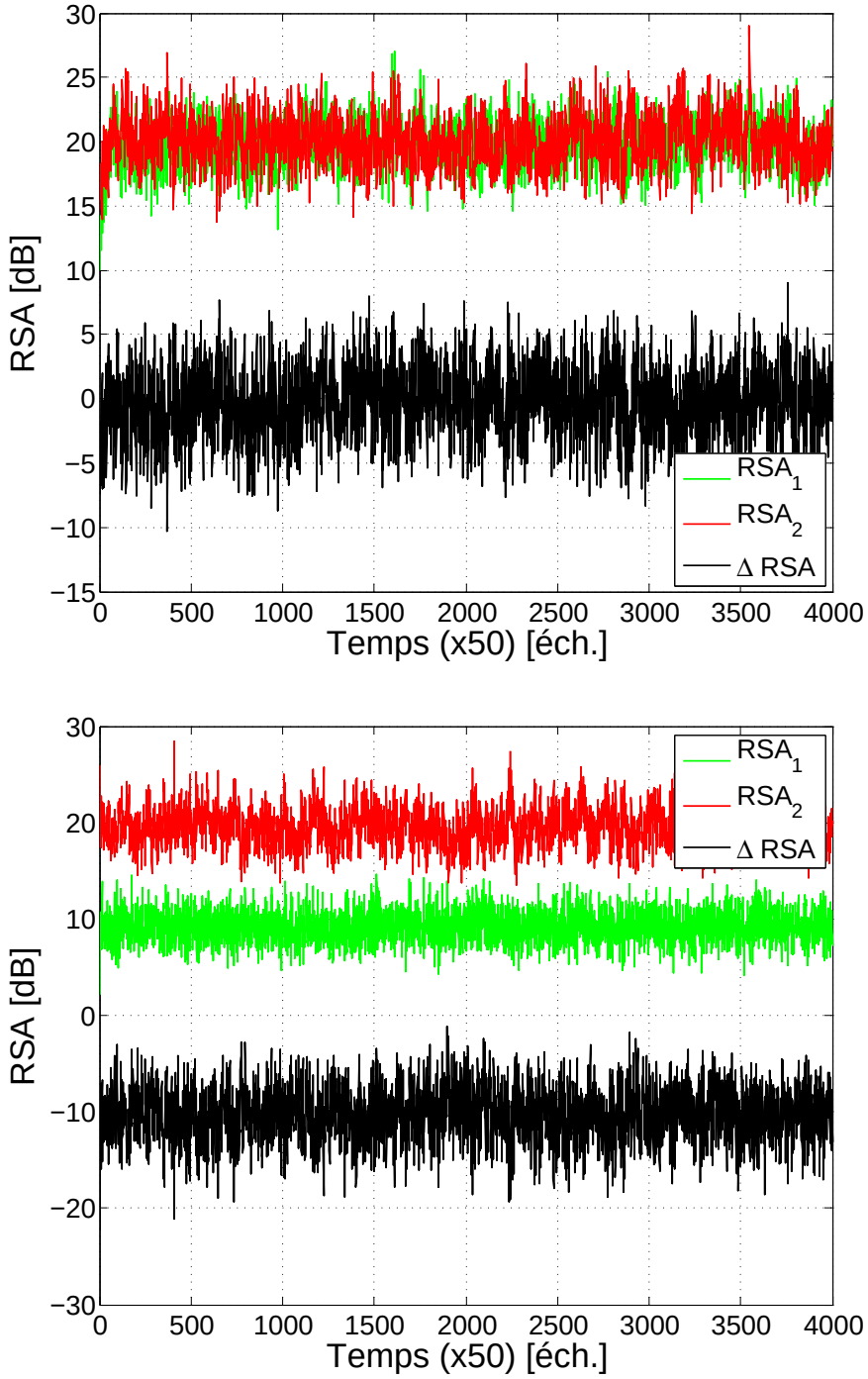


Figure 5.20 – Comportement des RSA dans le cas où une seule source se déplace. La différence d'énergie entre les chemins croisés est de $\Delta = 5\ dB$. En haut, le RSA obtenu avec des sources de même puissance. En bas, celui obtenu pour une puissance de s_1 de 1 dB et de s_2 de 10 dB.

lier. Nous excluons dans ce constat la restriction de l'analyse au mélange simplifié

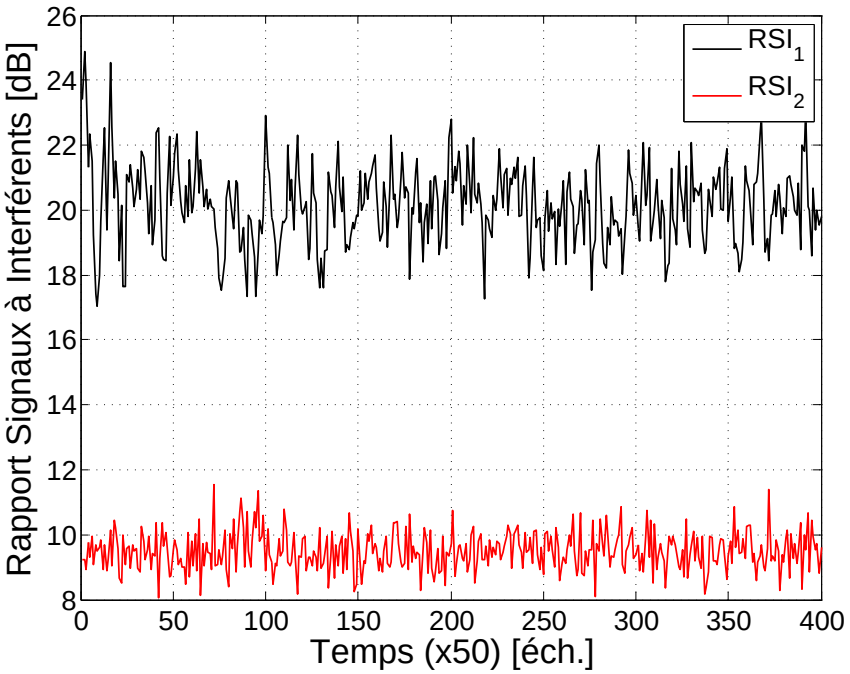


Figure 5.21 – Comportement des RSI expérimentaux dans le cas de sources disposées symétriquement (Cas 1). La puissance de la source s_1 est de 10 dB, celle de s_2 de 1 dB. Les RSI expérimentaux sont calculés à partir du calcul des énergies moyennées sur 30 tirages aléatoires du signal désiré et du signal interférent (signal résiduel présent sur la voie i et estimé sur la voie j) sur chaque voie.

ayant pour but de simplifier l'étude. Ainsi, en considérant différentes configurations des sources et différents mélanges, nous avons montré que le comportement global de la structure est régi par une relation d'équilibre issue de l'égalité des énergies des signaux d'erreurs ou artefacts sur les estimées et incluant l'énergie du signal de diaphonie et celle de la distorsion. Néanmoins, notre étude ne fournit pas de solutions explicites pour chacun des filtres, de même qu'elle ne prouve pas l'existence d'un point d'équilibre unique même si les expériences tendent à montrer que l'état stable identifié soit le seul. Enfin, bien que nous ayons mis explicitement en avant le comportement de la structure, cette connaissance n'apporte pas de moyens simples pour mieux la contrôler au sein de l'antenne que nous souhaitons concevoir. Par contre, nous apportons un éclairage nouveau sur l'importance du signal de diaphonie dans la convergence des filtres et également sur le Cas 3. En effet, en notant d'une part la symétrie des différentes branches du réseau décrit à la Figure 5.4 et d'autre part l'équivalence énergétique de ces différentes branches pour la configuration associée au Cas 3, nous mettons en évidence que celui-ci est bien plus qu'une configuration où les sources sont indissociables (cf. Chapitre 2). Il correspond en fait à un état

d'équilibre stable, et ce, à condition que les sources soient de même puissance et que les coefficients des filtres soient initialisés aux mêmes valeurs.

Chapitre 6

Extension au cas multicapteur : vers la définition d'un GSC récursif

6.1 Implémentations récursives du GSC 178
6.1.1 Rappel : GSC robuste (RGSC) 178
6.1.2 GSC robustes récursifs 179
6.1.3 Inconvénients de ces méthodes 181
6.2 GSC récursif à faible coût 182
6.3 Impact du chemin direct sur les estimées produites par le LCC-TRANC 183
6.4 Synchronisation dans les antennes récursives 185
6.5 Evaluation en conditions réelles 192
6.5.1 Description de l'environnement acoustique et de la prise de son 192
6.5.2 Description du corpus 193
6.5.3 Résultats expérimentaux 194
6.6 Synthèse du chapitre 198

Pour parvenir aux architectures récursives du GSC, ce dernier chapitre utilise les résultats et les analyses menées dans les trois chapitres précédents. Plus précisément, il s'agit ici d'étendre ou plutôt d'intégrer la cellule bicapteur du Chapitre 4 avec le nouvel algorithme proposé (LCCTRANC) dans le GSC, pour tenter de réduire plus efficacement la fuite du signal. Comme nous l'avons vu dans l'état de l'art, un grand nombre de méthodes existent et traitent le problème sous différentes approches. Bien que jusqu'ici nous n'ayons pas fait référence à des approches envisageant une mise en œuvre récursive du GSC, celles-ci existent. Cependant, elles sont très peu référencées dans la littérature et aucune étude aussi approfondie que celle de Reuven pour le GSC classique n'existe à leur sujet. Seuls deux articles [33, 194] et un brevet [34] en font état. Par conséquent, nous donnerons dans un premier temps une description des deux versions récursives du GSC proposée par Choi et al., puis, nous présenterons une nouvelle architecture à faible coût. Enfin, après une analyse des problèmes de synchronisation et une étude sur la longueur optimale des filtres adaptatifs pour un mélange non simplifié, nous donnerons les performances des implémentations récursives du GSC en environnement réel. Les travaux décrits dans ce chapitre ont été en partie publiés à IWAENC 2010, voir [115].

6.1 Implémentations récursives du GSC

6.1.1 Rappel : GSC robuste (RGSC)

Nous rappelons tout d'abord à la Figure 6.1 la structure du GSC robuste. Dans cette architecture, aucun élement n'est spécifiquement dédié à l'élimination du signal de fuite. En effet, aucun filtre adaptatif n'est clairement dédié à l'apprentissage de ce signal dans le but de le contrôler et le soustraire à l'estimation des bruits et des interférents. Or, bien que l'antenne fixe (FBF) permette de sélectionner spatialement la source d'intérêt, la directivité de ce module n'est pas suffisante, en particulier en basse fréquence. Ceci implique qu'une partie des signaux perturbateurs subsiste, y compris les réflexions du signal utile. Pour s'en convaincre, il suffit de reconsidérer la structure donnée à la Figure 6.2, déjà étudiée au Chapitre 2 et à la base de l'architecture du GSC robuste. Chacun des blocs adaptatifs (ABM et AIC) composant ce GSC est construit à partir de l'ANC standard piloté en fonction de la tâche que l'on souhaite lui voir accomplir : apprentissage du bruit ou du signal source. Or, cette méthode est efficace en conditions réelles uniquement si l'on garantit l'adaptation pendant les périodes d'absence/présence du signal source. Il faudra donc adjoindre à ce GSC un module complémentaire (cf. Chapitre 1) pour garantir l'adaptation aux instants opportuns.

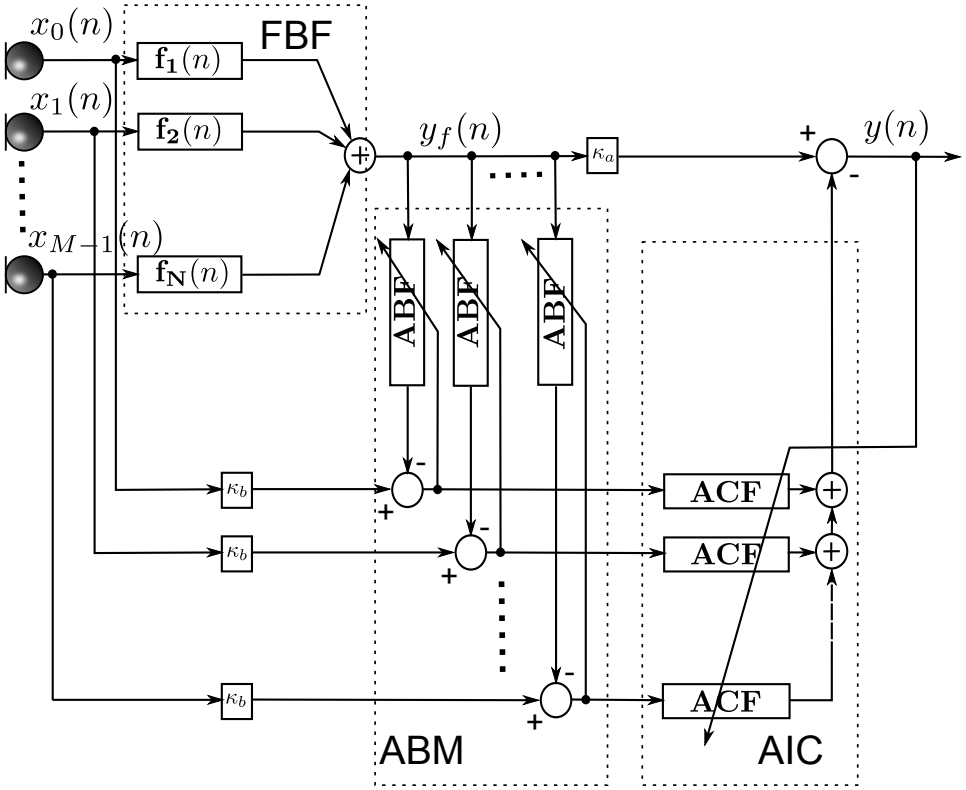


Figure 6.1 – Structure du GSC composé d'une antenne filtrage-sommation, une matrice de blocage adapative (ABM) et un annuleur d'interférents (AIC) multivoie.

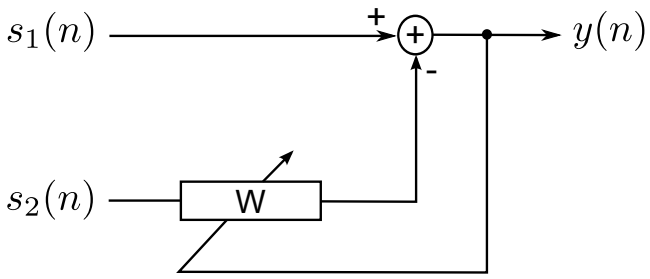


Figure 6.2 – Structure de l'annuleur de bruit adaptatif (ANC) classique.

6.1.2 GSC robustes récursifs

Deux méthodes sont proposées par Choi et al. [33, 34, 194], celles-ci sont décrites aux Figures 6.3 et 6.5. Contrairement à l'implémentation précédente, la mise en œuvre récursive des filtres nommés *Adaptive Blocking Filter* (ABF) et *Adaptive Canceling Filter* (ACF) rend la structure proposée par Choi et al. plus efficace vis-à-vis de la fuite du signal utile et donc, par là même, plus robuste aux erreurs de pointage. En effet, à chaque instant l'estimée du bruit est améliorée par la soustraction d'une estimée du signal utile produite en sortie afin de supprimer les composantes résiduelles de signal utile contenues dans les estimées du bruit. Cette chaîne

de retour sur les estimées permet ainsi de se prémunir de l'éventuelle suppression du signal. Dans ces antennes, l'utilisation des filtres duaux dont la structure est

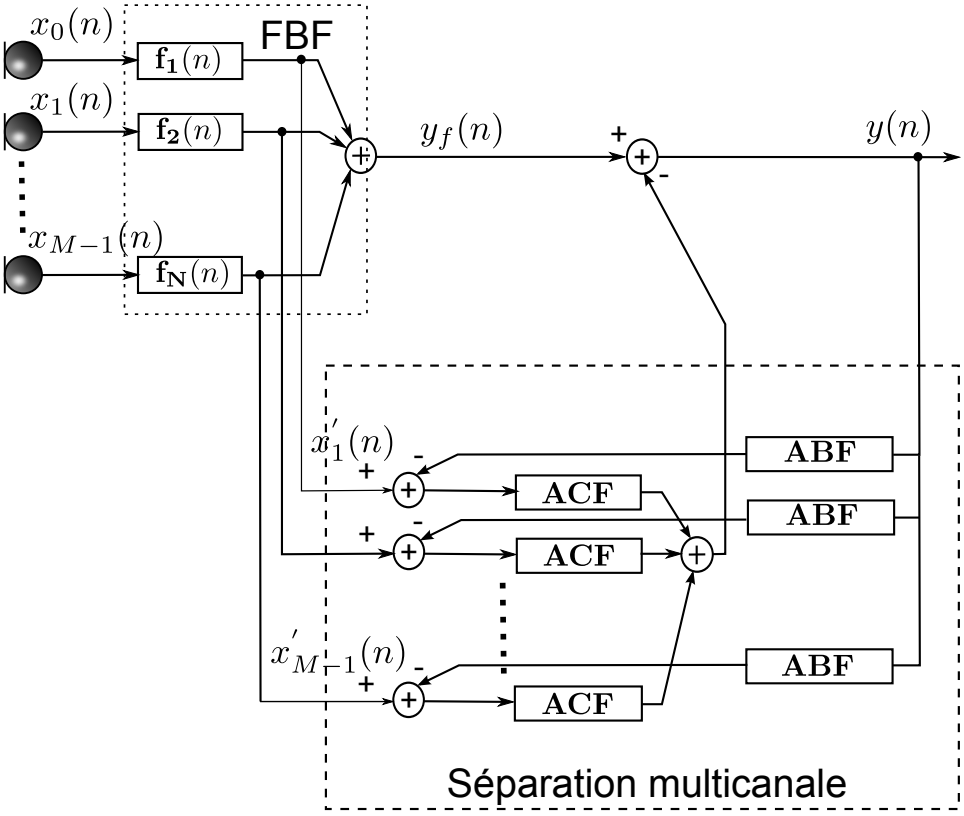


Figure 6.3 – Structure du GSC avec une configuration récursive des filtres ABF et ACF (Solution 1).

rappelée à la Figure 6.4, permet de former simultanément une voie de référence de bruit seul et une voie de parole seule par un double filtrage adaptatif croisé. Chaque filtre prend comme voie de référence la sortie de son dual. Le signal $x'_m(n)$ désigne le signal reçu par le $m^{ème}$ microphone après pointage et filtrage de voie du bloc de filtrage de l'antenne fixe, FBF. Par comparaison avec la structure de l'ANC classique

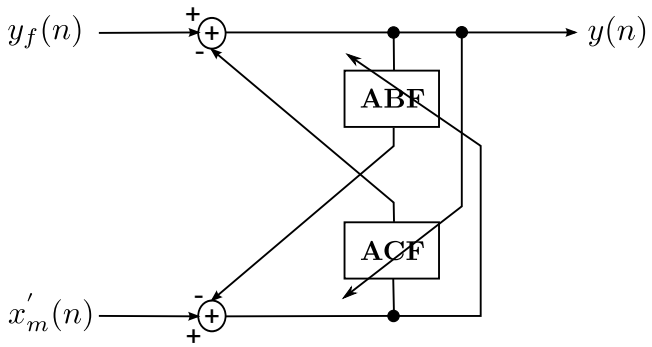


Figure 6.4 – Filtres duaux : estimation d'une voie de référence pour chaque type de signal.

donné à la Figure 6.2, la disposition des filtres selon la Figure 6.4 permet grâce au filtre ABF de soustraire le signal de fuite à l'estimation du bruit faite par le filtre ACF. Le signal de fuite est donc traité structurellement et non pas uniquement de manière logicielle à l'aide d'une DAV. Une configuration récursive des filtres profite

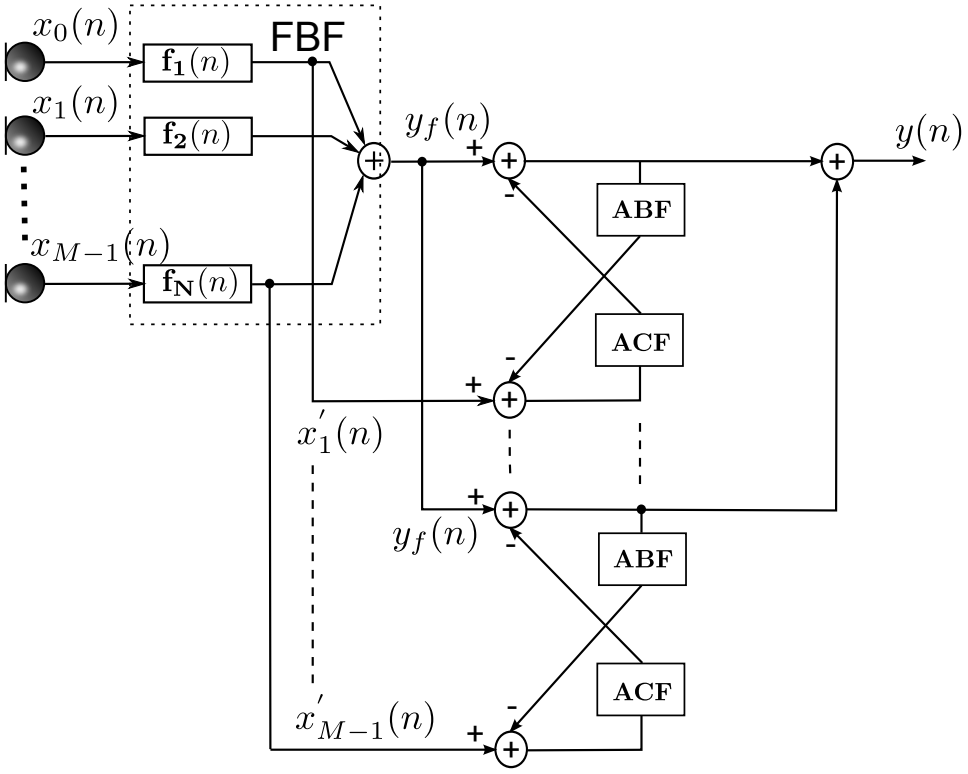


Figure 6.5 – Structure du GSC avec une configuration récursive des filtres ABF et ACF (Solution 2).

également au coût d'implémentation. En effet, il est possible de réduire la longueur des filtres ABF et ACF utilisés pour des performances de séparation identiques à celle du *Robust Generalized Sidelobe Canceller* (RGSC) utilisant des filtres nécessairement plus longs. Pour les deux structures récursives, l'ajout d'une DAV pour gérer les instants d'adaptation des filtres ABF et ACF est aussi un moyen d'en améliorer les performances.

6.1.3 Inconvénients de ces méthodes

La première de ces méthodes (RGSC) souffre de la sous-modélisation due à des filtres nécessairement plus courts que ceux qu'ils sont en charge d'estimer. En effet, pour la plupart des environnements réels, l'effet de salle impose une longueur de filtre supérieure à 1000 coefficients. Bien qu'il soit possible de réduire cette longueur tout en préservant la même précision de modélisation grâce à des filtres à Réponse Impulsionnelle Infinie (RII), les difficultés de mise en œuvre en particulier vis-à-

vis de la stabilité de ceux-ci font que les filtres RIF leurs sont préférés. De plus, l'apprentissage du signal de fuite n'est pas optimal, puisque celui-ci est réalisé à l'aide d'un ANC classique. S'il est possible de faire fonctionner efficacement cette méthode grâce à une DAV, utiliser des structures plus spécifiques à la gestion du signal de fuite et développées notamment dans les systèmes bivoies est une bonne alternative. Tout d'abord, dans des conditions de bruit fort (RSB < 5 dB) les DAV ne sont plus efficaces. Par ailleurs, le phénomène de fuite est mieux pris en compte au travers du second filtre. L'approche récursive déclinée sous deux solutions semble donc la plus attrayante. Néanmoins, celle-ci reste coûteuse en nombre de filtres avec une complexité égale au GSC classique, c'est-à-dire que le nombre de filtres qui composent la matrice de blocage et l'annuleur d'interférents est de $2 \times M_{mic}$, avec M_{mic} le nombre de micros utilisés.

6.2 GSC récursif à faible coût

Le GSC proposé repose sur la substitution de l'annuleur d'interférent multivoie (AIC) par la structure récursive bicapteur étudiée aux Chapitres 4 et 5 et dédiée à la gestion de la diaphonie. Nous donnons à voir ci-après, un schéma détaillé de

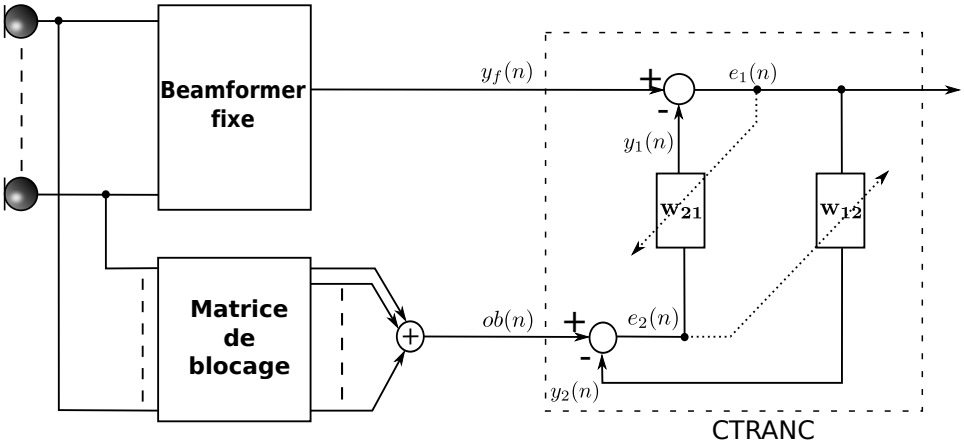


Figure 6.6 – Schéma bloc descriptif du GSC proposé.

l'antenne GSC proposée, avec des signaux de référence pris au niveau des microphones. L'utilisation des signaux après filtrage de voie est une autre mise en œuvre possible. L'avantage de la méthode proposée est son coût en terme du nombre de filtres requis, soit $M_{mic} + 2$ par comparaison avec les $2 \times M_{mic}$ filtres des méthodes précédentes. En ce sens, en plus de sa capacité à préserver le signal utile et réduire les perturbations de manière importante, le gain en complexité démarque clairement la solution proposée des méthodes de l'état de l'art.

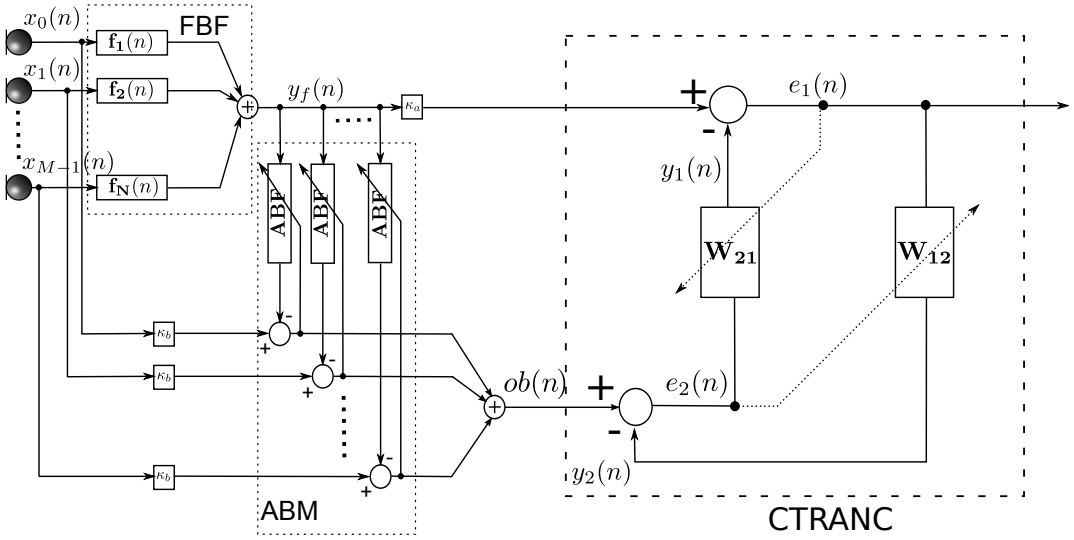


Figure 6.7 – Structure du GSC robuste proposé. Il est obtenu par la connexion des blocs FBF et ABM du RGSC avec la structure CTRANC.

Remarque 1 : *Le GSC proposé est de complexité équivalente vis-à-vis du nombre de filtres par rapport aux versions existantes, lorsque le système est composé de seulement deux microphones, soit* $M_{mic} = 2$.

6.3 Impact du chemin direct sur les estimées produites par le LCCTRANC

Avant de considérer les problèmes de synchronisation des antennes GSC considérées, nous revenons sur la structure à filtres duaux et son comportement dans le cas d'un mélange non simplifié. Dans les Chapitres 2 et 3, nous avons mis en avant la différence entre la structure directe et la structure récursive avec pour la première la nécessité de procéder à un postfiltrage tandis que pour la seconde, la capacité à estimer directement les chemins croisés. Ainsi, dans le cas d'un mélange simplifié et sans problème de sous-modélisation, les réponses des filtres adaptatifs obtenues après la convergence se superposent aux réponses impulsionnelles des chemins croisés. Ceci n'est plus vrai dès lors que les chemins directs ne sont pas réduits à des Dirac. En effet, chacun des filtres estime le chemin croisé qui lui est associé en cherchant le maximum de corrélation entre le signal à son entrée et le signal pris pour référence. Par conséquent, en se référant au schéma de la Figure 6.8 qui illustre de façon simple ce mécanisme, nous constatons que le maximum d'énergie des filtres estimés $\mathbf{w_{12}}$ et $\mathbf{w_{21}}$ sera toujours situé sur les premiers coefficients. Le retard de propagation associé à ces filtres correspond à la différence des temps de propagation entre le chemin direct et le chemin croisé, puisque chacun des filtres prend à son

entrée un signal retardé du temps de propagation associé aux chemins directs $\tau_{a_{11}}$ (respectivement $\tau_{a_{22}}$). Ainsi, le maximum de corrélation se situe dans une fenêtre centrée sur la différence entre les temps de propagation respectifs des chemins, par exemple $\tau_{a_{12}} - \tau_{a_{11}}$ qui permet de compenser le retard entre le signal $x_1(n)$ et $x_2(n)$. A travers cette analyse, nous justifions la troncature sur la longueur de la récursion de l'algorithme LCCTRANC décrit au Chapitre 4 qui s'appuie donc sur le fait que les premiers coefficients sont les plus significatifs. Cette situation se rencontre pleinement dans le cas où le mélange n'est pas simplifié, ce qui correspond à la plupart des cas pratiques.

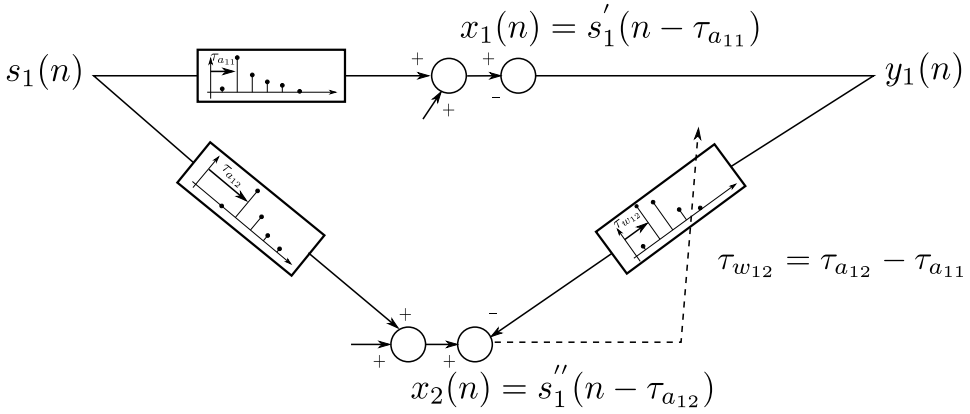


Figure 6.8 – Influence des chemins directs sur les filtres adaptatifs calculés. Illustration pour le filtre $\mathbf{w_{12}}$.

L'analyse précédente menée sur l'influence des chemins directs complète également l'analyse sur la physique du problème traitant des performances de la structure à filtres duaux par rapport à la position des sources. En effet, grâce à l'analyse précédente, nous constatons qu'il est possible de restreindre la longueur des filtres adaptatifs tout en s'assurant qu'ils comportent le maximum de corrélation. De ce fait, nous pouvons en déduire qu'il existe une contrainte sur la position des sources et la distance inter-microphone pour une taille des filtres donnée. En notant $L_{w_{21}}$ et $L_{w_{12}}$ la taille des filtres adaptatifs, la distance d entre les microphones devra garantir, pour des longueurs de filtre fixées, les contraintes suivantes :

$$L_{w_{12}} + \tau_{a_{11}}^{max} > \tau_{a_{12}}^{max} \quad (6.1)$$

$$L_{w_{21}} + \tau_{a_{22}}^{max} > \tau_{a_{21}}^{max} \quad (6.2)$$

Avec $\tau_{max} = \frac{d}{c} = \tau_{a_{12}}^{max} - \tau_{a_{11}}^{max} = \tau_{a_{21}}^{max} - \tau_{a_{22}}^{max}$, la latence maximale des filtres adaptatifs. En respectant ces contraintes, nous définissons le gabarit des filtres, Figure 6.9, en situant la position limite des maxima de leurs réponses impulsionnelles lorsqu'ils sont contraints à être de type RIF. En théorie, les retards de propagation de ces filtres seront nuls lorsque les sources se superposent sur le plan médian, mais nous

avons montré au Chapitre 2 que la structure est incapable de fonctionner dans ces conditions. Par conséquent, ceci ne peut être observé que de manière aléatoire car les filtres ont un fonctionnement non prédictibles dans cette configuration. A l'opposé, les retards de propagation maximaux seront observés lorsque les sources seront situées en *endfire*.

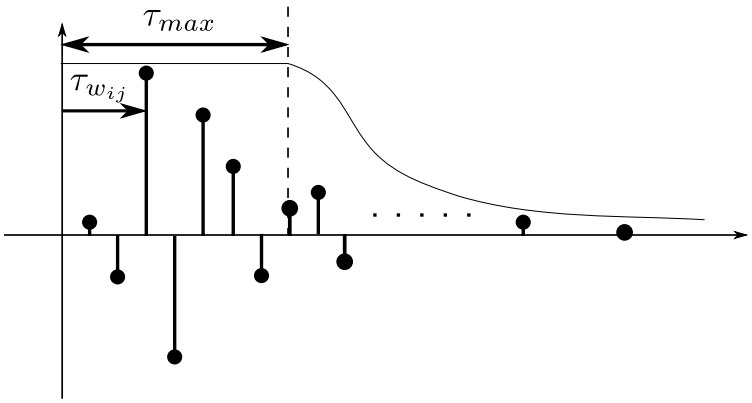


Figure 6.9 – Allure des réponses impulsionnelles des filtres adaptatifs $\mathbf{w_{ij}}$.

Nous appuyons le raisonnement à partir des résultats de simulation donnés à la Figure 6.10. Les filtres adaptatifs ont été calculés pour un mélange de deux bruits blancs gaussiens de même puissance. La configuration des sources correspond à la deuxième scène sonore synthétique décrite en Annexe D. Celles-ci sont donc placées de manière symétrique, l'une à 80 cm, l'autre à 2 m des microphones.

Les observations faites dans ce paragraphe vont nous permettre d'aborder la problématique de la synchronisation décrite dans la section ci-après.

6.4 Synchronisation dans les antennes récursives

Pour les différentes structures du GSC données dans ce mémoire, nous sommes en mesure de distinguer deux étages distincts qu'il nous faut synchroniser. Le premier étage est constitué des blocs FBF et ABM, le second possède à son entrée les deux signaux disponibles en sortie des blocs FBF et AIC. Cependant, il existe une différence fondamentale entre la problématique de la synchronisation en fonction du type de structure : directe ou récursive. Pour la première, le retard placé sur une voie n'a pas de répercussion sur la synchronisation de l'autre voie, ce qui est faux dans le cas de la structure récursive.
Considérons le premier étage décrit à la Figure 6.11, commun au GSC robuste et

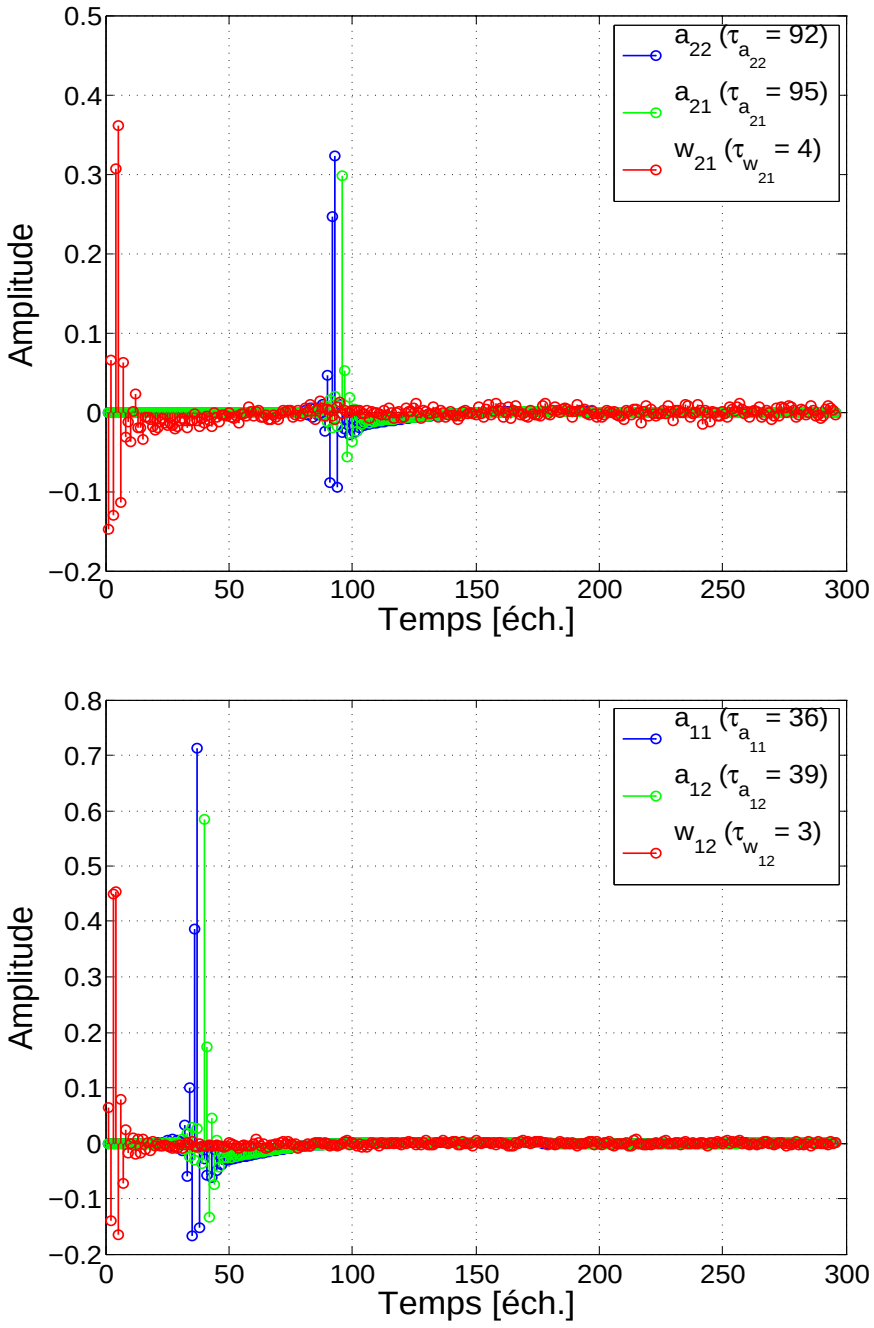


Figure 6.10 – Allure des réponses impulsionnelles des filtres estimés dans le cas d'un mélange de deux bruits blancs gaussiens placés en *broadside* et de manière symétrique par rapport au plan des microphones.

celui proposé. Deux options sont possibles pour la récupération des signaux destinés à la création de la référence de bruit.

1. Lorsque l'on travaille à partir des signaux captés par les microphones (confi-

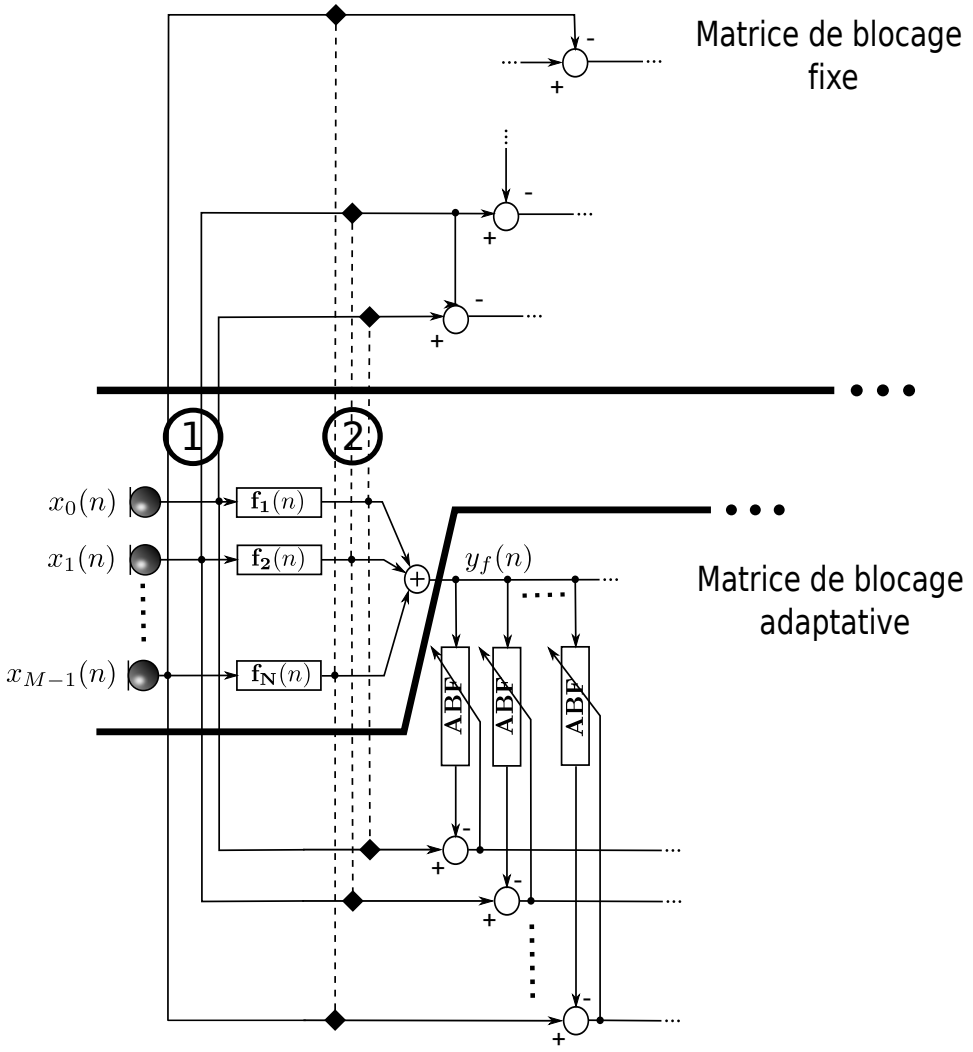


Figure 6.11 – Configurations possibles pour la création des signaux de référence de bruit. La configuration 1 utilise les signaux captés par les microphones. La configuration 2 correspond à la récupération des signaux après filtrage par l'antenne sur chacune des voies.

guration 1), nous profitons de la diversité spatiale du bruit environnant mais au risque d'avoir un niveau de bruit fort et moins corrélé avec le bruit perturbateur de la source. Avec cette option, le retard de synchronisation κ_b de la Figure 6.7 devra compenser le temps de propagation au travers le bloc FBF plus le retard de traitement lorsque l'on travaille dans le domaine fréquentiel.

2. La configuration 2 profite quant à elle du filtrage par l'antenne (pointage + filtrage de voies). Nous bénéficions donc d'une image locale du bruit autour de la source cible, cette image locale sera davantage corrélée avec le bruit perturbateur. Cependant, nous montrerons au paragraphe 6.5 que l'antenne peut avoir une influence sur les performances en raison du filtrage de voies réalisé

dans l'antenne à directivité contrôlée. En ce qui concerne la synchronisation, cette configuration permet d'avoir un retard plus court et servant uniquement à compenser le retard lié au traitement fréquentiel.

Cependant, à l'aide de l'analyse présentée au paragraphe 6.3, nous comprenons que les retards doivent être également ajustés pour garantir que le maximum de corrélation soit dans l'horizon d'observation des filtres adaptatifs utilisés dans les blocs ABM et ACF, horizon limité par la longueur choisie pour ceux-ci. Or, sachant que nous choisissons la configuration 2, compenser uniquement le retard associé au traitement fréquentiel implique que le maximum de corrélation soit observé sur le premier coefficient des filtres adaptatifs des blocs ABM (signaux synchrones) et, si les filtres sont de tailles suffisantes, à la période *pitch* du signal de parole. Mais cela pose un certains nombres de problèmes. En effet, la position des sources est entachée d'erreur, par conséquent le pointage ne peut être parfait. Nous avons vu également dans l'état de l'art (Chapitre 1) que les imperfections liées aux antennes acoustiques font que l'on ne sera jamais en mesure de pointer parfaitement le système sur les sources. Or, l'amélioration de la robustesse des antennes GSC que nous considérons est fondée notamment sur l'ajout de contraintes qui, en contraignant la réponse des filtres adaptatifs à un gabarit donné, créent des lobes d'observation plus larges afin de pallier les mouvements et les défauts de pointage de la source cible. Or, ces contraintes consistent à forcer le maximum de corrélation vu par les filtres à se situer sur leurs premiers coefficients et autour d'une position $n_0 \neq 0$, ce qui défini le centre et la largeur de la fenêtre d'observation spatiale.

Appliquons dans un premier temps ces contraintes sur le RGSC décrit à la Figure 6.1. Pour simplifier l'analyse, nous supposons que les filtres adaptatifs des blocs ABM et AIC ont la même la longueur, L. La position imposée n_0 sera fixée par le rapport L/X avec $X \in \mathbb{N}^* - \{1\}$. Ainsi, contraindre le maximum de la réponse impulsionnelle des filtres adaptatifs du bloc ABM à se situer en L/X à l'aide du délai κ_b, revient à trouver κ_b tel que l'intercorrélation entre les signaux $x_m^{'}(n-\kappa_b)$ et $y_f(n)$ soit maximale en L/X. Pour être tout à fait général, nous désignons par τ_1 la position du maximum de l'intercorrélation entre $x_m^{'}(n)$ et $y_f(n)$. L'expression de κ_b qui en résulte est donc :

$$\kappa_b = \frac{L}{X} - \tau_1 \tag{6.3}$$

Le même raisonnement peut être tenu pour les filtres adaptatifs du bloc AIC et le délai κ_a pour lequel nous pouvons écrire $\kappa_a = \frac{L}{X} - \tau_2$, où $\tau_2 = \kappa_b$. Dans la configuration 2, les signaux sont parfaitement synchronisés après passage par le bloc FBF, donc $\tau_1 = 0$, d'où $\kappa_a = 2L/X$ et si l'on prend $X = 2$, ces délais correspondent à $\kappa_b = L/2$ et $\kappa_a = L$. Appliquons maintenant le raisonnement au GSC proposé et

constitué du CTRANC. Pour maximiser l'énergie sur les L coefficients du filtre W_{12} et plus particulièrement autour de la position L/X, il faut ajuster κ_b tel que l'on maximise l'intercorrélation entre $ob(n)$ et $e_1(n)$ en cette position. Si τ_3 est la position du maximum avant l'application du retard, il faut donc que κ_b vérifie :

$$\kappa_b = \frac{L}{X} - \tau_3 \tag{6.4}$$

Or, selon l'analyse précédente sur le RGSC et applicable au bloc ABM et au filtre W_{21} du GSC proposé, les retards optimaux sont :

$$\kappa_b = \frac{L}{X} \quad \text{et} \quad \kappa_a = \frac{2L}{X}$$

Par conséquent, on en déduit la valeur $\tau_3 = -L/X$ et les deux relations contradictoires que doit vérifier κ_b, à savoir $\kappa_b = L/X$ et $\kappa_b = 2L/X$. Les contraintes de synchronisation sont donc plus délicates à ajuster lorsque le GSC possède une structure récursive. En fait, il existe des couples (κ_a, κ_b) optimaux et des couples théoriquement impossibles conduisant à des performances sous-optimales. Une étude de cas pour le GSC proposé est donnée ci-après, afin d'extraire les couples possibles et listés au tableau 6.1. Les résultats sont transposables aux versions proposées par Choi et al.

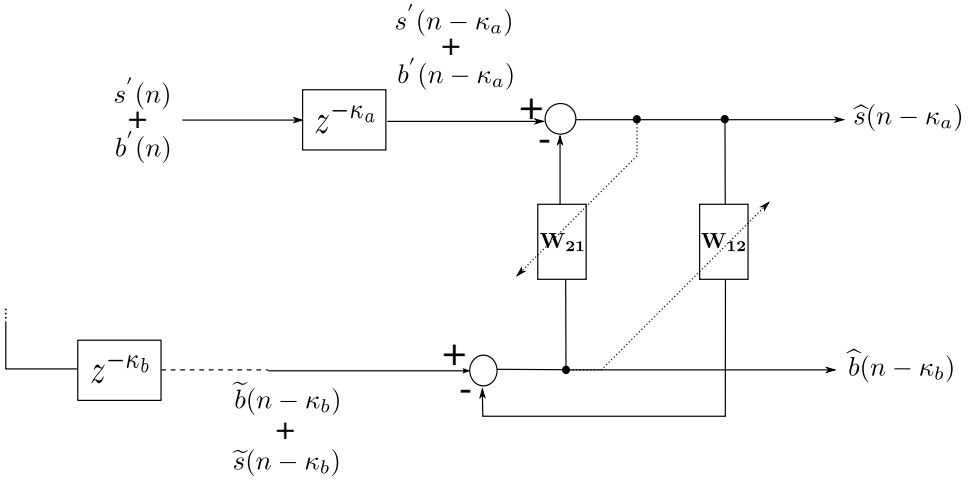


Figure 6.12 – Analyse de l'influence des retards sur la structure récursive.

A l'aide de la Figure 6.12, nous allons distinguer les trois modes possibles de fonctionnement pour le filtre $\mathbf{w_{12}}$ qui doit, selon la théorie de Wiener, maximiser la corrélation sur ses L coefficients entre $[\widehat{s}(n-\kappa_a)\ \widehat{s}(n-\kappa_a-1)\ \ldots\ \widehat{s}(n-\kappa_a-L+1)]$ et $\widetilde{s}(n-\kappa_b)$ et ce en fonction du délai κ_b.

Mode identification

Le filtre $\mathbf{w_{12}}$ fonctionnera en identification si $\kappa_b \in [\kappa_a, \kappa_a + 1, \ldots, \kappa_a + L - 1]$. Dans ce cas, le maximum de la réponse impulsionnelle se déplace dans la mémoire du filtre, autrement dit, sur l'un de ces L coefficients. Plus on éloigne la position de ce maximum des premiers coefficients, plus on augmente l'erreur d'estimation en perdant l'estimation de la queue des réponses impulsionnelles.

Mode prédiction "avant"

En fixant $\kappa_b < \kappa_a$, le filtre $\mathbf{w_{12}}$ fonctionne en prédiction avant de pas $P = \kappa_a - \kappa_b$ et $P > 0$.

Mode prédiction "arrière"

En fixant $\kappa_b > \kappa_a + L - 1$, le filtre $\mathbf{w_{12}}$ fonctionne en prédiction arrière de pas $P = \kappa_b - (\kappa_a + L - 1)$. Par exemple, prenons $\kappa_b = \kappa_a + L$. Dans ce cas, le filtre $\mathbf{w_{12}}$ devra prédire le signal $\widetilde{s}(n - \kappa_a - L)$ à partir de $[\widehat{s}(n - \kappa_a)\ \widehat{s}(n - \kappa_a - 1)\ \ldots\ \widehat{s}(n - \kappa_a - L + 1)]$, soit donc selon une prédiction arrière. De façon duale, le filtre $\mathbf{w_{21}}$ fonctionne dans ces conditions en prédiction avant, puisqu'il devra maximiser la corrélation entre $b'(n - \kappa_a)$ et $\left[\widehat{b}(n - \kappa_a - L)\ \widehat{b}(n - \kappa_a - L - 1)\ \ldots\ \widehat{b}(n - \kappa_a - 2L)\right]$.
Le même raisonnement vaut pour le filtre $\mathbf{w_{21}}$, les relations qui lui sont associées sont obtenues en inversant les rôles de κ_a et κ_b. Analysons maintenant les fonctionnements conjoints possibles pour ces deux filtres. Les différentes configurations possibles pour ce fonctionnement conjoint sont résumées au sein du tableau 6.1. Ces différentes configurations se distinguent suivant la configuration choisie pour chacun des filtres $\mathbf{w_{12}}$ et $\mathbf{w_{21}}$: mode d'identification, de prédiction avant ou de prédiction arrière.

Dans nos simulations, les GSC récursifs seront implémentés avec des filtres de longueur $L = 256$ coefficients contre $L = 512$ coefficients pour le RGSC. Le couple optimal (κ_b, κ_a) a été déterminé pour les différentes structures en évaluant pour chaque couple l'énergie de l'erreur en sortie des blocs ABM et AIC. Les couples optimaux étant ceux qui minimisent l'énergie des erreurs simultanément sur chacun des blocs. Ces valeurs sont données au tableau 6.2.

Mode de fonctionnement	Contraintes sur les retards	Observations
$\mathbf{w_{12}}$ et $\mathbf{w_{21}}$: identifications	$\kappa_a = \kappa_b$	Seule solution envisageable
$\mathbf{w_{12}}$: identification $\mathbf{w_{21}}$: prédiction avant	$\kappa_b \in [\kappa_a, \kappa_a + L - 1]$ $\kappa_a < \kappa_b$	Vrai dans $L-1$ cas. Le couple $\kappa_a = \kappa_b$ étant proscrit
$\mathbf{w_{12}}$: identification $\mathbf{w_{21}}$: prédiction arrière	$\kappa_b \in [\kappa_a, \kappa_a + L - 1]$ $\kappa_a > \kappa_b + L - 1$	Impossible
$\mathbf{w_{12}}$: prédiction avant $\mathbf{w_{21}}$: prédiction avant	$\kappa_a - \kappa_b > 0$ $\kappa_b - \kappa_a > 0$	Impossible
$\mathbf{w_{12}}$: prédiction avant $\mathbf{w_{21}}$: prédiction arrière	$\kappa_a > \kappa_b$ $\kappa_a > \kappa_b + L - 1$	OK
$\mathbf{w_{12}}$: prédiction avant $\mathbf{w_{21}}$: identification	$\kappa_a > \kappa_b$ $\kappa_a \in [\kappa_b, \kappa_b + L - 1]$	Vrai dans $L-1$ cas. Le couple $\kappa_a = \kappa_b$ étant proscrit
$\mathbf{w_{12}}$: prédiction arrière $\mathbf{w_{21}}$: identification	$\kappa_b > \kappa_a + L - 1$ $\kappa_a \in [\kappa_b, \kappa_b + L - 1]$	Impossible
$\mathbf{w_{12}}$: prédiction arrière $\mathbf{w_{21}}$: prédiction avant	$\kappa_b > \kappa_a + L - 1$ $\kappa_b - \kappa_a > 0$	OK
$\mathbf{w_{12}}$: prédiction arrière $\mathbf{w_{21}}$: prédiction arrière	$\kappa_b > \kappa_a + L - 1$ $\kappa_a > \kappa_b + L - 1$	Impossible

Table 6.1 – Lien entre les couples (κ_a, κ_b) et les modes de fonctionnement possibles pour la structure récursive.

Structure	κ_b	κ_a
RGSC	$L/8$	$L/4$
Solution 1	$L/8$	$L/8$
Solution 2	$L/8$	$L/8$
GSC proposé	$L/8$	$L/4$

Table 6.2 – Valeurs des retards κ_a et κ_b utilisées dans les simulations.

6.5 Evaluation en conditions réelles

6.5.1 Description de l'environnement acoustique et de la prise de son

La configuration que nous avons considérée, dont le plan sommaire est donné en Figure 6.12, est proche d'un scénario de visioconférence. Elle met en jeu deux sources, l'une sera prise pour source utile, l'autre comme source interférente. La prise de son effectuée dans cet exemple est relativement éloignée (1,80 m pour la source cible et 3 m pour la source interférente). La taille de la pièce correspond à celle d'un bureau. Son volume est de 61 m^3 et son temps de réverbération global est égal à 400 ms, celui-ci variant de 650 ms pour les basses fréquences à 250 ms pour les hautes fréquences. Notons de plus que la salle n'est pas traitée acoustiquement, seul le plafond présente des qualités absorbantes (dalles en polymère).

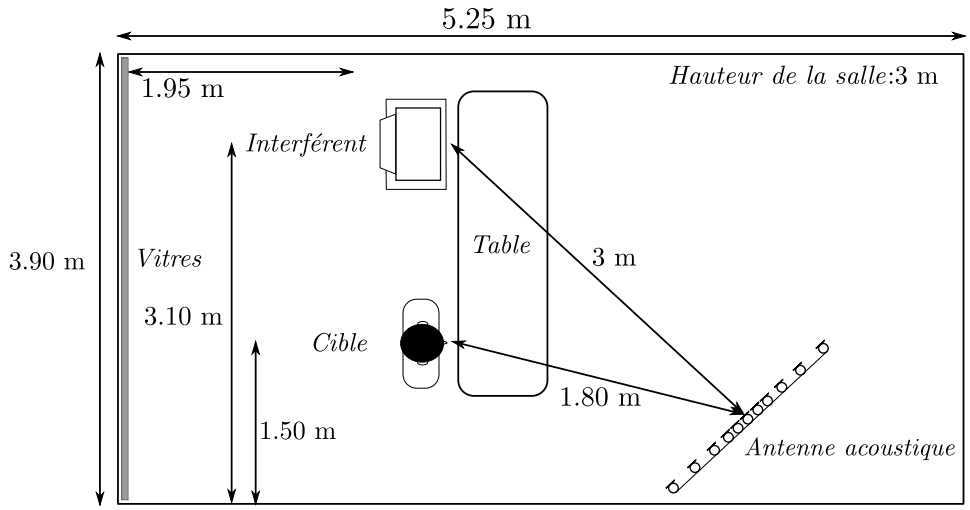


Figure 6.13 – Configuration des sources et de l'antenne acoustique dans les conditions de l'expérience.

Le frontal acoustique décrit en Figure 6.14, est une antenne rectiligne logarithmique composée de 11 microphones cardioïdes groupés en 4 sous-antennes. Pour garantir une ouverture et une hauteur des lobes secondaires identiques en fonction de la fréquence, le même nombre de microphones est utilisé dans chaque sous-bande. Chaque microphone est par ailleurs judicieusement réutilisé dans chaque sous-bande afin de limiter le nombre total de microphones. Le contrôle de la directivité est réalisé grâce aux filtres de limitation de bande g_i.

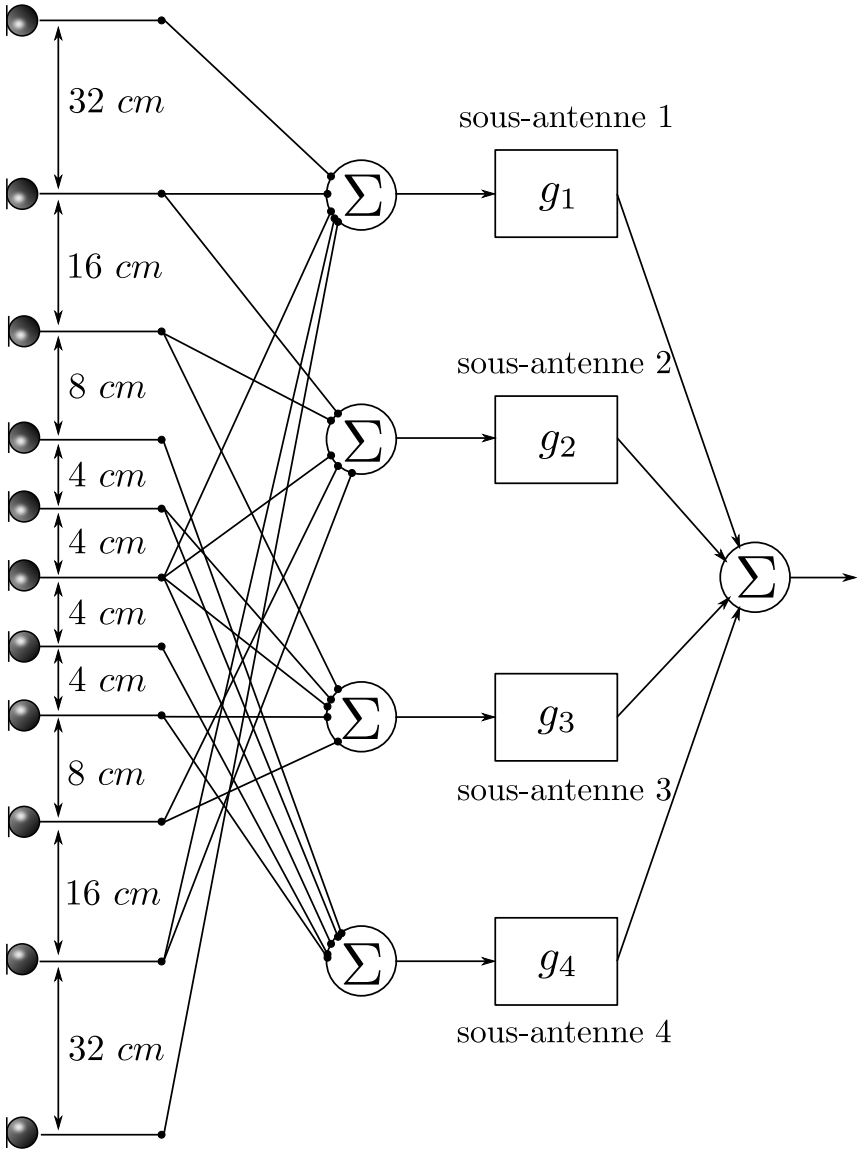


Figure 6.14 – Antenne rectiligne logarithmique composée de 11 microphones groupés en 4 sous-antennes. La distance inter-capteur n'est pas respectée sur ce schéma (doublement des distances d'une antenne à l'autre).

6.5.2 Description du corpus

Avant de donner les résultats des mesures objectives destinées à mesurer les performances de la réduction de bruit, nous décrivons le corpus de signaux utilisés. Ce corpus est constitué comme ceci :

- 4 locuteurs anglophones : 2 femmes et 2 hommes ;
- 4 doubles phrases de 8 s par locuteur ;
- 3 types de bruit : Bureau, Bureau plus bruit de clavier et Restaurant ;
- 5 valeurs de RSB : -5, 0, 5, 10 et 15 dB.

Une condition correspond à un type de bruit et à un RSB fixés, il existe donc 15 conditions différentes. Les résultats objectifs seront donc obtenus par la moyenne des 16 doubles phrases par condition. Ce corpus est extrait de la base de donnée de NTT (*database* 1994). Les signaux sont échantillonnés à 16 kHz et les séquences de parole propre (avant ajout de bruit) sont normalisées à -22 dB par rapport à la valeur de saturation pour en assurer un niveau d'écoute homogène lors de l'évaluation subjective. Cette normalisation est réalisée en utilisant l'outil *speech voltmeter* (SV56) suivant la recommandation ITU-T P.56. Les signaux de bruit qui constituent le corpus sont enregistrés séparément en condition réelle. Les signaux bruités qui constituent le corpus sont obtenus par addition des signaux de parole et de bruit pondérés de façon à obtenir le RSB souhaité. L'outil SV56 est également utilisé pour fixer les RSB au niveau des capteurs. De plus, les signaux bruités issus des doubles phrases sont répétés 4 fois (concaténation) de sorte à assurer la convergence des filtres. Les critères seront calculés sur la $4^{\text{ième}}$ et dernière séquence.
Les 3 bruits sont représentatifs des cas susceptibles d'être rencontrés. Le bruit Bureau est stationnaire et respecte en ce sens une hypothèse courante des techniques de réduction de bruit. Le bruit Bureau plus clavier et surtout Restaurant sont plus critiques dans la mesure où ils ne sont pas stationnaires.

6.5.3 Résultats expérimentaux

Les critères utilisés pour l'évaluation objective des méthodes sont : le facteur de réduction de bruit (**NR**), le gain en RSB (**G**) et la distance cepstrale (**DC**). Une description complète de ces mesures est donnée en Annexe C. Nous comparons les performances de 6 méthodes :

1. le RGSC mis en œuvre dans le domaine fréquentiel par Herbordt et Kellermann,
2. le GSC proposé que l'on désignera par Prop. avec l'algorihtme LCCTRANC,
3. les solutions 1 et 2 proposées par Choi et al. et qui utilisent l'algorithme information maximisation. Nous les désignons par Sol. 1 et Sol. 2,
4. ces mêmes structures où l'on remplace l'algorithme information maximisation par le LCCTRANC. Ces solutions sont estampillées Sol. 1+ et Sol. 2+.

Enfin, pour distinguer les performances de chacun des blocs constitutifs du GSC, nous donnons séparément les performances du frontal acoustique seul que l'on désigne par Ant. Pour chacune de ces méthodes, nous utilisons les pas adaptatifs basés TSNR décrits au Chapitre 3. Ces pas sont ajustés pour permettre un fonctionnement optimal de chaque méthode. Enfin, pour chaque algorithme, l'estimation de la DSP du bruit se fait à l'aide de la DAV idéale calculée pour chacune des sources de parole du corpus.

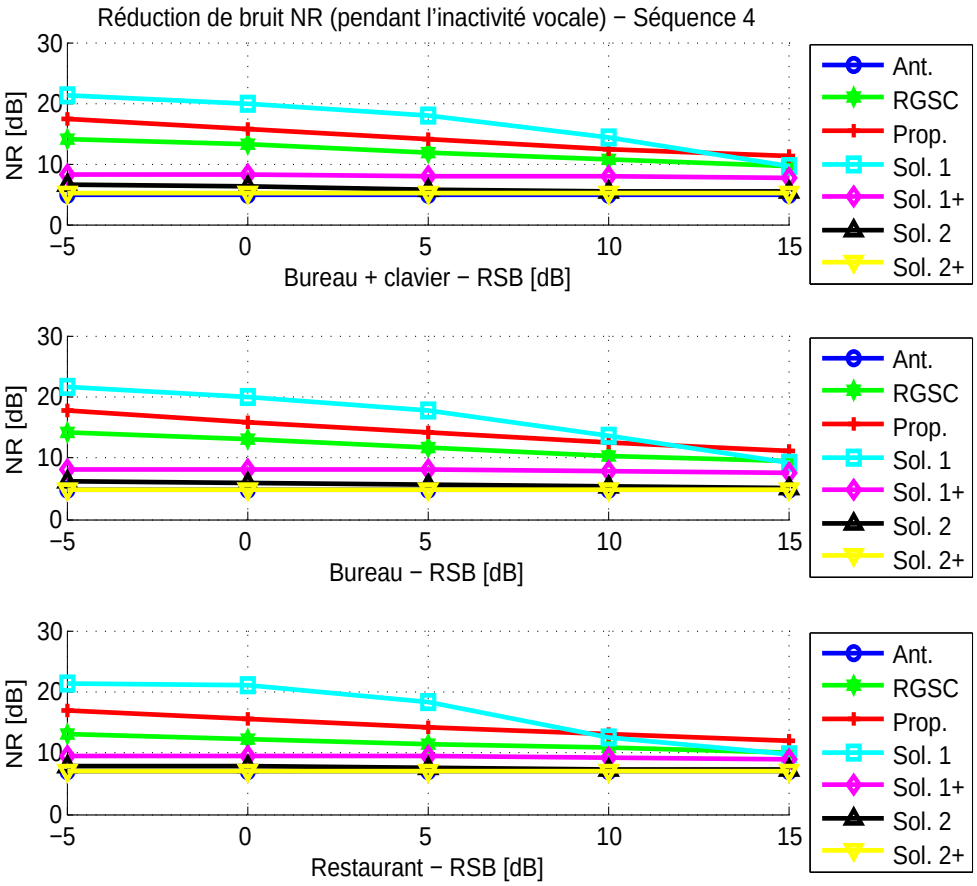


Figure 6.15 – Réduction de bruit.

L'analyse des courbes obtenues à la Figure 6.15 et relatives au facteur de réduction de bruit confirment la démarche développée depuis le Chapitre 2, puisque deux des trois structures récursives considérées surpassent les performances du RGSC. En effet, quel que soit le type de bruit ou la valeur du RSB, la réduction de bruit apportée par le GSC proposé ou la solution 1 est systématiquement au dessus de celle apportée par le RGSC. Notons par ailleurs que le GSC proposé tend à être le plus performant en terme de réduction de bruit dès lors que le RSB est supérieur à 10 dB. Néanmoins, cette forte réduction de bruit s'accompagne d'une dégradation du signal plus importante, ceci est rendu visible par les valeurs de la distance cepstrale tracées à la Figure 6.17. La dégradation est d'autant plus forte que le RSB est faible et inférieur à 5 dB. Cependant, dans de telles conditions, la qualité du signal n'est pas déterminante au regard de l'importance du niveau de bruit. De plus, les dégradations sur le signal de parole ne deviennent perceptible qu'à partir d'une distance

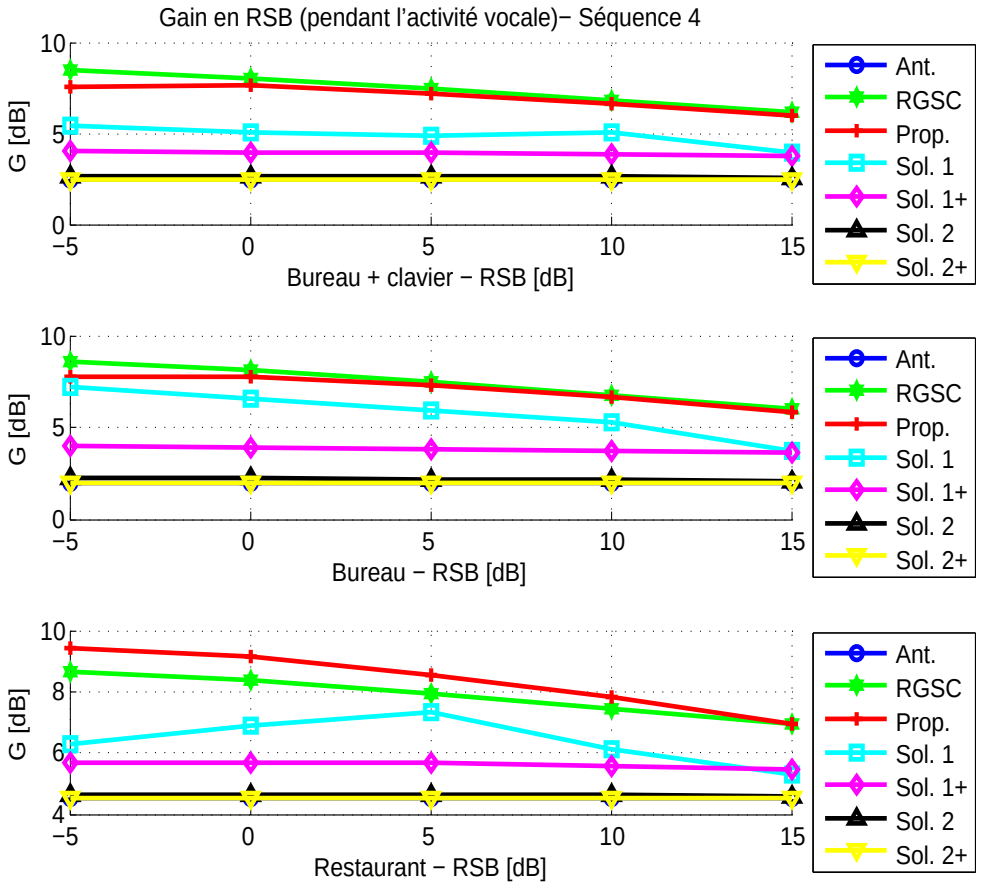


Figure 6.16 – Gain en RSB.

cepstrale supérieure à 0,5. Enfin, en tenant compte du coût réduit du GSC proposé par rapport au RGSC ainsi que des performances en terme de réduction de bruit et de gain en RSB, voir Figure 6.16, le GSC proposé s'affiche donc comme une solution efficace et très intéressante.

Par contre, les résultats obtenus avec Sol. 2, Sol. 2+ et Sol. 1+ sont décevants. En particulier pour Sol. 2 et Sol. 2+ qui sont complètement inefficaces, leurs performances sont réduites à celles du frontal acoustique (Ant.) dont les courbes sont systématiquement situées au plus bas pour les trois critères donnés. Cette constatation est surprenante, surtout si l'on compare les résultats pour Sol. 1+ à ceux que nous avons publiés dans [115]. Dans cet article, cette solution Sol. 1+ était la plus efficace comparée au RGSC, à Sol. 1 et au GSC proposé. En fait, grâce à des simulations sur une base de données plus conséquentes, nous pointons ici un problème qui n'est pas apparu dans l'article, à savoir la dépendance par rapport au contenu fréquentiel de la

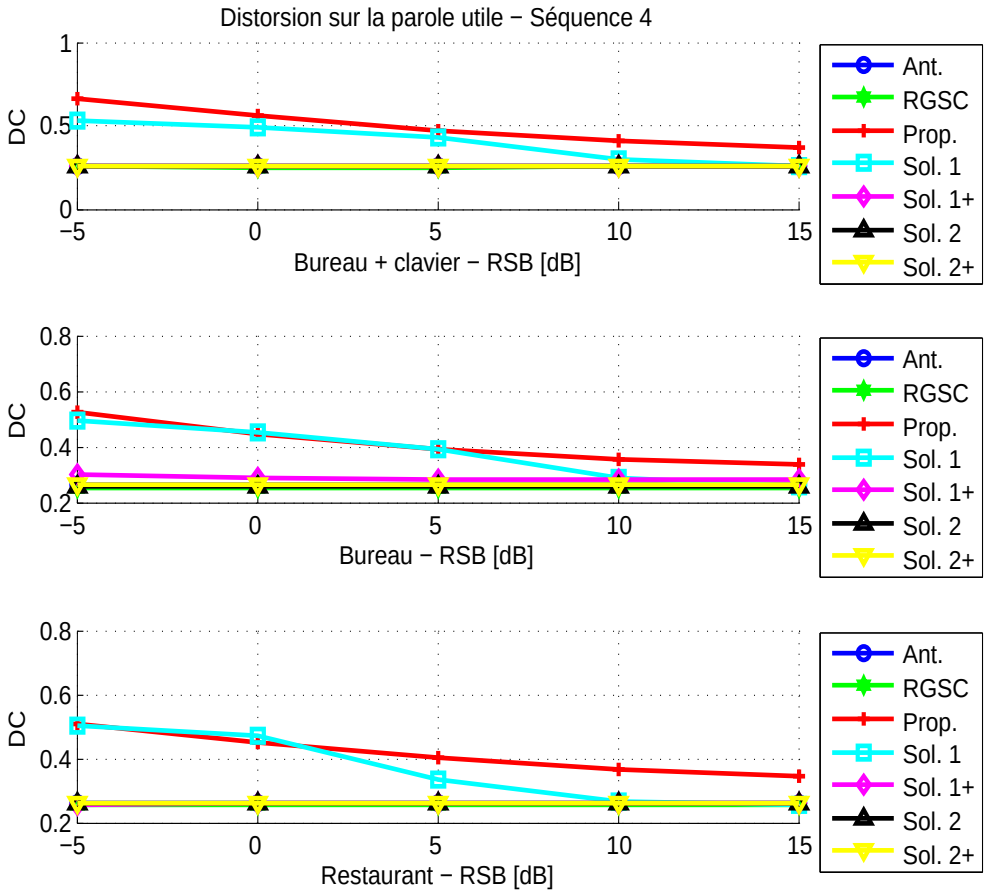


Figure 6.17 – Distance cepstrale.

source. En effet, pour garantir la stabilité des solutions 1 et 2 pour toutes les conditions de bruit testées, nous avons fortement réduit le taux d'adaptation des filtres. Les versions récursives du GSC étant très sensibles aux variations d'énergie rapides au point de sommation, nous avons diminué fortement les pas d'adaptations mais en moyenne seulement, et non pas par bandes de fréquences. Par conséquent, les résultats donnés dans ce chapitre, et en particulier pour Sol. 1+, sont sous-optimaux pour certaines conditions, exceptés pour Sol. 2 et Sol. 2+ qui n'ont jamais fourni de résultats satisfaisants quels que soient nos réglages. Ceci rejoint donc nos observations du Chapitre 3 sur la nécessité d'affiner les pas adaptatifs pour qu'ils fonctionnent précisément et efficacement dans chaque sous-bande. La sensibilité est par ailleurs accentuée par le frontal acoustique que nous utilisons, car les signaux de références de bruit, $x_m^{'}$, sont pris après filtrage de voies, ils subissent donc un filtrage passe-bas, passe-haut ou passe-bande. Ainsi, des pas ajustés en moyenne par une constante ne

permettent pas une adaptation fine des filtres qui tiennent compte de la répartition énergétique des composantes de la source. Des instabilités ont donc été observées dès lors que les bruits étaient non-stationnaires, mais aussi avec certaines voix d'hommes ou de femmes, y compris avec le bruit Bureau pourtant stationnaire.
Ce problème n'est pas apparu avec le GSC proposé puisqu'il utilise les signaux pleine bande, ob et y_f, à ses entrées. En résumé et en faisant exception des solutions Sol. 2 et Sol. 2+, l'amélioration des pas adaptatifs proposés au Chapitre 3 devrait permettre d'améliorer nettement les performances de Sol. 1+ et accentuer l'écart entre le GSC proposé et le RGSC, voir même de diminuer les dégradations sur le signal. Malheureusement, nous ne sommes pas parvenus à justifier plus précisément l'inefficacité des solutions Sol. 2 et Sol. 2+. Notons d'ailleurs qu'il n'existe pas d'article au sujet de cette structure et de son évaluation, elle a été décrite seulement dans un brevet [33].
Enfin, les algorithmes fréquentiels décrits dans ce mémoire sont basés sur la technique de reconstruction OLS introduisant un phénomène de clics qui peut se révéler audible lors d'une écoute au casque. Ce phénomène étant classique, nous avons choisi sciemment de l'ignorer dans un premier temps afin de concentrer notre travail sur la mise en œuvre récursive et adaptative du GSC pour en vérifier les possibilités. Aussi, dans le but d'améliorer la qualité du signal, il serait judicieux d'inclure un mécanisme d'interpolation des filtres entre les trames ou bien, décliner les algorithmes avec la méthode *Overlap-Add* (OLA) ou *Weighted Overlap-Add* (WOLA) en tirant parti des travaux de Moulines et al. [133] sur le *Generalized Multidelay Filter* (GMDF).

6.6 Synthèse du chapitre

Dans ce chapitre, nous avons proposé une nouvelle structure récursive pour le GSC, à faible coût et qui plus est efficace. Basée sur la structure à filtres duaux étudiée dans les chapitres précédents, nous avons montré l'apport des implémentations récursives du GSC qui gèrent de manière plus efficace la diaphonie et donc améliorent la réduction de bruit. De plus, la marge de progression est importante, en commençant notamment par améliorer les pas adaptatifs que nous avons définis. Cependant, les performances de la technique proposée présentent une marge de progression importante. Celle-ci peut être obtenue à travers une gestion plus fine des pas d'adaption des algorithmes proposés. Une analyse complémentaire des différents modes de fonctionnement des filtres adaptatifs en fonction des retards de synchronisation devrait également permettre d'améliorer les performances globales, notamment en tenant compte du contenu spectral de la source cible. En effet, nous avons vu qu'il était possible d'observer la période *pitch* de la cible, cela dépend donc de la source (homme ou femme) et de la longueur du filtre. Ainsi, en plus

des contraintes de robustesse, il serait intéressant de statuer sur la nécessité ou non d'observer la période *pitch* ainsi que sur le mode de fonctionnement optimal de la structure récursive en fonction des différentes possibilités offertes par le choix des délais. L'objectif serait de définir quel est le fonctionnement optimal entre prédiction directe, rétrograde ou identification. Enfin, nous n'avons pas présenté de résultats sur l'évaluation subjective des différentes méthodes, mais cette étape est incontournable, notamment pour aider à l'ajustement des paramètres des techniques de traitement par blocs utilisées dans les algorithmes fréquentiels mais probablement aussi pour le choix des retards optimaux. Des écoutes expertes informelles ont néanmoins permis de corroborer les résultats objectifs obtenus.

Conclusion et perspectives

Le but de ce travail de thèse consistait à proposer une solution efficace pour permettre une prise de son distante en environnement adverse et sans connaissance *a priori* sur celui-ci. Ainsi, pour être en mesure de s'adapter aux variations des conditions acoustiques, du nombre de sources et de leurs positions, nous nous sommes orientés vers les solutions adaptatives de type GSC. Cette antenne particulière est une solution intéressante pour le traitement de la parole dans le cas d'une prise de son éloignée, puisque, en reportant la complexité algorithmique sur l'architecture, elle permet,

1. d'augmenter le gain en RSB à mesure que l'on augmente le nombre de microphones ($10 * log(M_{mic})$ pour une antenne linéaire comportant M_{mic} microphones). L'utilisation d'une antenne plutôt qu'un simple microphone est incontournable lorsque qu'il est impossible d'équiper le locuteur d'un microphone ou que l'on ne souhaite pas le faire. Or, dans ces conditions, il est impossible d'obtenir un RSB aussi fort que lorsqu'un microphone est placé proche de la bouche du locuteur puisque l'éloignement source-microphone implique que toutes les modifications apportées au signal de parole par la salle sont captées par le microphone. De plus, l'énergie de la source captée par le microphone décroît de manière inversement proportionnelle au carré de la distance qui sépare la source du microphone, par conséquent, la prise de son dématérialisée distante nécessite l'utilisation d'une antenne acoustique. Enfin, l'antenne acoustique est également en mesure de réaliser un filtrage permettant de sélectionner spatialement les sources, ce qui permet de supprimer efficacement les sources interférentes et le signal d'écho,
2. de concevoir librement et facilement le frontal acoustique en fonction du type d'environnement ciblé (connaissance *a priori*) et indépendamment de l'algorithme de réduction de bruit adaptatif utilisé. En effet, en supposant que l'on vise une prise de son dans un environnement où la réverbération est très importante (par exemple un hall de gare ou un aéroport), le frontal acoustique sera préférentiellement conçu comme une antenne superdirective. En effet, le bruit diffus qui modélise ce genre d'environnement est pénalisant pour le GSC,

en particulier lorsqu'il est utilisé en *broadside*. Ainsi, pour améliorer la directivité de celui-ci, l'antenne est utilisée en *endfire* avec un espacement des microphones inférieur à $\lambda/2$. Dans ce cas, la directivité est proche de M_{mic}^2, contre M_{mic} pour l'antenne linéaire. Par ailleurs, pour améliorer la robustesse des antennes adaptatives vis-à-vis des erreurs de pointage et des imperfections des microphones, il est nécessaire d'ajouter un certain nombre de contraintes qui complexifient la mise en œuvre. Or, en séparant le frontal acoustique et le bloc d'annulation adaptative de bruit, l'implémentation des antennes de type LCMV sous la forme du GSC facilite l'ajout de ces contraintes,

3. de supprimer le bruit de manière adaptative. Par conséquent, en disposant d'un dispositif de pointage et de suivi de sources, le GSC est une solution autonome capable de fonctionner dans divers environnements. En effet, cette antenne crée, sous la contrainte de la géométrie de l'antenne et du nombre de microphones, des directivités spatiales complémentaires à celles du frontal acoustique qui vont s'ajuster en fonction de la position des sources utiles et de bruit. Nous disposons donc d'un nombre important de degrés de liberté pour ajuster indépendamment les directivités du frontal acoustique et celles du module de réduction de bruit.

Cependant, les solutions adaptatives sont plus sensibles aux imperfections du frontal acoustique, au mauvais pointage des sources et à la réverbération qui diminuent de façon significative leurs performances. Dans le cas du GSC, déjà très sensible à la réverbération, le mauvais pointage entraîne l'auto-annulation du signal utile en sortie à cause de la fuite de ce signal dans l'étage d'estimation du bruit. Pour corriger ce défaut majeur, nous proposons dans ce travail de thèse de repenser l'architecture du GSC à partir de la méthode récursive bicapteur de gestion de la diaphonie, CTRANC, étendue ensuite au cas multicapteur. En ce sens, une partie de ce document a été consacrée à l'étude complète des performances et du comportement de la structure récursive bicapteur, puis ensuite, à son intégration dans le GSC.

Dans un premier temps, nous nous sommes consacré, au Chapitre 2, à l'étude de la stabilité de la structure récursive ainsi qu'à son champ d'application en donnant les conditions requises à son bon fonctionnement. Nous avons par ailleurs montré que l'algorithme de décorrélation couramment utilisé pour la mise à jour des filtres adaptatifs est inadapté à cette structure, ce qui nous a conduit à proposer deux nouvelles solutions. La première, décrite au Chapitre 3, consiste à modifier l'architecture du CTRANC pour permettre l'utilisation des algorithmes d'annulation de bruit classiques. De plus, ce chapitre a été l'occasion de définir de nouveaux pas variables non paramétriques que nous utilisons ensuite dans les versions du GSC analysées.

La deuxième solution que nous proposons, décrite au Chapitre 4, est fondée sur les travaux de Zinser et Mirchandani [202]. A partir de l'algorithme optimal proposé par ces derniers pour la structure récursive, nous déclinons un nouvel algorithme adapté au débruitage de la parole et permettant de réduire la complexité de l'algorithme initial, de faciliter l'implémentation dans le domaine fréquentiel et enfin d'améliorer l'estimation des filtres et donc les performances globales du CTRANC en comparaison avec les techniques basées sur la décorrélation.

Dans un second temps, nous donnons au Chapitre 5 nos analyses et nos résultats concernant l'étude de la convergence et de l'existence d'un point d'équilibre stable pour la structure récursive. Ce travail qui fait suite à l'étude de la stabilité était rendu nécessaire afin de mieux appréhender le comportement et éventuellement le contrôle des GSC récursifs analysés et proposés. La force et la pertinence de notre approche pour traiter ce problème délicat du filtrage adaptatif tient dans le fait que notre analyse s'appuie uniquement sur les relations liant les signaux estimés et leurs artéfacts. Les résultats obtenus sont donc indépendants du type d'algorithme utilisé, de la nature des sources et de la nature des chemins de propagation.

Enfin, le dernier chapitre de la thèse présente l'extension de la méthode bicapteur analysée dans les trois chapitres précédents au cas du GSC. Après avoir tout d'abord exposé les deux seuls et uniques travaux à notre connaissance qui envisagent une implémentation récursive du GSC, nous avons décliné une nouvelle architecture récursive moins coûteuse que les deux précédentes. Après une analyse de l'influence des chemins acoustiques sur la structure récursive en lien avec les problèmes de synchronisation dans les antennes, nous avons montré le potentiel et la supériorité par rapport au GSC classique, des implémentations récursives du GSC qui prennent mieux en compte la fuite du signal et améliorent le facteur de réduction de bruit et le gain en RSB. Cependant, hormis la solution proposée, nous avons vu que ces structures pouvaient être instables, ceci en fonction du type de sources et des conditions de bruit.

En ce qui concerne les perspectives de ce travail, deux axes peuvent être envisagés. Pour améliorer les performances du GSC proposé et celles de l'algorithme développé pour les structures récursives, les efforts doivent se concentrer dans un premier temps sur le raffinement dans l'application des pas variables proposés. Ceux-ci doivent être ajustés par l'ajout de seuils qui doivent être fixés par bande de fréquence et éventuellement probabilisés. En effet, en repartant de la base de la conception du GSC, nous n'avons pas intégré de modèle probabiliste qui devrait profiter à l'estimation des filtres, à condition également de tirer profit de la récursivité. Par ailleurs, pour

pallier les erreurs de pointage et rendre le système final plus robuste et autonome, il serait intéressant d'envisager l'adjonction d'un module de pointage et de suivi de sources basé sur un couplage audio-vidéo. De nos jours, ces techniques commencent à se démocratiser dans les domaines de la robotique et de la sécurité. Comparées aux méthodes de déréverbération qui sont bien souvent moins performantes en pratique, les techniques qui mettent en œuvre un couplage audio-vidéo se révèlent certes plus complexes et emcombrantes mais également plus robuste à la variabilité des environnements acoutiques et plus efficaces dans la sélection des sources utiles compétitives.

Références bibliographiques

[1] « Software tools for speech and audio coding standardization ». ITU-T Software Tool Library 2009 UserŠs Manual / Chap. 17 - ITU-T Stereo processing tool, March 2010. (cité page 266)

[2] Aboulnasr (T.) et Mayyas (K.), « A robust variable step-size lms-type algorithm : analysis and simulations », *IEEE Transactions on Signal Processing*, **45**(3), March 1997, p. 631–639. (cité page 90)

[3] Affes (S.) et Grenier (Y.). « Test of adaptive beamformers for speech acquisition in cars ». Dans *Conference on Signal Processing Applications and Technology*, volume 1, p. 154–169, 1994. (cité page 41)

[4] Affes (S.), Gazor (S.) et Grenier (Y.). « Robust adaptive beamforming via lms-like target tracking ». Dans *IEEE International Conference on Acoustics Speech and Signal Processing*, volume 4. Institue of Electrical Engineers, 1994. (cité page 40)

[5] Affes (S.), Gazor (S.) et Grenier (Y.), « An algorithm for multisource beamforming and multitarget tracking », *IEEE Transactions on Signal Processing*, **44**(6), June 1996, p. 1512–1522. (cité page 40)

[6] Affes (S.) et Grenier (Y.). « A source subspace tracking array of microphones for double talk situations ». Dans *IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 2, p. 909–912, May 1996. (cité pages 41 et 237)

[7] Affes (S.) et Grenier (Y.), « A signal subspace tracking algorithm for microphone array processing of speech », *IEEE Transactions on Speech and Audio Processing*, **5**(5), September 1997, p. 425–437. (cité page 237)

[8] Affès (S.), Gazor (S.) et Grenier (Y.). « Wideband robust adaptive beamforming via target tracking ». Dans *IEEE Seventh SP Workshop on Statistical Signal and Array Processing*, p. 141–144, 1994. (cité page 40)

[9] Agrawal (M.) et Prasad (S.), « Robust adaptive beamforming for wideband, moving, and coherent jammers via uniform linear arrays », *IEEE Transactions on Antennas and Propagation*, **47**(8), August 1999, p. 1267–1275. (cité page 41)

[10] Al-Kindi (M.) et Dunlop (J.), « Improved adaptive noise cancellation in the presence of signal leakage on the noise reference channel », *Signal Processing*, **17**(3), July 1989, p. 241–250. (cité pages 69, 71, 75, 87 et 93)

[11] Arslan (L.), McCree (A.) et Viswanathan (V.). « New methods for adaptive noise suppression ». Dans *Conference on Acoustics, Speech , and Signal Processing*, volume 1, p. 812–815, May 1995. (cité page 243)

[12] Backman (J.) et Oyj (N.). « Microphone array beam forming for multichannel recording ». Dans *AES Convention paper*, March 2003. (cité page 22)

[13] Bell (K. L.), Ephraim (Y.) et Trees (H. V.). « Robust adaptive beamforming under uncertainty in source direction-of-arrival ». Dans *Workshop Statistics and Signal Array Process*, 1996. (cité page 36)

[14] Benesty (J.), Chen (J.), Huang (Y. A.) et Dmochowski (J.), « On microphone-array beamforming from a mimo acoustical signal processing perspective », *IEEE Transactions on Audio, Speech, and Language Processing*, **15**(3), March 2007, p. 1053–1065. (cité page 31)

[15] Benesty (J.) et Gay (S. L.). « An improved pnlms algorithm ». Dans *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, volume 2, p. 1881–1884, 2002. (cité page 90)

[16] Benesty (J.), Huang (Y.) et Chen (J.). *Springer Handbook of Speech Processing*, chapitre 6 - Wiener and adaptive filters, p. 101–120. Springer-Verlag New York, 2007. (cité page 90)

[17] Benesty (J.), Rey (H.), Vega (L. R.) et Tressens (S.), « A nonparametric vss nlms algorithm », *IEEE Signal Processing Letters*, **13**(10), October 2006, p. 581–584. (cité page 90)

[18] Bensaid (S.) et Slock (D.). « Single microphone blind audio source separation using short+long term ar modeling ». Dans *International Workshop on Acoustic Echo and Noise Control*, Tel-Aviv, September 2010. (cité page 6)

[19] Bitzer (J.) et Simmer (K. U.). *Microphone Arrays : Signal Processing Techniques and Applications*, chapitre 2 - Superdirective Microphone Arrays, p. 19–38. Springer, 2001. (cité pages 9 et 27)

[20] Bitzer (J.), Simmer (K. U.) et Kammeyer (K.-D.). « Multichannel noise reduction : algorithms and theoretical limits ». Dans *European Signal Processing Conference*, p. 105–108, 1998. (cité pages 34 et 49)

[21] Bitzer (J.), Simmer (K. U.) et Kammeyer (K.-D.). « Theoretical noise reduction limits of the generalized sidelobe canceller (gsc) for speech enhancement ». Dans *IEEE International Conference on Acoustics, Speech and Signal Processing*, volume 5, p. 2965–2968, March 1999. (cité page 49)

[22] Booker (A.) et Ong (C.-Y.), « Multiple constraint adaptive filtering », *Geophysics*, **36**, 1971, p. 498. (cité page 35)

[23] Brandstein (M.) et Ward (D.), *Microphone Arrays : Signal Processing Techniques and Applications*, Digital signal processing. Springer, 2001. (cité page 16)

[24] Broman (H.), Lindgren (U.), Sahlin (H.) et Stoica (P.). « Source separation : A tito system identification approach ». Rapport technique, Chalmers University of Technology, 1998. (cité page 61)

[25] Buckley (K. M.) et Griffiths (L. J.), « An adaptive generalized sidelobe canceller with derivative constraints », *IEEE Transactions on antennas and propagation*, **34**(3), March 1986, p. 311–319. (cité page 35)

[26] Campbell (D. R.), Palomäki (K. J.) et Brown (G. J.), « Roomsim, a matlab simulation of "shoebox" room acoustics for use in teaching and research », *http ://media. paisley. ac. uk/campbell/Roomsim*, **Online**, 2005, p. 1–4. (cité pages 72 et 256)

[27] Carlson (B. D.), « Covariance matrix estimation errors and diagonal loading in adaptive arrays », *IEEE Transactions on Aerospace and Electronic Systems*, **24**(4), 1988, p. 397–401. (cité pages 31 et 37)

[28] Charkani (N.). *Séparation auto-adaptative de sources pour des mélanges convolutifs. Application à la téléphonie mains-libres dans les voitures.* Thèse de doctorat, Institut National Polytechnique de Grenoble, November 1996. (cité page 173)

[29] Charkani (N.) et Deville (Y.), « Self-adaptive separation of convolutively mixed signals with a recursive structure. part i : stability analysis and optimization of asymptotic behaviour », *Signal Processing*, **73**(3), March 1999, p. 225–254. (cité pages 144 et 146)

[30] Charkani (N.) et Deville (Y.), « Self-adaptive separation of convolutively mixed signals with a recursive structure. part ii : theoretical extensions and application to synthetic and real signals », *Signal Processing*, **75**(2), 1999, p. 117–140. (cité pages 144 et 146)

[31] Chen (J.), Phua (K.), Shue (L.) et Sun (H.), « Performance evaluation of adaptive dual microphone systems », *Speech Communication*, **51**(12), June 2009, p. 1180–1193. (cité pages 34 et 49)

[32] Chen (T.). « Unified eigenfilter approach : with applications to spectral/spatial filtering ». Dans *IEEE International Symposium on Circuits and Systems*, p. 331–334, 1993. (cité page 27)

[33] Choi (C.), Kim (J.) et Kong (D.). « Adaptive beamforming method and apparatus using feedback structure », 2008. (cité pages 178, 179 et 198)

[34] Choi (C.), Kong (D.), Kim (J.) et Bang (S.). « Speech enhancement and recognition using circular microphone array for service robots ». Dans *International Conference on Intelligent Robots and Systems*, p. 3516–3521, Las Vegas, October 2003. (cité pages 178 et 179)

[35] Claesson (I.) et Nordholm (S.), « A spatial filtering approach to robust adaptive beaming », *IEEE Transactions on Antenna and Propagation*, **40**(9), September 1992, p. 1093–1096. (cité pages 41 et 45)

[36] Claudio (E. D. D.) et Parisi (R.). *Microphone Array : Signal Processing Techniques and Applications*, chapitre 9 - Multi-Source Localization Strategies, p. 181–201. Springer, 2001. (cité page 17)

[37] Cohen (I.), « Analysis of two channel generalized sidelobe canceller (gsc) with post-filtering », *IEEE Transactions on Speech and Audio Processing*, **11**(6), November 2003, p. 684–696. (cité page 18)

[38] Cohen (I.), « Noise spectrum estimation in adverse environments : improved minima controlled recursive averaging », *IEEE Transactions on Speech andAudio Processing*, **11**(5), September 2003, p. 466–475. (cité page 244)

[39] Cohen (I.) et Berdugo (B.). « Microphone array post-filtering for non-stationary noise suppression ». Dans *IEEE International Conference on Acoustics, Speech and Audio Processing*, volume 1, p. 901–904, May 2002. (cité page 18)

[40] Cohen (I.) et Berdugo (B.), « Noise estimation by minima controlled recursive averaging for robust speech enhancement », *IEEE Signal Processing Letters*, **9**(1), January 2002, p. 12–15. (cité page 244)

[41] Cox (H.), « Resolving power and sensitivity to mismatch of optimum array processor », *The Journal of the Acoustical Society of America*, **54**(3), January 1973, p. 771–785. (cité pages 73 et 231)

[42] Cox (H.), Zesking (R. M.) et Owen (M. M.), « Robust adaptive beamforming », *IEEE Transactions on Acoustics, Speech, and Signal Processing*, **ASSP-35**(10), October 1987, p. 1365–1376. (cité pages 31 et 37)

[43] Cui (J.), Naylor (P. A.) et BRown (D. T.). « An improved ipnlms algorithm for echo cancellation in packet-switched networks ». Dans *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, volume 4, p. 141–144, 2004. (cité page 90)

[44] Dibiase (J. H.), Silverman (H. F.) et Brandstein (M. S.). *Microphone Arrays : Signal Processing Techniques and Applications*, chapitre 8 - Robust Localization in Reverberant Rooms, p. 157–180. Springer, 2001. (cité page 17)

[45] Djendi (M.), Gilloire (A.) et Scalart (P.). « New frequency domain postfilters for noise cancellation using two closely spaced microphones ». Dans *Proceedings of EURASIP Conference EUSIPCO*, p. 218–221, Poznan, Poland, September 2007. (cité pages 108 et 112)

[46] Doclo (S.). *Multi-microphone noise reduction and dereverberation techniques for speech applications.* Thèse de doctorat, Katholieke Universiteit Leuven, May 2003. (cité pages 9, 10, 16 et 35)

[47] Doclo (S.) et Moonen (M.), « Design of broadband beamformers robust against gain and phase errors in the microphone array characteristics », *IEEE Transactions on Signal Processing*, **51**(10), 2003, p. 2511–2526. (cité page 37)

[48] Doclo (S.) et Moonen (M.), « Design of far-field and near-field broadband beamformers using eigenfilters », *Signal Processing*, **83**(12), 2003, p. 2641–2673. (cité pages 26 et 37)

[49] Doclo (S.) et Moonen (M.), « Multimicrophone noise reduction using recursive gsvd-based optimal filtering with anc postprocessing stage », *IEEE Transactions on Speech and Audio Processing*, **13**(1), January 2005, p. 53–69. (cité page 10)

[50] Doclo (S.), Spriet (A.), Wouters (J.) et Moonen (M.), « Frequency-domain criterion for the speech distorsion weighted multichannel wiener filter for robust noise reduction », *Speech Communication*, **49**(7-8), July-August 2007, p. 636–656. (cité pages 11 et 48)

[51] Duttweiler (D. L.), « Proportionate normalized least mean square adaptation in echo cancellers », *IEEE Transactions on Speech and Audio Processing*, **8**(5), 2000, p. 508–518. (cité page 90)

[52] Edo Hulsebos (D. d. V. Thomas Schuurmans) et Boone (R.). « Circular microphone array for discrete multichannel audio recording ». Dans *AES Convention Paper 5716*, March 2003. (cité page 22)

[53] El-Keyi (A.), Kirubarajan (T.) et Gershman (A. B.), « Robust adaptive beamforming based on the kalman filter », *IEEE Transactions on Signal Processing*, **53**(8), August 2005, p. 3032–3041. (cité page 39)

[54] Elko (G. W.). *Audio Signal Processing for Next-Generation Multimedia Communication Systems*, chapitre 2 - Differential microphone arrays, p. 11–65. Springer, 2004. (cité page 9)

[55] Elko (G. W.) et Meyer (J.). *Springer Handbook of Speech Processing*, chapitre 50 - Microphone arrays, p. 1019–1041. Springer-Verlag New York, 2007. (cité pages 9, 16 et 24)

[56] Ephraim (Y.) et Malah (D.), « Speech enhancement using a minimum mean-square error log-spectral amplitude estimator », *IEEE Transactions on Acoustics, Speech and Signal Processing*, **33**(2), April 1985, p. 443–445. (cité page 249)

[57] Ephraim (Y.) et Mallah (D.), « Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator », *IEEE Transactions on Acoustics, Speech and Signal Processing*, **32**(6), December 1984, p. 1109–1121. (cité pages 101 et 249)

[58] Er (M. H.) et Cantoni (A.), « Derivative constraints for broad-band element space antenna array processors », *IEEE Transactions on Acoustics, Speech and Signal Processing*, **31**(6), December 1983, p. 1378–1393. (cité page 35)

[59] Er (M. H.) et Cantoni (A.), « An unconstrained partitioned realization for derivative constrained broad-band antenna », *IEEE Transactions on Acoustics, Speech, and Signal Processing*, **ASSP-34**(6), 1986, p. 1376–1379. (cité page 35)

[60] Eweda (E.), « A new approach for analyzing the limiting behavior of the normalized lms algorithm under weak assumptions », *Signal Processing*, **89**(11), November 2009, p. 2143–2151. (cité page 147)

[61] Frost (O. L.). « An algorithm for linearly constrained adaptive array processing ». Dans *Proceedings of the IEEE*, volume 60, p. 926–935, August 1972. (cité pages 9, 31, 32 et 34)

[62] Fudge (G. L.) et Linebarger (D. A.), « A calibrated generalized sidelobe canceller for wideband beamforming », *IEEE Transactions on Signal Processing*, **42**(10), October 1994, p. 2871–2875. (cité page 40)

[63] Fudge (G. L.) et Linebarger (D. A.). « Steered response control of the generalized sidelobe canceller ». Dans *IEEE International Conference on Acoustics, Speech and Signal Processing*, volume 5, p. 3623–3626, May 1995. (cité page 41)

[64] Fudge (G. L.) et Linebarger (D. A.), « Spatial blocking filter derivative constraints for the generalized sidelobe canceller and music », *IEEE Transactions on Signal Processing*, **44**(1), January 1996, p. 51–61. (cité page 41)

[65] Gannot (S.). *Array Processing of Nonstationary Signals with Application to Speech*. Thèse de doctorat, Tel-Aviv University, 2000. (cité pages 41 et 46)

[66] Gannot (S.), Burnstein (D.) et Weinstein (E.), « Signal enhancement using beamforming and nonstationarity with applications to speech », *IEEE Transactions on Signal Processing*, **49**(8), August 2001, p. 1614–1626. (cité pages 41, 72, 227 et 238)

[67] Gannot (S.) et Cohen (I.), « Speech enhancement based on the general transfer function gsc and postfiltering », *IEEE Transactions on Speech and Audio Processing*, **12**(6), November 2004, p. 561–571. (cité page 18)

[68] Gannot (S.) et Cohen (I.). *Springer Handbook of Speech Processing*, chapitre 47 - Adaptive beamforming and postfiltering, p. 945–978. Springer-Verlag New York, 2007. (cité pages 16, 41, 43, 44, 238 et 240)

[69] Gerven (S. V.). *Adaptive noise cancellation and signal separation with application to speech enhancement*. Thèse de doctorat, Katholieke Universiteit Leuven, 1996. (cité pages 4, 57 et 67)

[70] Gerven (S. V.) et Compernolle (D. V.), « Signal separation by symmetric adaptive decorrelation : stability, convergence and uniqueness », *IEEE Transactions on Signal Processing*, **43**(7), July 1995, p. 1602–1612. (cité pages 64, 75, 76, 146 et 156)

[71] Gerven (S. V.) et Compernolle (D. V.). « Feedforward and feedback in symmetric adaptive noise canceller : stability analysis in a simplified case ». Dans *European Signal Processing Conference*, p. 1081–1084, Brussels, Belgium, August 1992. (cité pages 75 et 144)

[72] Gilbert (E.) et Morgan (S.), « Optimum design of directive antenna arrays subject to random variations », *The Bell System Technical Journal*, **34**, May 1955, p. 636–663. (cité page 29)

[73] Griffiths (L. J.) et Jim (C. W.). « An alternative approach to linearly constrained adaptive beamforming ». Dans *IEEE Transactions on Antennas and Propagation*, volume 30, p. 27–34, January 1982. (cité pages 9, 31 et 227)

[74] Guillaume (M.). *Analyse et synthèse de champs sonores.* Thèse de doctorat, ENST Paris, 2006. (cité page 23)

[75] Guérin (A.). *Rehaussement de la parole pour les communications mains-libres. Réduction de bruit et annulation d'écho non linéaire.* Thèse de doctorat, Université de Rennes 1, April 2002. (cité pages 101 et 173)

[76] Hamid (M. E.), Ogawa (K.) et Fukabayashi (T.), « Improved single-channel noise reduction method of speech by blind source separation », *Acoustical Science and Technology*, **28**(3), 2007, p. 153–164. (cité page 6)

[77] Haneda (Y.), Makino (S.) et Kaneda (Y.), « Common acoustical pole and zero modeling of room transfer functions », *IEEE Transactions on Speech and Audio Processing*, **2**(2), April 1994, p. 320–328. (cité page 61)

[78] Harmanci (K.), Tabrikian (J.) et Krolik (J. L.), « Relationships between adaptive minimu vairance beamforming and optimal source localization », *IEEE Transactions on Signal Processing*, **48**(1), 2000, p. 1–12. (cité page 37)

[79] Hassani (A. N. C. E.). *Séparation auto-adaptative de sources pour des mélanges convolutifs. Application à la téléphonie mains-libres dans les voitures.* Thèse de doctorat, Institut National Plolytechnique de Grenoble, November 1996. (cité page 65)

[80] Herbordt (W.) et Kellermann (W.). « Efficient frequency-domain realization of robust generalized sidelobe cancellers ». Dans *IEEE fourth Workshop on Multimedia Signal Processing*, p. 377–382, October 2001. (cité page 44)

[81] Herbordt (W.) et Kellermann (W.), « Frequency-domain integration of acoustic echo cancellation and a generalized sidelobe canceller with impro-

ved robustness », *European transactions on telecommunications*, **13**(2), 2002, p. 123–132. (cité page 44)

[82] Hirano (A.) et Nakayama (K.), « A low-distortion noise canceller and its learning algorithm in presence of crosstalk », *IEICE Transactions on Fundamentals*, **E84-A**(2), February 2001, p. 414–421. (cité page 58)

[83] Hirsch (H.) et Ehrlicher (C.). « Noise estimation techniques for robust speech recognition ». Dans *Conference on Acoustics, Speech, and Signal Processing*, volume 1, p. 153–156, May 1995. (cité page 243)

[84] Hoshuyama (O.), Begasse (B.), Sugiyama (A.) et Hirano (A.), « A real-time robust adaptive microphone array controlled by an snr estimate », *Threshold*, **1**, 1998, p. 3605–3608. (cité page 47)

[85] Hoshuyama (O.) et Sugiyama (A.). *Microphone Arrays : Signal Processing Techniques and Applications*, chapitre 5 - Robust Adaptive Beamforming, p. 87–109. Springer, 2001. (cité pages 34, 44 et 47)

[86] Hoshuyama (O.), Sugiyama (A.) et Hirano (A.). « A robust adaptive beamformer for microphone arrays with a blocking matrix using constrained adaptive filters ». Dans *IEEE Transactions on Signal Processing*, volume 47, p. 2677–2684, October 1999. (cité page 44)

[87] Huang (Y.), Benesty (J.) et Chen (J.). *Springer Handbook of Speech Processing*, chapitre 51 - Time Delay Estimation and Source Localization, p. 1043–1063. Verlag New York, 2007. (cité page 17)

[88] Ikeda (S.) et Sugiyama (A.), « An adaptive noise canceller with low signal-distorsion in the presence of crosstalk », *IEICE Transactions on Fundamentals*, **E82-A**(8), August 1999, p. 1517–1525. (cité page 90)

[89] Ikeda (S.) et Sugiyama (A.), « An adaptive noise canceller with low signal distortion for speech codecs », *IEEE Transactions on Signal Processing*, **47**(3), March 1999, p. 665–674. (cité page 90)

[90] Ito (N.), Ono (N.) et Sagayama (S.). « A blind noise decorrelation approach with crystal arrays on designing post-filters for diffuse noise suppression ». Dans *International Conference on Acoustics, Speech, and Signal Processing*, p. 317–320, 2008. (cité page 23)

[91] Jablon (N. K.), « Adaptive beamforming with the generalized sidelobe canceller in the presence of array imperfections », *IEEE Transactions on Antennas and Propagation*, **AP-34**(8), August 1986, p. 996–1012. (cité page 45)

[92] Jan (E.-E.) et Flanagan (J.). « Microphone arrays for speech processing ». Dans *URSI International Symposium on Signals, Systems, and Electronics*, p. 373–376, 1995. (cité page 31)

[93] Jan (E.-E.) et Flanagan (J.). « Sound capture from spatial volumes : Matched-filter processing of microphone arrays having randomly-distributed sensors ». Dans *IEEE International Conference on Acoustics, Speech and Signal Processing*, p. 917–920, Atlanta, 1996. (cité page 31)

[94] Joerg Bitzer (K. U. S.) et Kammeyer (K.-D.). « Multi-microphone noise reduction by post-filter and superdirective beamformer ». Dans *International Workshop on Acoustic Echo and Noise Control (IWAENC)*, p. 100–103, 1999. (cité pages 18 et 34)

[95] Joerg Bitzer (K. U. S.) et Kammeyer (K.-D.). « Theoretical noise reduction limits of the generalized sidelobe canceller for speech enhancement ». Dans *IEEE International Conference on Acoustics, Speech and Signal Processing*, volume 5, p. 2965–2968, Mars 1999. (cité page 34)

[96] Johnson (D. H.) et Dudgeon (D. E.), *Array signal processing : Concepts and Techniques*. Prentice Hall, 1993. (cité page 16)

[97] Johnson (D. H.) et Dudgeon (D. E.). *Array Signal Processing : Concepts and Techniques*, chapitre 8 - Tracking, p. 424–470. Prentice Hall, 1993. (cité page 17)

[98] Johnson (D. H.) et Dudgeon (D. E.). *Array signal processing : Concepts and Techniques*, chapitre 3 - Apertures and Arrays, p. 59–110. Prentice Hall, 1993. (cité page 73)

[99] Junfeng Li (M. A.), « A noise reduction system based on hybrid noise estimation technique and post-filtering in arbitrary noise environments », *Speech communication*, **48**, Février 2006, p. 111–126. (cité page 18)

[100] Junfeng Li (X. L.) et Akagi (M.). « A noise reduction system in arbitrary noise environments and its application to speech enhancement and speech recognition ». Dans *IEEE International Conference on Acoustics, Speech, and Signal Processing*, p. 277–280, 2005. (cité page 18)

[101] Kato (M.), Serizawa (M.), Toki (N.), Mori (U.), Morishita (Y.) et Hayashi (K.), « Noise suppression with high speech quality based on weighted noise estimation for 3g handsets », *NEC research & development*, **44**(4), 2003, p. 340–348. (cité page 244)

[102] Kennedy (R. A.), Abhayapala (T. D.) et Ward (D. B.), « Broadband near-field beamforming using a radial beampattern transformation », *IEEE Transactions on Signal Processing*, **46**(8), August 1998, p. 2147–2156. (cité page 26)

[103] Korompis (D.), Yao (K.) et Lorenelli (F.). « Broadband maximum energy array with user imposed spatial and frequency constraints ». Dans *Proceedings of the Acoustics, Speech, and Signal Processing*, volume 4, p. 529–532, 1994. (cité page 27)

[104] Krim (H.) et Viberg (M.), « Two decades of array signal processing research », *IEEE Signal Processing Magazine*, **13**(4), July 1996, p. 67–94. (cité page 16)

[105] Krolik (J. L.), « The performance of matched-field beamformers with mediterranean vertical array data », *IEEE Transactions on Signal Processing*, **44**(10), 1996, p. 2605–2611. (cité pages 35 et 36)

[106] Kuo (S. M.) et Peng (W. M.), « Principle and applications of asymmetric crosstalk-resistant adaptive noise canceler », *Journal of the Franklin Institute*, **337**(1), January 2000, p. 57–71. (cité page 72)

[107] Kwong (R. H.) et Johnston (E. W.), « A variable step size lms algorithm », *IEEE Transactions on Signal Processing*, **40**(7), July 1992, p. 1633–1642. (cité page 90)

[108] Lamberti (R.). *Aperçu sur les méthodes de filtrage des réseaux d'antennes*. Institut National des Télécommunications, 1995. (cité page 16)

[109] Laroche (J.). « Traitement des signaux audio-fréquence ». Télécom Paris, February 1995. (cité pages 24 et 25)

[110] Lee (C.-C.) et Lee (J.-H.), « Robust adaptive array beamforming under steering vector errors », *IEEE Transactions on Antennas and Propagation*, **45**(1), January 1997, p. 168–175. (cité page 36)

[111] Lefkimmiatis (S.) et Maragos (P.), « A generalized estimation approach for linear and nonlinear microphone array post-filters », *Speech Communication*, **49**(7-8), July-August 2007, p. 657–666. (cité page 18)

[112] Lepauloux (L.), Scalart (P.) et Marro (C.). « Low distortion decoupled crosstalk resistant adaptive noise canceller ». Dans *International Workshop on Acoustic Echo and Noise Control*, September 2008. (cité pages 84 et 93)

[113] Lepauloux (L.), Scalart (P.) et Marro (C.). « Annulation de bruit à faible distorsion », 2009. (cité page 84)

[114] Lepauloux (L.), Scalart (P.) et Marro (C.). « An efficient low-complexity algorithm for crosstalk-resistant adaptive noise canceller ». Dans *European Signal Processing Conference*, p. 204–208, August 2009. (cité page 116)

[115] Lepauloux (L.), Scalart (P.) et Marro (C.). « Computationally efficient and robust frequency-domain gsc ». Dans *Internationa Workshop on Acoustic Echo and Noise Control*, Tel-Aviv, Israel, September 2010. (cité pages 178 et 196)

[116] Li (J.), Stoica (P.) et Wang (Z.), « On robust capon beamforming and diagonal loading », *IEEE Transactions on Signal Processing*, **51**(7), January 2003, p. 1702–1715. (cité page 39)

[117] Lindgren (U.) et Broman (H.), « Source separation using a criterion based on second-order statistics », *IEEE Transactions on Signal Processing*, **46**(7), July 1998, p. 1837–1850. (cité pages 65 et 69)

[118] Liu (Y.), Zou (Q.) et Lin (Z.). « Generalized sidelobe cancellers with leakage constraints ». Dans *IEEE International Symposium on Circuits and Systems*, p. 3741–3744, 2005. (cité page 35)

[119] Loellman (H.), Yilmaz (E.), Jeub (M.) et Vary (P.). « An improved algorithm for blind reverberation time estimation ». Dans *International Workshop on Acoustic Echo and Noise Control*, Tel-Aviv, September 2010. (cité page 31)

[120] Lorenz (R. G.) et Boyd (S. P.), « Robust minimum variance beamforming », *IEEE Transactions on Signal Processing*, **53**(5), May 2005, p. 1684–1696. (cité page 38)

[121] Mabande (E.), Schad (A.) et Kellermann (W.). « Robust superdirectional beamforming for hands-free speech capture in cars ». NAG/DAGA, 2009. (cité page 39)

[122] Mabande (E.), Schad (A.) et Kellermann (W.). « Design of robust superdirective beamformers as a convex optimization problem ». Dans *International Workshop on Acoustic Echo and Noise Control*, Tel-Aviv, 2010. (cité page 39)

[123] Marro (C.). *Traitements de déréverbération et de débruitage pour le signal de parole dans des contextes de communication interactive*. Thèse de doctorat, Université de Rennes 1, 1996. (cité pages 18, 24, 26, 31, 62 et 252)

[124] Marro (C.), Mahieux (Y.) et Simmer (K. U.), « Analysis of noise reduction and dereverberation techniques based on microphone arrays with postfilte-

ring », *IEEE Transactions on Speech and Audio Processing*, **6**(3), May 1998, p. 240–259. (cité page 18)

[125] Mathews (V. J.) et Xie (Z.), « A stochastic gradient adaptive filter with gradient adaptive step size », *IEEE Transactions on Signal Processing*, **41**(6), June 1993, p. 2075–2087. (cité page 90)

[126] McCowan (I.). « Microphone arrays : a tutorial », April 2001. (cité pages 16 et 24)

[127] McCowan (I.), Lincoln (M.) et Himawan (I.), « Microphone array shape calibration in diffuse noise fields », *IEEE Transactions on Audio, Speech, and Language Processing*, **16**(3), March 2008, p. 666–670. (cité page 40)

[128] McCowan (I.), Moore (D.) et Sridharan (S.). « Speech enhancement using near-field superdirectivity with an adaptive sidelobe canceler and post-filter ». Dans *Australian International Conference on Speech Science and Technology*, p. 268–273, 2000. (cité page 34)

[129] McCowan (I. A.) et Bourlard (H.). « Microphone array post-filter for diffuse noise field ». Dans *Proceedings of the 2002 IEEE International Conference on Acoustics, Speech and Signal Processing*, 2002. (cité page 18)

[130] McCowan (I. A.) et Bourlard (H.). « Microphone array post-filter based on noise field coherence ». Dans *IEEE Transactions on Speech and Audio Processing*, volume 11, p. 709–716, November 2003. (cité page 18)

[131] Mei (T.) et Yin (F.), « Blind separation of convolutive mixtures by decorrelation », *Signal Processing*, **84**(12), December 2004, p. 2297–2313. (cité pages 66, 69 et 77)

[132] Mirchandani (G.), Zinser (R. L.), Jr. et Evans (J. B.), « A new adaptive noise cancellation scheme in the presence of crosstalk », *IEEE Transactions on circuits and systems II : Analog and Digital Signal Processing*, **39**(10), October 1992, p. 681–694. (cité pages 69, 77 et 116)

[133] Moulines (E.), Amrane (O. A.) et Grenier (Y.), « The generalized multidelay adaptive filter : structure and convergence analysis », *IEEE Transactions on Signal Processing*, **43**(1), January 1995, p. 14–28. (cité page 198)

[134] Mourjopoulos (J.) et Paraskevas (M. A.), « Pole and zero modeling of room transfer functions », *Journal of Sound and Vibration*, **146**(146), 1991, p. 281–302. (cité pages 60 et 61)

[135] Mourjopoulos (J. N.), « Digital equalization of room acoustics », *Journal of the Audio Engineering Society*, **42**(11), November 1994, p. 884–900. (cité page 61)

[136] Neely (S. T.) et Allen (J. B.), « Invertibility of a room impulse response », *Journal of the Acoustical Society of America*, **6**(1), July 1979, p. 165–169. (cité page 61)

[137] Nordholm (S.), Claesson (I.) et Dahl (M.), « Adaptive microphone array employing calibration signals : an analytical evaluation », *IEEE transactions on Speech and Audio processing*, **7**(3), May 1999, p. 241–252. (cité page 40)

[138] Nordholm (S.), Dam (H. Q.), Grbi ? (N.) et Low (S. Y.). *Speech Enhancement*, chapitre 9 - Adaptive microphone array employing spatial quadratic soft constraints and spectral shaping, p. 229–246. Springer, 2005. (cité page 35)

[139] Ono (N.), Ito (N.) et Sagayama (S.). « Five class of crystal arrays for blind decorrelation of diffuse noise ». Dans *International Workshop on Sensor Array and Multichannel Signal Processing (SAM)*, p. 151–154, Darmstadt, July 2008. (cité pages 22 et 24)

[140] Park (P.), Chang (M.) et Kong (N.), « Scheduled-stepsize nlms algorithm », *IEEE Signal Processing Letters*, **16**(12), December 2009, p. 1055–1058. (cité pages 88 et 90)

[141] Plapous (C.). *Traitements pour la réduction de bruit. Application à la communication parlée.* Thèse de doctorat, Université de Rennes 1, 2005. (cité pages 101, 102 et 242)

[142] Plapous (C.), Marro (C.) et Scalart (P.). « A two-step noise reduction technique ». Dans *IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 1, 2004. (cité page 101)

[143] Plapous (C.), Marro (C.) et Scalart (P.). « Speech enhancement using harmonic regeneration ». Dans *IEEE International Conference on Acoustics, Speech, and Signal Processing*, Philadelphie, 2005. (cité page 18)

[144] Proakis (J. G.) et Manolakis (G.). *Digital signal processing - Principles, algorithms, and applications*, chapitre 5, p. 354–358. Pearson Prentice Hall, 4 édition, 2007. (cité pages 61 et 62)

[145] Rabinkin (D.), Renomeron (R.), Flanagan (J.) et Macomber (D.). « Optimal truncation time for matched filter array processing ». Dans *IEEE In-*

ternational Conference on Acoustics, Speech, and Signal Processing, volume 6, 1998. (cité page 31)

[146] Rafaely (B.). « Super-resolution spherical microphone arrays ». Dans *IEEE Convention of Electrical and Electronics Engineers in Israel*, September 2004. (cité page 22)

[147] Rafaely (B.), « Analysis and design of spherical microphone arrays », *IEEE Transactions on Speech and Audio Processing*, **13**(1), January 2005, p. 135–143. (cité page 22)

[148] Rafaely (B.), « Phase-mode versus delay-and-sum spherical microphone array processing », *IEEE signal Processing Letters*, **12**(10), October 2005, p. 713–716. (cité page 22)

[149] Reuven (G.), Gannot (S.) et Cohen (I.), « Joint noise reduction and acoustic echo cancellation using the transfer-function generalized sidelobe canceller », *Speech Communication*, **49**(7-8), July-August 2007, p. 623–635. (cité page 227)

[150] Reuven (G.), Gannot (S.) et Cohen (I.), « Performance analysis of dual source transfer-function generalized sidelobe canceller », *Speech Communication*, **49**(7-8), July-August 2007, p. 602–622. (cité page 49)

[151] Reuven (G.), Gannot (S.) et Cohen (I.), « Dual-source transfer-function generalized sidelobe canceller », *IEEE Transaction on Audio, Speech and Language Processing*, **16**(4), May 2008, p. 711–727. (cité pages 227, 238 et 239)

[152] Rombouts (G.) et Moonen (M.), « Fast qrd-lattice-based unconstrained optimal filtering for acoustic noise reduction », *IEEE Transactions on Speech and Audio Processing*, **13**(6), 2005, p. 1130–1143. (cité page 10)

[153] Sahlin (H.) et Broman (H.). « Signal separation applied to real world signals ». Dans *The International Workshop on Acoustic Echo and Noise Control*, p. 73–76, 1997. (cité page 65)

[154] Saruwatari (H.), Kajita (S.), Takeda (K.) et Itakura (F.), « Speech enhancement using nonlinear microphone array based on complementary beamforming », *IEICE Transactions on Fundamentals*, **E82-A**(8), August 1999, p. 1501–1510. (cité page 41)

[155] Saruwatari (H.), Kajita (S.), Takeda (K.) et Itakura (F.). « Speech enhancement using nonlinear microphone array with noise adaptive complementary beamforming ». Dans *IEEE International conference on Acoustics, Speech, and Signal Processing*, volume 2, p. 1049–1052, June 1999. (cité page 41)

[156] Sato (M.), Sugiyama (A.) et Ohnaka (S.), « An adaptive noise canceller with low signal-distortion based on variable stepsize subfilters for human-robot communication », *IEICE Transactions on Fundamentals*, **E88-A**(8), August 2005, p. 2055–2061. (cité page 91)

[157] Scalart (P.) et Filho (J. V.). « Speech enhancement based on a priori signal to noise estimation ». Dans *IEEE International Conference on Acoustics, Speech and Signal Processing*, volume 2, p. 629–632, May 1996. (cité page 243)

[158] Scalart (P.) et Lepauloux (L.). « On the convergence behavior of recursive adaptive noise cancellation structure in the presence of crosstalk ». Dans *Sensor Signal Processing for Defence conference*, London, September 2010. (cité page 144)

[159] Schutz (A.) et Slock (D.). « Blind audio source separation using short+long term ar source models and iterative itakura-saito distance minimization ». Dans *International Workshop on Acoustic Echo and Noise Control*, Tel-Aviv, September 2010. (cité page 6)

[160] Sehr (A.), Habets (E.), Maas (R.) et Kellermann (W.). « Towards a better understanding of the effect of reverberation on speech recognition performance ». Dans *International Workshop on Acoustic Echo and Noise Control*, Tel-Aviv, September 2010. (cité page 7)

[161] Sehr (A.) et Kellermann (W.). « On the statistical properties of reverberant speech feature vector sequences ». Dans *International Workshop on Acoustic Echo and Noise Control*, Tel-Aviv, September 2010. (cité page 7)

[162] Shalvi (O.) et Weinstein (E.), « System identification using nonstationary signals », *IEEE Transactions on Signal Processing*, **44**(8), 1996, p. 2055–2063. (cité pages 238, 239 et 240)

[163] Shimizu (H.), Ono (N.), Matsumoto (K.) et Sagayama (S.). « Isotropic noise suppression in the power spectrum domain by symmetric microphone arrays ». Dans *International Workshop on Applications of Signal Processing to Audio and Acoustics*, p. 54–57, New Paltz, October 2007. (cité page 22)

[164] Shynk (J. J.), « Adaptive iir filtering », *ASSP Magazine*, **6**(2), April 1989, p. 4–21. (cité page 146)

[165] Shynk (J. J.), « Frequency-domain and multirate adaptive filtering », *IEEE Signal Processing Magazine*, **9**(1), January 1992, p. 14–37. (cité pages 108 et 139)

[166] Simmer (K. U.), Bitzer (J.) et Marro (C.). *Microphone Arrays : Signal Processing Techniques and Applications*, chapitre 3 - Post-filtering Techniques, p. 39–60. Springer, 2001. (cité page 10)

[167] Slavakis (K.), Yukawa (M.) et Yamada (I.). « Robust capon beamforming by the adaptive projected subgradient method ». Dans *IEEE International Conference on Acoustics, Speech and Signal Processing*, p. 1005–1008, May 2006. (cité page 36)

[168] Spriet (A.). *Adaptive filtering techniques for noise reduction and acoustic feedback cancellation in hearing aids*. Thèse de doctorat, Katholieke Universiteit Leuven, September 2004. (cité pages 11 et 48)

[169] Spriet (A.), Moonen (M.) et Wouters (J.), « A multi-channel subband gsvd approach to speech enhancement », *European Transactions on telecommunications*, **13**(2), March 2002, p. 149–158. (cité page 10)

[170] Spriet (A.), Moonen (M.) et Wouters (J.), « Spatially pre-processed speech distortion weighted multi-channel wiener filtering for noise reduction », *Signal Processing*, **84**(12), September 2004, p. 2367–2387. (cité page 48)

[171] Spriet (A.), Moonen (M.) et Wouters (J.), « Robustness analysis of multichannel wiener filtering and generalized sidelobe cancellation for multimicrophone noise reduction in hearing aid applications », *IEEE Transactions on Speech and Audio Processing*, **13**(4), July 2005, p. 487–503. (cité pages 10 et 48)

[172] Spriet (A.), Moonen (M.) et Wouters (J.), « Stochastic gradient-based implementation of spatially preprocessed speech distortion weighted multichannel wiener filtering for noise reduction in hearing aids », *IEEE Transactions on Signal Processing*, **53**(3), March 2005, p. 911–925. (cité page 48)

[173] Sugiyama (A.), Hua (T. P.), Jeannes (R. L. B.) et Faucon (G.), « A comparative study of adaptation-mode control for generalized sidelobe cancellers in human-robot communications », *Hands-Free Speech Communication and Microphone Arrays*, **NC**, May 2008, p. 13–16. (cité page 47)

[174] Tashev (I.). « Statistical modeling of the speech signal ». Dans *International Workshop on Acoustic Echo and Noise Control*, Tel-Aviv, September 2010. (cité page 7)

[175] Thi (H.-L. N.), Gerven (S. V.), Jutten (C.) et Compernolle (D. V.), « Stability study for source separation in convolutive mixtures of two sources », *Signal Processing*, **62**(2), October 1997, p. 163–171. (cité page 68)

[176] Thi (H.-L. N.) et Jutten (C.), « Blind source separation for convolutive mixtures », *Signal Processing*, **45**, 1995, p. 209–229. (cité page 64)

[177] Tonelli (M.) et Davies (M.). « A blind multichannel dereverberation algorithm based on the natural gradient ». Dans *International Workshop on Acoustic Echo and Noise Control*, Tel-Aviv, September 2010. (cité page 31)

[178] Trees (H. L. V.), *Optimum array processing*, volume 4. Wiley-Interscience, 2002. (cité page 16)

[179] Tseng (C.-Y.) et Griffiths (L. J.), « A unified approach to the design of linear constraints in minimum variance adaptive beamformers », *IEEE Transactions on antennas and propagation*, **40**(12), December 1992, p. 1533–1542. (cité page 34)

[180] Tseng (C.-Y.) et Griffiths (L. J.). « Adaptive broadband beamforming using phase-independent derivative constraint techniques ». Dans *Proceedings of the Acoustics, Speech, and Signal Processing*, volume 4, p. 536–539. IEEE Computer Society, 1993. (cité page 35)

[181] van de Kerkhof (L. M.) et Kitzen (W. J. W.), « Tracking of a time-varying acoustic impulse response by an adaptive filter », *IEEE Transactions on Signal Processing*, **40**(6), June 1992, p. 1285–1294. (cité page 88)

[182] Veen (B. D. V.) et Buckley (K. M.), « Beamforming : a versatile approach to spatial filtering », *ASSP Magazine*, **5**(2), April 1988, p. 4–24. (cité page 16)

[183] Vorobyov (S. A.), Gershman (A. B.) et Luo (Z.-Q.), « Robust adaptive beamforming using worst-case performance optimization : a solution to the signal mismatch problem », *IEEE Transactions on Signal Processing*, **51**(2), February 2003, p. 313–324. (cité pages 37 et 39)

[184] Wang (D. L.) et Lim (J. S.). « The unimportance of phase in speech enhancement ». Dans *IEEE Transactions on Acoustics, Speech, and Signal Processing*, volume 30, p. 679–681, August 1982. (cité page 242)

[185] Ward (D. B.), Kennedy (R. A.) et Williamson (R. C.), « Fir filter design for frequency invariant beamformers », *IEEE Signal Processing Letters*, **3**(3), March 1996, p. 69–71. (cité page 26)

[186] Ward (D. B.), Kennedy (R. A.) et Williamson (R. C.). *Microphone Array : Signal Processing Techniques and Applications*, chapitre 1 - Constant Directivity Beamforming, p. 3–17. Springer, 2001. (cité pages 9 et 26)

[187] Ward (D. B.). *Theory and Applications of Broadband Frequency Invariant Beamforming*. Thèse de doctorat, Australian National University, 1996. (cité pages 9 et 26)

[188] Weinstein (E.), Feder (M.) et Oppenheim (A. V.), « Multi-channel signal separation by decorrelation », *IEEE Transactions on Speech and Audio Processing*, **1**(4), October 1993, p. 405–413. (cité pages 46, 62, 75 et 173)

[189] Widrow (B.), Glover (J. R.), Mccool (J. M.), Kaunitz (J.), Williams (C. S.), Hearn (R. H.), Zeidler (J. R.), Dong (E.) et Goodlin (R. C.). « Adaptive noise cancelling : principles and applications ». Dans *Proc. IEEE*, volume 63, p. 1692–1716, August 1975. (cité pages 9 et 56)

[190] Widrow (B.), McCool (J. M.), Larimore (M. G.), Jonhson (C. R.) et JR.. « Stationary and nonstationary learning characteristics of the lms adaptive filter ». Dans *Proceedings of the IEEE*, volume 64, August 1976. (cité page 88)

[191] Widrow (B.) et Streams (S. D.), *Adaptive signal processing*. Prentice Hall, 1985. (cité pages 9 et 87)

[192] Wolff (T.) et Buck (M.). « A generalized view on microphone array postfilters ». Dans *International Workshop on Acoustic Echo and Noise Control*, Tel-Aviv, September 2010. (cité page 18)

[193] Wolff (T.) et Buck (M.). « Influence of blocking matrix design on microphone array postfilters ». Dans *International Workshop on Acoustic Echo and Noise Control*, 2010. (cité pages 19 et 50)

[194] Woo (S.), Lee (S.) et Jeong (H.). « Combined architecture of adaptive beamforming and blind source separation for speech recognition of intelligent service robots ». Dans *IEEE International Conference on Intelligent Pervasive Computing*, p. 214–219, 2007. (cité pages 178 et 179)

[195] Yan (S.) et Ma (Y.), « Robust supergain beamforming for circular array via second-order cone programming », *Applied acoustics*, **66**(9), September 2005, p. 1018–1032. (cité pages 29, 37 et 38)

[196] Yan (S.), Ma (Y.) et Hou (C.), « Optimal array pattern synthesis for broadband arrays », *The Journal of the Acoustical Society of America*, **122**(5), November 2007, p. 2686–2696. (cité page 39)

[197] Yu (T.) et Hansen (J. H.). « Robust auto-focusing wideband bayesian beamforming ». Dans *Hands-Free Speech Communication and Microphone Arrays*, p. 61–64, June 2008. (cité page 36)

[198] Yu (Z. L.) et Er (M. H.). « A robust beamformer with new uncertainty constraint on steering vector ». Dans *IEEE International Conference on Acoustics, Speech and Signal Processing*, p. 1029–1032, May 2006. (cité page 36)

[199] Zelinski (R.). « A microphone array with adaptive post-filtering for noise reduction in reverberant rooms ». Dans *IEEE International Conference on Acoustics, Speech and Signal Processing*, volume 5, p. 2578–2581, April 1988. (cité page 18)

[200] Zhang (M.) et Er (M. H.), « Robust adaptive beamforming for broadband arrays », *Circuits, Systems, and Signal Processing*, **16**(2), February 1997, p. 207–216. (cité page 36)

[201] Zhang (M.) et Lang (H.). « Noise canceler system with adaptive cross-talk filters », March 2004. (cité page 91)

[202] Zinser (R. L.) et Evans (J. B.). « Some experimental and theoretical results using a new adaptive filter structure for noise cancellation in the presence of crosstalk ». Dans *IEEE International Conference on Acoustics, Speech and Signal Processing*, volume 10, p. 1253–1256, April 1985. (cité pages 12, 77, 84, 89, 116 et 203)

Annexes

Annexe A

Generalized Sidelobe Canceller

A.1 GSC généralisé

La description donnée dans cette partie est tirée des articles de Reuven et al. [149, 151] qui s'inscrivent dans la suite des travaux initialement proposés par Gannot et al. [66]. Il s'agit d'obtenir les expressions des différents blocs du GSC lorsque l'on remplace les simples retards pris pour modèle de propagation par l'estimation des Fonction de Transfert Acoustiques (FTA) reliant les sources aux microphones. Il s'agit donc d'une généralisation du GSC donné par [73] où le modèle trop réducteur des simples retards pour la construction du formateur de voie fixe et de la matrice de blocage est remplacé par une prise en compte des réponses impulsionnnelles réelles. Le modèle de signal considéré comprend bien évidemment un signal source localisé non-stationnaire S_1, un signal interférent également localisé et non-stationnaire S_2, et un signal de bruit stationnaire qui peut être localisé, diffus ou non corrélé $\mathbf{B}$. La formulation vectorielle du signal reçu sur les M microphones de l'antenne dans le domaine de la Transformée de Fourier à Court Terme (TFCT) est de la forme :

$$\mathbf{X}(k,l) = \mathbf{A}^{S_1}(k)S_1(k,l) + \mathbf{A}^{S_2}(k)S_2(k,l) + \mathbf{B}(k,l) \tag{A.1}$$

avec l l'indice de trame, k la composante fréquentielle d'indice k et les quantités

- $\mathbf{X}(k,l)$: vecteur des observations ;

$$\mathbf{X}(k,l) = [X_1(k,l)\ X_2(k,l) \cdots X_M(k,l)]$$

- $\mathbf{A^{S_1}}(k)$: vecteur des réponses inpulsionnelles entre le signal source et chacun des micros ;

$$\mathbf{A}^{S_1}(k) = \left[A_1^{S_1}(k)\ A_2^{S_1}(k) \cdots A_M^{S_1}(k)\right]$$

– $\mathbf{A}^{S_2}(k)$: vecteur des réponses inpulsionnelles entre le signal interférent et chacun des micros ;

$$\mathbf{A}^{S_2}(k) = \left[A_1^{S_2}(k)\ A_2^{S_2}(k) \cdots A_M^{S_2}(k)\right]$$

– $\mathbf{B}(k,l)$: vecteur du bruit reçu sur chacun des micros.

$$\mathbf{B}(k,l) = [B_1(k,l)\ B_2(k,l) \cdots B_M(k,l)]$$

L'égalité dans l'équation (A.1) n'est vraie que dans le cas où la longueur de la TFCT est supérieure à la longueur des réponses impulsionnelles acoustiques. Cependant, comme le souligne les auteurs, l'égalité est une approximation raisonnable. Nous considérons un traitement par trame pour lequel nous noterons que les variations des réponses impulsionnelles sont considérées négligeables au regard de l'intervalle d'observation défini par la durée de la fenêtre d'analyse choisie. A partir de l'observation bruitée $\mathbf{X}(k,l)$, il s'agit maintenant d'être capable de reconstruire le signal de parole désiré $S_1(k,l)$ à l'aide du filtrage par le vecteur $\mathbf{W}(k,l)$ contenant les filtres placés sur chacune des voies de l'antenne filtrage-sommation décrite à la Figure A.1. Dans les développements qui suivent, les Réponses Impulsionnelles (RI) ou Fonctions de Transfert Acoustiques (FTA) $\mathbf{A}^{S_1}(k)$ et $\mathbf{A}^{S_2}(k)$ sont supposées connues afin de se concentrer dans un premier temps sur l'obtention de l'algorithme du GSC.

A.1.1 Solution optimale sous contraintes - Frost

Tout d'abord, explicitons le signal en sortie d'antenne après filtrage par le vecteur $\mathbf{W}(k,l) = [W_1^*(k,l)\ W_2^*(k,l) \cdots W_M^*(k,l)]$.

$$\begin{aligned} Y(k,l) &= \mathbf{W}^H(k,l)\mathbf{X}(k,l) \\ &= \mathbf{W}^H(k,l)\mathbf{A^{S_1}}(k)S_1(k,l) + \mathbf{W}^H(k,l)\mathbf{A^{S_2}}(k)S_2(k,l) + \mathbf{W}^H(k,l)\mathbf{B}(k,l) \\ &= Y_{S_1}(k,l) + Y_{S_2}(k,l) + Y_B(k,l) \end{aligned} \tag{A.2}$$

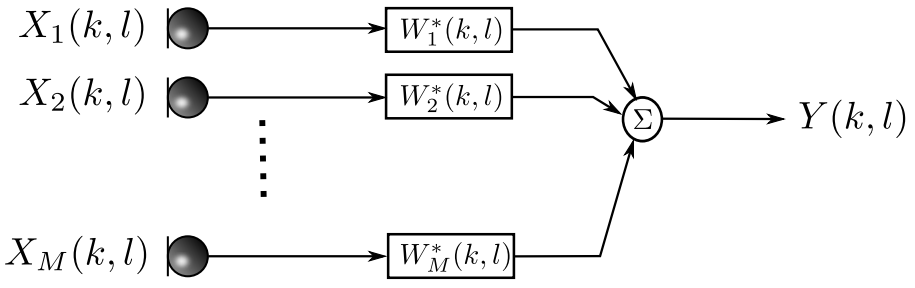


Figure A.1 – Antenne filtrage-sommation.

où Y_{S_1}, Y_{S_2} et Y_B représentent respectivement les versions filtrées du signal désiré, du signal interférent et du bruit stationnaire. La puissance du signal de sortie a pour expression :

$$\begin{aligned} E\left[Y(k,l)Y^*(k,l)\right] &= E\left[\mathbf{W}^H(k,l)\mathbf{X}(k,l)\mathbf{X}^H(k,l)\mathbf{W}(k,l)\right] \\ &= \mathbf{W}^H(k,l)\mathbf{\Phi_{XX}}(k,l)\mathbf{W}(k,l) \end{aligned} \tag{A.3}$$

avec $\mathbf{\Phi_{XX}}(k,l) = E\left[\mathbf{X}(k,l)\mathbf{X}^H(k,l)\right]$. Les filtres optimaux sont obtenus par la minimisation de la puissance en sortie, sous les contraintes de préservation du signal utile et de suppression de l'interférent. Celles-ci s'écrivent selon le formalisme suivant.

$$\begin{aligned} Y_{S_1}(k,l) &= \mathbf{W}^H(k,l)\mathbf{A}^{S_1}(k)S_1(k,l) \\ &= \mathcal{F}^*(k)S_1(k,l) \\ Y_{S_2}(k,l) &= \mathbf{W}^H(k,l)\mathbf{A}^{S_2}(k)S_2(k,l) = 0 \end{aligned} \tag{A.4}$$

$\mathcal{F}(k)$ est la réponse d'un filtre arbitraire choisi qui peut, par exemple, correspondre à un simple retard. Le problème de minimisation à résoudre pour la définition de $\mathbf{W}(k,l)$ est donc

$$\min\left\{\mathbf{W}^H(k,l)\mathbf{\Phi_{XX}}(k,l)\mathbf{W}(k,l)\right\}$$

$$\begin{gathered} \text{s.c.q } \mathbf{W}^H(k,l)\mathbf{A}^{S_1}(k) = \mathcal{F}^*(k) \\ \text{et } \mathbf{W}^H(k,l)\mathbf{A}^{S_2}(k) = 0 \end{gathered}$$

Par application du Lagrangien évalué dans le domaine des complexes et dérivé par rapport à $\mathbf{W}^*$ on obtient :

$$\mathbf{W}(k,l) = -\mathbf{\Phi_{XX}}(k,l)\left[\lambda_1\mathbf{A}^{S_1}(k) + \lambda_2\mathbf{A}^{S_2}(k)\right]$$

où la détermination des multiplicateurs de Lagrange λ_1 et λ_2 se fait en ajoutant les contraintes données en (A.4). Ceci permet d'aboutir à la solution optimale "fixe" suivante :

$$\mathbf{W_{opt}}(k,l) = \mathcal{F}^*(k)\Phi^{-1}_{\mathbf{XX}}(k,l) \times \frac{\frac{\mathbf{A}^{S_1}(k)}{\left\|\mathbf{A}^{S_1}(k)\right\|^2_\Phi} - \rho_\Phi(k)\frac{\mathbf{A}^{S_2}(k)}{\left\|\mathbf{A}^{S_1}(k)\right\|_\Phi\left\|\mathbf{A}^{S_2}(k)\right\|_\Phi}}{1-\left|\rho_\Phi(k)\right|^2} \tag{A.5}$$

Dans cette expression, deux quantités apparaissent. Une norme pondérée $\|\mathbf{Z}(k)\|^2_\Phi \triangleq \mathbf{Z}^H(k)\Phi^{-1}_{\mathbf{XX}}(k,l)\mathbf{Z}(k)$ et un produit scalaire pondéré (A.6) discuté plus loin. De plus, en ce qui concerne l'inversibilité de la matrice des interspectres $\Phi_{\mathbf{XX}}$, celle-ci est garantie par la présence continuelle d'un minimum de bruit décorrélé sur chacun des microphones (bruit électronique des microphones notamment).

$$\rho_\Phi(k) \triangleq \frac{\mathbf{A}^{S_2^H}(k)\Phi^{-1}_{\mathbf{XX}}(k,l)\mathbf{A}^{S_1}(k)}{\sqrt{\mathbf{A}^{S_1^H}(k)\Phi^{-1}_{\mathbf{XX}}(k,l)\mathbf{A}^{S_1}(k)}\sqrt{\mathbf{A}^{S_2^H}(k)\Phi^{-1}_{\mathbf{XX}}(k,l)\mathbf{A}^{S_2}(k)}} \tag{A.6}$$

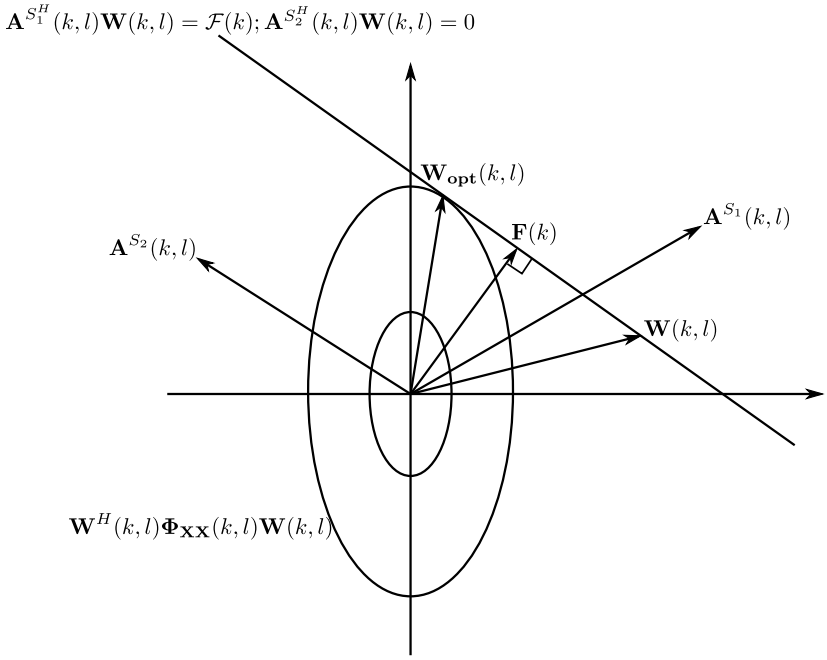


Figure A.2 – Interprétation géométrique du problème de minimisation sous contrainte. Les ellipses décrivent les iso-surfaces de puissance. La droite donne l'hyperplan des contraintes. $\mathbf{W}(k,l)$ est une solution à la trame l alors que $\mathbf{W_{opt}}(k,l)$ est la solution optimale.

Pour obtenir une mise en œuvre adaptative de la solution (A.5) permettant de suivre les variations de l'environnement, il suffit d'appliquer un algorithme de plus grande pente sur le Lagrangien associé au problème d'optimisation précédent.

$$\begin{aligned}\mathbf{W}(k,l+1) &= \mathbf{W}(k,l) - \mu\nabla_{\mathbf{W}^*}\mathcal{L}(k) \\ &= \mathbf{W}(k,l) - \mu\left[\Phi^{-1}_{\mathbf{XX}}(k,l)\mathbf{W}(k,l) + \lambda_1\mathbf{A}^{S_1}(k) + \lambda_2\mathbf{A}^{S_2}(k)\right] \quad \text{(A.7)}\end{aligned}$$

Les expressions des multiplicateurs de Lagrange de (A.7) en fonction des données du problème sont calculées à partir des contraintes, ce qui conduit à l'équation de mise à jour des filtres (A.8).

$$\mathbf{W}(k,l+1) = P(k)\mathbf{W}(k,l) - \mu P(k)\Phi_{\mathbf{XX}}(k,l)\mathbf{W}(k,l) + \mathbf{F}(k) \quad \text{(A.8)}$$

Les expressions de P et $\mathbf{F}$ peuvent être écrites en faisant apparaître cette fois un produit scalaire non pondéré entre les vecteurs des réponses impulsionnelles (FTA) :

$$\rho(k) = \frac{\mathbf{A}^{S_2^H}(k)\mathbf{A}^{S_1}(k)}{\|\mathbf{A}^{S_1}(k)\|\,\|\mathbf{A}^{S_2}(k)\|} \quad \text{(A.9)}$$

$$\begin{aligned}P(k) &= I - \frac{1}{1-|\rho(k)|^2} \times \left[\frac{\mathbf{A}^{S_1}(k)\mathbf{A}^{S_1^H}(k)}{\|\mathbf{A}^{S_1}(k)\|^2} - \rho^*(k)\frac{\mathbf{A}^{S_1}(k)\mathbf{A}^{S_2^H}(k)}{\|\mathbf{A}^{S_1}(k)\|\,\|\mathbf{A}^{S_2}(k)\|}\right. \\ &- \left.\rho(k)\frac{\mathbf{A}^{S_2}(k)\mathbf{A}^{S_1^H}(k)}{\|\mathbf{A}^{S_1}(k)\|\,\|\mathbf{A}^{S_2}(k)\|} + \frac{\mathbf{A}^{S_2}(k)\mathbf{A}^{S_2^H}(k)}{\|\mathbf{A}^{S_2}(k)\|^2}\right] \quad \text{(A.10)}\end{aligned}$$

$$\mathbf{F}(k) = \frac{\frac{\mathbf{A}^{S_1}(k)}{\left\|\mathbf{A}^{S_1}(k)\right\|^2} - \rho(k)\frac{\mathbf{A}^{S_2}(k)}{\left\|\mathbf{A}^{S_1}(k)\right\|\left\|\mathbf{A}^{S_2}(k)\right\|}}{1-\left|\rho(k)\right|^2}\mathcal{F}(k) \tag{A.11}$$

L'interprétation géométrique de cette solution est donnée à la Figure A.2. Le terme $\mathbf{F}$ représente la composante non adaptative de la solution. Nous justifierons dans la partie associée au GSC le fait que celle-ci est orthogonale au plan des contraintes et prise pour l'initialisation de l'agorithme adaptatif. La solution optimale est donc atteinte au point où l'hyperplan des contraintes est tangent aux contours d'équipuissance. Pour cette valeur, on vérifie $\mathbf{W}_{opt}^H(k,l)\mathbf{A}^{S_2}(k) = 0$ et $\mathbf{W}_{opt}^H(k,l)\mathbf{A}^{S_1}(k) = \mathcal{F}^*(k)$.
Par ailleurs, nous constatons que, dans sa version adaptative, le produit scalaire devient indépendant de la puissance des observations. En d'autres termes, ce sont les caractéristiques liées à la position des sources dans l'environnement de propagation qui sont importantes et ce indépendamment des propriétés statistiques du signal source. De plus, l'apparition d'un produit scalaire dans l'expression du filtre optimal est à relier directement avec la résolution d'une antenne acoustique ou plus simplement sa capacité à discriminer spatialement deux sources selon leurs proximités. Nous pouvons en ce sens faire référence aux travaux de Cox [41] sur la notion de cône de confusion, qui montre qu'en dessous d'un certain angle d'ouverture du cône, il devient impossible de séparer deux sources. A contrario, si les vecteurs $\mathbf{A}^{S_1}(k)$ et $\mathbf{A}^{S_2}(k)$ sont orthogonaux, ce qui en pratique signifie que les sources S_1 et S_2 sont situées à des endroits distincts et distants, alors $\rho(k) = 0$. Dans ce cas, le filtre $\mathbf{W}_0 = \mathbf{F}$ correspond au cas où nous traiterions qu'une seule source, simplifiant ainsi les expressions (A.10) et (A.11). Ceci correspond géométriquement au cas de la Figure A.3.

Le développement présenté dans cette partie relate l'obtention de la solution avec des contraintes algorithmiques permettant de définir la réponse spatio-fréquentielle de l'antenne. Dans la partie qui suit nous allons exposer la manière de reporter ces contraintes au niveau matériel et passer à un problème d'optimisation non-contraint.

A.1.2 Solution optimale non contrainte - GSC

Pour obtenir cette solution qui définit l'algorithme et la structure du GSC, la théorie s'appuie sur le théorème de l'algèbre linéaire. Considérons les définitions suivantes :

– $\mathcal{N}(k)$ l'espace nul ou noyau de $\left[\mathbf{A}^{S_1}(k)|\mathbf{A}^{S_2}(k)\right]$, tel que

$$\mathcal{N}(k) \triangleq \left\{\mathbf{W} \mid \left[\mathbf{A}^{S_1}(k)|\mathbf{A}^{S_2}(k)\right]^H \mathbf{W}(k) = [0|0]\right\} \tag{A.12}$$

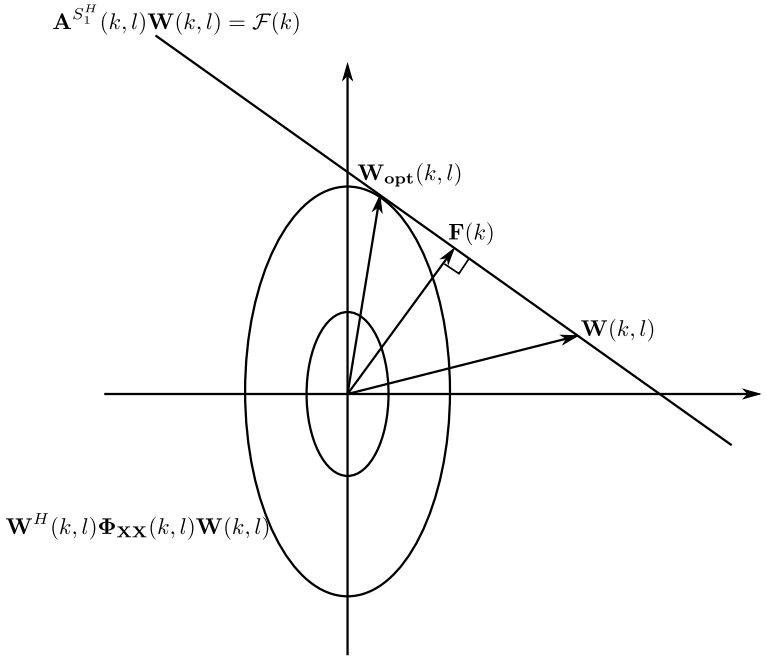


Figure A.3 – Interprétation géométrique du problème de minimisation sous contrainte. Cas où les FTA sont orthogonales.

– $\Lambda(k)$ l'hyperplan des contraintes parallèle à $\mathcal{N}(k)$

$$\Lambda(k) \triangleq \left\{ \mathbf{W} \mid \left[\mathbf{A}^{S_1}(k) | \mathbf{A}^{S_2}(k) \right]^H \mathbf{W}(k) = [\mathcal{F}((k))|0] \right\} \tag{A.13}$$

– $\mathcal{R}(k)$ l'espace des colonnes ou image de $\left[\mathbf{A}^{S_1}(k) | \mathbf{A}^{S_2}(k) \right]$, tel que

$$\mathcal{R}(k) \triangleq \left\{ \kappa_1 \mathbf{A}^{S_1}(k) + \kappa_2 \mathbf{A}^{S_2}(k) \mid \forall (\kappa_1, \kappa_2) \in \mathbb{R}^* \right\} \tag{A.14}$$

D'après le théorème fondamental de l'algèbre linéaire nous savons que $\mathcal{R}(k) \perp \mathcal{N}(k)$. Or, de manière évidente $\mathbf{F}(k) \in \mathcal{R}(k)$. Par ailleurs, il est simple de montrer par le calcul que $\mathbf{A}^{S_1^H}(k)\mathbf{F}(k) = \mathcal{F}(k)$ et $\mathbf{A}^{S_2^H}(k)\mathbf{F}(k) = 0$. Par conséquent, $\mathbf{F}(k) \in \Lambda(k)$. Or, $\mathbf{F}(k) \perp \mathcal{N}(k)$ et $\mathcal{N}(k) \parallel \Lambda(k)$ d'où $\mathbf{F}(k) \perp \Lambda(k)$. Donc $\mathbf{F}(k)$ est la perpendiculaire de l'origine à l'hyperplan des contraintes. Par le calcul, nous pouvons montrer que $P(k)\mathbf{A}^{S_1}(k) = 0$ et $P(k)\mathbf{A}^{S_2}(k) = 0$. Ce qui signifie que $P(k)$ correspond à la matrice de projection orthogonale sur l'espace nul de $\left[\mathbf{A}^{S_1}(k) | \mathbf{A}^{S_2}(k) \right]$. Grâce à cette analyse sur les espaces nous constatons qu'il est possible de décomposer le vecteur contenant les filtres de formations de voies en deux vecteurs de la manière suivante :

$$\mathbf{W}(k,l) = \mathbf{W}_0(k,l) - \mathbf{V}(k,l) \tag{A.15}$$

où $\mathbf{W}_0(k,l) \in \mathcal{R}(k)$ et $-\mathbf{V}(k,l) \in \mathcal{N}(k)$. Par définition de $\mathcal{N}(k)$, $\mathbf{V}(k,l)$ peut être écrit tel que :

$$\mathbf{V}(k,l) = \mathcal{H}(k)\mathbf{G}(k,l) \tag{A.16}$$

avec $\mathcal{H}(k)$ une matrice de rang plein de dimension $M \times (M-2)$ telle que les colonnes de $\mathcal{H}(k)$ couvrent l'espace nul de $\left[\mathbf{A}^{S_1}(k)|\mathbf{A}^{S_2}(k)\right]$, c'est-à-dire

$$\begin{aligned} \mathbf{A}^{S_1^H}(k,l)\mathcal{H}(k) &= 0 \\ \mathbf{A}^{S_2^H}(k,l)\mathcal{H}(k) &= 0 \end{aligned} \tag{A.17}$$

Cette matrice $\mathcal{H}(k)$ est appelée matrice de blocage. Le vecteur $\mathbf{G}(k,l)$ contient, quant à lui, $M-2$ filtres adaptatifs. De l'interprétation géométrique de l'algorithme de Frost donnée en A.1.1, nous en déduisons la valeur de la première composante de $\mathbf{W}$.

$$\mathbf{W}_0(k,l) = \mathbf{F}(k) = \frac{\frac{\mathbf{A}^{S_1}(k)}{\left\|\mathbf{A}^{S_1}(k)\right\|^2} - \rho(k)\frac{\mathbf{A}^{S_2}(k)}{\left\|\mathbf{A}^{S_1}(k)\right\|\left\|\mathbf{A}^{S_2}(k)\right\|}}{1-|\rho(k)|^2}\mathcal{F}(k) \tag{A.18}$$

Ainsi, à partir des équations (A.2), (A.15) et (A.16), le signal en sortie d'antenne s'écrit :

$$\begin{aligned} Y(k,l) &= Y_{FBF}(k,l) - Y_{ANC}(k,l) \\ &= \mathbf{W}_0(k,l)\mathbf{X}(k,l) - \mathbf{G}^H(k,l)\mathcal{H}(k)\mathbf{X}(k,l) \end{aligned} \tag{A.19}$$

La sortie du GSC est donc la différence de deux termes. Le rôle de $\mathbf{W}_0(k,l)$ est de pointer l'antenne dans la direction de S_1 tout en assurant un zéro dans la direction de l'interférent S_2. Ce premier bloc est appelé *Matched Beamformer*. Il est complètement défini dès lors que l'on connaît les FTA. Il dépend donc des propriétés du signal, comme expliqué ultérieurement, et ne possède plus de caractère adaptatif. Pour la matrice $\mathcal{H}(k)$, il s'agit de bloquer les deux sources localisées (reliées aux microphones par des FTA) tout en laissant passer le bruit (stationnaire ou non), afin de créer la référence de bruit. Sous forme mathématique, cela revient à dire que :

$$\mathbf{U}(k,l) = \mathcal{H}^H(k)\mathbf{X}(k,l) = \mathcal{H}(k)\mathbf{B}(k,l) \tag{A.20}$$

Nous nous sommes donc ramené à un problème d'annulation de bruit multicanal standard qui peut être résolu à l'aide d'un filtrage de Wiener, dont la solution adaptative donnée par Widrow au travers l'algorithme des moindres carrés (LMS) permet d'ajuster correctement les filtres $\mathbf{G}(k,l)$. Le terme "correctement" signifie que ces filtres doivent créer une image du bruit contenu dans le signal utile extrait par $\mathbf{W}_0(k,l)\mathbf{X}(k,l)$ afin de le soustraire à celui-ci. Les équations (A.19) et (A.20) identifient les grandeurs constituant les différents blocs du GSC, à savoir : $\mathbf{W}_0$ le formateur de voies fixe, $\mathcal{H}$ la matrice de blocage et $\mathbf{G}$ la matrice des filtres adaptatifs de l'annuleur de bruit.

A.1.3 Construction de la matrice de blocage

Pour ce faire, les auteurs supposent que la matrice $\mathcal{H}(k)$ correspond à mise en cascade de deux matrices de blocage, $\mathcal{H}(k) = \mathcal{H}_1(k)\mathcal{H}_2(k)$. $\mathcal{H}_1(k)$ est conçue de manière à bloquer les composantes du signal utile, pendant que $\mathcal{H}_2(k)$ bloque les composantes du signal interférent après rotation par la matrice $\mathcal{H}_1(k)$. Un choix possible consiste à prendre comme valeurs celles décrites ci-après.

$$\mathcal{H}_1(k) = \begin{bmatrix} \frac{-A_2^{S_1^*}(k)}{A_1^{S_1^*}(k)} & \frac{-A_3^{S_1^*}(k)}{A_1^{S_1^*}(k)} & \dots & \frac{-A_M^{S_1^*}(k)}{A_1^{S_1^*}(k)} \\ 1 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \\ 0 & 0 & \dots & 1 \end{bmatrix}$$

Cette matrice $\mathcal{H}_1(k)$ permet effectivement de bloquer le signal utile puisque nous pouvons vérifier que $\mathbf{A}^{\mathbf{S_1^*}}(k)\mathcal{H}_1(k) = 0$.

$$\begin{aligned} \mathbf{A}^{S_2^H}(k)\mathcal{H}(k) &= \mathbf{A}^{S_2^H}(k)(\mathcal{H}_1(k)\mathcal{H}_2(k)) = (\mathbf{A}^{S_2^H}(k)\mathcal{H}_1(k))\mathcal{H}_2(k) \\ &= \left[A_1^{S_2^*}(k)A_2^{S_2^*}(k)\dots A_M^{S_2^*}(k)\right] \bullet \begin{bmatrix} \frac{-A_2^{S_1^*}(k)}{A_1^{S_1^*}(k)} & \frac{-A_3^{S_1^*}(k)}{A_1^{S_1^*}(k)} & \dots & \frac{-A_M^{S_1^*}(k)}{A_1^{S^*(k)}} \\ 1 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \\ 0 & 0 & \dots & 1 \end{bmatrix} \\ &= \begin{bmatrix} -A_1^{S_2^*}(k)\frac{A_2^{S_1^*}(k)}{A_1^{S_1^*}(k)} + A_2^{S_2^*}(k) \\ -A_1^{S_2^*}(k)\frac{A_3^{S_1^*}(k)}{A_1^{S_1^*}(k)} + A_3^{S_2^*}(k) \\ \vdots \\ -A_1^{S_2^*}(k)\frac{A_M^{S_1^*}(k)}{A_1^{S_1^*}(k)} + A_M^{S_2^*}(k) \end{bmatrix} \end{aligned} \tag{A.21}$$

En accord avec le rôle attendu de la matrice $\mathcal{H}(k)$, la multiplication par $\mathcal{H}_2(k)$ dans l'équation (A.21) doit donner le vecteur nul. Ainsi, en résolvant ce problème pour $m = 3, \dots, M$ avec $\mathcal{H}_2(k)$ de la forme

$$\mathcal{H}_2(k) = \begin{bmatrix} L_3(k) & L_4(k) & \dots & L_M(k) \\ 1 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \\ 0 & 0 & \dots & 1 \end{bmatrix}$$

nous en déduisons que

$$L_m(k) = -\frac{\frac{A_m^{S_1^*}(k)}{A_1^{S_1^*}(k)} - \frac{A_m^{S_2^*}(k)}{A_1^{S_2^*}(k)}}{\frac{A_2^{S_1^*}(k)}{A_1^{S_1^*}(k)} - \frac{A_2^{S_2^*}(k)}{A_1^{S_2^*}(k)}} \tag{A.22}$$

Or, $\mathcal{H}(k)$ est définie comme la mise en cascade de $\mathcal{H}_1(k)$ et $\mathcal{H}_2(k)$. Ainsi, après réarrangements des termes, la structure de la matrice de blocage estimée à chaque trame est donnée par :

$$\mathcal{H}(k) = \begin{bmatrix} Q_3(k) & Q_4(k) & \dots & Q_M(k) \\ L_3(k) & L_4(k) & \dots & L_M(k) \\ 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \\ 0 & 0 & \dots & 1 \end{bmatrix}$$

avec

$$Q_m(e^{j\omega}) = -\frac{\frac{A_2^{S_1^*(k)}}{A_1^{S_1^*(k)}} \cdot \frac{A_m^{S_2^*(k)}}{A_1^{S_2^*(k)}} - \frac{A_2^{S_2^*(k)}}{A_1^{S_2^*(k)}} \cdot \frac{A_m^{S_1^*(k)}}{A_1^{S_1^*(k)}}}{\frac{A_2^{S_1^*(k)}}{A_1^{S_1^*(k)}} - \frac{A_2^{S_2^*(k)}}{A_1^{S_2^*(k)}}} \tag{A.23}$$

La matrice de blocage ainsi définie vérifie les contraintes imposées. En supposant que l'estimation des RI est parfaite, elle doit donc nous permettre de créer une référence de bruit optimale. Or, en analysant l'expression de $\mathbf{U}$ donnée en (A.20) sur une voie m choisie, nous constatons que la connaissance des Fonctions de Transfert Relatives (FTR) $\frac{A_m^{S_1^*}(k)}{A_1^{S_1^*}(k)}$ et $\frac{A_m^{S_2^*}}{A_1^{S_2^*}(k)}$ suffit à l'obtention du bruit perturbant le signal. En effet,

$$U_m(k,l) = Q_m(k)X_1(k,l) + L_m(k)X_2(k,l) + X_m(k,l) \tag{A.24}$$

Cela signifie, qu'en disposant des estimées des FTR, nous sommes capables de connaître le bruit stationnaire contenu dans le signal et donc de le soustraire à la source estimée par le formateur de voies fixe.

A.1.4 Pointage de la source - *Matched Beamformer*

Nous avons vu que le filtre $\mathbf{W}_0$ permettait de réaliser un pointage optimal de la source utile, même dans un milieu réverbérant, grâce à l'expression construite à partir des FTA du milieu. En utilisant les FTR plutôt que les FTA dans (A.18), il est possible de réécrire l'équation du filtre comme ceci :

$$\mathbf{W_0}(k) = \frac{\frac{\left\|\mathbf{A}^{S_2}(k)\right\|^2}{\left|A_1^{S_2}(k)\right|}\frac{\mathbf{A}^{S_1}(k)}{A_1^{S_1}(k)} - \frac{\mathbf{A}^{S_2}(k)}{A_1^{S_2}(k)}\frac{\mathbf{A}^{S_2^H}(k)}{A_1^{S_2}(k)}\frac{\mathbf{A}^{S_1}(k)}{A_1^{S_1}(k)}}{\frac{\left\|\mathbf{A}^{S_1}(k)\right\|^2}{\left|A_1^{S_1}(k)\right|^2}\frac{\left\|\mathbf{A}^{S_2}(k)\right\|^2}{\left|A_1^{S_2}(k)\right|^2} - \frac{\mathbf{A}^{S_1^H}(k)}{A_1^{S_1}(k)}\frac{\mathbf{A}^{S_2}(k)}{A_1^{S_2}(k)}\frac{\mathbf{A}^{S_2^H}(k)}{A_1^{S_2}(k)}\frac{\mathbf{A}^{S_1}(k)}{A_1^{S_1}(k)}}\mathcal{F}(k) \tag{A.25}$$

Cette nouvelle expression assure toujours le blocage du signal interférent et introduit une distorsion du signal due à $A_1^{S_1}(k)$. Ceci est rendu visible en calculant $\mathbf{W}_0\mathbf{A}^{S_1}(k) = A_1^{S_1}(k)\mathcal{F}^*(k)$.

A.1.5 Annuleur de bruit multicanal

Le calcul de $\mathbf{G}(k,l)$ se fait selon l'algorithme des moindres carrés classiques (LMS) pour lequel l'objectif est de minimiser l'Erreur Quadratique Moyenne (EQM) définie par :

$$E\left[\left|Y_{FBF}(k,l) - \mathbf{G}^H(k,l)\mathbf{U}(k,l)\right|^2\right] \tag{A.26}$$

Posons $\phi_{\mathbf{U}Y}(k,l) = E\left[\mathbf{U}(k,l)Y^*(k,l)\right]$ et $\phi_{\mathbf{UU}}(k,l) = E\left[\mathbf{U}(k,l)\mathbf{U}^H(k,l)\right]$, la solution de Wiener multicanale donne

$$\mathbf{G}(k,l) = \phi_{\mathbf{UU}}^{-1}(k,l)\phi_{\mathbf{U}Y}(k,l) \tag{A.27}$$

Afin de suivre en permanence l'évolution des statistiques du bruit due aux variations de l'environnement, une mise en oeuvre adaptative à l'aide de l'algorithme LMS fréquentiel est

$$Y(k,l) = Y_{FBF}(k,l) - \mathbf{G}^H(k,l)\mathbf{U}(k,l) \tag{A.28}$$

L'application du principe d'orthogonalité (A.29) permet d'aboutir à l'équation de mise à jour (A.30).

$$E\left[\mathbf{U}(k,l)Y^*(k,l)\right] = 0 \tag{A.29}$$

$$\mathbf{G}(k,l+1) = \mathbf{G}(k,l) + \mu\mathbf{U}(k,l)Y^*(k,l) \tag{A.30}$$

Pour des raisons de stabilité et d'indépendance par rapport à la puissance des signaux reçus, la version normalisée (NLMS) est préférentiellement utilisée. Celle-ci est donnée ci-après.

$$\mathbf{G}_m(k,l+1) = \mathbf{G}_m(k,l) + \mu\frac{U_m(k,l)Y^*(k,l)}{P_{est}(k,l)} \quad m = 2,\cdots,M \tag{A.31}$$

avec $P_{est}(k,l) = \eta P_{est}(k,l-1) + (1-\eta)\sum_m |U_m(k,l)|^2$ est l'estimée de la Densité Spectrale de Puissance (DSP) du bruit. En général, les valeurs choisies pour le facteur d'oubli sont telles que $0.8 < \eta < 1$. Une DAV est souvent utilisée pour se prémunir des fuites du signal utile qui biaise l'estimée du bruit. En effet, $\mathbf{G}(k,l)$ doit être mis

à jour que lorsque le signal cible est inactif.

Les étapes de l'algorithme associé à cette version du GSC appelée *DTF-GSC* sont décrites ci-après.

1. Extraction du signal utile - Matched Beamformer

$$Y_{FBF}(k,l) = \mathbf{W}_0^H(k)\mathbf{X}(k)$$

2. Calcul de la référence de bruit

$$\mathbf{U}(k,l) = \mathcal{H}^H(k)\mathbf{X}(k)$$

3. Calcul de la sortie

$$Y(k,l) = Y_{FBF}(k,l) - \mathbf{G}^H(k)\mathbf{U}(k,l)$$

4. Mise à jour des filtres pour $m = 3, \cdots, M$

$$\widetilde{G}_m(k,l+1) = G_m(k,l) + \mu\frac{\mathbf{U}(k,l)Y^*(k,l)}{P_{est}(k,l)}$$
$$G_m(k,l+1) \overset{RIF}{\leftarrow} \widetilde{G}_m(k,l+1)$$

   L'opération $\overset{RIF}{\leftarrow}$ fait référence à la méthode de conversion du filtre $\widetilde{G}_m$ en un filtre RIF adéquat. Ceci se fait en trois étapes. Tout d'abord, le filtre est ramené dans le domaine temporel par transformée de Fourier inverse. Puis, par une symétrie centrale sur les échantillons obtenus, la structure temporelle du filtre est extraite et tronquée de manière symétrique. Enfin, le filtre RIF obtenu est ramené dans le domaine fréquentiel par transformée de Fourier.

5. Conservation des échantillons non entachés d'erreurs dues au repliement spectral.

Les simples retards pour la synchronisation des signaux ont donc été remplacés par les FTA, ou plutôt les FTR. Ces dernières sont finalement un bon compromis car en réalité il est difficile d'estimer les réponses impulsionnelles acoustiques qui peuvent être très longues. Les FTR, nécessairement plus courtes, seront en ce sens une bon compromis entre coût de calcul et modélisation.

A.2 Estimation des Fonctions de Transfert FTA et FTR

A.2.1 Méthode du sous-espace signal

Affes et Grenier ont démontré qu'il était possible d'estimer les FTA en recherchant le sous-espace signal [6, 7]. La clé de la méthode tient dans le constat que

l'énergie totale contenue dans la réponse impulsionnelle reliant une source aux microphones est quasiment constante sur toutes les fréquences et ce quelle que soit la position d'un locuteur situé proche d'un microphone de référence. Ainsi, en supposant que le signal est perturbé par un bruit spatiallement blanc de matrice de covariance $\sigma_B^2\mathbf{I}$, les auteurs proposent de calculer la décomposition en valeurs propres de la matrice de covariance des observations. Les FTA seront ensuite estimées en identifiant les lignes associées aux valeurs propres maximales telles que $\lambda_{max} > \sigma_B^2$, soit donc le sous-espace signal. Le cas plus général est décrit dans [68].

A.2.2 Méthode utilisant la non-stationnarité du signal

L'estimation des FTR $\frac{\mathbf{A}^{S_1}(k)}{A_1^{S_1}(k)}$ et $\frac{\mathbf{A}^{S_2(k)}}{A_1^{S_2}(k)}$ se fait à partir des observations $X_m(k,l)$, c'est-à-dire à partir du signal bruité. La méthode proposée [66, 151, 162] s'appuie sur deux hypothèses.

- $(\mathcal{H}_1)$: les FTR varient peu au cours du temps comparé aux variations temporelles des signaux utiles et interférents ;
- $(\mathcal{H}_2)$: les statistiques du bruit varient faiblement relativement aux variations des statistiques des signaux utile et interférents.

Pour le calcul de $\mathbf{W}_0$ défini en (1.20), la fenêtre d'analyse utilisée pour le traitement par TFCT va être subdivisée en I sous trames pour lesquelles la stationnarité des sources utile et interférentes sera supposée vérifiée. Par ailleurs, l'objectif du *Matched Beamformer* étant de sélectionner la seule source cible parmis N, l'estimation des FTR relatives à cette source doit être précise et surtout à l'image de la propagation de cette seule source dans l'environnement. En d'autres termes, l'estimation des FTR relative à la source utile ne doit pas être faite lorsqu'un autre signal non-stationnaire (ici, la composante interférente) est actif. Posons $\widetilde{A}_m \triangleq \frac{A_m^{S_1}(k)}{A_1^{S_1}(k)}$. En supposant la composante interférente inactive, il vient :

$$X_m(k,l) = A_m^{S_1}(k)S_1(k,l) + B_m(k,l) \quad m = 1,\cdots,M \tag{A.32}$$

$$U_m(k,l) = X_m(k,l) - \frac{A_m(k)}{A_1^{S_1}(k)}X_1(k,l) \quad m = 2,\cdots,M \tag{A.33}$$

Avec (A.32) on tire $X_m(k,l) = \widetilde{A}_m(k)X_1(k,l) + U_m(k,l)$. Ainsi, pour la i^{ieme} trame, nous pouvons écrire :

$$\Phi^{(i)}_{X_mX_1}(k) = H_m(k)\Phi^{(i)}_{X_1X_1}(k) + \Phi_{U_mX_1}(k) \quad i = 1,\cdots,I \tag{A.34}$$

Or, d'après (A.32) et (A.33) nous avons

$$X_1(k,l) = A_1^{S_1}(k)S_1(k,l) + B_1(k,l) \quad (A.35)$$

$$U_m(k,l) = B_m(k,l) - \widetilde{A}_m(k)B_1(k,l) \quad (A.36)$$

Puisque B_m est supposé stationnaire et indépendant de S_1, alors $\Phi_{U_mX_1}$ est indépendant de l'indice de trame i. De ce fait, en s'appuyant sur (A.34), l'estimée des densités interspectrales de puissance à mettre en œuvre dans l'algorithme est

$$\widehat{\Phi}^{(i)}_{X_mX_1}(k) = \widetilde{A}_m(k)\widehat{\Phi}^{(i)}_{X_1X_1}(k) + \Phi_{U_mX_1}(k) + \epsilon_m^{(k)}(k) \quad (A.37)$$

où $\epsilon_m^{(i)}(k) = \widehat{\Phi}^{(i)}_{U_mX_1}(k) - \Phi_{U_mX_1}(k)$ est l'erreur d'estimation sur les interspectres entre l'observation bruitée et la $m^{\text{ième}}$ estimée du bruit à la $i^{\text{ième}}$ trame. Sur l'ensemble des I trames de la fenêtre d'analyse, on obtient donc le système d'équation suivant :

$$\begin{bmatrix} \widehat{\Phi}^{(1)}_{X_mX_1}(k) \\ \widehat{\Phi}^{(2)}_{X_mX_1}(k) \\ \vdots \\ \widehat{\Phi}^{(I)}_{X_mX_1}(k) \end{bmatrix} = \begin{bmatrix} \widehat{\Phi}^{(1)}_{X_1X_1}(k) & 1 \\ \widehat{\Phi}^{(2)}_{X_1X_1}(k) & 1 \\ \vdots & \vdots \\ \widehat{\Phi}^{(I)}_{X_1X_1}(k) & 1 \end{bmatrix} \times \begin{bmatrix} \widetilde{A}_m(k) \\ \Phi_{U_mZ_1}(k) \end{bmatrix} + \begin{bmatrix} \widehat{\epsilon}_m^{(1)}(k) \\ \widehat{\epsilon}_m^{(2)}(k) \\ \vdots \\ \widehat{\epsilon}_m^{(I)}(k) \end{bmatrix} \quad (A.38)$$

Une estimée non biaisée de $\widetilde{A}_m$ peut donc être obtenue en résolvant ce système par un critère de moindres carrés [162]. La détermination de $\frac{\mathbf{A}^{S_2}(k)}{A_1^{S_2}(k)}$ pourra être faite de la même manière. Considérons à présent les expressions de L_m et Q_m données respectivement en (A.22) et (A.23). D'après ce qui vient d'être dit sur la manière d'obtenir une estimée des FTR, nous pouvons directement obtenir les valeurs de la matrice de blocage en substituant les quantités correspondantes. Cependant, une autre méthode peut aussi être envisagée. En effet, Reuven et al. [151] décrivent également un moyen d'estimer cette matrice de blocage durant les instants de double parole (source utile et source interférente activent aux mêmes instants). Cette technique ne vaut que pour l'obtention de $\mathcal{H}(k)$ puisque son but est de bloquer les sources ponctuelles non localisées.

$$X_m(k,l) = -Q_m(k)X_1(k,l) - L_m(k)X_2(k,l) + U_m(k,l) \quad (A.39)$$

En découpant de nouveau la fenêtre d'analyse en I trames, la densité interspectrale de puissance théorique à la $i^{\text{ième}}$ trame vaut :

$$\Phi^{(i)}_{X_mX_1}(k) = -Q_m(k)\Phi^{(i)}_{X_1X_1}(k) - L_m(k)\Phi_{X_2X_1}(k) + \Phi_{U_mX_1}(k) \quad i = 1, \cdots, I \quad (A.40)$$

Les estimées correspondantes sont donc :

$$\widehat{\Phi}^{(i)}_{X_mX_1}(k) = -Q_m(k)\widehat{\Phi}^{(i)}_{X_1X_1}(k) - L_m(k)\widehat{\Phi}_{X_2X_1}(k) + \widehat{\Phi}_{U_mX_1}(k) + \widehat{\epsilon}_m^{(i)}(k) \quad i = 1, \cdots, I \quad (A.41)$$

Ainsi, selon la même démarche que précédemment, on obtiendra une estimée directe de Q_m et L_m en résolvant le système d'équations qui suit pour $m = 3, \cdots, M$.

$$\begin{bmatrix} \widehat{\Phi}^{(1)}_{X_m X_1}(k) \\ \widehat{\Phi}^{(2)}_{X_m X_1}(k) \\ \vdots \\ \widehat{\Phi}^{(I)}_{X_m X_1}(k) \end{bmatrix} = \begin{bmatrix} \widehat{\Phi}^{(1)}_{X_1 X_1}(k) & \widehat{\Phi}^{(1)}_{X_2 X_1}(k) & 1 \\ \widehat{\Phi}^{(2)}_{X_1 X_1}(k) & \widehat{\Phi}^{(2)}_{X_2 X_1}(k) & 1 \\ \vdots & \vdots & \vdots \\ \widehat{\Phi}^{(I)}_{X_1 X_1}(k) & \widehat{\Phi}^{(I)}_{X_2 X_1}(k) & 1 \end{bmatrix} \times \begin{bmatrix} -Q_m(k) \\ -L_m(k) \\ \Phi_{U_m X_1}(k) \end{bmatrix} + \begin{bmatrix} \widehat{\epsilon}^{(1)}_m(k) \\ \widehat{\epsilon}^{(2)}_m(k) \\ \vdots \\ \widehat{\epsilon}^{(I)}_m(k) \end{bmatrix} \tag{A.42}$$

Il est particulièrement important de remarquer que, dans cette méthode d'estimation, il est impératif de disposer d'une DAV efficace afin de garantir une parfaite sélection de la source utile et une parfaite suppression de la composante interférente. Or, en considérant les erreurs d'estimations, les approximations faites notamment au travers les hypothèses et la faible efficacité des DAV dans un contexte très bruité, nous sommes à même de voir que la fuite du signal de parole ne peut être résolue complètement par cette méthode. Un contexte typique où il sera difficile de faire la distinction entre la source utile et la source interférente est une prise de son réalisée dans des environnements tels qu'un restaurant ou une foule qui se caractérisent par un bruit non-stationnaire qui peut varier rapidement. Enfin, dans [68] les auteurs proposent une amélioration afin de corriger le biais de la méthode des moindres carrés proposées par Shalvi et Weinstein [162]. Celle-ci consiste à prendre en compte la probabilité de présence du signal afin d'estimer $\widetilde{A}_m(k)$ et $\Phi_{U_m X_1}(k)$ aux moments opportuns.

Annexe B

Réduction de bruit par atténuation spectrale à court terme

B.1 Introduction **242**
B.2 Estimation du bruit et calcul de RSB **242**
B.2.1 Estimation du bruit avec DAV 243
B.2.2 Estimation continue 243
B.2.3 Calcul du RSB par l'approche décision-dirigée (DD) 244
B.3 Règles d'atténuation spectrale **246**
B.3.1 Approches non statistiques 246
B.3.2 Approches statistiques 248

L'atténuation spectrale vise à estimer le spectre à court terme du signal de parole $\widehat{S(k,l)}$, où l est l'indice de trame courante et k l'indice du bin de fréquence, sous l'hypothèse qu'il est toujours possible d'estimer la Densité Spectrale de Puissance (DSP) du bruit, soit pendant les périodes de silence ou de manière continue. En l'absence de connaissance *a priori* sur la DSP du bruit, il est alors impossible de concevoir une technique de réduction de bruit monovoie. Ce que nous allons présenter dans cette partie est une synthèse des travaux de thèse de Cyril Plapous [141]. En ce qui concerne la phase, celle-ci est directement réutilisée. La phase du signal de parole est supposée être identique à celle du signal bruité reçu sur le micro. Cette hypothèse s'appuie sur les caractéristiques physiologiques de l'oreille humaine, très peu sensible aux variations de la phase. Cependant, nous noterons que cela reste vrai sauf pour les basses fréquences et en deça d'un RSB de 5 dB [184] où les variations de celle-ci deviennent perceptibles. Par conséquent, Plapous souligne qu'il serait intéressant de réfléchir à un traitement pour son évaluation.

B.1 Introduction

De manière générale, l'estimée du signal par atténuation spectrale peut se mettre sous la forme : $\widehat{S(k,l)} = G(k,l)X(k,l)$ où $G(k,l)$ est un gain appliqué au signal bruité $X(k,l)$ et dépendant du Rapport Signal à Bruit (RSB) *a priori* et *a posteriori*. En notant $B(k,l)$ la composante de bruit, ces deux RSB ont pour expression :

$$RSB_{post}(k,l) = \frac{|X(k,l)^2|}{\gamma_b(k)} = \frac{|X(k,l)^2|}{E\left[|B(k,l)|^2\right]} \tag{B.1}$$

$$RSB_{prio}(k) = \frac{\gamma_s(k)}{\gamma_b(k)} = \frac{E\left[|S(k,l)|^2\right]}{E\left[|B(k,l)|^2\right]} \tag{B.2}$$

La possibilité du traitement et son efficacité étant conditionnées par l'estimation de la DSP du bruit mais également par le choix de l'estimateur du RSB intervenant dans le calcul du gain, nous allons dans un premier temps faire un point sur les méthodes permettant de les obtenir. Ensuite nous présenterons les différentes règles d'atténuation spectrale.

B.2 Estimation du bruit et calcul de RSB

Deux grandes classes de méthode d'estimation du bruit se distinguent selon qu'une Détection de l'Activité Vocale (DAV) est utilisée ou non. Lorsqu'une DAV

est mise en œuvre, il faut que le contexte d'utilisation ne soit pas trop adverse, car la robustesse n'est garantie que dans des conditions où le RSB n'est pas trop faible (RSB $> 5\ dB$).

B.2.1 Estimation du bruit avec DAV

Grâce à l'utilisation de la DAV, nous allons être capable de mettre à jour l'estimation de la DSP du bruit avec une mémoire du passé choisie. En effet, le principe consiste à ré-estimer la DSP, $\widehat{\Phi_{BB}}(k,l)$, à chaque période de silence grâce à un lissage exponentiel. Cet estimateur est ensuite réutilisé sans aucune autre forme de calcul pendant les périodes de parole [157].

Le lissage exponentiel quant à lui consiste à ajuster un modèle paramétrique de façon locale, c'est-à-dire dont les paramètres évoluent au cours du temps. L'intérêt de cette méthode réside dans la facilité de mise à jour de cette prévision lors de l'acquisition d'une nouvelle donnée. La nouvelle prévision est égale à la précédente pondérée par la dernière erreur effectuée, ce qui fait directement allusion au filtre de Kalman.

$$\widehat{\Phi_{BB}}(k,l) = \begin{cases} \lambda_B \widehat{\Phi_{BB}}(k,l-1) + (1-\lambda_B)\,|X(k,l)|^2 & \text{bruit seul} \\ \widehat{\Phi_{BB}}(k,l-1) & \text{sinon} \end{cases}$$

Où λ_B est appelé le facteur d'oubli choisi en général tel que $0,5 \leq \lambda_B \leq 0,99$ ou alors variant en fonction du temps et défini par $\lambda_B = exp(-R/(F_e\tau_b))$. La grandeur R représente le nombre de points de recouvrement entre deux trames successives, F_e la fréquence d'échantillonnage et τ_b une constante de temps fixée selon le niveau de suivi désiré de la stationnarité du bruit.

B.2.2 Estimation continue

En ce qui concerne l'estimation en continu de la DSP du bruit, certaines méthodes utilisent également la technique du lissage exponentiel vue au paragraphe précédent mais en y introduisant des bornes limitant les variations maximales de la DSP du bruit pendant l'activité vocale, [11]. Typiquement, l'estimée de la DSP du bruit ne peut pas augmenter de plus de 3 dB par seconde et diminuer de plus de 12 dB par seconde, ce qui permet une augmentation lente pendant les instants de parole et un retour rapide possible à une valeur correcte pendant les périodes de bruit seul. Une autre méthode toujours basée sur le lissage exponentiel effectue la comparaison du module de l'observation par rapport à un seuil, [83]. Ainsi, toutes les composantes fréquentielles qui contribuent à un dépassement de ce seuil sont ignorées car considérées comme de la parole. Dans le cas où le seuil n'est pas franchi, la

DSP du bruit est mise à jour à l'aide d'un lissage exponentiel. L'inconvénient de la première méthode est que l'on introduit de la distorsion dans le signal de parole due à une estimation biaisée. Une amélioration de la technique du lissage exponentiel est proposée par Hirsch en considérant les amplitudes maximales de la distribution obtenues par l'analyse de l'histogramme des valeurs passées de $|X(k,l)|$. Il existe également des méthodes basées sur la poursuite des minimas du spectre du signal bruité. Ces minimas sont considérés comme des représentants du spectre lissé du bruit. Cette méthode, appelée *minimum statistics*, possède l'avantage de bien suivre les non-stationnarités du bruit. Inspiré par la méthode de recherche de minimum de la DSP, Cohen propose une approche basée sur le modèle à deux états du signal de parole, introduit initialement par McAulay. L'expression pour le calcul de l'estimée est identique à celle du lissage exponentiel à ceci près, le paramètre λ_B est remplacé par son estimée $\widehat{\lambda_B}(k,l)$ calculée à partir de λ_B et de la probabilité conditionnelle de présence de parole qui, elle aussi, doit être estimée [40] et [38]. Notons que l'introduction de paramètres variant en temps et en fréquence permet de suivre la forte corrélation du signal de parole entre chaque bin fréquentiel sur des trames consécutives.
L'estimation de la DSP peut enfin être obtenue par la pondération de son spectre sur plusieurs trames à l'aide du RSB *a posteriori*, cette solution proposée par Kato [101] a été intégrée dans un mobile 3G.

B.2.3 Calcul du RSB par l'approche décision-dirigée (DD)

L'approche DD, *decision-directed*, proposée par Ephraïm et Malah se décompose en trois étapes. Tout d'abord, une estimation de la DSP du bruit grâce à $\lambda_B\widehat{\Phi_{BB}}(k,l-1)+(1-\lambda_B)\left|X(k,l)\right|^2$ permet de calculer les deux estimateurs des RSB suivants :

$$\widehat{RSB_{post}}(k,l)=\frac{\left|X(k,l)^2\right|}{\widehat{\Phi_{BB}}(k,l)} \tag{B.3}$$

$$\widehat{RSB_{prio}}(k,l)=\beta\frac{\left|\widehat{S}(k,l-1)\right|^2}{\widehat{\Phi_{BB}}(k,l)}+(1-\beta)max(\widehat{RSB_{post}}(k,l)-1,0) \tag{B.4}$$

où λ varie strictement entre 0 et 1 et β prend typiquement la valeur de 0,98. Si cette approche permet de supprimer efficacement le bruit musical il en ressort qu'elle introduit un effet de réverbération, car cet estimateur est en fait biaisé vis-à-vis des composantes de parole. Cette propriété peut être mise en évidence à partir de l'équation (B.4) où l'on constate que l'estimation est faite à partir de $\widehat{S}(k,l-1)$, ce qui rend cet estimateur plus adapté au traitement de la trame précédente que celle courante. Partant de cette observation, Plapous analyse que l'idéal serait de

combiner les avantages respectifs de ces RSB de tel sorte à aboutir au comportement suivant :

$$\widehat{RSB^{id}_{prio}}(k,l) = \begin{cases} \widehat{RSB_{post}}(k,l) - 1 & \text{pour les composantes de parole} \\ \widehat{RSB^{DD}_{post}}(k,l) & \text{pour les composantes de bruit} \end{cases}$$

Cette théorie s'appuie sur l'étude de résultats de simulation qui permettent de dégager ce qui suit sur les tendances de ces deux estimateurs.

RSB *a posteriori* :

- Laisse un bruit musical important,
- Fourni une estimation non biaisée du signal de parole.

RSB *a priori* :

- Limite efficacement le bruit musical,
- Fourni une estimation biaisée de la parole, ce qui amène un effet de réverbération.

Dans le but d'éliminer le biais de l'estimateur DD, Plapous définit un estimateur en deux passes, *Two-Step Noise Reduction* (TSNR), qui calcule le gain à la trame $l+1$ et l'applique à la trame l. Celui-ci n'est en fait rien d'autre que l'estimateur du RSB *a priori* de l'approche DD réévalué pour la trame $l+1$ et pour lequel β vaut 1. Grâce à cela il n'est pas nécessaire de connaître la trame future et le RSB et le gain s'expriment à présent comme suit :

$$\widehat{RSB^{TSNR}_{prio}}(k,l) = \beta \frac{\left|\widehat{S_{DD}}(k,l)^2\right|}{\widehat{\gamma_b(k)}} \tag{B.5}$$

$$G_{TSNR}(k,l) = \frac{RSB^{TSNR}_{prio}(k,l)}{1 + RSB^{TSNR}_{prio}(k,l)} \tag{B.6}$$

Les performances sont comme nous pouvions l'espérer une suppression du biais et une limitation du bruit musical par rapport à l'approche DD. Les attaques et extinctions de parole sont préservées ainsi que l'effet de réverbération. Néanmoins, il reste des distorsions liées à l'estimation de la DSP du bruit et l'influence de la phase.

Une deuxième méthode a été développée dans le but d'améliorer la qualité du signal restauré en régénérant les harmoniques des sons voisés qui ont été supprimées, y compris pour l'approche TSNR, d'où le nom *Harmonic Regeneration Noise*

Reduction (HRNR). Sachant que 80% des sons prononcés sont voisés, nous percevons l'importance que prend une telle méthode dans les techniques de postfiltrage. Le principe permettant de regénérer l'harmonicité est relativement simple et provient des méthodes de l'électronique : le redressement mono-alternance. En effet, en utilisant la fonction $max()$, fonction non linéaire, les harmoniques perdues sont restituées avec un positionnement exact mais à un facteur d'amplitude près. Le RSB *a priori* de cette méthode est estimé grâce à :

$$\widehat{RSB}_{prio}^{HRNR}(k,l) = \frac{\rho(k,l)\left|\widehat{S_{TSNR}}(k,l)\right|^2 + (1-\rho(k,l))\left|\widehat{S_{harmo}}(k,l)\right|^2}{\widehat{\gamma_b(k)}} \tag{B.7}$$

où $\rho(k,l) = G_{TSNR}(k,l)$ tout simplement pour s'assurer que celui-ci aura le comportement suivant :

$$\begin{aligned} \rho(k,l) &\longrightarrow 1 \text{ si l'estimation de } \widehat{S_{TSNR}}(k,l) \text{ est fiable} \\ \rho(k,l) &\longrightarrow 0 \text{ sinon} \end{aligned}$$

De cette manière, les distorsions harmoniques inhérentes au biais sur l'amplitude des harmoniques "ré-générées" sont évitées avec une méthode simple et robuste.

B.3 Règles d'atténuation spectrale

Comme pour les techniques d'estimation du bruit, il est de nouveau possible d'extraire des classes parmi les règles d'atténuation spectrale à court terme. Ici, elles sont aux nombres de trois avec pour distinction la nécessité ou non d'un modèle statistique des signaux traités ou l'utilisation d'un modèle psychoacoustique (non présenté ici).

B.3.1 Approches non statistiques

Soustraction spectrale

Il existe quatre variantes parmi les règles de soustraction spectrale introduitent par Boll, Lim, Vary puis Berouti que nous résumerons à partir de celle fournie par ce dernier, à savoir la soustraction spectrale généralisée. L'objectif de cette généralisation proposée par Berouti est d'inclure des paramètres qui permettent d'ajuster au mieux la méthode afin de réduire le bruit musical introduit. L'estimation du spectre du signal utile s'exprime tel que :

$$\left|\widehat{S(k,l)}\right|^2 = \begin{cases} D(k,l)^{\frac{1}{\gamma}} & \text{si } D(k,l)^{\frac{1}{\gamma}} \geq \beta \cdot E\left[|B(k,l)|^2\right] \\ \beta \cdot E\left[|B(k,l)|^2\right] & \text{sinon} \end{cases}$$

avec

$$D(k,l) = (|X(k,l)|^2)^\gamma - \alpha(E\left[|B(k,l)|^2\right])^\gamma$$

où $\alpha \geq 1$, $0 \leq \beta \leq 1$ et $\gamma \geq 0$. Cette reformulation permet ainsi de réduire l'excursion spectrale grâce à :

- α, **coefficient de surestimation** : il permet de réduire le bruit musical résiduel au détriment d'une dégradation du signal de parole. Pour palier cet effet, celui-ci peut être rendu adaptatif en fonction de la fréquence et suivant le type de bruit ;
- β, **plancher spectral** : il limite la réduction du bruit de fond qui assure d'une part un certain confort lors de l'écoute et d'autre part masque le bruit musical. Si β est faible, le bruit de fond est réduit mais le bruit musical reste audible et inversement, lorsque β est grand le bruit musical est totalement supprimé mais le bruit original reste très présent ;
- γ, **agressivité** : il fixe le type de soustraction spectrale utilisée (Soustraction Spectrale en Amplitude (SSA), en Puissance (SSP), etc ...).

La reformulation de la soustraction spectrale généralisée (SSG) sous la forme spécifique des méthodes d'atténuation donne

$$\left|\widehat{S(k,l)}\right| = G_{SSG}(k,l)X(k,l)$$

avec

$$G_{SSG}(k,l) = \begin{cases} \left(1 - \alpha \frac{1}{RSB^\gamma_{post}(k,l)}\right)^{\frac{1}{2\gamma}} & \text{si } (RSB^\gamma_{post}(k,l) - \alpha)^{\frac{1}{\gamma}} \geq \beta \\ \sqrt{\frac{\beta}{RSB_{post}(k,l)}} & \text{sinon} \end{cases}$$

Pour retrouver les autres variantes citées précédement, soit la SSA ou SSP, il suffit de prendre $\alpha = 1$, $\beta = 0$ ce qui conduit avec $\gamma = \frac{1}{2}$ à la SSA et à la SSP lorsque $\gamma = 1$. L'inconvénient majeur de ces deux méthodes est qu'elles introduisent du bruit musical. Nous noterons qu'une approche paramétrique de la SSG connue sous le nom SSGP a été développée par Sim pour tenter d'obtenir un compromis entre un gain suffisamment sélectif pour supprimer le bruit sans être trop abrupte afin ne pas dégrader le signal de parole.

Filtrage de Wiener

Le filtrage de Wiener consiste à rechercher un filtre linéaire optimal dans le cas de signaux stationnaires. L'optimalité est déduite du critère de minimisation de l'erreur quadratique moyenne (EQM), c'est-à-dire :

$$E\left[(S(k,l) - G_W(k,l)X(k,l))^2\right] \tag{B.8}$$

Sous les hypothèses d'orthogonalité ou de décorrélation des signaux $s(n)$ et $b(n)$ on obtient une première expression du gain du filtre.

$$G_W = \frac{E\left[|S(k,l)|^2\right]}{E\left[|S(k,l)|^2\right] + E\left[|B(k,l)|^2\right]} = \frac{RSB_{prio}}{1 + RSB_{prio}} \tag{B.9}$$

Une autre écriture du gain du filtre de Wiener est obtenue en estimant la DSP du signal utile à partir de la SSP. Il s'agit alors du filtre pseudo-Wiener pour lequel nous obtenons :

$$G_W(k,l) = G_{SSP}^2(k,l) = 1 - \frac{1}{RSB_{post}(k,l)} \tag{B.10}$$

B.3.2 Approches statistiques

Méthode de MacAulay et Malpass

Dans cette méthode, un moèle reposant sur les "deux états" du signal parole va être considéré. Les auteurs proposent donc une règle de suppression du type détection combinée à un problème d'estimation avec pour hypothèses :

- H_0 : absence de parole $\longrightarrow x(n) = b(n)$
- H_1 : présence de parole $\longrightarrow x(n) = b(n) + s(n)$

L'estimateur au sens du minimum de l'EQM est donné par :

$$\begin{aligned} \left|\widehat{S(k,l)}\right| &= E\left[|S(k,l)|\,|\,|X(k,l)|\,, H_0\right] \cdot p(H_0|\,|X(k,l)|) \\ &+ E\left[|S(k,l)|\,|\,|X(k,l)|\,, H_1\right] \cdot p(H_1|\,|X(k,l)|) \end{aligned}$$

où $p(H_i|\,|X(k,l)|$ désigne la probabilité conditionnelle. Or, l'espérance du module étant nulle sous H_0, l'estimateur se simplifie et donne

$$\left|\widehat{S(k,l)}\right| = E\left[|S(k,l)|\,|\,|X(k,l)|\,, H_1\right] \cdot p(H_1|\,|X(k,l)|) \tag{B.11}$$

où la probabilité *a posteriori* $p(H_1|\,|X(k,l)|)$ est donnée par :

$$p(H_1|\,|X(k,l)|) = \frac{\Lambda(k,l)}{1 + \lambda(k,l)}$$

avec

$$\Lambda(k,l) = e^{-RSB_{prio}} I_0\left[2\sqrt{RSB_{prio} RSB_{post}(k,l)}\right] \tag{B.12}$$

Dans cette expression (B.12), I_0 représente la fonction de Bessel modifiée du premier genre et d'ordre zéro. Cette méthode ressemble à la SSG avec l'avantage de ne pas

pâtir d'une valeur biaisée du RSB *a posteriori*. Le gain associé mis sous sa forme classique est donné par :

$$G_{STSA}^{SD}(k,l) = \frac{1}{2}\left[1 + \sqrt{1 - \frac{1}{RSB_{post}(k,l)}}\right] p(H_1 | |X(k,l)|) \tag{B.13}$$

Méthodes d'Ephraïm et Malah

Les auteurs proposent une estimée de l'amplitude du spectre du signal de parole optimale au sens de l'erreur quadratique moyenne. Pour y parvenir ils approximent le spectre à court terme par un mélange de gaussiennes aléatoires et indépendantes. De plus, le bruit est également supposé gaussien et indépendant du signal de parole. Cet estimateur, appelé estimateur MMSE STSA pour *Minimum Mean Square Error Short Time Spectral Amplitude* [57] et [56], est donné par l'expression suivante :

$$\left|\widehat{S(k,l)}\right| = E\left[|S(k,l)| \, | \, |X(k,l)|\right] = E\left[A(k,l) | \, |X(k,l)|\right] \tag{B.14}$$

Le gain associé écrit sous la forme caractéristique des méthodes d'atténuation est :

$$G_{STSA}(k,l) = \frac{\sqrt{\pi}}{2}\sqrt{\left(\frac{1}{RSB_{post}(k,l)}\right)\left(\frac{RSB_{prio}(k,l)}{1+RSB_{prio}(k,l)}\right)} M(\nu(k,l)) \tag{B.15}$$

avec

$$\nu(k,l) = \frac{RSB_{prio}(k,l)}{1+RSB_{prio}(k,l)} RSB_{post}(k,l)$$

où M est la fonction hypergéométrique confluente. En tenant compte de la sensibilité de l'oreille plus marquée pour les variations du logarithme de l'amplitude que de l'amplitude elle-même, les auteurs proposent d'estimer le logarithme de l'amplitude du signal de parole comme ci-après.

$$\left|\widehat{S(k,l)}\right| = exp\left(E\left[ln(|S(k,l)|) | \, |X(k,l)|\right]\right) \tag{B.16}$$

Nous parlerons ainsi de l'estimateur MMSE LSA pour *Minimum Mean Square Error Log-Spectral Amplitude* dont le gain devient :

$$G_{LSA}(k,l) = \frac{RSB_{prio}(k,l)}{1+RSB_{prio}(k,l)} exp\left[\frac{1}{2}\int_{\nu(k,l)}^{\infty} \frac{exp(-t)}{t} dt\right] \tag{B.17}$$

Pour ces deux estimateurs, les auteurs proposent aussi un calcul à l'aide de l'aspect deux états du signal de parole présenté précédement. Ce qu'il faut en retenir c'est que l'adjonction de l'approche deux états permet de corriger un filtre globalement dominé par le RSB *a priori* par un facteur de pondération globalement dominé par le RSB *a posteriori*.

Annexe C

Evaluation objective

C.1 Description de la méthode d'évaluation **252**
C.2 Critères utilisés **252**

Cette annexe décrit la procédure mise en œuvre pour l'évaluation objective des techniques proposées. La méthode d'évaluation utilisée permet de mesurer à la fois la réduction des perturbations ainsi que la dégradation apportée au signal utile. Plus de détails au sujet de l'évaluation des performances des systèmes et y compris des antennes acoustiques sont donnés dans [123].

C.1 Description de la méthode d'évaluation

La première phase de la mesure consiste à faire fonctionner le traitement à évaluer en considérant en entrée le signal perturbé, $\mathbf{a}^s * s(n) + \mathbf{a}^b * b(n)$. Durant cette phase, nous autorisons l'adaptation des filtres adaptatifs pour produire le signal estimé $y(n)$. Dans la seconde phase, l'adaptation des filtres est gelée. Les coefficients des filtres obtenus pendant la première phase ont été stockés trame par trame. Ceux-ci sont recopiés et appliqués indépendamment au signal utile pour produire le signal $y_s(n)$ et à la perturbation pour fournir le signal $y_b(n)$. Dans le cas d'un système de réduction de bruit idéal, le signal correspondant à la perturbation est complètement atténué, alors que le signal utile ne subit aucune modification.

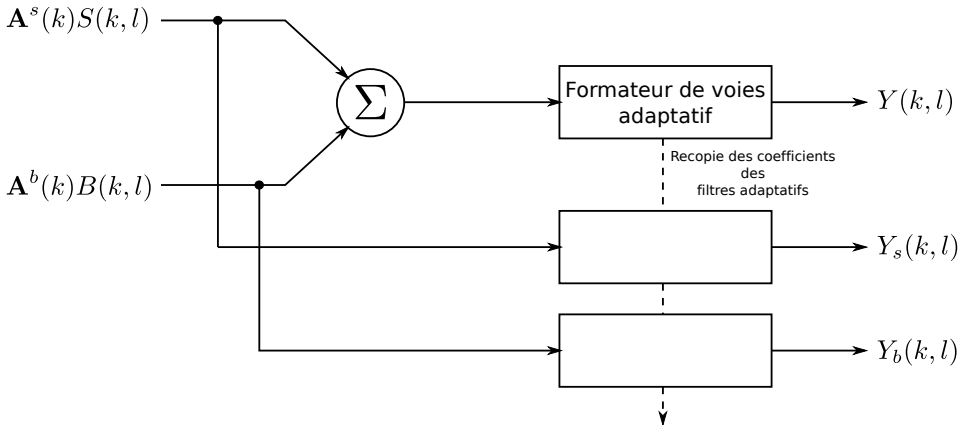


Figure C.1 – Procédure expérimentale pour évaluer la réduction des perturbations dans le cas d'un traitement multicapteur.

C.2 Critères utilisés

La mesure du gain en RSB apporté par l'antenne et calculé sur les instants de parole est donné par :

$$\mathbf{G} = \frac{\sum_{n=1}^{N} y_s^2(n) \sum_{m=1}^{M_{mic}} \sum_{n=1}^{N} \widetilde{b}_m^2(n)}{\sum_{n=1}^{N} y_b^2(n) \sum_{m=1}^{M_{mic}} \sum_{n=1}^{N} \widetilde{s}_m^2(n)} \tag{C.1}$$

où $\widetilde{s}_m = \mathbf{a}^s * s(n)$ et $\widetilde{b}_m = \mathbf{a}^b * b(n)$ sont respectivement les signaux désirés et les signaux de bruit associés à chaque microphone et $y_s(n)$ et $y_b(n)$ les signaux correspondants mesurés en sortie d'antenne. Les termes intégrant une double somme proviennent du calcul d'une moyenne des énergies des signaux captés sur les microphones. Ceci permet d'éviter le biais d'une mesure employant un seul microphone pour référence de signal non traité. En effet, la fonction de transfert d'une seule réponse impulsionnelle peut laisser apparaître un zéro à une fréquence donnée alors que ce phénomène n'apparaît pas nécessairement sur les autres canaux acoustiques. D'autre part, l'énergie parvenant sur chaque microphone est différente selon les positions relatives de l'antenne et du locuteur. Il est donc plus juste d'utiliser une moyenne énergétique sur tous les capteurs. Enfin, lorsque le gain en RSB défini à l'équation (C.1) est exprimé en décibel, il correspond alors à la différence entre le RSB en sortie et en entrée. Or, le RSB est une mesure à long terme qui se prête peu au signal de parole qui est par essence non-stationnaire. Pour l'évaluation des techniques de réduction de bruit, on lui préfère le RSB segmental qui est calculé sur des trames d'une durée de quelques dizaines de milliseconde. Le RSB segmental permet donc de suivre l'évolution du RSB au cours des trames.

La réduction du bruit apportée par l'antenne et mesurée sur les instants de silence est calculée comme suit.

$$\mathbf{NR} = \frac{\sum_{m=1}^{M_{mic}} \sum_{n=1}^{N} \widetilde{b}_m^2(n)}{\sum_{n=1}^{N} y_b^2(n)} \tag{C.2}$$

Cette expression du critère pour le calcul de la réduction du bruit correspond au critère du gain en RSB, mais calculé sur les instants où le signal utile est absent. En ce sens, ces deux critères, **G** et **NR**, sont tous deux un calcul de la réduction des perturbations, l'un durant les périodes d'activité vocale, l'autre pendant les périodes de bruit seul.

Pour évaluer de manière objective la qualité du signal en sortie de traitement, nous utiliserons la distance cepstrale qui est une mesure de la différence des cepstres des signaux d'entrée $s_{in}(n)$ et de sortie $y_s(n)$. Ce critère est plus en corrélation avec les tests subjectifs d'évaluation de la qualité du signal que les deux critères donnés précédemment. Dans notre implémentation, la suite notée $c_n(p)$, des coefficients cepstraux $c_i(p)$ obtenus pour la trame $x_p(n)$ est calculée comme ceci :

$$c_n(p) = TFDI\left(ln|TFD(x_p(n))|\right)$$

Où TFD et TFDI désignent respectivement la transformée de Fourier Discrète et son inverse. Le signal que nous prendrons pour référence correspond au son direct

capté par le microphone, $s_{in}(n)$. Il sera mesuré à partir du microphone placé au centre de l'antenne. La distance cepstrale, notée **DC**, est par conséquent calculée pour chaque trame p à partir des coefficients cepstraux $c_i(p)$ et $c_i^{'}(p)$ correspondant respectivement aux signaux $s_{in}(n)$ et $y_s(n)$ selon la formule suivante :

$$\mathbf{DC}(p) = 2\sum_{i=1}^{M}(c_i(p) - c_i^{'}(p))^2 \tag{C.3}$$

Notons que le terme $(c_0(p) - c_0^{'}(p))^2$ qui correspond à une amplification n'est pas pris en compte dans le calcul. Dans nos résultats, nous utiliserons $M = 11$ coefficients et donnerons la valeur moyenne de la distance cepstrale.

Annexe D

Roomsim

D.1 Présentation des possibilités de *Roomsim* **256**
D.2 Scène synthétique n°1 . . . . . **256**
D.3 Scène synthétique n°2 . . . . . **257**

Pour créer des scènes sonores synthétiques, nous utilisons l'outil *Roomsim* [26] mis à disposition dans le domaine public par Campbell (licence GNU) et qui permet de générer des réponses impulsionnelles avec des degrés de liberté vis-à-vis des conditions de propagation. Bien sûr, les réponses générées ne sont pas en adéquation parfaite avec la réalité, il s'agit d'une modélisation basée sur la méthode image proposée par Allen et Berckley et qui a pour le moins l'avantage de pouvoir se placer dans des conditions proches d'une propagation en milieu clos. De plus, cet outil étant dans le domaine public, il facilite la reproductibilité des résultats en offrant à tout à chacun la possibilité de reproduire les expériences dans des conditions identiques.

D.1 Présentation des possibilités de *Roomsim*

En supposant que la propagation s'effectue dans une pièce dont la géométrie correspond à un parallélépipède rectangle (boîte à chaussure), le programme génère les réponses impulsionnelles entre les sources omnidirectionnelles et les microphones dans des conditions où nous pouvons entre autre fixer :

1. la fréquence d'échantillonnage ;
2. les conditions physiques la propagation (taux d'humidité, température de l'air, avec ou sans absorption, avec ou sans atténuation en fonction de la distance, le nombre de réflexions) ;
3. la géométrie de la salle (longueur, largeur, hauteur) et les coefficients d'absorption des différents murs ;
4. la longueur de la réponse impulsionnelle ;
5. le nombre de microphones (1 ou 2), leurs directivités, leurs positions sur l'antenne ainsi que leur orientation dans l'espace et la position de l'antenne dans la pièce ;
6. la position et le nombre de sources ainsi que leurs orientations ;

En plus de cette liste non exhaustive de possibilités, il offre des modes de visualisation 2D ou 3D, la possibilité de créer des scénarios de type *cocktail party*, ainsi que des opérations de filtrage, de calcul de corrélation, etc. Enfin, il peut être aisément agrémenté de fonctions supplémentaires en fonction des besoins de l'utilisateur.

D.2 Scène synthétique n°1

La première scène sonore décrite à la Figure D.1 correspond à une propagation dans un milieu anéchoïque lorsque les deux sources sont placées à égales distances des microphones, soit 40 *cm* et de manière symétrique. L'antenne est constituée de deux

microphones placés à une hauteur de 1,20 m et espacés de 15 cm. La Figure D.3 donne l'ensemble des paramètres fixés pour cette configuration avec les positions exactes des sources et de l'antenne.

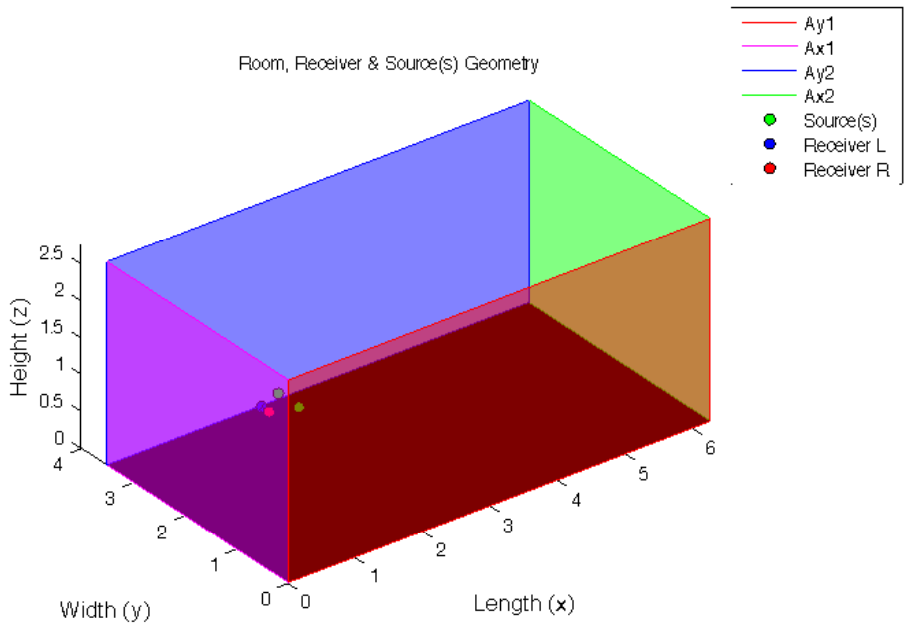


Figure D.1 – Position des sources et de l'antenne.

D.3 Scène synthétique n°2

La seconde scène sonore décrite à la Figure D.4 correspond à deux types de propagation, l'une anéchoïque, l'autre échoïque. La source utile est placée à 80 cm des microphones tandis que la source perturbatrice est située à 2 m. La disposition symétrique est conservée. L'antenne est constituée de deux microphones placés à une hauteur de 1,20 m et espacés de 15 cm. Les Figures D.6 et D.10 donnent l'ensemble des paramètres fixés pour ces deux configurations avec les positions exactes des sources et de l'antenne. Les Figures D.8 et D.9 décrivent également l'évolution des coefficients d'absorption des murs ainsi que celle du temps de réverbération en fonction de la fréquence.

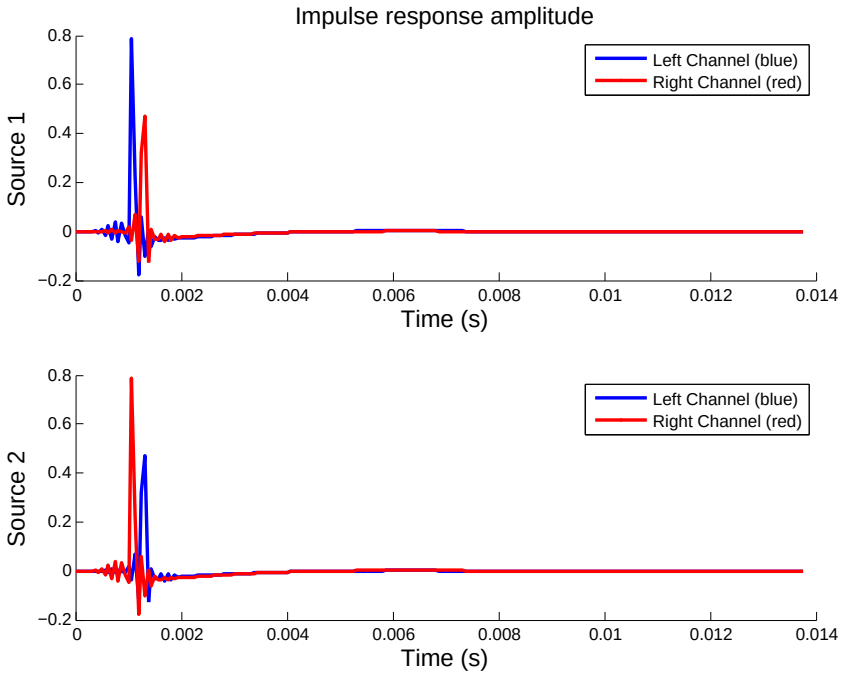


Figure D.2 – Réponses impulsionnelles générées.

```
Parameter       Value                   % Comment
Fs              16000                   % Sampling frequency (Hz)
humidity        40                      % Relative humidity of air (%) (Used to calculate air absorption coefficient "m", valid range 20<h <70)
TEMP            20                      % Temperature of air (deg C) (Used to calculate speed of sound c (m/s))
order           5                       % If -ve then a value will be computed, else value supplied here is used (limits order of reflections computed)
H_length        256                     % Length of Impulse Response. If -ve then a value will be computed, else value supplied here is used"
H_filename      H_ROOM_2M_2S_SYM_ANEC_SYNTH     % Filename for impulse response.
air_F           1                       % 0 = no absorption due to air, 1 = air absorption is present.
dist_F          1                       % 0 = no distance attenuation applied (ie no 1/R effect), 1= distance attenuation applied
Fc_HP           100.00                  % 0 = no high-pass filter. If scalar value supplied for cut-off frequency Fc_HP then high-pass filter applied.
smooth_F        1                       % 0 = no smoothing filter applied, 1= smoothing filter used. (Not used for MIT or CIPIC)
plot_F2         1                       % 0 = no plot, 1 = 2D-plan, shows image rooms and sources on constant z plane."
plot_F3         1                       % 0 = no plot, 1 = 3D-plot, shows image sources. Rotatable in MATLAB"
alpha_F         0                       % 0 = fixed transparent surfaces for Room Geometry plot, 1 = (surface opacity = reflectivity)

Lx              6.25                    % Room Length (m) x direction
Ly              3.50                    % Room Width (m) y direction
Lz              2.75                    % Room Height (m) z direction
xp              1                       % Receiver x co-ordinate (m)
yp              1.75                    % Receiver y co-ordinate (m)
zp              1.2                     % Receiver z co-ordinate (m) e.g. 1.2m is a typical height above floor of ears of seated human subject
Receiver        two_mic                 % Receiver, copy one of these:   one_mic     two_mic     mithrir cipicir NB all strings must be same length for later logical tests.
sensor_space    0.15                    % Sensor separation in 'two_mic' case (0.145 m is CIPIC average)
MIT_root        MIT_HRTF                % The root of MIT Kemar data base file
subdir1         Kemar                   % MIT sub-directory1 name
subdir2         compact                 % MIT sub-directory2 name
filename        hrir_final.mat          % MIT filename for HRIR
CIPIC_root      CIPIC_HRTF              % The root of the CIPIC data base files
subdir1         standard_hrir_database  % CIPIC sub-directory1 name
subdir2         subject_                % CIPIC sub-directory2 name (program combines this with subject number)
S_No            021                     % CIPIC subject number, format '&&&' (e.g. '021' is the Kemar with small pinnae)
filename        hrir_final.mat          % CIPIC filename for HRIR
receiver_yaw    0                       % Yaw (Azimuth) offset of receiver system (degrees) +ve slew left
receiver_pitch  0                       % Pitch (Elevation) offset of receiver system (degrees) +ve nose up
receiver_roll   0                       % Roll offset of receiver system (degrees) +ve right wing down
%-------------------------------------------------------------------------------------------------
        Set the room surface absorptions
F_abs 125   250   500   1000  2000  4000  % Standard measurement frequencies (Hz)
Ax1   1     1     1     1     1     1     % Absorption of wall in x=0 plane (behind Receiver in plan)
Ax2   1     1     1     1     1     1     % Absorption of wall in x=Lx plane (front in plan)
Ay1   1     1     1     1     1     1     % Absorption of wall in y=0 plane (right in plan)
Ay2   1     1     1     1     1     1     % Absorption of wall in y=Ly plane (left in plan)
Az1   1     1     1     1     1     1     % Absorption of floor i.e. z=0 plane
Az2   1     1     1     1     1     1     % Absorption of ceiling i.e. z=Lz plane
%-------------------------------------------------------------------------------------------------
Directionality  Single/Left     Right_(if present)
azim_off        30              -30             % ]azimuth offset for sensor (degrees)
elev_off        30              30              % ]elevation offset for sensor
roll_off        0               0               % ]roll offset for sensor
SENSOR_root     SENSOR          SENSOR          % The root of the SENSOR data base file
subdir1         Types           Types           % SENSOR sub-directory1 name
filename        cardoid.mat     cardoid.mat     % Sensor filenames for Impulse Response
%-------------------------------------------------------------------------------------------------
SOURCES R_s (m)     alpha (deg)     beta (deg)      % R_s = Radial distance(s) of source(s) from head (m).
1       0.40        30              0               % alpha = Azimuth(s) of sources -180<alpha < 180 (deg) NB +ve is ACW on xy plane.
2       0.40        -30             0               % beta = Elevation(s) of sources -90<beta < 90 (deg).
```

Figure D.3 – Fichier de configuration pour le calcul des réponses impulsionnelles.

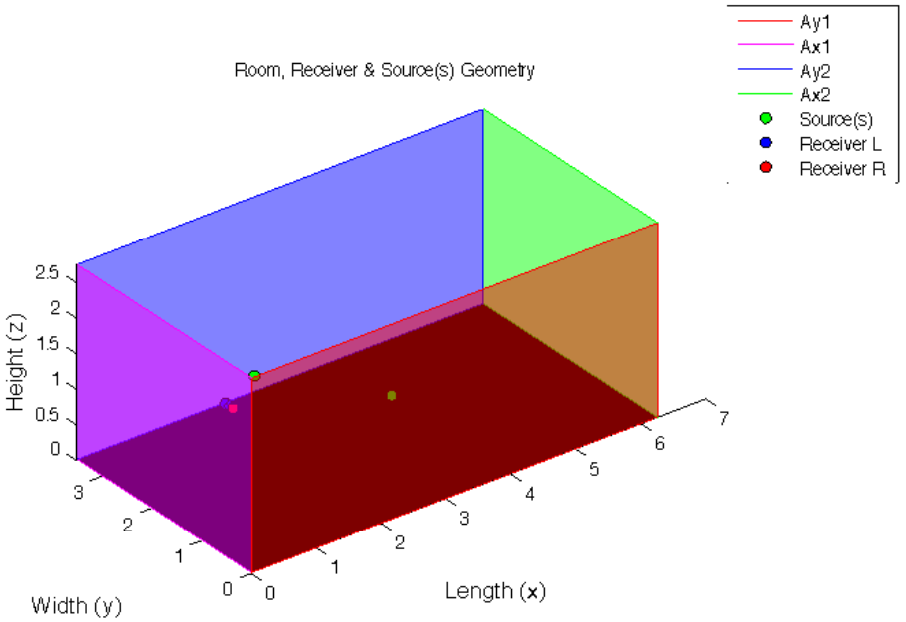


Figure D.4 – Position des sources et de l'antenne.

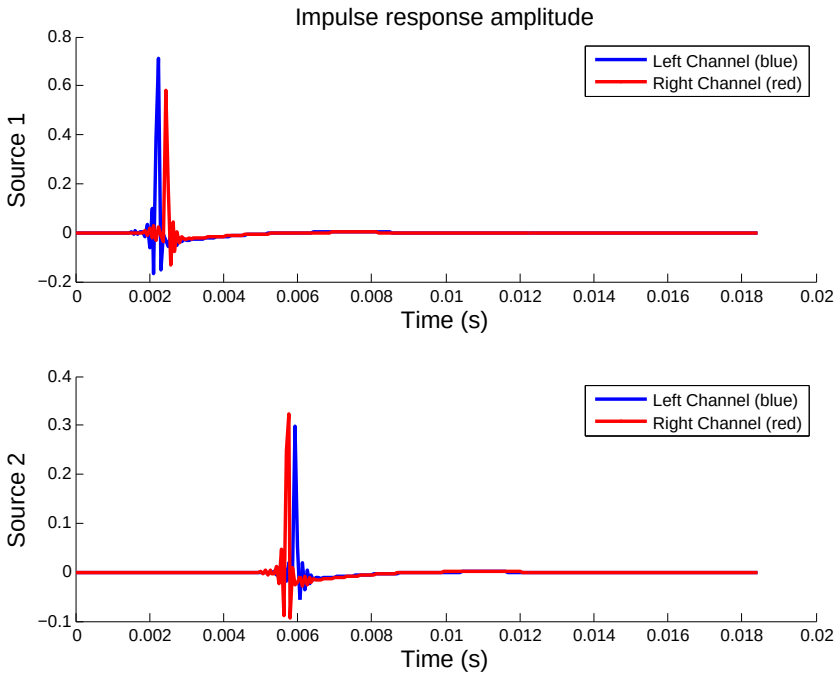


Figure D.5 – Réponses impulsionnelles générées dans le cas anéchoïque.

```
Parameter       Value                   % Comment
Fs              16000                   % Sampling frequency (Hz)
humidity        40                      % Relative humidity of air (%) (Used to calculate air absorption coefficient "m", valid range 20<h <70)
TEMP            20                      % Temperature of air (deg C) (Used to calculate speed of sound c (m/s))
order           5                       % If -ve then a value will be computed, else value supplied here is used (limits order of reflections computed)
H_length        256                     % Length of Impulse Response. If -ve then a value will be computed, else value supplied here is used"
H_filename      H_ROOM_2M_2S_SYM_ANEC   % Filename for impulse response.
air_F           1                       % 0 = no absorption due to air, 1 = air absorption is present.
dist_F          1                       % 0 = no distance attenuation applied (ie no 1/R effect), 1= distance attenuation applied
Fc_HP           100.00                  % 0 = no high-pass filter. If scalar value supplied for cut-off frequency Fc_HP then high-pass filter applied.
smooth_F        1                       % 0 = no smoothing filter applied, 1= smoothing filter used. (Not used for MIT or CIPIC)
plot_F2         1                       % 0 = no plot, 1 = 2D-plan, shows image rooms and sources on constant z plane."
plot_F3         1                       % 0 = no plot, 1 = 3D-plot, shows image sources. Rotatable in MATLAB"
alpha_F         0                       % 0 = fixed transparent surfaces for Room Geometry plot, 1 = (surface opacity = reflectivity)

Lx              6.25                    % Room Length (m) x direction
Ly              3.50                    % Room Width (m) y direction
Lz              2.75                    % Room Height (m) z direction
xp              1                       % Receiver x co-ordinate (m)
yp              1.75                    % Receiver y co-ordinate (m)
zp              1.2                     % Receiver z co-ordinate (m) e.g. 1.2m is a typical height above floor of ears of seated human subject
Receiver        two_mic                 % Receiver, copy one of these:   one_mic   two_mic   mithrir cipicir NB all strings must be same length for later logical tests.
sensor_space    0.15                    % Sensor separation in 'two_mic' case (0.145 m is CIPIC average)
MIT_root        MIT_HRTF                % The root of MIT Kemar data base file
subdir1         Kemar                   % MIT sub-directory1 name
subdir2         compact                 % MIT sub-directory2 name
filename        hrir_final.mat          % MIT filename for HRIR
CIPIC_root      CIPIC_HRTF              % The root of the CIPIC data base files
subdir1         standard_hrir_database  % CIPIC sub-directory1 name
subdir2         subject_                % CIPIC sub-directory2 name (program combines this with subject number)
S_No            021                     % CIPIC subject number, format '&&&' (e.g. '021' is the Kemar with small pinnae)
filename        hrir_final.mat          % CIPIC filename for HRIR
receiver_yaw    0                       % Yaw (Azimuth) offset of receiver system (degrees) +ve slew left
receiver_pitch  0                       % Pitch (Elevation) offset of receiver system (degrees) +ve nose up
receiver_roll   0                       % Roll offset of receiver system (degrees) +ve right wing down
%-------------------------------------------------------------------------------------------------
        Set the room surface absorptions
F_abs 125  250  500  1000  2000  4000   % Standard measurement frequencies (Hz)
Ax1   1    1    1    1     1     1      % Absorption of wall in x=0 plane (behind Receiver in plan)
Ax2   1    1    1    1     1     1      % Absorption of wall in x=Lx plane (front in plan)
Ay1   1    1    1    1     1     1      % Absorption of wall in y=0 plane (right in plan)
Ay2   1    1    1    1     1     1      % Absorption of wall in y=Ly plane (left in plan)
Az1   1    1    1    1     1     1      % Absorption of floor i.e. z=0 plane
Az2   1    1    1    1     1     1      % Absorption of ceiling i.e. z=Lz plane
%-------------------------------------------------------------------------------------------------
Directionality  Single/Left     Right_(if present)
azim_off        30              -30             % ]azimuth offset for sensor (degrees)
elev_off        30              30              % ]elevation offset for sensor
roll_off        0               0               % ]roll offset for sensor
SENSOR_root     SENSOR          SENSOR          % The root of the SENSOR data base file
subdir1         Types           Types           % SENSOR sub-directory1 name
filename        cardoid.mat     cardoid.mat     % Sensor filenames for Impulse Response
%-------------------------------------------------------------------------------------------------
SOURCES  R_s (m)   alpha (deg)  beta (deg)   % R_s = Radial distance(s) of source(s) from head (m).
1        0.80      30           0            % alpha = Azimuth(s) of sources -180< alpha <180 (deg) NB +ve is ACW on xy plane.
2        2.00      -30          0            % beta = Elevation(s) of sources -90<beta <90 (deg).
```

Figure D.6 – Fichier de configuration pour le calcul des réponses impulsionnelles dans le cas anéchoïque.

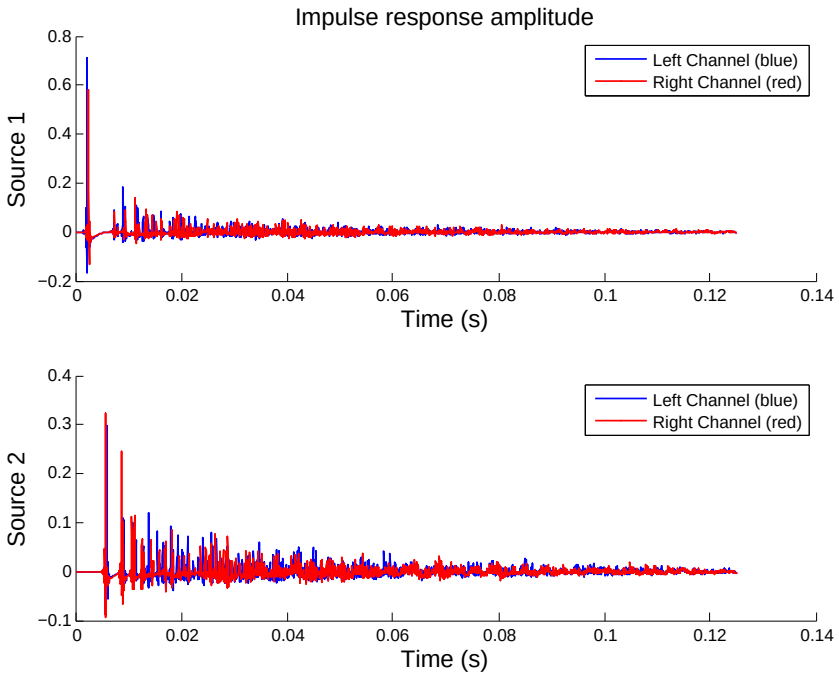


Figure D.7 – Réponses impulsionnelles générées dans le cas échoïque.

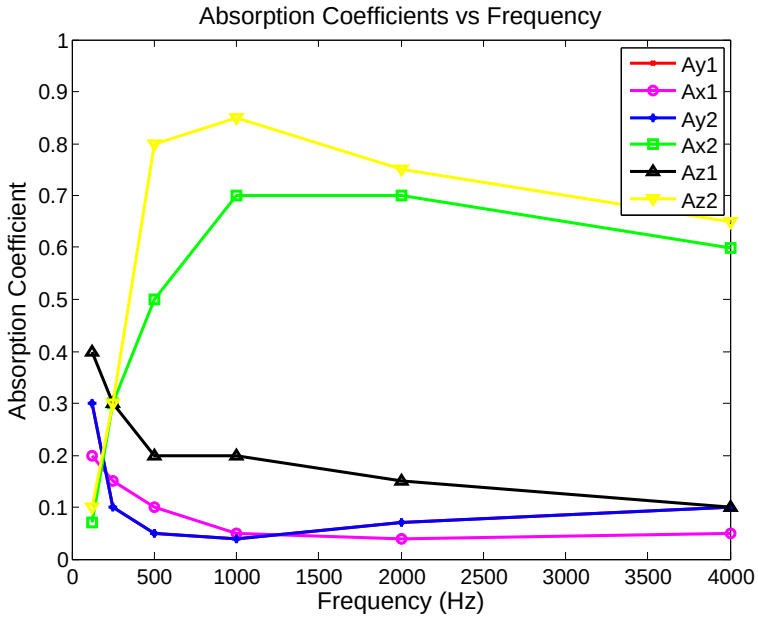


Figure D.8 – Evolution des coefficients d'absorption en fonction de la fréquence.

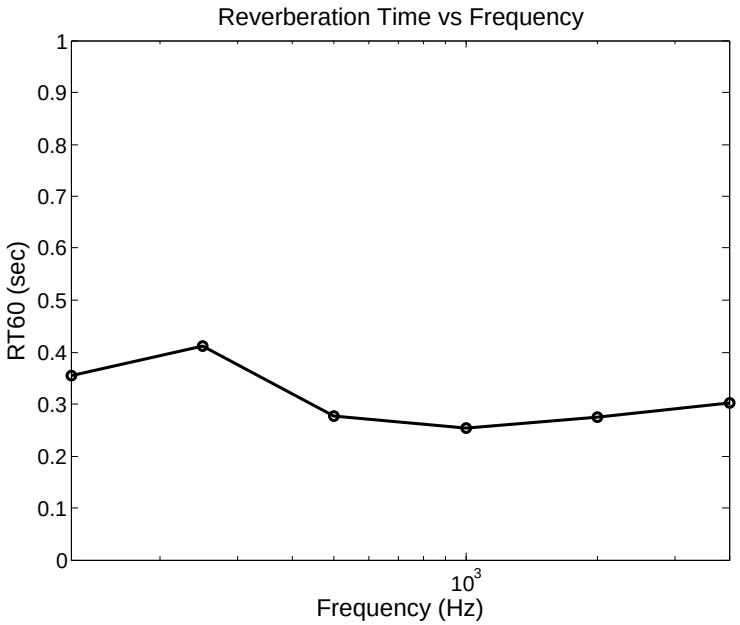


Figure D.9 – Evolution du temps de réverbération en fonction de la fréquence.

```
Parameter       Value                   % Comment
Fs              16000                   % Sampling frequency (Hz)
humidity        40                      % Relative humidity of air (%) (Used to calculate air absorption coefficient "m", valid range 20<h <70)
TEMP            20                      % Temperature of air (deg C) (Used to calculate speed of sound c (m/s))
order           5                       % If -ve then a value will be computed, else value supplied here is used (limits order of reflections computed)
H_length        2000                    % Length of Impulse Response. If -ve then a value will be computed, else value supplied here is used"
H_filename      H_ROOM_2M_2S_SYM        % Filename for impulse response.
air_F           1                       % 0 = no absorption due to air, 1 = air absorption is present.
dist_F          1                       % 0 = no distance attenuation applied (ie no 1/R effect), 1= distance attenuation applied
Fc_HP           100.00                  % 0 = no high-pass filter. If scalar value supplied for cut-off frequency Fc_HP then high-pass filter applied.
smooth_F        1                       % 0 = no smoothing filter applied, 1= smoothing filter used. (Not used for MIT or CIPIC)
plot_F2         1                       % 0 = no plot, 1 = 2D-plan, shows image rooms and sources on constant z plane."
plot_F3         1                       % 0 = no plot, 1 = 3D-plot, shows image sources. Rotatable in MATLAB"
alpha_F         0                       % 0 = fixed transparent surfaces for Room Geometry plot, 1 = (surface opacity = reflectivity)

Lx              6.25                    % Room Length (m) x direction
Ly              3.50                    % Room Width (m) y direction
Lz              2.75                    % Room Height (m) z direction
xp              1                       % Receiver x co-ordinate (m)
yp              1.75                    % Receiver y co-ordinate (m)
zp              1.2                     % Receiver z co-ordinate (m) e.g. 1.2m is a typical height above floor of ears of seated human subject
Receiver        two_mic                 % Receiver, copy one of these:   one_mic   two_mic   mithrir cipicir NB all strings must be same length for later logical tests.
sensor_space    0.15                    % Sensor separation in 'two_mic' case (0.145 m is CIPIC average)
MIT_root        MIT_HRTF                % The root of MIT Kemar data base file
subdir1         Kemar                   % MIT sub-directory1 name
subdir2         compact                 % MIT sub-directory2 name
filename        hrir_final.mat          % MIT filename for HRIR
CIPIC_root      CIPIC_HRTF              % The root of the CIPIC data base files
subdir1         standard_hrir_database  % CIPIC sub-directory1 name
subdir2         subject_                % CIPIC sub-directory2 name (program combines this with subject number)
S_No            021                     % CIPIC subject number, format '&&&' (e.g. '021' is the Kemar with small pinnae)
filename        hrir_final.mat          % CIPIC filename for HRIR
receiver_yaw    0                       % Yaw (Azimuth) offset of receiver system (degrees) +ve slew left
receiver_pitch  0                       % Pitch (Elevation) offset of receiver system (degrees) +ve nose up
receiver_roll   0                       % Roll offset of receiver system (degrees) +ve right wing down
%-------------------------------------------------------------------------------------------------
        Set the room surface absorptions
F_abs 125   250   500   1000  2000  4000  %Standard measurement frequencies (Hz)
Ax1   0.2   0.15  0.1   0.05  0.04  0.05  %Absorption of wall in x=0 plane (behind Receiver in plan)
Ax2   0.07  0.3   0.5   0.7   0.7   0.6   %Absorption of wall in x=Lx plane (front in plan)
Ay1   0.3   0.1   0.05  0.04  0.07  0.1   %Absorption of wall in y=0 plane (right in plan)
Ay2   0.3   0.1   0.05  0.04  0.07  0.1   %Absorption of wall in y=Ly plane (left in plan)
Az1   0.4   0.3   0.2   0.2   0.15  0.1   %Absorption of floor i.e. z=0 plane
Az2   0.1   0.3   0.8   0.85  0.75  0.65  %Absorption of ceiling i.e. z=Lz plane
%-------------------------------------------------------------------------------------------------
Directionality  Single/Left     Right_(if present)
azim_off        30              -30             % ]azimuth offset for sensor (degrees)
elev_off        30              30              % ]elevation offset for sensor
roll_off        0               0               % ]roll offset for sensor
SENSOR_root     SENSOR          SENSOR          % The root of the SENSOR data base file
subdir1         Types           Types           % SENSOR sub-directory1 name
filename        cardoid.mat     cardoid.mat     % Sensor filenames for Impulse Response
%-------------------------------------------------------------------------------------------------
SOURCES  R_s (m)    alpha (deg)  beta (deg)   % R_s = Radial distance(s) of source(s) from head (m).
1        0.80       30           0            % alpha = Azimuth(s) of sources -180<alpha <180 (deg) NB +ve is ACW on xy plane.
2        2.00       -30          0            % beta = Elevation(s) of sources -90<beta <90 (deg).
```

Figure D.10 – Fichier de configuration pour le calcul des réponses impulsionnelles dans le cas échoïque.

Annexe E

Prise de son binaurale

E.1 Introduction **266**
E.2 Description de l'environnement acoustique **266**

Cette annexe présente les réponses impulsionnelles stéréo qui ont été mesurées dans le cadre de la phase de qualification de l'extension G.729.1/G.718 *Superwideband* stéréo. Ces nouvelles réponses impulsionnelles sont vouées à être intégrées dans l'outil de réverbération du logiciel de l'ITU-T.

E.1 Introduction

Afin de simuler les effets de la réverbération dans les salles de vidéoconférence, un outil de réverbération a été créé pour la STL2005 [1] et des réponses impulsionnelles de salles de réunion de type mono ont également été mesurées et incluses dans la STL2005. Au cours de l'extension stéréo du G.729.1/G.718 SWB, il a été décidé que des réponses impulsionnelles stéréos étaient nécessaires. Ainsi, lors de la dernière réunion SG12, une liste de scénarii a été réalisée afin d'évaluer l'extension stéréo du G.729.1/G.718 SWB. Deux scénarii ont été introduits pour l'évaluation de la qualité et des traitements, et France Télécom a proposé de fournir toutes les réponses impulsionnelles nécessaires à cette phase de qualification. Dans ce document, nous présentons uniquement le contexte dont nous nous servons dans ce mémoire de thèse pour l'évaluation de nos méthodes bicapteurs.

E.2 Description de l'environnement acoustique

Le scénario correspond à une petite salle de conférence dont on donne un aperçu à la Figure E.1. La pièce comporte 7 sièges qui peuvent être disposés selon 7 positions possibles dans une zone de +/- 45 degrés. La Figure E.2 montre la configuration des microphones en fonction des différents types de prise de son qui ont été réalisées (MS pour *Mid Sid*, AB pour couple AB et Binaurale). Cependant, seule la prise de son binaurale nous intéresse. Les microphones utilisés sont deux DPA 4060 omnidirectionnels insérés dans les oreilles d'une tête artificielle.

Figure E.1 – Salle de vidéoconférence.

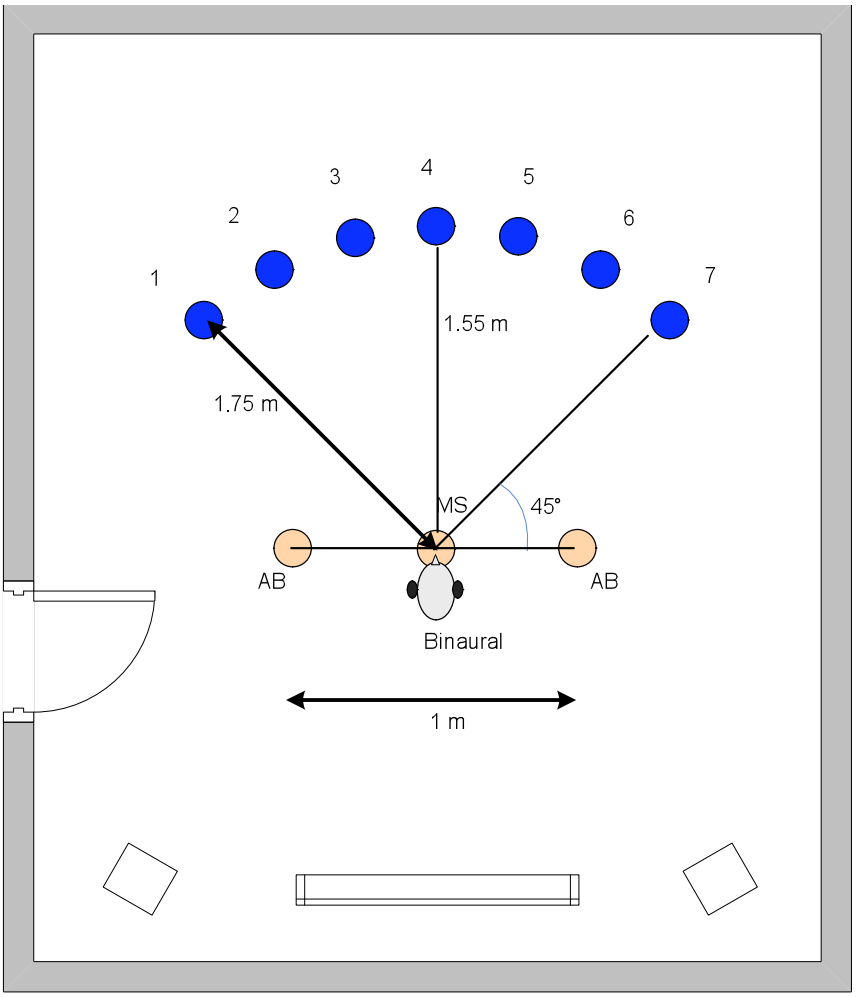


Figure E.2 – Description schématique des différentes configurations de la prise de son.