

**UNIVERSITE BLIDA 1**  
**Faculté de Technologie**  
Département d'Electronique

**Thèse de Doctorat**  
En Electronique

METHODES AUTOMATIQUES ET ROBUSTES POUR LE  
REHAUSSEMENT DU SIGNAL DE PAROLE APPLIQUEES A LA  
TELEPHONIE MAINS-LIBRES

Par

**Meriem ZOULIKHA**

Devant le jury composé de:

Z.BENSELAMA	Professeur, U.Blida1	Président
A. MELLIT	Professeur, U. de Jijel	Examineur
M.SOLTANE	MCA, U. de Médéa	Examineur
L.MITICHE	Professeur,U. de Djelfa	Examineur
M.DJENDI	Professeur, U.Blida1	Directeur de thèse
A.GUESSOUM	Professeur, U.Blida1	Co-Directeur de thèse

Blida, le 09/10/2021

## Résumé

La technologie de la communication mains-libres en voiture offre la sécurité et la flexibilité de communiquer tout en conduisant. Malheureusement, dans les conversations par ce type de système, 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. L'objectif des travaux effectués dans le cadre de cette thèse est de fournir des méthodes efficaces de réduction de bruit et du rehaussement du signal de parole, en utilisant deux microphones. Nous nous sommes intéressés à la structure de séparation aveugle de sources (SAS) de type directe et récursive. Ces deux structures ont comme limitation pratique, l'utilisation d'un système de détection d'activité vocale manuel (DAVM) afin de contrôler l'adaptation des filtres de séparation. Dans cette thèse nous proposons trois types de méthodes consacrées au contrôle automatique d'adaptation. Le premier type correspond aux versions améliorées des structures classiques de SAS. Ces méthodes utilisent un système de DAVA automatique basé sur l'estimation des rapports signal à bruit (*RSB*) sur les deux voies de la structure de SAS. Le deuxième type correspond aux solutions structurelles qui reposent sur la simplification du système de DAVA des méthodes améliorées. Le nouveau système de DAVA simplifié permet de contrôler automatiquement la mise à jour des filtres adaptatifs à faible complexité par rapport aux méthodes améliorées. Le troisième type correspond aux solutions algorithmiques, ce type de méthode est basée sur un nouvel algorithme de filtrage adaptatif à pas variable utilisé avec la structure de SAS directe afin d'ajuster les filtres de séparation et contrôler automatiquement leur adaptation. Les résultats expérimentaux obtenus montrent les bonnes performances de ces propositions. Une étude comparative entre ces trois types de méthodes est présentée.

**Mots-clés :** Rehaussement de la parole, Réduction de bruit, Séparation Aveugle de Sources (SAS), filtrage adaptatif, détection d'activité vocale (DAV).

## Abstract

Hands-free communication technology in car offers the security and flexibility to communicate while driving. Unfortunately, in conversations with this type of system, it is difficult to guarantee the speech signal quality where the recorded microphone signals are corrupted by noise. The main objective of the works reported in this thesis is to provide an efficient noise reduction and speech enhancement method, using two microphones. We are interested in the forward and backward Blind Source Separation (BSS) structures. The practical limitation of these two structures is the use of a Manual Voice Activity Detection system (MVAD) to control the separation filters adaptation's. In this thesis we propose three types of methods dedicated to automatic adaptation control. The first type corresponds to the improved versions of the classical BSS structures. These methods use an Automatic Voice Activity Detection (AVAD) system based on the signal-to-noise ratios (*SNR*) estimation on both separation structure channels. The second type corresponds to the structural solutions which are based on the simplification of the improved methods AVAD system. The new simplified AVAD system allows to control the separation filters adaptation's automatically with low complexity compared to the improved methods. The third type corresponds to an algorithmic solutions, this type of method is based on a new variable step sizes algorithm used with the forward BSS structure in order to adjust the separation filters and control their adaptation automatically. The experimental results obtained show the good performances of these propositions. A comparative study between these three types of methods is presented.

**Keywords :** Speech enhancement, Noise reduction, Blind source separation (BSS), Adaptive filtering, Voice activity detection (VAD).

## ملخص

توفر تقنية الاتصال بدون استخدام اليدين داخل السيارة الأمان والمرونة أثناء القيادة. لكن في المحادثات بهذا النوع من الأنظمة، يصعب ضمان جودة إشارة الكلام أثناء إلتقاط الصوت عن بعد، لان المكروفونات تلتقط بالإضافة الي الإشارة المطلوبة الضوضاء. الهدف من العمل المنفذ في إطار هذه الأطروحة هو توفير طرق فعالة للحد من الضوضاء وتعزيز إشارة الكلام باستخدام مكرفونين. لقد اهتمنا بالهيكليين (BSS) المباشر والرجعي لفصل المصادر. العائق الرئيسي لهذين الهيكليين هو استخدام نظام (VAD) يدوي للكشف عن فترات وجود الإشارة الصوتية. نقترح في هذه الاطروحة، ثلاثة انواع من الطرق المخصصة للتحكم التلقائي في تكيف مرشحات الفصل. النوع الأول يتوافق مع الإصدارات المحسنة لهياكل BSS الكلاسيكية. تستخدم هذه الطرق نظام AVAD التلقائي الذي يستند على تقدير نسبة الإشارة إلى الضوضاء (SNR). النوع الثاني يتوافق مع الحلول الهيكلية التي تعتمد على تبسيط نظام AVAD للطرق المحسنة. يتحكم نظام AVAD الجديد والمبسط تلقائيًا في تحديث المرشحات التكيفية بتعقيد منخفض مقارنة مع الطرق المحسنة. النوع الثالث يتوافق مع الحلول الخوارزمية ، ويستند هذا النوع على استخدام خوارزمية الترشيح التكيفي بخطوة متغيرة مع هيكلي BSS من اجل ضبط مرشحات الفصل والتحكم التلقائي في تكيفها. اظهرت النتائج التجريبية التي تم الحصول عليها الأداء الجيد لهذه المقترحات . كما قدمنا مقارنة بين هذه الأنواع الثلاثة من الأساليب.

**الكلمات المفتاحية :** تعزيز إشارة الكلام، الحد من الضوضاء، الفصل الأعمى للمصادر (BSS) ، الترشيح التكيفي، الكشف عن فترات وجود الإشارة الصوتية (VAD).

## **DEDICACE**

Je dédié ce modeste travail

à mon très cher père **Daoud ZOULIKHA**,

à ma mère et ma tante Djamila,

à ma sœur et mes frères, et

à tous mes amis en particulier Houda BENIELLES, Aouda BOUNIF,  
Khadidja MOKHNACHE, Meriem SAFI et Houcine GUESMI.

## REMERCIEMENTS

Je tiens à remercier en tout premier lieu Dieu tout puissant, de m'avoir permis de poursuivre mes objectifs et surmonter toutes les dures épreuves et moments difficiles pour aboutir à la fin de ce travail.

J'adresse mes plus vifs remerciements à mon directeur de thèse, Monsieur Mohamed DJENDI, Professeur à l'université de Blida 1, pour ses conseils précieux, sa bonne humeur, sa grande qualité scientifique et pédagogique, et pour avoir encadré mes premiers pas dans la recherche lors de mon Ingéniorat, mon Magister et tout au long de ce travail de thèse de doctorat. Je le remercie pour sa patience infinie, ainsi que son suivi intensif malgré sa charge académique et professionnelle.

Mes remerciements s'adressent aussi à mon co-directeur de thèse, Monsieur Abderrezak GUESSOUM, Professeur à l'université de Blida 1 et responsable du laboratoire LATSI, pour ses conseils positifs et ses encouragements.

Je tiens à remercier sincèrement Monsieur Zoubir BENSELAMA, Professeur à l'université de Blida 1, d'avoir accepté de présider le jury de ma thèse.

Mes remerciements vont également aux membres du jury, Monsieur Adel MELLIT, Professeur à l'université de Jijel, Monsieur Mohamed SOLTANE, Maître de conférences (A) à l'université de Médéa et Monsieur Lahcene MITICHE Professeur à l'université de Djelfa , qui ont bien voulu lire, corriger, et valoriser ma modeste thèse.

Enfin, j'exprime toute ma gratitude à ceux qui, de près ou de loin, chacun à sa manière, ont contribué à la concrétisation de cette thèse.

## TABLE DES MATIERES

<b>RESUME</b> .....	01	
<b>DEDICACE</b> .....	04	
<b>REMERCIEMENTS</b> .....	05	
<b>TABLE DES MATIERES</b> .....	06	
<b>LISTE DES ILLUSTRATIONS, GRAPHIQUES ET TABLEAUX</b> .....	10	
<b>INTRODUCTION</b> .....	14	
<b>CHAPITRE 1 : ETAT DE L'ART DE LA REDUCTION DE BRUIT</b>		
1.1	Introduction.....	18
1.2	Généralités sur la parole, le bruit et le filtrage adaptatif.....	18
1.2.1	Définition de la parole.....	18
1.2.2	Production du signal de parole.....	19
1.2.3	Classification des sons de la parole.....	20
1.2.3.1	Les sons voisés.....	20
1.2.3.2	Les sons non-voisés.....	20
1.2.3.3	Les silences.....	21
1.2.4	Perception de la parole.....	21
1.2.5	Nature et caractéristiques du bruit.....	21
1.2.6	Principe du filtre de Wiener.....	22
1.2.7	Algorithme du gradient stochastique LMS.....	26
1.2.8	Algorithmes du gradient stochastique normalisé NLMS.....	27
1.2.9	Filtrage adaptatif.....	28
1.3	Réduction du bruit .....	29
1.3.1	Techniques de réduction du bruit mono-capteur.....	29
1.3.1.1	Soustraction spectrale.....	31
1.3.1.2	Filtrage de Wiener.....	32
1.3.2	Techniques de réduction du bruit multi-capteurs.....	33

1.3.2.1	Annulation adaptative du bruit (ANC).....	33
1.3.2.2	Séparation aveugle de sources (SAS).....	36
1.4	Conclusion.....	37

## **CHAPITRE 2 : REDUCTION DE BRUIT PAR SEPARATION AVEUGLE DE SOURCES (SAS)**

2.1	Introduction.....	39
2.2	Modèle et problème de la séparation de sources.....	39
2.2.1	Modélisation du mélange.....	40
2.2.1.1	Les différents types de mélanges .....	41
2.2.1.2	Modèle de mélange convolutif d'une prise de son .....	43
2.2.2	Solution de la séparation de sources .....	45
2.2.2.1	La structure réursive ou Backward .....	45
2.2.2.2	La structure directe ou Forward .....	47
2.3	Contrôle des deux structures directe et réursive par un système de DAV .....	49
2.4	Limitation des structures de SAS classiques .....	51
2.5	Description des signaux et des critères de comparaison utilisés en simulations.....	52
2.5.1	Description des signaux.....	52
2.5.2	Description des chemins acoustiques.....	53
2.5.3	Détection d'activité vocale (DAV).....	56
2.5.4	Description des critères de comparaisons .....	56
2.5.4.1	Critères objectifs de qualité.....	57
2.5.4.2	Critères subjectifs de qualité.....	59
2.5.4.3	Analyse de la complexité.....	59
2.6	Conclusion.....	60

## **CHAPITRE 3 : AMELIORATION DES TECHNIQUES DE SEPARATION AVEUGLE DE SOURCES (DIRECTE ET RECURSIVE)**

3.1	Introduction.....	61
3.2	Structures améliorées.....	61
3.2.1	Structure réursive améliorée.....	62
3.2.2	Principe de fonctionnement des structures améliorées.....	62
3.2.2.1	Estimation des rapports signaux à bruit RSB .....	63



3.2.2.2	Contrôle des pas d'adaptations des filtres principaux.....	64
3.2.3	Structure directe améliorée.....	67
3.3	Résultats expérimentaux .....	68
3.3.1	Simulation de la structure directe améliorée (AFBSS).....	69
3.3.2	Etude des performances de la structure directe améliorée AFBSS .....	74
3.3.2.1	Effet de la taille du filtre adaptatif $L$ .....	75
3.3.2.2	Effet du pas d'adaptation $\mu$ .....	75
3.3.2.3	Effet du $RSB$ du mélange .....	76
3.3.3	Comparaison entre les structures AFBSS et FBSS.....	77
3.3.3.1	Représentation du signal rehaussé.....	77
3.3.3.2	Évaluation de la distance cepstrale (DC).....	78
3.3.3.3	Évaluation du rapport signal à bruit (RSB).....	79
3.3.3.4	Évaluation du système mismatch (SM).....	80
3.3.3.5	Evaluation de la complexité.....	81
3.4	Conclusion .....	82

## **CHAPITRE 4 : SOLUTIONS STRUCTURELLES ET ALGORITHMIQUES POUR LA SEPARATION AVEUGLE DE SOURCES**

4.1	Introduction.....	83
4.2	Solutions structurelles proposées.....	83
4.2.1	Solution 1 : Structure automatique de séparation de sources directe à complexité réduite (LC-AFBSS).....	84
4.2.1.1	Etage 1 : SAS.....	85
4.2.1.2	Etage 2 : Bloc de contrôle automatique de la DAV (DAVA).....	86
4.2.2	Solution 2 : Structure automatique de séparation de sources récursive à complexité réduite (LC-ABBSS).....	88
4.3	Evaluation de la première solution structurelle LC-AFBSS.....	90
4.4	Comparaison entre les structures LC-AFBSS et AFBSS.....	92
4.5	Solution algorithmique proposée.....	96
4.5.1	Solution 3 : Algorithme régularisé pour la FBSS (RFBSS).....	96
4.5.1.1	Méthode de Newton .....	97
4.5.1.2	Dérivation de l'algorithme RFBSS proposé .....	98
4.6	Evaluation de la solution algorithmique RFBSS.....	101

4.7	Comparaison entre les structures RFBSS et LC-AFBSS.....	104
4.8	Conclusion.....	108
	<b>CONCLUSION.....</b>	<b>110</b>
	<b>APPENDICE A : LISTE DES ABBREVIATIONS.....</b>	<b>112</b>
	<b>APPENDICE B : LISTE DES SYMBOLES.....</b>	<b>114</b>
	<b>REFERENCES.....</b>	<b>116</b>
	<b>PUBLICATIONS ET CONFERENCES.....</b>	<b>122</b>

## LISTE DES ILLUSTRATIONS, GRAPHIQUES ET TABLEAUX

### LISTE DES FIGURES

Figure 1	Communication téléphonique mains-libres en automobile. Chaîne de traitement pour l'amélioration de la parole : la Réduction de bruit (Noise Reduction) (NR), l'Annulation d'écho acoustique (Acoustic Echo Cancellation) (AEC) et la Déréverbération (DR) .....	16
----------	---	----

#### Chapitre 1

Figure 1.1	Représentation de l'appareil phonatoire humain.....	19
Figure 1.2	Modèle simple de production de la parole .....	20
Figure 1.3	Représentation schématique d'un filtre de Wiener.....	23
Figure 1.4	Schéma général d'un système de filtrage adaptatif .....	29
Figure 1.5	Modèle de débruitage mono-capteur .....	30
Figure 1.6	Structure d'annulation adaptative de bruit (ANC) .....	35

#### Chapitre 2

Figure 2.1	Structure générale du problème de la séparation aveugle de sources .....	40
Figure 2.2	Schéma général de la relation sources-observations en présence de bruit .....	40
Figure 2.3	Exemple d'un mélange instantané de sources.....	42
Figure 2.4	Illustration d'un mélange convolutif acoustique à deux sources et deux capteurs.....	42
Figure 2.5	Modèle général du mélange convolutif .....	44
Figure 2.6	Modèle simplifier du mélange convolutif .....	44
Figure 2.7	Structure de séparation de sources récursive (BBSS) .....	45
Figure 2.8	Structure de séparation de sources directe (FBSS) .....	47
Figure 2.9	Schéma bloc du principe général des structures de séparation de sources classiques .....	50
Figure 2.10	Les étapes de la structure FBSS classique .....	51
Figure 2.11	Signal de parole utilisé, $F_e = 16$ KHz.....	53
Figure 2.12	Signal du bruit Blanc utilisé, $F_e = 16$ KHz.....	53

Figure 2.13 Exemples d'une réponse impulsionnelle simulée ( $h_{12}(n)$ ), cas des capteurs espacés.....	55
Figure 2.14 Exemples d'une réponse impulsionnelle simulée ( $h_{21}(n)$ ), cas des capteurs espacés.....	55
Figure 2.15 Signal de parole utilisé avec une segmentation (DAV) manuelle .....	56
Figure 2.16 Organigramme des différents critères utilisés pour une étude comparative.....	57

### Chapitre 3

Figure 3.1 Schéma bloc du principe des structures de séparation de sources améliorées.....	61
Figure 3.2 Structure récursive améliorée (ABBSS).....	63
Figure 3.3 Illustration de la variation des pas par rapport aux $RSB$ en fonction de leurs seuils maximum et minimum sur chacune des voies.....	66
Figure 3.4 Diagramme d'activités du calcul des pas d'adaptations.....	67
Figure 3.5 Structure directe améliorée (AFBSS).....	68
Figure 3.6 Signaux sources (parole originale $s(n)$ , bruit Blanc $b(n)$ ), signaux de mélange ( $p_1(n)$ , $p_2(n)$ ) et signaux de sortie ( $u_1(n)$ , $u_2(n)$ ) de la structure AFBSS.....	70
Figure 3.7 Signal de parole original (en haut), le $RSB_1$ estimé au niveau de l'entrée primaire de la structure AFBSS (au milieu) et le pas d'adaptation $\mu_{w_{21}}(n)$ du filtre principal $w_{21}(n)$ (en bas).....	71
Figure 3.8 Signal de parole original (en haut), le $RSB_2$ estimé au niveau de l'entrée de référence de la structure AFBSS (au milieu) et le pas d'adaptation $\mu_{w_{12}}(n)$ du filtre principal $w_{12}(n)$ (en bas).....	72
Figure 3.9 Comportement des pas d'adaptations: $\mu_{w_{12}}(n)$ (en magenta) et $\mu_{w_{21}}(n)$ (en cyan).....	73
Figure 3.10 Signal de parole utilisé avec les deux segmentations (manuelle et automatique (AFBSS)), $F_e = 16$ KHz.....	74
Figure 3.11 Effet de la taille du filtre sur la structure directe améliorée AFBSS.....	75
Figure 3.12 Effet du pas d'adaptation sur la structure directe améliorée AFBSS.....	76
Figure 3.13 Effet du $RSB$ sur la structure directe améliorée AFBSS.....	77
Figure 3.14 ( $a_1$ , $a_2$ ) Signal de parole original $s(n)$ , ( $b_1$ , $b_2$ ) signal de mélange $p_1(n)$ , ( $c_1$ , $c_2$ ) signal de sortie $u_1(n)$ de la structure (AFBSS), ( $d_1$ , $d_2$ ) signal de sortie de la structure (FBSS).....	78

Figure 3.15 Evolution du critère de la distance cepstrale des deux structures AFBSS (en cyan) et FBSS (en bleu).....	79
Figure 3.16 Evolution du critère du rapport signal à bruit (RSB) pour les deux structures AFBSS (en cyan) et FBSS (en bleu).....	80
Figure 3.17 Comparaison du désajustement obtenu avec les deux structures AFBSS (en cyan) et FBSS (en bleu).....	81

#### Chapitre 4

Figure 4.1 Schéma bloc du principe des solutions structurelles .....	84
Figure 4.2 La première solution structurelle proposée LC-AFBSS .....	85
Figure 4.3 La deuxième solution structurelle proposée LC-ABBSS .....	88
Figure 4.4 Diagramme d'activités de la gestion des pas d'adaptations du 2 <sup>ème</sup> étage de la structure LC-ABBSS.....	89
Figure 4.5 Signaux sources (parole originale $s(n)$ , bruit Blanc $b(n)$ ), signaux de mélange ( $p_1(n)$ , $p_2(n)$ ) et signaux de sortie ( $u_1(n)$ , $u_2(n)$ ) de la structure LC- AFBSS.....	91
Figure 4.6 Signal de parole originale (en haut), l'énergie estimée $E_e$ au niveau de l'entrée primaire (LC-AFBSS) (au milieu) et le pas d'adaptation $\mu_{w_{21}}(n)$ du filtre principal $w_{21}(n)$ (en bas).....	92
Figure 4.7 Signal de parole originale $s(n)$ (à gauche), signal de sortie $u_1(n)$ de la structure LC-AFBSS (en haut), signal de sortie de la structure AFBSS (en bas).....	93
Figure 4.8 Evolution du critère de la distance cepstrale des deux structures LC-AFBSS (en gris) et AFBSS (en cyan).....	94
Figure 4.9 Evolution du critère du rapport signal à bruit (RSB) pour les deux structures LC-AFBSS (en gris) et AFBSS (en cyan).....	94
Figure 4.10 Comparaison du désajustement obtenu avec les deux structures LC-AFBSS (en gris) et AFBSS (en cyan) .....	95
Figure 4.11 Représentation graphique de la méthode de Newton.....	97
Figure 4.12 Signaux sources (parole originale $s(n)$ , bruit Blanc $b(n)$ ), signaux de mélange ( $p_1(n)$ , $p_2(n)$ ) et signaux de sortie ( $u_1(n)$ , $u_2(n)$ ) de la structure RFBSS....	102
Figure 4.13 L'évolution du pas d'adaptation $\mu_{w_{21}}(n)$ du filtre principal $w_{21}(n)$ pour la solution algorithmique RFBSS.....	103

Figure 4.14	Signal de parole originale $s(n)$ (à gauche), signal de sortie $u_1(n)$ de la structure RFBSS (en haut), signal de sortie $u_1(n)$ de la structure LC-AFBSS (en bas).....	104
Figure 4.15	Comparaison des évolutions des pas d'adaptations $\mu_{w21}(n)$ pour les deux méthodes LC-FBSS et RFBSS.....	105
Figure 4.16	Evolution du critère de la distance cepstrale des deux structures RFBSS (en rouge) et LC-AFBSS (en gris).....	106
Figure 4.17	Evolution du critère du rapport signal à bruit (RSB) pour les deux structures RFBSS (en rouge) et LC-AFBSS (en gris).....	106
Figure 4.18	Comparaison du désajustement obtenu avec les deux structures RFBSS (en rouge) et LC-AFBSS (en gris).....	107

#### LISTE DES TABLEAUX

Tableau 1.1	Différentes propriétés du bruit acoustique .....	22
Tableau 1.2	Les différentes grandeurs intervenant dans le filtre de Wiener .....	23
Tableau 1.3	Complexité algorithmique de l'algorithme LMS.....	27
Tableau 1.4	Complexité algorithmique de l'algorithme NLMS .....	28
Tableau 2.1	Echelle de qualité utilisée lors du test MOS [38].....	59
Tableau 3.1	Paramètres de simulations des structures (Mélange convolutif, FBSS, AFBSS).....	69
Tableau 3.2	Comparaison de la complexité pour les structures AFBSS et FBSS.....	81
Tableau 4.1	Les équations du premier étage de la solution structurelle LC-ABBSS .....	89
Tableau 4.2	Les valeurs optimales des paramètres utilisées dans la solution structurelle LC-AFBSS.....	91
Tableau 4.3	La complexité de la solution structurelle LC-AFBSS .....	95
Tableau 4.4	La complexité de la solution algorithmique RFBSS .....	107
Tableau 4.5	Comparaison subjective des différentes méthodes de la SAS .....	108

## INTRODUCTION

La communication orale est un mode de transmission d'information incontournable, que ce soit dans sa forme la plus simple, comme la conversation face-à-face (*i.e.* le locuteur et l'auditeur se parlent en se faisant face) ou au moyen d'un système de télécommunication qui permet d'établir une conversation entre deux individus situés loin l'un de l'autre (*i.e.* communication à distance). Quel que soit le mode de communication utilisé, l'information véhiculée dans le signal de parole doit être de bonne qualité afin de faciliter sa compréhension.

Dans la dernière décennie, la communication vocale a vu un développement croissant dans différentes applications, telles que les communications mobiles, la téléconférence, les opérateurs mains-libres, les aides auditives, etc. Les communications mobiles sont devenues un moyen indispensable afin d'assurer une conversation vocale entre deux personnes distantes. En raison de la flexibilité offerte par le réseau téléphonique mobile, les utilisateurs peuvent communiquer à tout moment et n'importe quel point, donc il est devenu possible de téléphoner de partout (ou presque) dans des environnements aussi divers et variés que la rue, une gare ou bien encore une voiture.

Une contrainte de sécurité est apparue dans un grand nombre d'applications téléphoniques notamment dans l'environnement véhicule, puisqu'il est nécessaire de tenir un combiné téléphonique dans la main. En effet, le fait d'immobiliser une main du conducteur, réduit dangereusement la concentration et expose le conducteur à un danger certain. La téléphonie mains-libres à l'intérieur de l'habitacle des véhicules en est un exemple typique qui présente l'avantage aux automobilistes de converser de manière plus naturelle et plus libre lorsqu'ils conduisent leur véhicule, sans avoir à tenir un combiné. Le système mains-libres répond donc à un besoin de confort, de sécurité et à un besoin juridique, car dans de nombreux pays, la loi interdit de manipuler un téléphone portable en conduisant : « *Téléphoner ou conduire, il faut choisir* ». Pour répondre à ce besoin de sécurité, ce système doit fournir une qualité de communication assez bonne pour ne pas perturber le conducteur, qui doit rester concentrer sur sa conduite.

Les éléments principaux susceptibles qui altèrent la qualité d'une communication téléphonique mains libres à l'intérieur d'un véhicule sont :

**Le bruit ambiant** qui provient de multiples sources, telles que le moteur, les turbulences environnantes, le roulement du pneu sur la route ;

**L'écho acoustique** provoqué par le fait que le système d'acquisition acoustique (un ou plusieurs microphones) capte, en plus de la voix du conducteur, la voix du locuteur distant ;

**La réverbération** elle résulte de la déformation du signal reçu au niveau du microphone à cause des réflexions, diffractions, atténuations subies par le signal émis.

Ces effets indésirables perturbent surtout le locuteur distant, et pas directement le conducteur. Mais une meilleure compréhension par le locuteur distant permet une conversation plus fluide, et plus confortable pour les deux utilisateurs.

L'utilisation de la téléphonie mains-libres dans les voiture requiert donc la résolution de ces trois problèmes à travers une chaîne de traitement constituée de plusieurs modules :

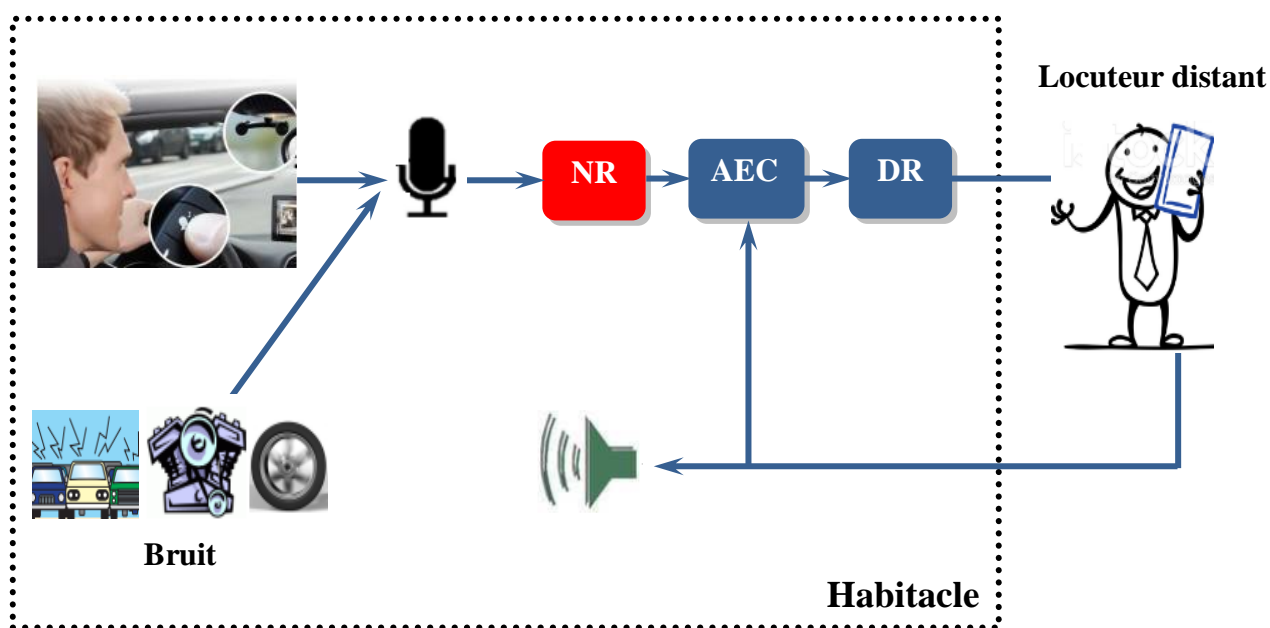
- Module de débruitage,
- Module d'annulation d'écho,
- Module de déréverbération.

La figure 1 illustre une communication téléphonique mains-libres en automobile. Dans les travaux présentés dans cette thèse, seule la réduction de bruit fera l'objet de notre étude.

Le problème de débruitage de la parole a été très largement abordé dans la littérature et se base généralement sur un système simple constitué d'un microphone. A partir d'une seule et unique observation constituée d'un mélange de parole et de bruit, le système doit être en mesure de séparer ces signaux en vue de restituer un signal moins bruité. Ceci est rendu réalisable par la formulation d'hypothèses sur les signaux : le bruit est généralement supposé stationnaire, ce qui permet de distinguer des composantes dynamiques de la parole.

L'objectif de ce travail réside dans l'amélioration de la qualité du traitement de débruitage, en se basant sur une observation supplémentaire par ajout d'un second microphone. Parmi les structures qui utilisent deux microphones dans leur fonctionnement, on trouve la structure de séparation aveugle de sources de type directe et récursive.





**Figure 1** : Communication téléphonique mains-libres en automobile. Chaîne de traitement pour l'amélioration de la parole : la Réduction de bruit (Noise Reduction) (NR), l'Annulation d'écho acoustique (Acoustic Echo Cancellation) (AEC) et la Déréverbération (DR).

A partir des travaux de la littérature [1-3], ces deux structures incluent une étape de contrôle d'adaptation dans leur fonctionnement mise en œuvre à partir d'une détection d'activité vocale (DAV) manuelle, malheureusement cette technique n'est pas réalisable en pratique.

Dans le cadre de ce travail de thèse, nous portant notre effort principal sur la proposition des nouvelles techniques de DAV automatique pour le contrôle d'adaptation des structures de séparation aveugle de sources (SAS).

La présente thèse est constituée de quatre chapitres organisés comme suit :

Le premier chapitre introduit des généralités sur le signal de parole, le bruit et le filtrage adaptatif, ainsi qu'un état de l'art des techniques mono/multi-capteurs utilisées dans le domaine du débruitage de la parole. Cette présentation est accompagnée d'une analyse systématique des avantages et les limites de chaque méthode.

Le deuxième chapitre présente, en premier lieu, une étude détaillée des structures de séparation de source classiques directe (FBSS) et récursive (BBSS) ainsi qu'une analyse de

leurs solutions dans le contexte d'annulation du bruit acoustique. La limitation pratique de ces structures accompagne cette présentation. En deuxième lieu, nous décrivons les signaux sources, le modèle de mélange convolutif et les critères de comparaison utilisés dans les simulations.

Le troisième chapitre a pour objet de présenter deux nouvelles structures AFBSS et ABBSS qui sont des versions améliorées des deux structures de séparation de sources (direct et récursive). Cette proposition a pour but de s'affranchir de l'utilisation de la DAV manuelle dans le fonctionnement de la structure de SAS. Les résultats de simulation obtenus par la méthode améliorée de type directe (AFBSS) sont comparés avec sa version classique de la littérature.

Le quatrième chapitre est réservé à la proposition d'une variété de solutions (structurelle et algorithmique) qui permettent, au prix d'une faible complexité, de contrôler automatiquement les deux structures de SAS. Des résultats de simulation sont présentés pour valider l'efficacité de nos solutions

Enfin, nous terminerons notre thèse par une conclusion.

## CHAPITRE 1

### ETAT DE L'ART DE LA REDUCTION DE BRUIT

#### 1.1 Introduction

Le domaine de la réduction de bruit est un domaine de recherche très actif et présent dans de nombreux champs d'applications du traitement de signal (téléphonie, restauration d'enregistrements audio anciens, analyse des signaux biomédicaux, etc...). L'objectif des techniques de réductions de bruit est de restaurer un signal utile (la parole) à partir d'observations corrompues par un bruit. Dans les problématiques de cette thèse, qui concernent en particulier la réduction du bruit de signaux vocaux, nous faisons toujours appel à des algorithmes de filtrage adaptatif.

Dans ce chapitre, nous allons présenter dans un premier temps, les principales caractéristiques de la parole et du bruit. Nous présenterons aussi un rappel théorique sur le filtrage adaptatif en commençant par le filtre de Wiener, puis les algorithmes de types gradient stochastique, et ensuite le principe de base du filtrage adaptatif.

Dans un deuxième temps, nous présentons quelques méthodes de débruitage de la parole. La classification que nous adoptons pour ces méthodes repose sur le nombre d'observations disponibles (lié directement au nombre de microphones utilisés).

#### 1.2 Généralités sur la parole, le bruit et le filtrage adaptatif

##### 1.2.1 Définition de la parole

La parole est un moyen de communication très efficace et naturel de l'humain; elle se distingue des autres sons par ses caractéristiques acoustiques qui ont leur origine dans les mécanismes de production. Elle apparaît physiquement comme une variation de la pression de l'air causée et émise par le système articulatoire. L'unité de parole de plus petite taille est un phonème (voyelle ou consonne). Le nombre de phonèmes est toujours très limité, normalement inférieur à cinquante. Par exemple : la langue française comprend 36 phonème [4]. La phonétique acoustique étudie ce signal en le transformant dans un premier temps en signal électrique grâce à un transducteur approprié : le microphone. De nos jours, le signal électrique résultant est le plus souvent numérisé.

### 1.2.2 Production du signal de parole

La parole est un signal complexe résultant de mouvements coordonnés de plusieurs muscles du corps humain [5]. Elle est produite par le système articulaire présenté à la figure 1.1 :

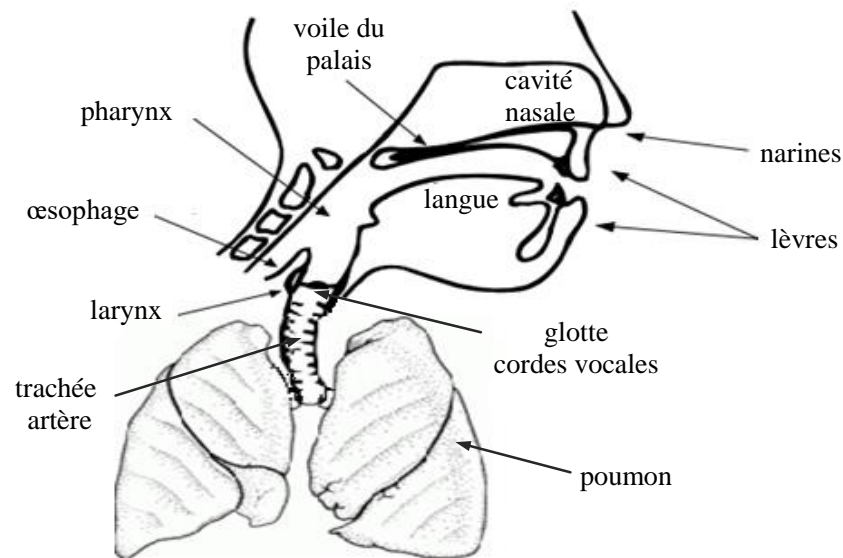


Figure 1.1 : Représentation de l'appareil phonatoire humain.

L'appareil respiratoire fournit l'air qui est expiré par la trachée-artère. L'air passe ensuite par le larynx où sa pression est modulée grâce aux cordes vocales qui déterminent la taille de l'ouverture (la glotte) par laquelle il peut passer. Finalement, l'air transite par le conduit vocal qui s'étend du pharynx aux lèvres pour devenir le signal vocal qui est émis par le locuteur.

La parole peut être *modélisée par le* résultat de l'excitation du conduit vocal par un train d'impulsions ou un bruit donnant lieu respectivement aux sons voisés et non voisés (voir figure 1.2). Dans le cas des sons voisés, l'excitation est une vibration périodique des cordes vocales suite à la pression exercée par l'air provenant de l'appareil respiratoire. Ce mouvement vibratoire correspond à une succession de cycles d'ouverture et de fermeture de la glotte. Le nombre de ces cycles par seconde correspond à la fréquence fondamentale. Quant aux signaux non-voisés, l'air passe librement à travers la glotte sans provoquer de vibration des cordes vocales [6].

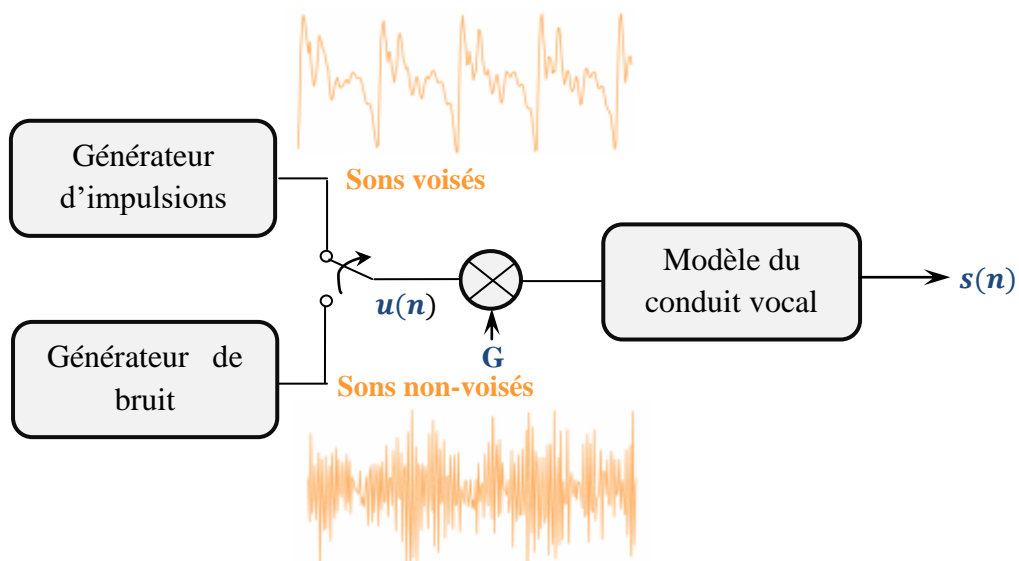


Figure 1.2 : Modèle simple de production de la parole.

### 1.2.3 Classification des sons de la parole

La parole se distingue par sa complexité, due à des modes de production variables, engendrant des signaux fortement non-stationnaires. Cependant, du fait de l'inertie des muscles de l'appareil phonatoire, le signal vocal ne peut être considéré comme quasi-stationnaire que sur des intervalles de temps de durée limitée [7] donc le traitement du signal de parole se fait sur des fenêtres d'analyse de l'ordre de 20 à 40 ms. Les sons de la parole peuvent être classés en trois catégories :

#### 1.2.3.1 Les sons voisés

Un son voisé est par définition un signal quasi-périodique très riche en harmoniques d'une fréquence fondamentale, appelée « pitch ». Ce qui lui donne un caractère assez facilement prévisible. Il est de forte énergie [8]. Toutes les voyelles en français présentent les caractéristiques des sons voisés sans oublier une bonne partie des consonnes /b/, /d/ et /g/.....

#### 1.2.3.2 Les sons non-voisés

Un son non voisé ne présente pas de structure périodique tel que les consonnes /f/ et /s/..., il peut être considéré comme un bruit blanc filtré par le conduit vocal.

Son spectre ne possède donc pas de structure particulière et c'est souvent dans les hautes fréquences que le spectre est le plus énergétique [8].

### 1.2.3.3 Les silences

Les silences sont tout simplement des intervalles où le signal utile est absent. En pratique il s'agit de bruits, d'origines diverses, d'énergie négligeable devant celle du signal utile.

### 1.2.4 Perception de la parole

Le signal de parole est un vecteur acoustique porteur d'informations d'une grande complexité, variabilité et redondance. Les caractéristiques de ce signal sont appelées traits acoustiques. Chaque trait acoustique a une signification sur le plan perceptuel [6].

Le premier trait est la **fréquence fondamentale**, fréquence de vibration des cordes vocales. Ses variations définissent le pitch qui constitue la perception de la hauteur (où les sons s'ordonnent de grave à aigu). Seuls les sons quasi-périodiques (voisés) engendrent une sensation de hauteur tonale bien définie. La fréquence fondamentale peut varier selon le genre (masculin ou féminin) et l'âge du locuteur. Cette dernière peut varier comme suit :

- de 80 à 200 Hz pour une voix masculine,
- de 150 à 400 Hz pour une voix féminine,
- de 200 à 600 Hz pour une voix d'enfant.

Le deuxième trait est le **spectre fréquentiel** dont dépend principalement le **timbre** de la voix. Le timbre est une caractéristique permettant d'identifier une personne à la simple écoute de sa voix.

Le dernier trait acoustique est l'**énergie** correspondant à l'intensité sonore. Elle est habituellement plus forte pour les segments voisés de la parole que pour les segments non-voisés.

### 1.2.5 Nature et caractéristiques du bruit

On appelle bruit tout signal nuisible qui se superpose au signal utile en un point quelconque d'une chaîne de mesure ou d'un système de transmission. Il constitue donc une gêne dans la compréhension du signal utile, qui est dans notre cas, la parole. En physique, en acoustique et en traitement du signal, bien que le bruit soit, par nature, aléatoire, il possède

certaines caractéristiques statistiques, spectrales ou spatiales. Le tableau 1.1 représente les différentes classes auxquelles un bruit peut appartenir [6].

Tableau 1.1 : Différentes propriétés du bruit acoustique.

<b>Propriétés du bruit</b>	
Structure	: Continu / Impulsif / Périodique
Type d'interaction	: Additif / Multiplicatif/ Convolutif
Comportement temporel	: Stationnaire / Non-stationnaire
Bande de fréquence	: Etroite / Large
Dépendance	: Corrélé / Décorrélé
Propriétés statistiques	: Dépendants / Indépendants
Propriétés spatiales	: Cohérents / Incohérents

Comme notre but est orienté vers le débruitage de la parole pour des applications de téléphonie mains libres à l'intérieur d'un habitacle de voiture, on se limite dans notre étude aux mélanges convolutifs utilisant différents type de bruits, stationnaires, faiblement et fortement corrélés.

### 1.2.6 Principe du filtre de Wiener

La figure (1.3) montre une représentation schématique d'un filtre de Wiener  $w$  dans le contexte d'estimation d'une séquence désirée  $d(n)$  basée sur un signal de référence  $x(n)$ . La sortie du filtre est  $y(n)$ , l'erreur d'estimation est  $e(n)$  avec  $L$  étant la taille du filtre. Quand l'erreur d'estimation tend vers zéro, la sortie du filtre tend vers la séquence désirée.

Pour obtenir la configuration optimale du filtre, une méthode directe consiste à choisir une fonction appropriée de l'erreur d'estimation (parfois appelée fonction de performance ou fonction coût) qui permet d'obtenir les coefficients du filtre ( $w$ ) de telle sorte à optimiser cette fonction dans un certain sens. Plusieurs fonctions coût permettent d'achever des objectifs différents. Pour un processus déterministe, on peut minimiser l'erreur des moindres carrés (least square error ou LSE), alors que pour un processus stochastique, on minimise l'erreur quadratique moyenne (EQM ou MSE en anglais) [9].

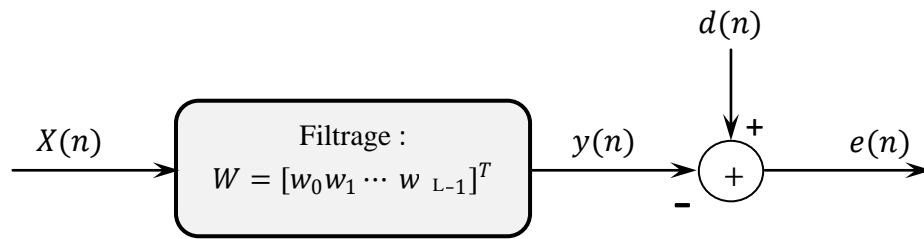


Figure 1.3 : Représentation schématique d'un filtre de Wiener.

Les différentes grandeurs utilisées dans le filtre de Wiener sont représentées sur le tableau 1.2.

Tableau 1.2 : Les différentes grandeurs intervenant dans le filtre de Wiener.

Grandeur	Filtre de Wiener
Entrée du filtre	$\mathbf{x}(n) = \begin{bmatrix} x(n) \\ x(n-1) \\ \vdots \\ x(n-L+1) \end{bmatrix}$
Coefficients du filtre	$\mathbf{w} = \begin{bmatrix} w_0 \\ w_1 \\ \vdots \\ w_{L-1} \end{bmatrix}$
Sortie du filtre	$y(n) = \sum_{k=0}^{L-1} w_k x(n-k) = \mathbf{w}^T(n)\mathbf{x}(n)$
Erreur d'estimation	$e(n) = d(n) - y(n)$

Dans le cas du filtre optimal de Wiener, la fonction coût, dénommée  $J$ , est donnée par le carré de l'erreur d'estimation. La fonction  $J$  est aussi appelée erreur quadratique moyenne.

$$J = E[e(n)^2] \quad (1.1)$$

où le signal d'erreur  $e(n)$  est donné par :

$$e(n) = d(n) - y(n) = d(n) - \mathbf{w}^T(n)\mathbf{x}(n) = d(n) - \mathbf{x}(n)^T\mathbf{w}(n) \quad (1.2)$$

avec  $(.)^T$  étant l'opérateur de la transposée.



En faisant l'hypothèse que le signal  $\{x\}$  est stationnaire, et si on introduit l'équation (1.2) dans l'équation (1.1), on peut écrire la fonction coût de la manière suivante [10]:

$$\begin{aligned}
 J &= E[(d(n) - \mathbf{w}^T(n) \mathbf{x}(n))^2] \\
 J &= E[d(n)^2 - 2 \mathbf{w}^T(n) \mathbf{x}(n) d(n) + \mathbf{w}^T(n) \mathbf{x}(n) \mathbf{x}(n)^T \mathbf{w}(n)] \\
 J &= E[d(n)^2] - 2 \mathbf{w}^T(n) E[\mathbf{x}(n) d(n)] + \mathbf{w}^T(n) E[\mathbf{x}(n) \mathbf{x}(n)^T] \mathbf{w}(n) \\
 J &= E[d(n)^2] - 2 \mathbf{w}^T(n) \mathbf{P}_{xd} + \mathbf{w}^T(n) \mathbf{R}_{xx} \mathbf{w}(n) \tag{1.3}
 \end{aligned}$$

où  $\mathbf{R}$  est une matrice d'autocorrélation du signal d'entrée  $\mathbf{x}(n)$  de taille  $(L \times L)$ , elle est symétrique et définie positive.

$$\mathbf{R}_{xx} = E[\mathbf{x}(n) \mathbf{x}(n)^T] \tag{1.4}$$

et  $\mathbf{P}$  est un vecteur d'intercorrélacion entre le signal d'entrée  $\mathbf{x}(n)$  et le signal désiré  $d(n)$  de taille  $M$  définie par :

$$\mathbf{P}_{xd} = E[\mathbf{x}(n) d(n)] \tag{1.5}$$

Les coefficients du filtre optimal ( $\mathbf{w}^{opt}$ ) sont obtenus en annulant le gradient de (1.1) par rapport aux coefficients du filtre [9].

$$\begin{aligned}
 \nabla J &= \frac{\partial J}{\partial \mathbf{w}(n)} \\
 \nabla J &= 2E[e(n) \frac{\partial e(n)}{\partial \mathbf{w}(n)}] \\
 \nabla J &= -2E[e(n) \mathbf{x}(n)] \tag{1.6}
 \end{aligned}$$

où  $E[.]$  et  $\nabla$  désignent respectivement la moyenne statistique et l'opérateur gradient.

La solution optimale correspond à une erreur quadratique minimale, notée  $e_{min}(n)$ , telle que:

pour  $\nabla J = 0$ , on aura

$$E[e_{min}(n) \mathbf{x}(n)] = 0 \tag{1.7}$$

C'est le *principe d'orthogonalité* signifiant que toutes les entrées  $\mathbf{x}(n)$  sont décorréliées de l'erreur  $e_{min}(n)$  [11].

En développant cette dernière équation on obtient :

$$E[\mathbf{x}(n)\mathbf{x}(n)^T]\mathbf{w}^{opt}(n) = E[\mathbf{x}(n) d(n)] \quad (1.8)$$

On peut la réécrire sous la forme :

$$\mathbf{R}_{xx}\mathbf{w}^{opt}(n) = \mathbf{P}_{xd} \quad (1.9)$$

L'expression (1.9) permet d'aboutir au filtre optimal de Wiener  $\mathbf{w}^{opt}$  donné par (1.10) plus connue sous le nom représentation matricielle de l'équation de Wiener-Hopf [12],

$$\mathbf{w}^{opt}(n) = \mathbf{R}_{xx}^{-1}\mathbf{P}_{xd} \quad (1.10)$$

avec  $\mathbf{w}^{opt}$  étant le vecteur optimum des coefficients.

Une première approche pour résoudre l'équation de Wiener-Hopf (1.10) consiste à utiliser l'algorithme du gradient déterministe. Afin de minimiser la puissance du signal d'erreur  $E[e(n)^2]$ , on utilise la méthode de la plus grande pente, qui consiste à rechercher à chaque itération la direction de la descente maximale (ou plus grande pente), pour obtenir les coefficients du filtre  $\mathbf{w}$  [9].

La méthode du gradient consiste à [11] :

- Choisir un vecteur initial de coefficients  $\mathbf{w}(0)$ .
- Obtenir, à partir d'un vecteur  $\mathbf{w}(n)$  donné, un vecteur  $\mathbf{w}(n + 1)$  par incrémentation de  $\mathbf{w}(n)$  dans la direction opposée du gradient de la fonction coût  $J$ .

$$\mathbf{w}(n + 1) = \mathbf{w}(n) - \frac{1}{2}\mu \nabla J|_{\mathbf{w}=\mathbf{w}(n)} \quad (1.11)$$

où  $\mu$  est une constante positive qui contrôle le taux de convergence et communément dénommée facteur de convergence ou pas d'adaptation.

Le gradient est égal à :

$$\nabla J|_{\mathbf{w}=\mathbf{w}(n)} = 2\mathbf{R}_{xx}\mathbf{w}(n) - 2\mathbf{P}_{xd} \quad (1.12)$$

La forme explicite de l'algorithme du gradient déterministe est donc :

$$\mathbf{w}(n + 1) = \mathbf{w}(n) + \mu (\mathbf{P}_{xd} - \mathbf{R}_{xx}\mathbf{w}(n)) \quad (1.13)$$

### 1.2.7 Algorithme du gradient stochastique LMS

L'algorithme du gradient stochastique est une approximation de l'algorithme du gradient déterministe. Cet algorithme adaptatif est connu sous le nom de Least-Mean-Square (LMS) en anglais. La mise en œuvre de cet algorithme repose sur la simplification du gradient de  $J$ ,  $\nabla J|_{\mathbf{w}=\mathbf{w}(n)} = 2\mathbf{R}_{xx}\mathbf{w}(n) - 2\mathbf{P}_{xd}$ , obtenue en remplaçant les estimations des coefficients de  $\mathbf{R}_{xx}$  et  $\mathbf{P}_{xd}$  fait à l'aide de moyennes statistiques par une estimation instantanée [13]:

$$\hat{\mathbf{R}}_{xx} = \mathbf{x}(n)\mathbf{x}(n)^T \quad (1.14)$$

$$\hat{\mathbf{P}}_{xd} = \mathbf{x}(n) d(n) \quad (1.15)$$

Remplaçant (1.14) et (1.15) dans l'expression du gradient déterministe qui est donnée par (1.13), on obtient l'algorithme du gradient stochastique :

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \mu e(n)\mathbf{x}(n) \quad (1.16)$$

où  $e(n)$  est donnée par la relation (1.2).

D'après [13], La condition nécessaire et suffisante de convergence en moyenne de l'algorithme (LMS) est :

$$0 < \mu < \frac{2}{\lambda_{max}} \quad (1.17)$$

où  $\lambda_{max}$  est la plus grande valeur propre de la matrice d'autocorrélation  $\mathbf{R}_{xx}$  donnée par (1.4).

En pratique, on utilise souvent, pour le choix de la constante  $\mu$ , une condition plus restrictive:

$$0 < \mu \leq \frac{2}{\text{Trace}(\mathbf{R}_{xx})} = \frac{2}{M\sigma_x^2} \quad (1.18)$$

où  $\text{Trace}(\mathbf{R}_{xx})$  représente la somme des éléments de la diagonale de la matrice  $\mathbf{R}_{xx}$  et  $\sigma_x^2$  désigne l'énergie du signal d'entrée  $\mathbf{x}(n)$ .

La complexité algorithmique de l'algorithme LMS est connue : chaque itération  $n$  comporte  $(2L+1)$  multiplications et  $(2L)$  additions. Le tableau (1.3) résume les opérations nécessaires pour l'algorithme LMS à chaque itération  $n$  [10].

Tableau 1.3 : Complexité algorithmique de l'algorithme LMS.

Etape de l'algorithme	Nombre d'additions	Nombre de multiplications
$e(n) = d(n) - \mathbf{x}(n)^T \mathbf{w}$	$L$	$L$
$\mathbf{w}(n+1) = \mathbf{w}(n) + \mu e(n)\mathbf{x}(n)$	$L$	$L+1$
Total par itération	$2L$	$2L+1$

### 1.2.8 Algorithme du gradient stochastique normalisé NLMS

L'algorithme NLMS (Normalized Least Mean Square) consiste à normaliser le pas d'adaptation  $\mu$  dans l'algorithme LMS par rapport à l'énergie du signal d'entrée pour réduire au minimum l'effet de variation de la puissance du signal d'entrée et de rendre ainsi la convergence plus au moins uniforme en passant d'une étape d'adaptation à une autre. Dans l'équation (1.16), le pas d'adaptation  $\mu$  est alors remplacé par un pas d'adaptation  $\mu_n$  défini à chaque itération par [10]:

$$\mu_n = \frac{\mu}{\mathbf{x}(n)^T \mathbf{x}(n)} \quad (1.19)$$

La mise à jour des coefficients du filtre adaptatif par l'algorithme NLMS est alors donnée par:

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \frac{\mu}{\mathbf{x}(n)^T \mathbf{x}(n) + \beta} e(n)\mathbf{x}(n) \quad (1.20)$$

où  $\beta$  est un facteur permettant de suivre plus ou moins rapidement les variations d'énergie dans le signal d'entrée  $\{x\}$ .

La convergence de cet algorithme est garantie pour un pas d'adaptation  $0 < \mu < 2$ .

L'intérêt de l'algorithme NLMS par rapport à l'algorithme LMS est de rendre l'algorithme indépendant de la variance du signal d'entrée. Cependant, la distribution des valeurs propres de la matrice d'autocorrélation du signal d'entrée n'est en rien modifiée. Ceci implique précisément la même dépendance, dans les deux cas, de la convergence vis à vis de la statistique du signal d'entrée.

Pour les signaux stationnaires tels que le bruit blanc ou non-stationnaires tels que la parole, l'algorithme NLMS apporte une amélioration significative sur le taux de convergence par rapport au LMS grâce à la normalisation du pas d'adaptation.

Cet algorithme peut être plus complexe que l'algorithme LMS mais il reste toujours l'un des algorithmes les plus simples à mettre en application. Un des inconvénients de l'algorithme NLMS par rapport au LMS est l'augmentation de la complexité de calcul au niveau du nombre de multiplications. Le tableau (1.4) résume le nombre d'opérations nécessaires dans l'algorithme NLMS à chaque itération  $n$  [10].

Tableau 1.4 : Complexité algorithmique de l'algorithme NLMS.

Etape de l'algorithme	Nombre d'additions	Nombre de multiplications
$e(n) = d(n) - \mathbf{x}(n)^T \mathbf{w}$	$L$	$L$
$\mathbf{w}(n+1) = \mathbf{w}(n) + \mu \frac{e(n)\mathbf{x}(n)}{\mathbf{x}(n)\mathbf{x}(n)^T + \beta}$	$L+1$	$2L+1$
Total par itération	$2L+1$	$3L+1$

### 1.2.8 Filtrage adaptatif

Un filtre adaptatif est un filtre numérique dont les coefficients se modifient eux-mêmes en fonction des signaux extérieurs. Il est utilisé chaque fois qu'un environnement est mal connu ou changeant, ou pour supprimer des perturbations situées dans le domaine des fréquences du signal utile, ce que les filtres classiques ne peuvent pas faire [14].

Le filtre numérique peut être de type filtre à réponse impulsionnelle fini (RIF) ou filtre à réponse impulsionnelle infinie (IIR). Dans notre étude, les filtres adaptatifs que nous allons utiliser sont de type RIF.

Le rôle primordial d'un filtre adaptatif est d'ajuster un vecteur de paramètre  $\mathbf{w}$  pour un objectif bien défini (minimisation de l'EQM : erreur quadratique moyenne). La figure (1.4) schématise le principe d'un filtre adaptatif bouclé par un algorithme d'adaptation.

La technique de filtrage adaptatif se décompose classiquement en deux étapes [10] :

1. Une étape de filtrage qui permet d'obtenir une estimation du signal inconnu en convoluant le signal d'entrée  $\mathbf{x}(n)$  avec les coefficients du filtre adaptatif  $\mathbf{w}$ . L'erreur d'estimation  $e(n) = d(n) - y(n)$  est ensuite utilisée dans la partie adaptation pour mettre à jour les coefficients du filtre.
2. Une étape d'adaptation qui permet d'ajuster les coefficients du filtre adaptatif  $\mathbf{w}(n)$  suivant un algorithme donné.

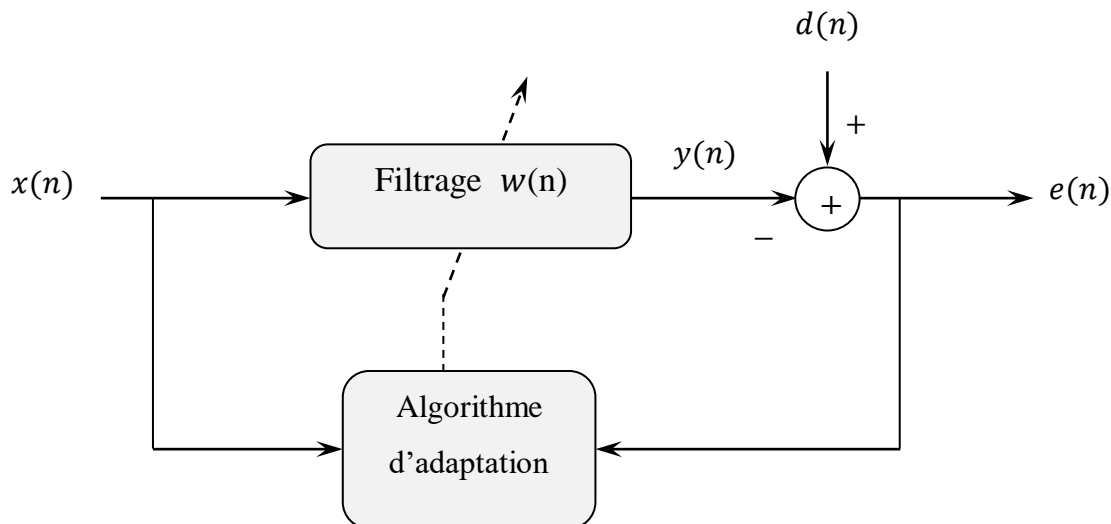


Figure 1.4 : Schéma général d'un système de filtrage adaptatif.

### 1.3 Réduction de bruit

#### 1.3.1 Techniques de réduction débruitage mono-capteur

Le domaine du rehaussement de la parole dans le cas où un seul microphone est utilisé pour capter le signal de parole dégradé par le bruit est riche en solutions et fait appel à de nombreux concepts de traitement du signal. Les traitements associés utilisent donc un seul signal pour assurer le débruitage. Plusieurs méthodes à base du traitement mono-capteur ont été proposées dans la littérature, on cite dans ce chapitre les plus couramment utilisées :

- Soustraction spectrale ;
- Filtrage de Wiener.

La configuration générale d'un système de débruitage mono-capteur du signal de parole utilisé pour ces méthodes est illustré sur la figure 1.5. Pour ce système, les signaux captés par le microphone sont :

- la parole prononcée par le locuteur, appelée par la suite signal utile et notée  $s(n)$ ;
- le bruit, noté  $b(n)$ , correspondant à l'ensemble des sources sonores captées par le microphone en dehors du signal utile.

L'objectif, dans le cas mono-capteur, consiste à estimer le signal utile  $s(n)$  qui est noyé dans le bruit additif  $b(n)$  supposé indépendant, à partir de la seule observation du signal bruité  $p(n)$  qui s'écrit comme la somme des deux composantes décrites jusqu'à présent :

$$p(n) = s(n) + b(n) \quad (1.21)$$

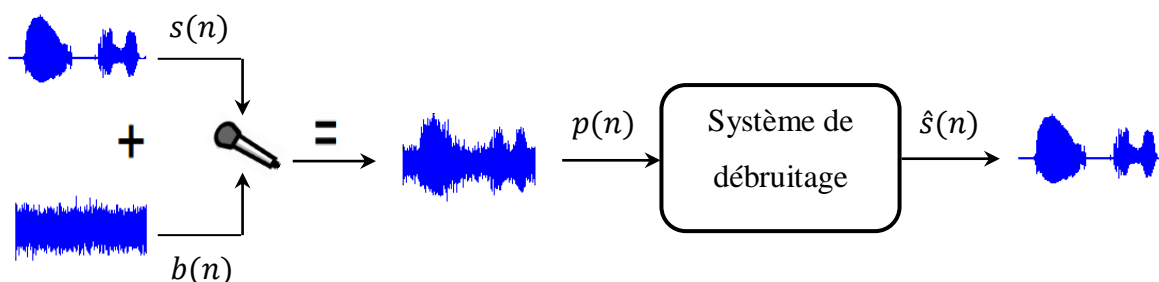


Figure 1.5 : Modèle de débruitage mono-capteur.

Nous désignons par  $\hat{s}(n)$  le signal rehaussé en sortie qui est une estimation de  $s(n)$ .

Sans hypothèse ou connaissance *a priori* sur les signaux, le problème est mal posé et ne peut être résolu. Il faut aussi des connaissances sur l'un ou l'autre des signaux, voire sur les deux, afin d'être en mesure de discriminer le signal utile du bruit ambiant. Les méthodes décrites dans ce paragraphe opèrent dans le domaine de Fourier. En supposant les signaux stationnaires et indépendants, la relation (1.1) s'écrit [15] :

$$\gamma_p(f) = \gamma_s(f) + \gamma_b(f) \quad (1.22)$$

où  $\gamma_p(f)$ ,  $\gamma_s(f)$  et  $\gamma_b(f)$  sont les densités spectrales de puissances (DSP) respectives des signaux  $p(n)$ ,  $s(n)$  et  $b(n)$ . Cette représentation sous forme de DSP n'est malheureusement pas exploitable en raison de la non-stationnarité du signal de parole. En effet, s'il est acceptable de considérer le bruit stationnaire, la parole ne peut être considérée comme telle que sur de courtes durées. Il devient alors possible d'exploiter la quasi-stationnarité de la parole sur des trames d'une durée de l'ordre de 20 à 40 ms. C'est pour cette raison que l'on considère plutôt les spectres instantanés, obtenus à l'aide de la transformée de Fourier à court terme (TFCT) sur des trames successives. Chaque trame du signal temporel  $p(n)$ , peut donc être représentée dans le domaine fréquentiel par son module  $|P(f, k)|$  et sa phase associée  $\Phi_p(f, k)$ , où  $f$  représente la fréquence et  $k$  l'indice de la trame courante [8].

La relation entre les composantes fréquentielles des signaux s'écrit :

$$P(f, k) = S(f, k) + B(f, k) = |P(f, k)|e^{j\Phi_p(f, k)} \quad (1.23)$$

Remarque 1 :

En tenant compte du fait que la perception auditive humaine est essentiellement guidée par le spectre d'amplitude alors que la connaissance de la phase n'apporte rien pour ce type de traitement [16].

### 1.3.1.1 Soustraction spectrale

La méthode de STEVEN F.BOLL [17] constitue le principe de base de la méthode de soustraction spectrale. Elle est basée sur la connaissance *a priori* de la densité spectrale de puissance du bruit. Cette technique a pour but d'obtenir une estimation de l'amplitude du spectre à court terme du signal utile, en soustrayant explicitement l'amplitude du spectre de bruit à partir de celui du signal bruité. Cela suppose évidemment que le bruit et la parole sont non corrélés. En outre, étant donné que les techniques de soustraction spectrale nécessitent une estimation du bruit pendant les pauses (phases de silence), on suppose que les caractéristiques du bruit changent lentement (c.-à-d. le bruit est supposé stationnaire) [18]. L'amplitude estimée du spectre du signal débruité est donnée par l'équation suivante :

$$|\hat{S}(f, k)| = |P(f, k)| - \sqrt{E[|B(f, k)|^2]} \quad (1.24)$$

C'est la *soustraction spectrale d'amplitude* (SSA). On ne peut pas garantir que la valeur de  $|\hat{S}(f, k)|$  soit toujours positive car on soustrait une valeur moyennée  $\sqrt{E[|B(f, k)|^2]}$  au module du spectre instantané du signal bruité  $|P(f, k)|$  dont la variance est beaucoup plus importante. Toutefois, le module du spectre du signal estimé doit rester positif ou nul, une valeur négative n'ayant pas de signification physique, cette contrainte est satisfaite par un simple seuillage [8] :

$$|\hat{S}(f, k)| = \begin{cases} |P(f, k)| - \sqrt{E[|B(f, k)|^2]} & \text{si } |P(f, k)| \geq \sqrt{E[|B(f, k)|^2]} \\ 0 & \text{sinon.} \end{cases} \quad (1.25)$$

Si on détermine  $H(f, k)$  comme une fonction de transfert, le signal estimé  $\hat{S}(f, k)$  peut s'exprimer ainsi :

$$\hat{S}(f, k) = H(f, k)P(f, k) \quad (1.26)$$

Ici, on a :

$$H_{SSA}(f, k) = \begin{cases} 1 - \frac{\sqrt{E[|B(f, k)|^2]}}{|P(f, k)|} & \text{si } |P(f, k)| \geq \sqrt{E[|B(f, k)|^2]} \\ 0 & \text{sinon.} \end{cases} \quad (1.27)$$



D'autres règles basées sur le principe de la soustraction spectrale ont été proposées, l'une des plus connues étant la *soustraction spectrale en puissance* (SSP) [8] :

$$|\hat{S}(f, k)|^2 = |P(f, k)|^2 - E [|B(f, k)|^2] \quad (1.28)$$

De la même façon que pour la SSA, les valeurs négatives du module carré  $|\hat{S}(f, k)|^2$  sont à proscrire ce qui amène à contraindre le résultat :

$$|\hat{S}(f, k)|^2 = \begin{cases} |P(f, k)|^2 - E [|B(f, k)|^2] & \text{si } |P(f, k)|^2 \geq E [|B(f, k)|^2] , \\ 0 & \text{sinon.} \end{cases} \quad (1.29)$$

En utilisant l'expression de la réponse fréquentielle  $H_{SSP}(f, k)$ , on obtient le gain apporté au spectre à court-terme sous la forme :

$$H_{SSP}(f, k) = \begin{cases} \sqrt{1 - \frac{E [|B(f, k)|^2]}{|P(f, k)|^2}} & \text{si } |P(f, k)|^2 \geq E [|B(f, k)|^2] \\ 0 & \text{sinon.} \end{cases} \quad (1.30)$$

L'avantage de la SSP (et de la soustraction spectrale en général) est sans aucun doute la simplicité de sa mise en œuvre, cependant cette approche possède un inconvénient majeur car elle engendre un bruit résiduel dit "bruit musical" qui se révèle très gênant à l'écoute. Ce bruit musical est dû au fait que le spectre moyen d'énergie du bruit  $E [|B(f, k)|^2]$  est soustrait à un spectre d'énergie instantané  $|P(f, k)|^2$  possédant une forte variance par rapport à cette moyenne (en période de bruit seul) [8].

### 1.3.1.2 Filtrage de Wiener

Le filtre de Wiener est une technique populaire qui a été utilisée dans de nombreuses méthodes de rehaussement du signal de parole. Il est défini comme le filtre fournissant l'estimateur linéaire optimal d'un signal utile  $s(n)$  à partir d'une observation  $p(n)$ . Le critère d'optimalité est la minimisation de l'erreur quadratique moyenne (EQM) [15] entre le signal  $s(n)$  et le résultat  $\hat{s}(n)$  du filtrage de  $p(n)$ . Le filtre minimisant  $E[(s(n) - \hat{s}(n))^2]$  est donné dans le domaine fréquentiel par :

$$H_w(f, k) = \frac{\gamma_{sp}(f, k)}{\gamma_{pp}(f, k)} \quad (1.31)$$

où  $\gamma_{sp}(f, k)$  et  $\gamma_{pp}(f, k)$  sont respectivement la densité inter-spectrale de puissance entre le signal utile  $s(n)$  et l'observation  $p(n)$ , et la densité auto-spectrale de puissance de

l'observation  $p(n)$ . Si nous suivons l'hypothèse que le bruit  $b(n)$  est décorrélé du signal de parole  $s(n)$ , alors la densité inter-spectrale est nulle et on peut écrire :

$$\gamma_{sp}(f, k) = \gamma_s(f, k) + \gamma_{sb}(f, k) = \gamma_s(f, k) \quad (1.32)$$

$$\gamma_{pp}(f, k) = \gamma_{ss}(f, k) + \gamma_{sb}(f, k) + \gamma_{bs}(f, k) + \gamma_{bb}(f, k) = \gamma_s(f, k) + \gamma_b(f, k) \quad (1.33)$$

d'où

$$H_W(f, k) = \frac{\gamma_s(f, k)}{\gamma_s(f, k) + \gamma_b(f, k)} \quad (1.34)$$

En utilisant l'équation (1.22) de la densité spectrale de puissance du signal bruité, le filtre de Wiener s'exprime aussi sous la forme :

$$H_W(f, k) = 1 - \frac{\gamma_b(f, k)}{\gamma_p(f, k)} \quad (1.35)$$

Si on utilise les expressions de transformé de Fourier, on obtient alors :

$$H_W(f, k) = 1 - \frac{E[|B(f, k)|^2]}{|P(f, k)|^2} \quad (1.36)$$

On peut remarquer que  $H_W(f, k) = H_{SSP}^2(f, k)$  attestant du fait que ce filtre fait partie de la famille des méthodes de soustraction spectrale.

### 1.3.2 Techniques de réduction de bruit multi-capteurs

*Le rôle principal des techniques multi-capteurs dans la réduction de bruit est d'améliorer la qualité du signal de parole estimé en utilisant plusieurs capteurs. Pour ces techniques, nous limitons notre étude à deux voies dans le cadre de ce travail.*

Contrairement aux méthodes présentées dans la partie précédente qui permettent de filtrer le bruit d'un signal enregistré avec un seul microphone, les méthodes bi-capteur nécessitent l'utilisation de deux microphones pour supprimer le bruit. Un des microphones, appelé entrée primaire et l'autre appelé entrée de référence.

Dans cette partie, nous présentons les principales techniques de débruitage bi-voies les plus couramment employées dans le domaine de réduction de bruit et le rehaussement du signal de parole.

### 1.3.2.1 Annulation adaptative du bruit (ANC)

Les premiers travaux d'annulation du bruit adaptatif [19] ont été effectués par Howells et Applebaum et leurs collègues de la General Electric Company entre les années 1957 et 1960. Ils ont conçu un système d'annulation pour une antenne utilisant une entrée de référence reçue d'une antenne auxiliaire et d'un simple filtre adaptatif [19]. Depuis 1965, l'annulation adaptative du bruit a été appliquée avec succès à un certain nombre de problèmes, y compris l'électrocardiographie, l'élimination de l'interférence périodique en générale [20] et l'élimination des échos sur les lignes de transmission téléphonique à longue distance [21-22]. Le développement de l'algorithme LMS (Least Mean Square) par Widrow et Hoff [23] a permis l'élaboration de cette technique sous sa forme actuelle.

#### ❖ Principe de base de l'ANC

Considérons une observation bruitée  $p(n)$  d'un signal  $s(n)$  c.à.d :

$$p(n) = s(n) + b_0(n) \quad (1.37)$$

où  $b_0(n)$  est une contribution décorrélée du signal utile  $s(n)$ .  $p(n)$  représente le signal d'observation sur la voie primaire. Supposons que l'on dispose au niveau d'un deuxième capteur d'un bruit  $b(n)$  fortement corrélé à  $b_0(n)$  mais décorrélé de  $s(n)$ .  $b(n)$  fournit un signal secondaire d'observation sur la voie de référence au système d'ANC. La relation entre  $b_0(n)$  et  $b(n)$  est supposée linéaire, c.à.d :

$$b_0(n) = h(n) * b(n) \quad (1.38)$$

Où (\*) représente le produit de convolution et  $h(n)$  représente la réponse impulsionnelle du canal de fonction de transfert  $H(f)$ .

Le principe de la technique d'annulation adaptative de bruit (Adaptive Noise Cancellation ou ANC) est de filtrer la référence  $b(n)$  de manière adéquate afin d'obtenir la meilleure estimation de  $b_0(n)$  (au sens de l'erreur quadratique moyenne) ce qui apporte une réduction du bruit en sortie [24]. Une illustration graphique de ce principe est présentée sur la figure 1.6.

En tenant compte de la relation linéaire entre les bruits dans les deux observations, le signal de sortie peut s'écrire sous la forme :

$$e(n) = s(n) + (h(n) - w(n)) * b(n) \quad (1.39)$$

La solution d'un débruitage idéal  $e(n) = s(n)$  est donc équivalente à  $w^{opt}(n) = h(n)$ . Une manière de déterminer le filtre  $w^{opt}(n)$  sans connaissance ni du signal  $s(n)$  ni du filtre  $h(n)$  consiste à considérer la solution de la minimisation de l'énergie du signal estimé  $e(n)$ . En effet, compte tenu de la décorrélation entre  $b_0(n)$  et  $s(n)$ , on a d'après (1.39) :

$$E [e^2(n)] = E [s^2(n)] + E [(h(n) - w(n)) * b(n)]^2 \quad (1.40)$$

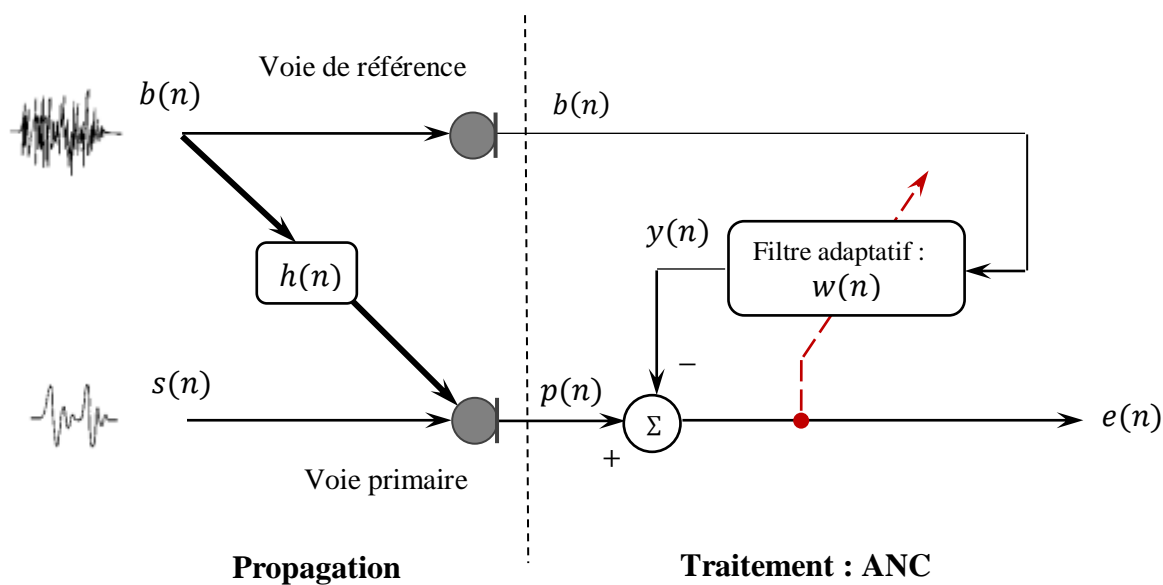


Figure 1.6 : Structure d'annulation adaptative de bruit (ANC).

La valeur minimale de  $E [e^2(n)]$  est alors atteinte pour  $w^{opt}(n) = h(n)$ . La solution de ce problème correspond au filtre de Wiener ; cette dernière est décrite ci-dessous.

#### ❖ Solution optimale de Wiener au problème d'annulation du bruit

Dans ce paragraphe, nous présentons la solution optimale de Wiener au problème statistique non contraint d'annulation du bruit. Le schéma du filtre de Wiener est déjà donné par la figure 1.3. Le filtre  $W(f)$  est linéaire, il est optimisé afin de minimiser la puissance de l'erreur  $e(n) = d(n) - y(n)$  suivant le critère du minimum d'erreur quadratique moyenne.

La fonction de transfère du filtre de Wiener est donnée par :

$$W^{opt}(f) = \frac{\gamma_{dx}(f)}{\gamma_{xx}(f)} \quad (1.41)$$

avec

$$\gamma_{dx}(f) = TF \{ \mathbf{r}_{dx}(k) \} \quad (1.42)$$

$$\gamma_{xx}(f) = TF \{ \mathbf{r}_{xx}(k) \} \quad (1.43)$$

où  $\gamma_{dx}(f)$  représente la densité inter-spectrale de puissance entre le signal d'entrée et le signal désiré et  $\gamma_{xx}(f)$  est la densité auto-spectrale de puissance du signal d'entrée. Avec la fonction d'inter-corrélation entre l'entrée  $x(n)$  et le signal désiré  $d(n)$  donnée par :

$$\mathbf{r}_{dx}(k) = E [ d(n)x^*(n - k) ] \quad (1.44)$$

et la fonction d'autocorrélation du signal d'entrée définie par :

$$\mathbf{r}_{xx}(k) = E [ x(n)x^*(n - k) ] \quad (1.45)$$

Appliquons la solution de Wiener donnée par l'équation (1.41) au cas de l'annulation adaptative du bruit suivant le schéma détaillé de la figure 1.6. Le signal désiré  $d(n)$  correspond au signal observé sur la voie primaire  $p(n)$  qui est donné par l'équation (1.37) (c.à.d.  $d(n) = p(n)$ ) et le signal à l'entrée du filtre  $x(n)$  correspond au signal d'observation sur la voie de référence ( $x(n) = b(n)$ ), cette solution s'exprime par :

$$W^{opt}(f) = \frac{\gamma_{pb}(f)}{\gamma_{bb}(f)} \quad (1.46)$$

La densité inter-spectrale de puissance  $\gamma_{pb}(f)$  entre le signal primaire et le signal de référence est donnée par l'expression suivante :

$$\gamma_{pb}(f) = \gamma_{bb}(f) H(f) \quad (1.47)$$

En remplaçant l'expression (1.47) dans l'expression du filtre de Wiener (1.46), on obtient :

$$W^{opt}(f) = H(f) \quad (1.48)$$

Le filtre de Wiener identifie alors parfaitement la réponse impulsionnelle  $H(f)$ .

### 1.3.2.2 Séparation aveugle de sources (SAS)

La technique de séparation de sources a suscité un intérêt croissant depuis son apparition au milieu des années 80. Cette technique est très vite devenue un domaine attractif du traitement du signal. Le problème consiste à retrouver des signaux utiles (par exemple signaux de parole ou des signaux émis par des téléphones portables), aussi appelés sources, à partir de mélanges de nature instantanés ou convolutifs, appelés observations. Généralement,

les observations sont des signaux obtenus à partir d'un ensemble de capteurs (microphones ou antennes par exemple). Dans un contexte aveugle, aucune connaissance a priori n'est disponible ni sur les sources, ni sur le processus de mélange (le contexte des observations), cette situation est alors appelée séparation aveugle de sources (SAS) (en anglais Blind Source Separation (BSS)). Pour résoudre ce problème, une solution possible consiste à ne faire qu'une seule hypothèse fondamentale : l'indépendance statistique mutuelle des sources [25]. La SAS se rencontre dans plusieurs applications tels que: radiocommunication, traitement des images, rehaussement de la parole, débruitage, etc...

Les techniques de SAS se plaçant dans le cadre de la réduction de bruit feront l'objet d'une étude approfondie dans le chapitre suivant.

#### 1.4 Conclusion

Dans ce chapitre, nous avons rappelé, dans une première partie, les propriétés et les caractéristiques de la parole et du bruit afin de mieux les traiter dans le cadre de cette thèse. Le principe du filtrage adaptatif et les algorithmes du gradient stochastique LMS et NLMS ont été présentés. Ces algorithmes sont caractérisés par leur stabilité numérique et simplicité d'implémentation. Dans le chapitre 2, l'algorithme NLMS sera appliqué au sein de la structure de séparation de sources bi-capteurs afin de mettre à jour les coefficients des filtres adaptatifs de cette structure.

Dans la deuxième partie de ce chapitre, nous avons présenté les différentes techniques de la réduction de bruit et le rehaussement du signal de parole. Ces techniques ont été différenciées par le nombre de microphone utilisés.

Nous avons vu que les méthodes classiques mono-capteur présentées dans ce chapitre, comme la soustraction spectrale ou le filtrage de Wiener, réussissent à réduire le bruit additif, mais en contrepartie, introduisent un bruit résiduel (bruit musical) gênant pour la perception humaine. Ces méthodes sont contraintes de faire l'hypothèse particulièrement pénalisante de stationnarité (à long terme) du bruit. Cette hypothèse, qui n'est pas valide dans la plupart des cas. L'apport d'informations fournies par une seconde voie d'observation permet d'envisager une discrimination des signaux de parole et de bruit sur des critères autre que celui de la stationnarité.

Pour les techniques bi-capteurs décrites dans ce chapitre, le grand intérêt de la structure d'annulation adaptative de bruit (ANC) est qu'elle ne nécessite pas de connaître a priori les statistiques du bruit ou de la parole. De plus, et contrairement à la soustraction spectrale, cette méthode est applicable à des bruits non stationnaires. Malheureusement, les performances de

l'ANC sont limitées dans les applications de la téléphonie mains-libres dans les voitures. Ceci est principalement lié à la difficulté de disposer d'un signal de référence « bruit seul » ne comportant aucune composante du signal utile de parole. La structure de séparation aveugle de sources (SAS) résout théoriquement ce problème, c'est pourquoi cette structure fera l'objet d'une étude complète qui sera présentée au sein du chapitre 2.

## CHAPITRE 2

### REDUCTION DE BRUIT PAR SEPARATION AVEUGLE DE SOURCES (SAS)

#### 2.1 Introduction

La séparation aveugle de sources est un domaine du traitement du signal qui a connu un essor majeur depuis une vingtaine d'années. Elle consiste à retrouver des signaux sources originaux, non observés, à partir de l'observation d'un ou plusieurs mélanges de ces sources. Notre cadre d'étude est le débruitage de la parole ; cette tâche est présentée dans la première partie de ce chapitre sous forme d'un problème de séparation de deux sources (la parole et le bruit) dans un contexte aveugle car l'information sur les sources ou plus particulièrement sur la source vocale qui nous intéresse est indisponible. Pour résoudre ce problème deux structures de SAS appelées directe et récursive, sont fréquemment utilisées pour séparer le signal de parole et le bruit à partir de la seule connaissance des signaux de mélange.

Dans la deuxième partie de ce chapitre, nous présentons les signaux et des critères de comparaison utilisés en simulation.

#### 2.2 Modèle et problème de la séparation de sources

On considère  $N$  sources notées  $s_i(t)$ , avec  $1 \leq i \leq N$ , regroupées en un vecteur  $S(t)$ . Soit  $M$  le nombre de capteurs utilisés,  $P(t)$  est le vecteur d'observations dont les composantes sont  $p_j(t)$ , avec  $1 \leq j \leq M$ .  $X(t)$  est le vecteur de des sources estimées. En général, le problème de la séparation aveugle de sources peut être divisé en deux parties (voir figure 2.1) [26]:

- **Modélisation du problème** : le mélange qui définit le lien entre les sources réelles et les observations.
- **Traitement du problème** : la séparation qui génère les sources estimées à partir des observations afin de résoudre le problème.



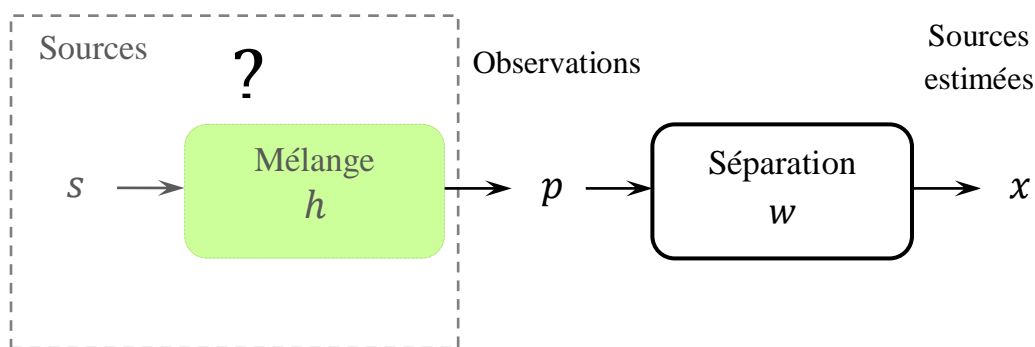


Figure 2.1 : Structure générale du problème de la séparation aveugle de sources.

### 2.2.1 Modélisation du mélange

Le processus de mélange entre les sources et les observations est modélisé en général par l'équation vectorielle suivante :

$$p(t) = h [S(t)] + b(t) \quad (2.1)$$

où

- $s(t) = [s_1(t), \dots, s_N(t)]^T$  est le vecteur des signaux sources que l'on cherche à estimer ; il est composé de  $N$  signaux.
- $p(t) = [p_1(t), \dots, p_M(t)]^T$  est le vecteur des observations mesurées par les capteurs ; il est constitué de  $M$  composants.
- $b(t) = [b_1(t), \dots, b_M(t)]^T$  est le vecteur de bruit additionnel qui modélise les erreurs de mesure des capteurs.
- $h$  est l'opérateur de mélange, il correspond au canal acoustique entre les sources et les microphones.

La figure (2.2) symbolise la relation sources-observations.

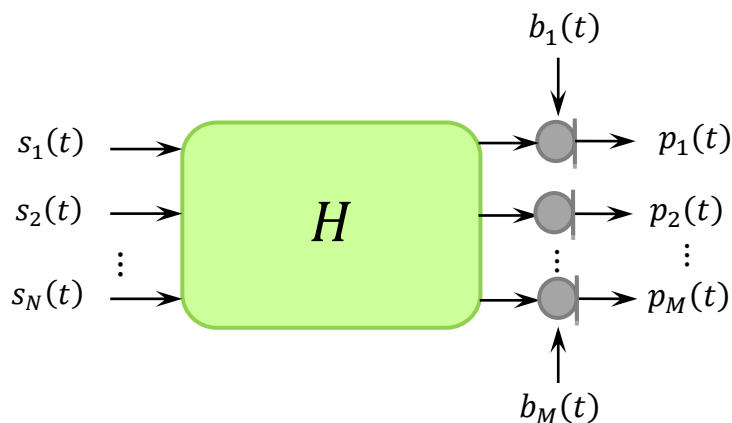


Figure 2.2 : Schéma général de la relation sources-observations en présence de bruit.

### 2.2.1.1 Les différents types de mélanges

Nous pouvons classer les mélanges en deux grandes catégories selon la nature de l'opérateur  $h$  [27]:

a) Les mélanges linéaires pour lesquels les observations s'écrivent :

$$p(t) = h(t) * s(t) + b(t) \quad (2.2)$$

où

- $h(t)$  est une matrice de réponses impulsionnelles de filtres.
- $*$  est l'opérateur de convolution continue.

Si les signaux mesurés sont à temps discret, ce qui est nécessairement le cas pour des signaux numériques enregistrés, la variable du temps  $t$  est remplacée par un indice temporel  $n$  entier et l'opérateur de convolution continue est remplacé par une convolution discrète. L'équation (2.2) devient alors :

$$p(n) = h(n) * s(n) + b(n) \quad (2.3)$$

Deux sous-catégories peuvent être déclinées dans ces mélanges linéaires :

- ❖ Mélange instantané : l'observation à un instant donné  $t$  est une combinaison linéaire des sources au même instant  $t$ . Ainsi,  $h$  se réduit à une matrice de mélange scalaire et l'opérateur de convolution devient une simple multiplication matricielle, dans ce cas de mélange l'équation (2.3) devient :

$$p(n) = h s(n) + b(n) \quad (2.4)$$

Un exemple du mélange instantané de sources est illustré sur la figure (2.3).

- ❖ Mélange convolutif : l'observation à l'instant donné  $t$  est une combinaison linéaire des valeurs des sources aux instants précédents. La fonction de transfert entre les sources et les observations peut être modélisée par une matrice de filtres linéaires.

La figure (2.4) présente un exemple du mélange convolutif de deux sources et deux observations.

b) Les mélanges non-linéaires sont les plus généraux mais aussi les plus difficiles à traiter. L'extrême complexité de ce cas limite son étude à des configurations bien particulières, principalement du type *mélanges post-non-linéaires* dans lesquels le

mélange est d'abord convolutif avant d'arriver sur les récepteurs qui introduisent des effets non linéaires. Au final, les observations issues des capteurs correspondent à la mise en cascade d'un mélange convolutif suivi d'une non linéarité propre à chaque récepteur. Une saturation en réception constitue un exemple simple de ce type de configuration [27].

Dans cette thèse, nous ne considérons que le cas de mélange linéaires alors la classe des mélanges non-linéaires n'est pas été étudiée dans la suite de ce document.

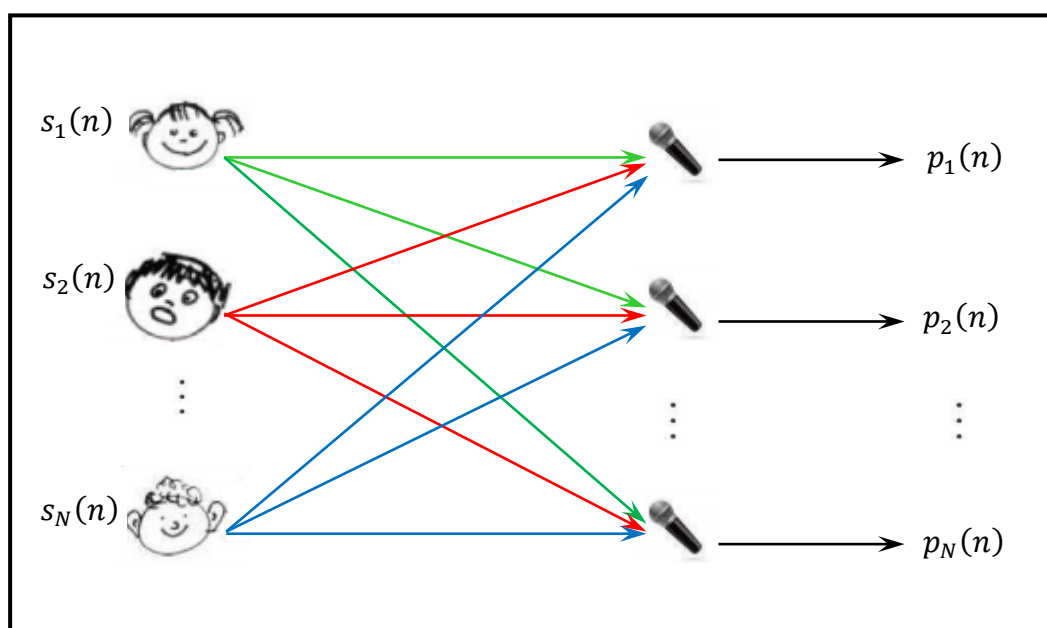


Figure 2.3 : Exemple d'un mélange instantané de sources.

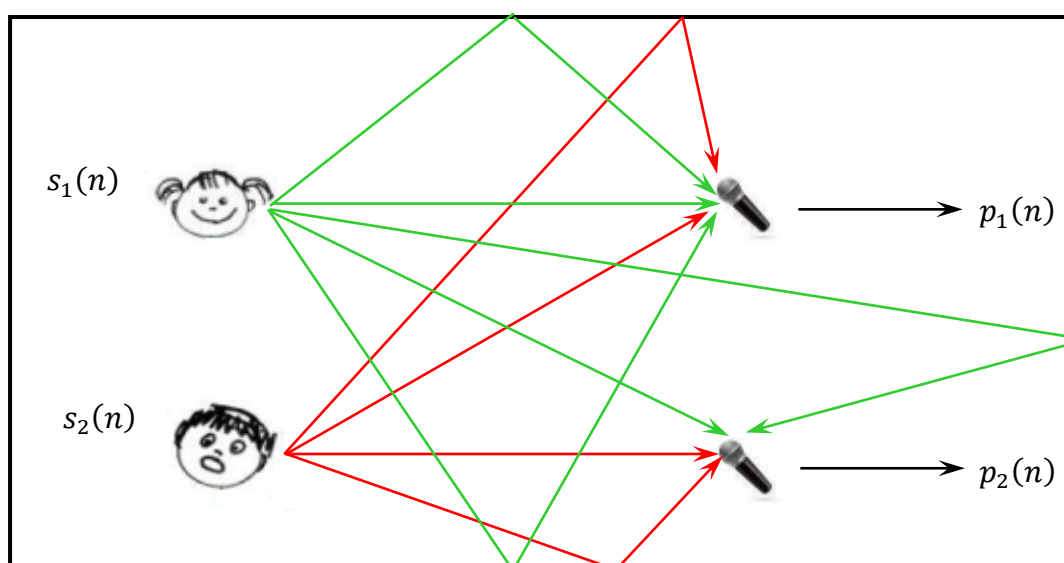


Figure 2.4 : Illustration d'un mélange convolutif acoustique à deux sources et deux capteurs.

Une autre classification des mélanges peut s'opérer suivant le nombre  $M$  d'observations considéré et le nombre  $N$  de sources, trois configurations de mélanges existent :

- ❖ S'il y a plus d'observations que de sources ( $M > N$ ), le mélange est dit **sur-déterminé**,
- ❖ S'il y a autant d'observations que de sources ( $M = N$ ), le mélange est dit **déterminé**,
- ❖ S'il y a plus de sources que d'observations ( $M < N$ ), le mélange est dit **sous-déterminé**.

Dans cette thèse, nous nous intéressons uniquement aux mélanges convolutifs déterminés d'ordre 2 ( $N = M = 2$ ).

### 2.2.1.2 Modèle de mélange convolutif d'une prise de son

#### - Modèle général

Considérons une situation de prise de son par deux microphones : une source de parole (locuteur) se situe près du 1<sup>er</sup> microphone, l'autre source de perturbation (un bruit ou un autre signal de parole) est près du 2<sup>ème</sup> microphone. A la sortie des microphones, on observe une superposition des signaux primitifs inconnus selon un mélange inconnu (figure 2.5). En général, c'est un mélange convolutif des signaux à large bande, qui dépend de la propagation des signaux dans le milieu, de la position des microphones et des sources, et des caractéristique de la salle. Les équations des signaux du mélange s'écrivent alors [28]:

$$p_1(n) = h_{11}(n) * s(n) + h_{21}(n) * b(n) + b_1(n) \quad (2.5)$$

$$p_2(n) = h_{22}(n) * b(n) + h_{12}(n) * s(n) + b_2(n) \quad (2.6)$$

où :  $s(n)$  et  $b(n)$  sont deux sources inconnues supposées indépendantes,  $p_1(n)$  et  $p_2(n)$  sont les deux signaux observés à la sortie des microphones,  $b_1(n)$  et  $b_2(n)$  sont les erreurs de mesure.  $h_{11}(n)$  et  $h_{22}(n)$  représentent les réponses impulsionnelles des canaux directs de couplage, et  $h_{12}(n)$  et  $h_{21}(n)$  représentent les effets du couplage mutuel entre les signaux.

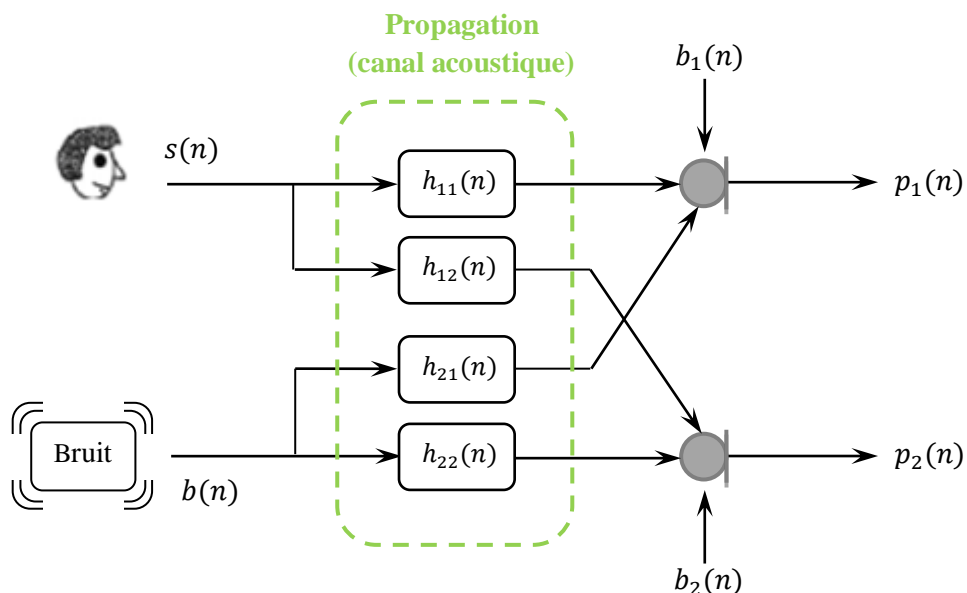


Figure 2.5 : Modèle général du mélange convolutif.

- Modèle convolutif simplifié

Pour simplifier le modèle de mélange convolutif qui est représenté dans la figure (2.5), certaines suppositions sont prise en compte tel que :

- un microphone est placé près du locuteur et l'autre près de la source de perturbation.
- Les cannaux directs  $h_{11}(n)$  et  $h_{22}(n)$  sont égaux à l'impulsion élémentaire  $\delta(n)$ .
- Les erreurs de mesure des microphones  $b_1(n)$  et  $b_2(n)$  sont négligeables.

Dans ces conditions un modèle simplifié du mélange convolutif modélisé par les filtres RIF causaux est présenté dans la figure (2.6).

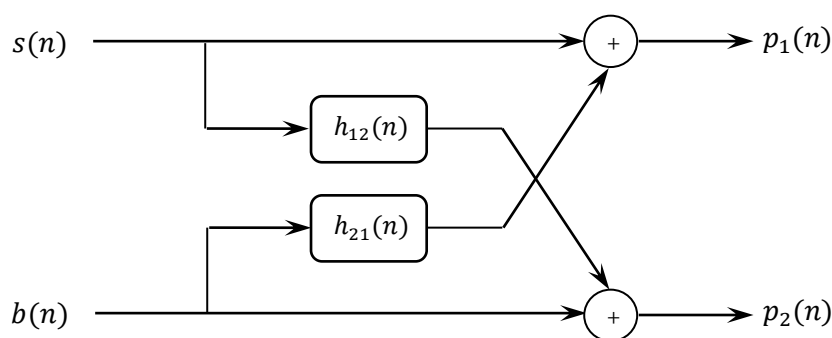


Figure 2.6 : Modèle simplifier du mélange convolutif.

Les équations des signaux du mélange deviennent :

$$p_1(n) = s(n) + h_{21}(n) * b(n) \quad (2.7)$$

$$p_2(n) = b(n) + h_{12}(n) * s(n) \quad (2.8)$$

où les sources  $s(n)$  et  $b(n)$  et les filtres  $h_{12}(n)$  et  $h_{12}(n)$  sont tous inconnus.

A partir des signaux observés  $p_1(n)$  et  $p_2(n)$ , comment peut-on retrouver les sources ?

### 2.2.2 Solution de la séparation de sources

Pour résoudre ce problème, deux structures de séparation de sources bien connues sont classiquement utilisées : la structure Forward [29] et la Backward [30].

#### 2.2.2.1 La structure réursive ou Backward

Elle s'applique lorsque chaque sortie du système de séparation est une combinaison linéaire d'une observation  $p_i(n)$  et de l'autre sortie  $v_j(n)$ , la structure réursive est représentée par la figure (2.7).

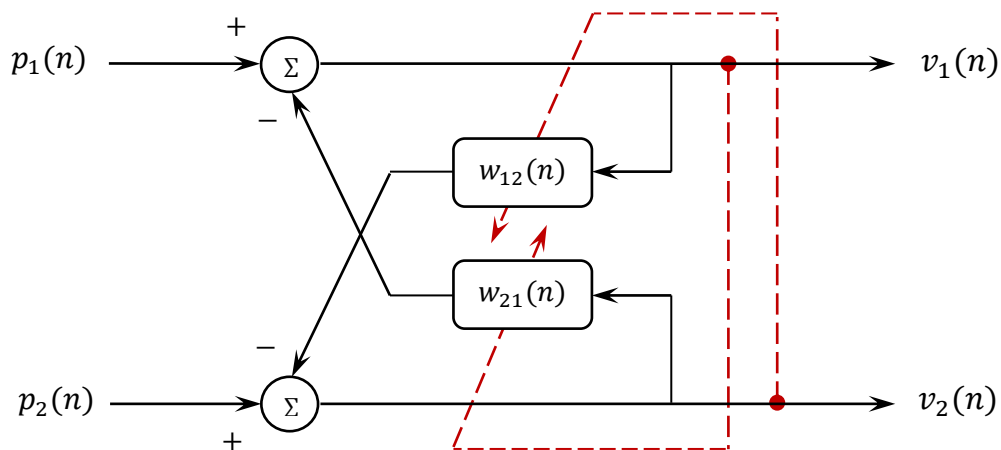


Figure 2.7 : Structure de séparation de sources réursive (BBSS).

Les signaux de sortie de mélange convolutif ( $p_1(n)$  et  $p_2(n)$ ), qui seront utilisés comme des entrées de la structure de séparation de sources réursive de la figure (2.6), sont donnés par :

$$p_1(n) = s(n) + h_{21}(n) * b(n) \quad (2.9)$$

$$p_2(n) = b(n) + h_{12}(n) * s(n) \quad (2.10)$$

La structure réursive conduit aux relations observations/sorties suivantes :

$$v_1(n) = p_1(n) - v_2(n) * w_{21}(n) \quad (2.11)$$

$$v_2(n) = p_2(n) - v_1(n) * w_{12}(n) \quad (2.12)$$

En remplaçant les expressions de  $p_1(n)$  et  $v_2(n)$  dans l'expression de  $v_1(n)$ , nous obtenons :

$$v_1(n) = s(n) + h_{21}(n) * b(n) - [p_2(n) - v_1(n) * w_{12}(n)] * w_{21}(n) \quad (2.13)$$

En remplaçant  $p_2(n)$  par son expression dans  $v_1(n)$ , on obtient :

$$v_1(n) = \frac{1}{\delta(n) - w_{12}(n) * w_{21}(n)} (s(n) * (\delta(n) - h_{12}(n) * w_{21}(n)) + b(n) * (h_{21}(n) - w_{21}(n))) \quad (2.14)$$

Et de même pour  $v_2(n)$ , en remplaçant les expressions de  $p_2(n)$  et  $v_1(n)$  ensuite de  $p_1(n)$  dans l'expression de  $v_2(n)$ , nous obtenons :

$$v_2(n) = \frac{1}{\delta(n) - w_{12}(n) * w_{21}(n)} (b(n) * (\delta(n) - h_{21}(n) * w_{12}(n)) + s(n) * (h_{12}(n) - w_{12}(n))) \quad (2.15)$$

Pour l'obtention du signal de parole sur la sortie  $v_1(n)$  et le bruit sur la sortie  $v_2(n)$ , une hypothèse d'optimalité est utilisée pour les deux filtres adaptatifs ( $w_{21}^{opt}(n) = h_{21}(n)$  et  $w_{12}^{opt}(n) = h_{12}(n)$ ).

Dans ce cas, les sorties  $v_1(n)$  et  $v_2(n)$  deviennent:

$$v_1(n) = s(n) \quad (2.16)$$

$$v_2(n) = b(n) \quad (2.17)$$

Nous notons que les coefficients des deux filtres de séparation  $w_{21}(n)$  et  $w_{12}(n)$  sont adaptés à partir de l'algorithme NLMS (Normalized Least Mean Square) décrit dans le chapitre 1, et les solutions optimales de ces filtres sont obtenues en utilisant le critère du minimum de l'erreur quadratique moyenne (MMSE ou bien EQMM).

L'adaptation des deux filtres adaptatifs  $w_{21}(n)$  et  $w_{12}(n)$  est donnée par les équations suivantes :

$$\mathbf{w}_{12}(n) = \mathbf{w}_{12}(n-1) + \mu_{12} \frac{v_2(n) \mathbf{k}_1(n)}{\mathbf{k}_1^T(n) \mathbf{k}_1(n) + \beta} \quad (2.18)$$

$$\mathbf{w}_{21}(n) = \mathbf{w}_{21}(n-1) + \mu_{21} \frac{v_1(n) \mathbf{k}_2(n)}{\mathbf{k}_2^T(n) \mathbf{k}_2(n) + \beta} \quad (2.19)$$

où

$$\mathbf{k}_1(n) = [v_1(n), v_1(n-1), \dots, v_1(n-L+1)]^T$$

$$\mathbf{k}_2(n) = [v_2(n), v_2(n-1), \dots, v_2(n-L+1)]^T$$

$\mu_{12}$  et  $\mu_{21}$  sont respectivement les pas d'adaptations des deux filtres adaptatifs  $w_{12}(n)$  et  $w_{21}(n)$ . Ils doivent être compris entre 0 et 2 afin d'assurer la stabilité de la structure.

Remarque 2 :

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 (autorisation d'adaptation).

➤ Limitation de la structure réursive

Pour pouvoir utiliser la structure réursive, un certain nombre de conditions doivent être satisfaites, au niveau de la stabilité et de la causalité des filtres de séparation ( $w_{21}(n)$  et  $w_{12}(n)$ ) tel que [31]:

- Chacun des filtres doivent être stable ;
- Les filtres  $w_{21}(n)$  et  $w_{12}(n)$  doivent être causaux. 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  $w_{21}(n)$  (respectivement  $w_{12}(n)$ ) est constituée de la sortie  $v_2(n)$  (respectivement  $v_1(n)$ ) du système, dont les valeurs futures ne sont pas encore disponibles à l'instant  $n$  car elles dépendent également des valeurs futures, donc encore inconnues des signaux de sortie du même dispositif de séparation.

### 2.2.2.2 La structure directe ou Forward

Connue sous le terme anglo-saxon Forward. Lorsque chaque sortie du système de séparation est une simple combinaison linéaire des observations, le schéma de la séparation que l'on qualifie de directe, est illustré par la figure (2.8).

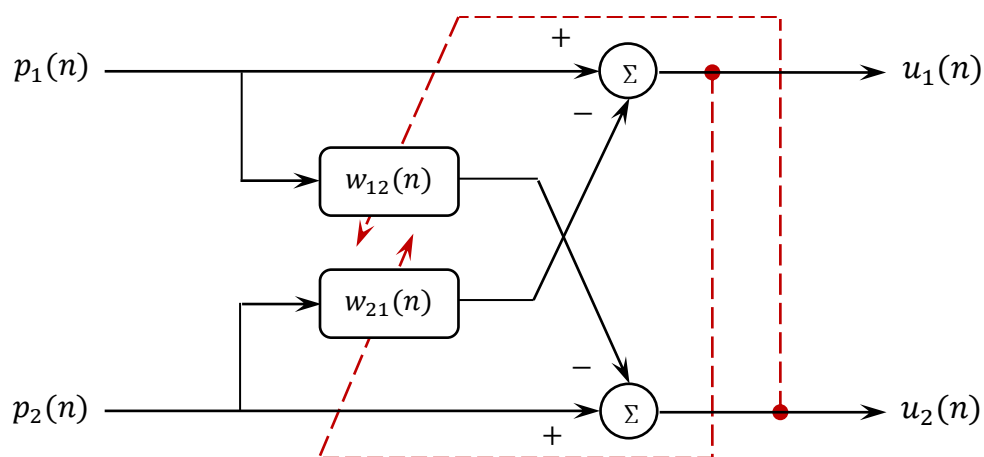


Figure 2.8 : Structure de séparation de sources directe (FBSS).



Pour ce modèle, les signaux de sortie  $u_1(n)$  et  $u_2(n)$  sont données par les formules suivantes :

$$u_1(n) = p_1(n) - p_2(n) * w_{21}(n) \quad (2.20)$$

$$u_2(n) = p_2(n) - p_1(n) * w_{12}(n) \quad (2.21)$$

Ces relations s'écrivent en fonction des sources ( $s(n)$  et  $b(n)$ ) :

$$u_1(n) = b(n) * [h_{21}(n) - w_{21}(n)] + s(n) * [\delta(n) - h_{12}(n) * w_{21}(n)] \quad (2.22)$$

$$u_2(n) = s(n) * [h_{12}(n) - w_{12}(n)] + b(n) * [\delta(n) - h_{21}(n) * w_{12}(n)] \quad (2.23)$$

si

$$w_{21}(n) = h_{21}(n) \text{ et } w_{12}(n) = h_{12}(n)$$

alors

$$u_1(n) = s(n) * [\delta(n) - h_{12}(n) * h_{21}(n)] \quad (2.24)$$

$$u_2(n) = b(n) * [\delta(n) - h_{21}(n) * h_{12}(n)] \quad (2.25)$$

La mise à jour des coefficients des deux filtres adaptatifs  $w_{21}(n)$  et  $w_{12}(n)$  est donnée par les expressions suivantes :

$$\mathbf{w}_{12}(n) = \mathbf{w}_{12}(n-1) + \mu_{12} \frac{u_2(n) \mathbf{m}_1(n)}{\mathbf{m}_1^T(n) \mathbf{m}_1(n) + \beta} \quad (2.26)$$

$$\mathbf{w}_{21}(n) = \mathbf{w}_{21}(n-1) + \mu_{21} \frac{u_1(n) \mathbf{m}_2(n)}{\mathbf{m}_2^T(n) \mathbf{m}_2(n) + \beta} \quad (2.27)$$

où

$$\mathbf{m}_1(n) = [p_1(n), p_1(n-1), \dots, p_1(n-L+1)]^T$$

$$\mathbf{m}_2(n) = [p_2(n), p_2(n-1), \dots, p_2(n-L+1)]^T$$

➤ Avantage de la structure directe

Contrairement au modèle récursif, la structure directe permet d'utiliser des filtres non-causaux de manière naturelle car elle ne fait pas intervenir les différentes sorties du système dans le calcul des sorties des filtres de séparation.

➤ Inconvénient de la structure directe

A partir des équations (2.24) et (2.25), nous remarquons que les signaux de sortie  $u_1(n)$  et  $u_2(n)$  de la structure directe sont distordus, c'est l'inconvénient majeur de cette structure. Pour éliminer ces distorsions, une solution a été proposée dans [31], consiste l'utilisation des post-filtres placés en sortie de la structure directe.

### 2.3 Contrôle des deux structures directe et récursive par un système de DAV

Les deux structures de séparation de sources classiques (directe et récursive) décrites précédemment, utilisent une hypothèse d'optimalité ( $w_{21}^{opt}(n) = h_{21}(n)$  et  $w_{12}^{opt}(n) = h_{12}(n)$ ) ; cette dernière est assurée en pratique grâce au contrôle de l'adaptation des deux filtres adaptatifs  $w_{21}(n)$  et  $w_{12}(n)$  par un système de détection d'activité vocale manuel DAVM (c-à-d segmentation manuelle du signal de parole), cette adaptation est contrôlée de la manière suivante : le filtre  $w_{21}(n)$  est mis à jour uniquement pendant les période de non-activité vocale (période de bruit seul), alors que le filtre  $w_{12}(n)$  est mis à jour uniquement pendant les périodes d'activité vocale (période de présence du signal de parole utile) [32], ce contrôle est résumé dans l'algorithme 1. Le principe général des structures de séparation de sources classique (directe ou récursive) est montré par la figure 2.9.

---

**Algorithme 1** Contrôle d'adaptation par un système de DAVM.

---

Début

1 : **si** DAVM = 1 **alors** // absence du signal de parole (zones de silence)

2 :  $\mathbf{w}_{21}(n) = \mathbf{w}_{21}(n-1) + \mu_{21} \frac{u_1(n) \mathbf{m}_2(n)}{\mathbf{m}_2^T(n) \mathbf{m}_2(n) + \beta}$

3 : **sinon** // présence du signal de parole

4 :  $\mathbf{w}_{12}(n) = \mathbf{w}_{12}(n-1) + \mu_{12} \frac{u_2(n) \mathbf{m}_1(n)}{\mathbf{m}_1^T(n) \mathbf{m}_1(n) + \beta}$

5 : **fin si**

Fin

---

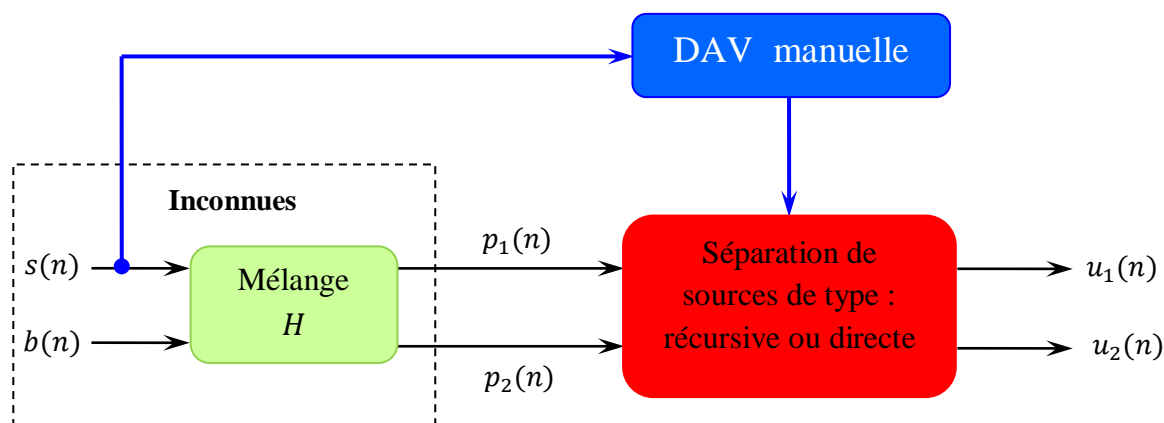


Figure 2.9 : Schéma bloc du principe général des structures de séparation de sources classiques.

En résumé les étapes de fonctionnement des structures de SAS classiques (FBSS et BBSS) par la figure 2.10 :

**1<sup>er</sup> étape :** Modèle de mélange convolutif ( $H$ ) qui détermine la relation entre les signaux sources réelles ( $s(n)$ ,  $b(n)$ ) et les signaux d'observations bruités (signaux captés par les microphones) ( $p_1(n)$ ,  $p_2(n)$ ).

**2<sup>ème</sup> étape :** Structure de séparation ( directe ou récursive) qui récupère les signaux sources estimés ( $u_1(n)$ ,  $u_2(n)$ ) à partir des signaux d'observations bruités.

**3<sup>ème</sup> étape :** Système de détection d'activité vocale manuel (DAVM) qui contrôle l'ajustement des filtres de séparation ( $w_{21}(n)$ ,  $w_{12}(n)$ ) dans les périodes appropriées pour chacun des filtres; le  $w_{21}(n)$  n'est actif que durant les périodes où la parole est absente (zones de silence) et le  $w_{12}(n)$  est actif que durant les périodes où la parole est présente.

**4<sup>ème</sup> étape :** Mise à jour des filtres qui permet d'ajuster les coefficients des filtres adaptatifs  $W$  selon l'algorithme NLMS.

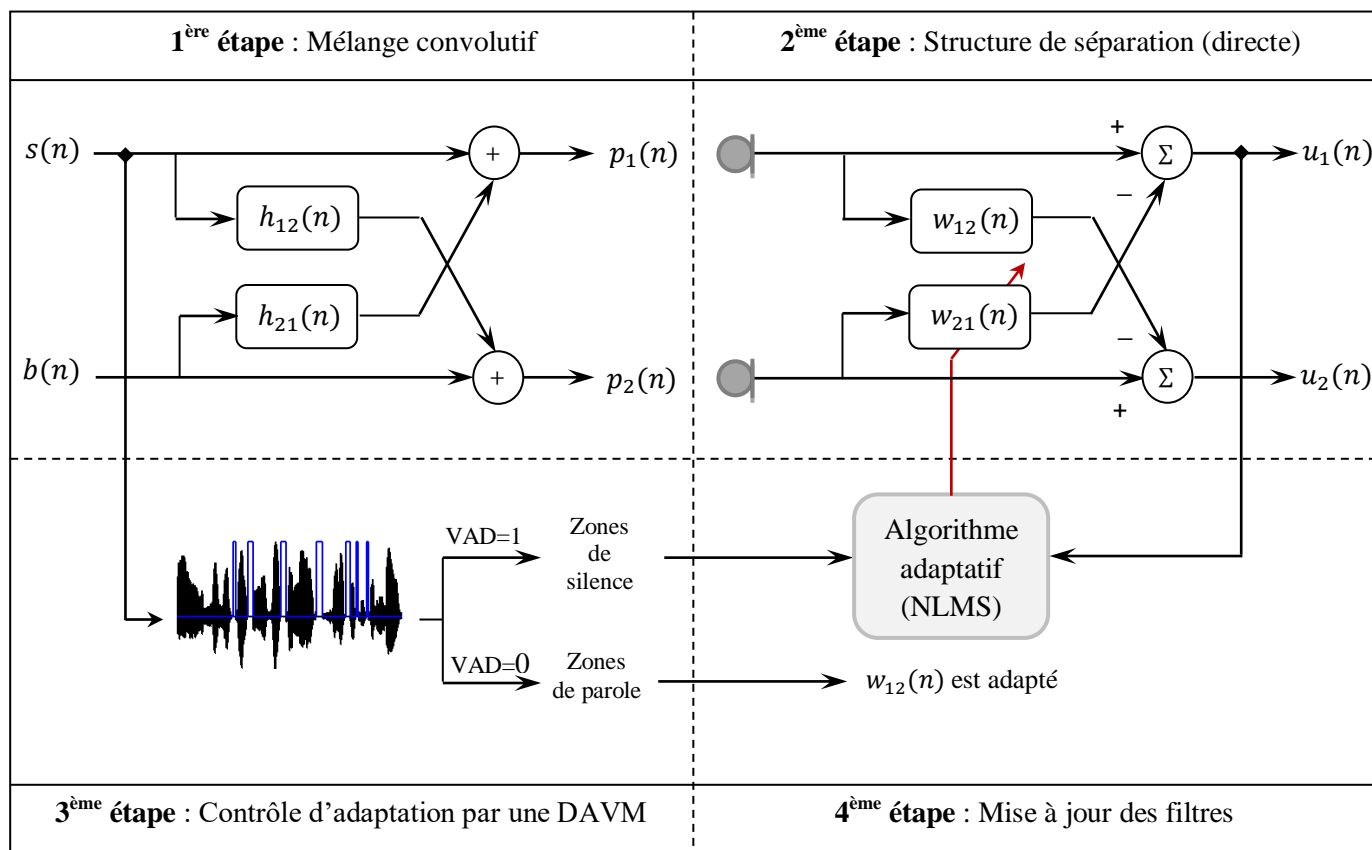


Figure 2.10 : Les étapes de la structure FBSS classique (c'est les même étapes pour une structure BBSS classique sauf que la structure directe de la 2<sup>ème</sup> étape est remplacée par une structure réursive de la figure 2.7).

## 2.4 Limitation des structures de SAS classiques

Dans la pratique, le contrôle d'adaptation des structures de SAS par un système de DAVM s'avère une tâche difficile car les sources sont inconnues et les signaux disponibles au niveau des capteurs de la téléphonie mains-libres sont des mélanges, donc la difficulté à discerner la source utile (signal de parole) de la source interférente (signal de bruit) à partir seulement de la connaissance des signaux de mélanges, nous à empêcher à segmenter le signal de parole manuellement autrement dit détecter les DAV manuellement, simplement parce qu'il n'existe pas d'instant pendant lequel le bruit peut être considéré comme absent dans les signaux de mélanges. Alors à partir des signaux de mélange, comment peut-on segmenter le signal de parole automatiquement afin de contrôler l'adaptation des filtres de séparation ?

En ce sens, les chapitres qui suivent vont se concentrer sur la proposition de nouvelles méthodes qui solutionnent cette problématique en utilisant des techniques de segmentations automatiques du signal de parole (c-à-d une DAV automatique) dans le fonctionnement de la SAS.

Notons que toutes ces méthodes que nous allons proposer sont basées sur la variation des pas d'adaptations, car c'est le seul paramètre maîtrisable dans l'équation de mise à jour des filtres, ce choix est justifié par la Remarque 2.

## 2.5 Description des signaux et des critères de comparaison utilisés en simulations

Afin de caractériser le comportement relatif à chaque technique de séparation de sources utilisées et proposées dans cette thèse, nous décrivons en ce qui suit les signaux et les critères de comparaison utilisés en simulation. Pour l'étape de programmation, on utilisera MATLAB de la compagnie Maths Works, outil logiciel très répandu dans le milieu du traitement de signal, qui offre un langage de programmation souple, facile et agréable.

### 2.5.1 Description des signaux

Les résultats présentés dans ce document sont obtenus à partir de l'utilisation de deux types de signaux source, un signal de parole et un bruit Blanc. Ces deux signaux sont échantillonnés à la fréquence  $F_e = 16\text{KHz}$ .

- Le signal de parole correspond à une voie d'homme parlant français (voir figure 2.11). Ce fichier provient de la base de données AURORA [33], qui est très populaire dans le domaine du traitement du signal de parole.
- Le bruit Blanc (bruit stationnaire) est une réalisation d'un processus aléatoire dans lequel la densité spectrale de puissance est la même pour toutes les fréquences, ce signal est représenté dans le domaine temporel par la figure 2.12.

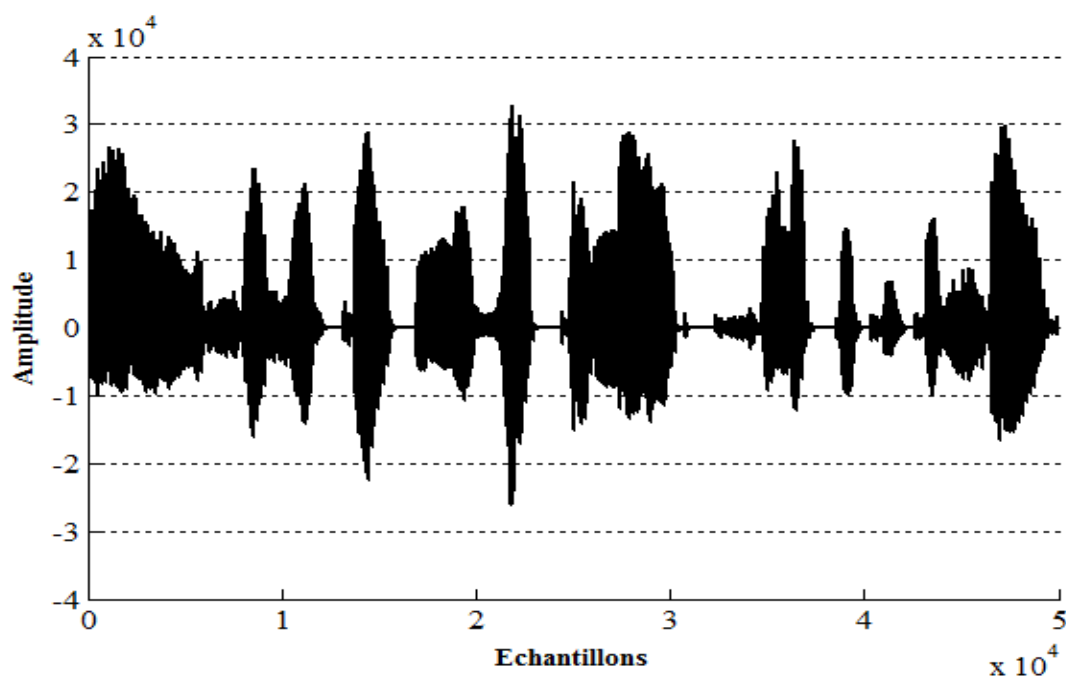


Figure 2.11 : Signal de parole utilisé,  $F_e = 16$  KHz.

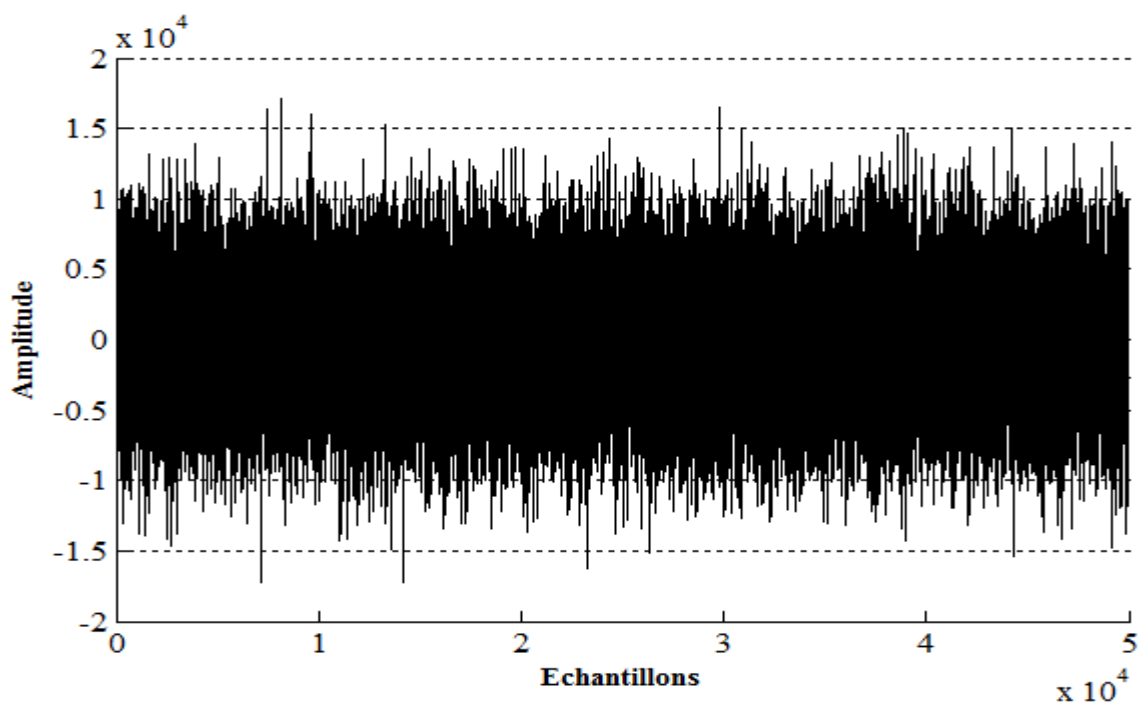


Figure 2.12 : Signal du bruit Blanc utilisé,  $F_e = 16$  KHz.

### 2.5.2 Description des chemins acoustiques

Nous utilisons dans cette partie une implémentation spécifique du modèle de mélange de la figure (2.6). L'idée que nous poursuivons vise à utiliser un modèle conforme à la physique

du problème et capable de prendre convenablement en compte l'effet de la distance entre les deux microphones tout en restant conforme avec les fonctions de transfert supposées unitaires entre chaque source ponctuelle de signal et l'ensemble des chemins acoustiques directs. Les réponses impulsionnelles du couplage croisé  $h_{12}(n)$  et  $h_{21}(n)$  simulées sont donc construites selon [34] comme suit :

$$h_{12}(n) = \delta(n) + h'_{12}(n) \quad (2.28)$$

$$h_{21}(n) = \delta(n) + h'_{21}(n) \quad (2.29)$$

où

- $\delta(n)$  est une impulsion unitaire localisée au début de la réponse impulsionnelle, qui représente le trajet acoustique direct de chaque source et le microphone de couplage croisé.
- $h'(n)$  est une queue de réponse impulsionnelle, qui représente la contribution des trajets multiples correspondant aux réflexions à l'intérieur de l'habitacle du véhicule. Cette queue de réponse impulsionnelle est obtenue à partir d'un bruit aléatoire dont l'enveloppe temporelle est contrôlée selon la fonction de pondération suivante :

$$f(n) = Ae^{-Bn} \quad (2.30)$$

où  $A$  est un facteur d'échelle (pris égal à 1), et  $B$  un facteur d'atténuation qui modélise l'absorption des ondes sonores sur les parois de la voiture et qui est donc lié au temps de réverbération  $n_r$ , ce facteur est donné par l'expression suivante :

$$B = 3 \ln(10) / n_r = 6.9078 / n_r \quad (2.31)$$

Les amplitudes des réponses impulsionnelles  $h_{12}(n)$  et  $h_{21}(n)$  sont ajustées à partir de la variance d'un bruit aléatoire.

Les figures (2.13) et (2.14) représentent un exemple typique de réponses impulsionnelles acoustiques générées suivant la méthode décrite ci-dessus. La variance du bruit aléatoire est égale à 0.5 [34]. Le facteur d'atténuation est pris égal à  $B = 0.028$  [34] ce qui correspond à une configuration de microphones relativement espacés et la période d'échantillonnage

$T_e = 625\mu s$ . Le temps de réverbération correspondant est de 30.8 ms [34]. La majeure partie de l'énergie contenue dans une réponse impulsionnelle correspond à un support temporel d'environ  $L = 64$  échantillons à la fréquence d'échantillonnage considérée.

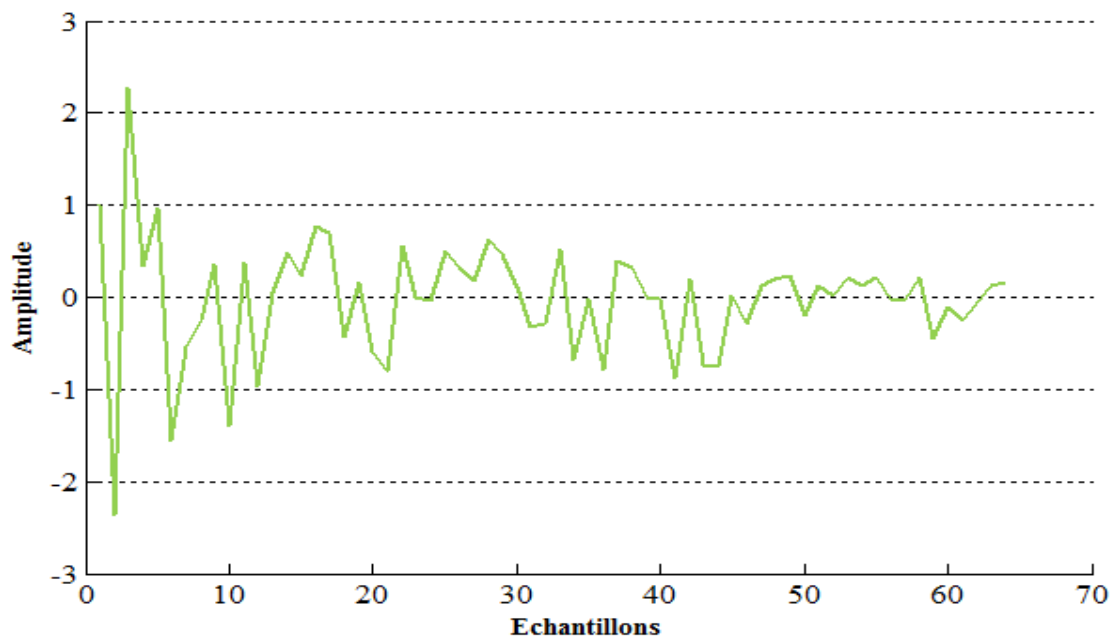


Figure 2.13 : Exemples d'une réponse impulsionnelle simulée ( $h_{12}(n)$ ), cas des capteurs espacés.

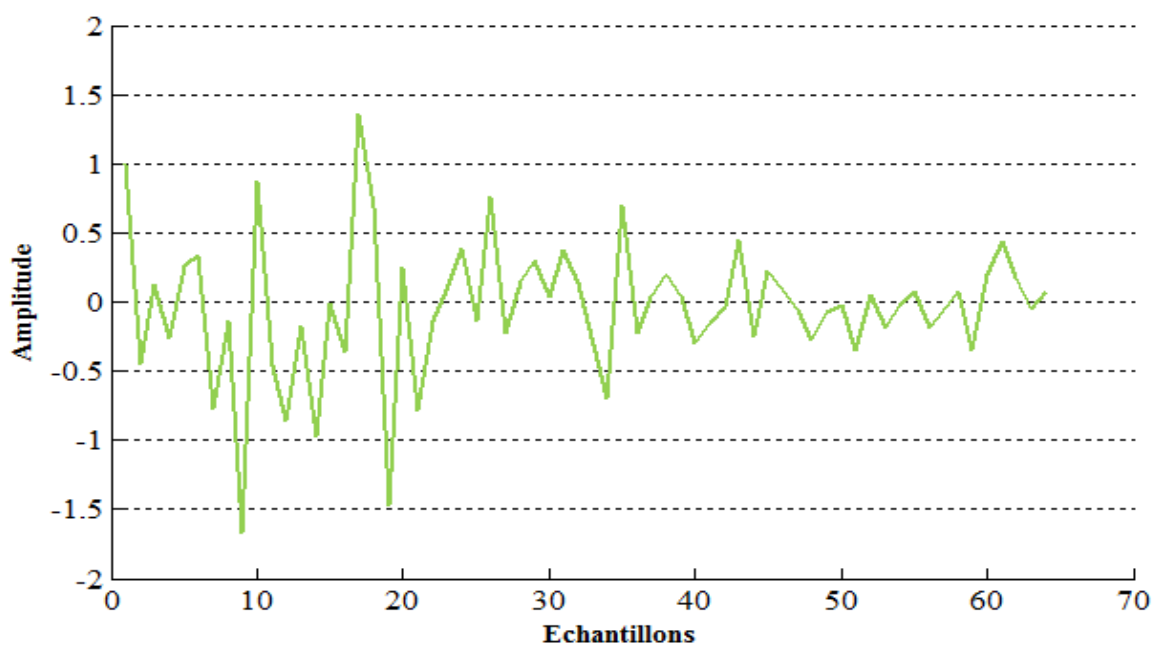


Figure 2.14 : Exemples d'une réponse impulsionnelle simulée ( $h_{21}(n)$ ), cas des capteurs espacés.



### 2.5.3 Détection d'activité vocale (DAV)

Il a déjà été noté que les deux structures classiques de séparation de sources directe (FBSS) et récursive (BBSS) sont contrôlées à partir d'un mécanisme de Détection d'Activité Vocale (DAV) manuel. En simulations et pour le signal de parole, nous avons créé un fichier correspondant de VAD manuelle, fait à l'aide du logiciel *Cool Edit* en signalant manuellement les zones actives et inactives du signal de parole. Cette DAVM est de nature binaire : elle vaut 1 lorsque la parole est absent (période d'adaptation), et 0 sur les phases de parole (adaptation bloquée) [32]. La figure (2.15) présente un exemple de DAV manuelle sur le signal de parole originale.

Remarque 3 :

En tenant avantage de la non-stationnarité du signal de parole c'est-à-dire de la propriété d'intermittence de ce dernier, nous pouvons segmenter le signal de parole manuellement.

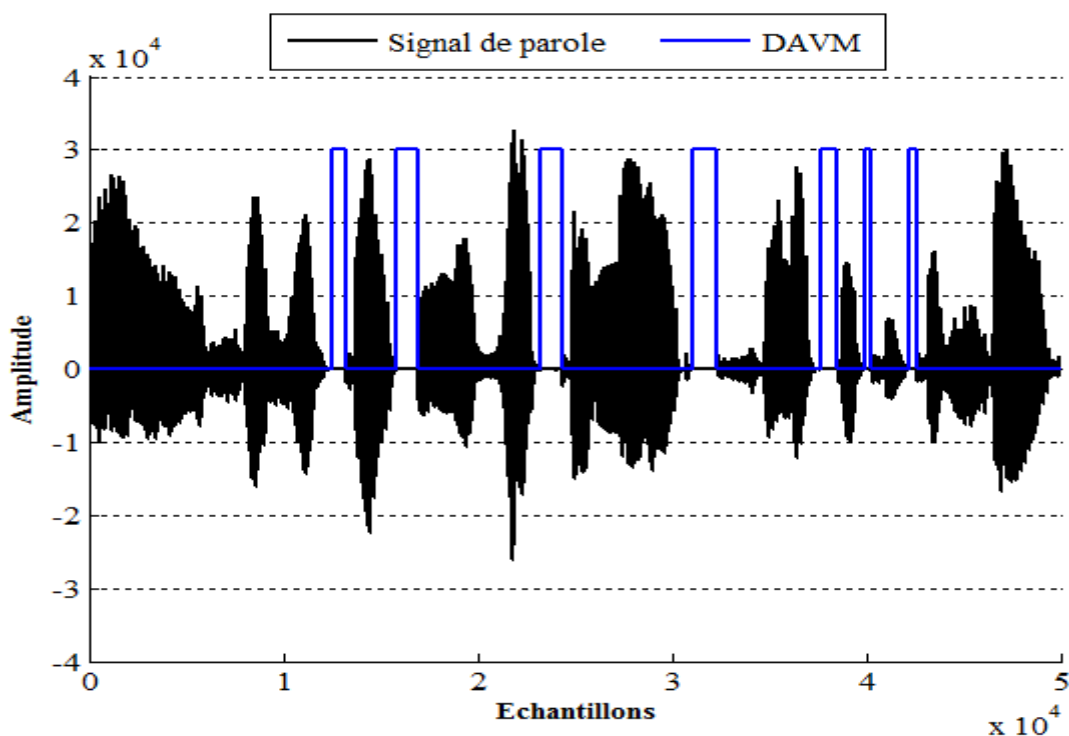


Figure 2.15 : Signal de parole utilisé avec une segmentation (DAV) manuelle.

### 2.5.4 Description des critères de comparaisons

Pour évaluer et comparer les performances des différentes techniques proposées pour la SAS dans les chapitres suivants, nous allons utiliser plusieurs critères tels que : l'évolution de

l'enveloppe temporelle et le contenu spectral des signaux de sortie pour chaque méthode, critères objectifs, critères subjectifs et l'évolution de la complexité. Pour chaque méthode de SAS proposée, nous validerons par des simulations son fonctionnement puis nous comparerons leurs performances avec celles de la méthode qui l'a précède selon l'organigramme présenté sur la figure 2.16. En ce qui concerne la comparaison subjective, nous la laisserons en dernier où nous comparerons toutes les techniques de SAS.

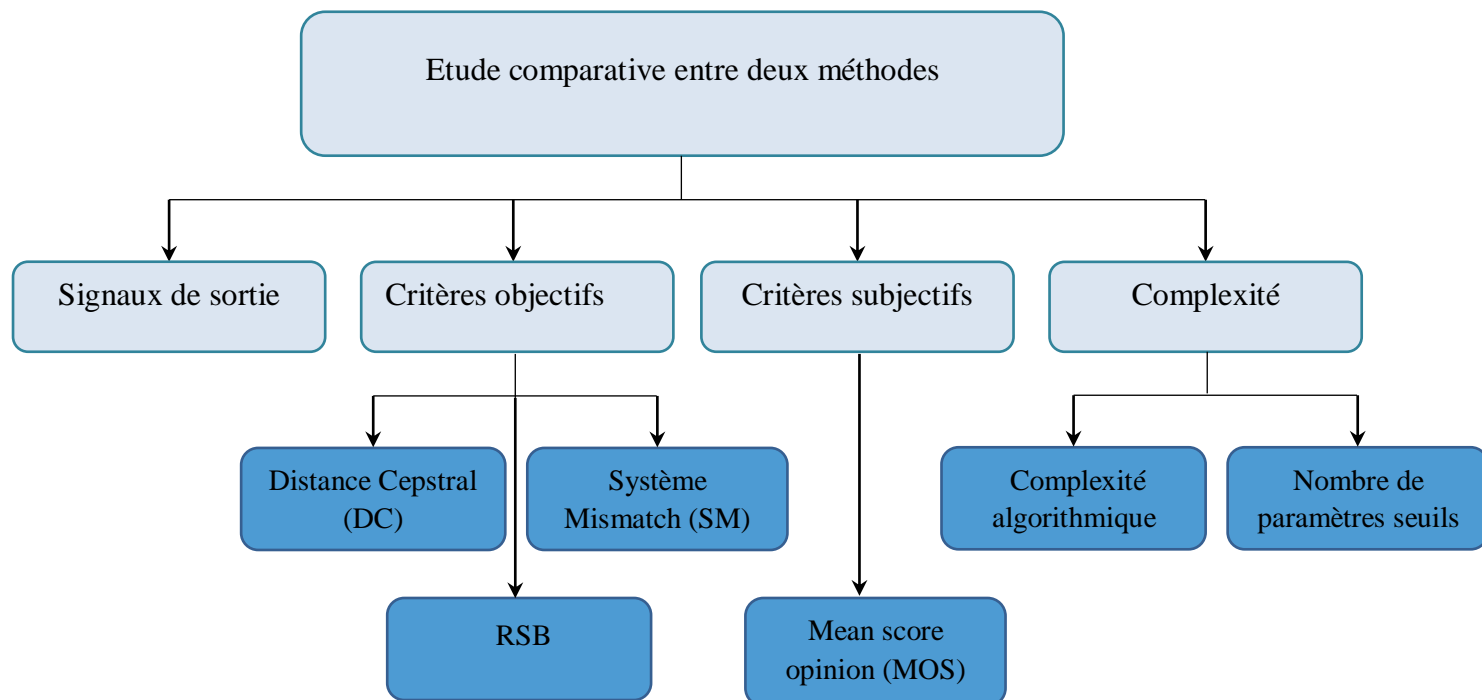


Figure 2.16 : Organigramme des différents critères utilisés pour une étude comparative.

#### 2.5.4.1 Critères objectifs de qualité

L'évaluation des performances des techniques de réduction de bruit représente une véritable problématique. En effet, c'est l'oreille humaine qui est le juge finale de la qualité d'un système de débruitage de la parole, jugement qui est donc par essence subjectif. L'évaluation subjective est peu utilisée dû à son coût et le temps qu'elle demande. C'est la raison pour laquelle des méthodes d'évaluation dites « objectives », de la qualité vocale sont proposées comme alternative aux méthodes subjectives. Ces méthodes sont basées sur des formules mathématiques et elles permettent de donner une évaluation quantitative. Les mesures objectives de qualité des signaux vocaux utilisées dans cette thèse sont :

➤ Le critère du Rapport Signal à Bruit (RSB)

Le critère du rapport signal à bruit permet de mesurer les pertes de puissance, il donne une information globale du rapport entre la puissance du signal de parole rehaussé et celle du bruit, suivant :

$$RSB = 10 \log_{10} \left\{ \frac{P_S(n)}{P_N(n)} \right\} \quad (2.32)$$

où  $P_S(n)$  et  $P_N(n)$  sont les estimées courantes à court-terme de la puissance des composantes de signal de parole et du bruit en cours d'étude.

➤ Le critère de la Distance Cepstrale (CD)

La distance cepstrale permet de mesurer les distortions, elle fournit une mesure de similarité entre les log-spectres (cepstres) de deux signaux. Pour deux signaux temporels  $s_1(n)$  et  $s_2(n)$ , la distance cepstrale, évaluée pour chaque trame  $k$  de  $N$  échantillons [35] :

$$CD(k) = \sum_{i=1}^{N-1} [IFFT (\log_{10} (|S_1(f_i, k)|)) - \log_{10} (|S_2(f_i, k)|)]^2 \quad (2.33)$$

où  $f_i$  est une fréquence (discrétisée) de la FFT. La fréquence nulle  $f_0$ , correspondant à la puissance des signaux, n'est pas prise en compte.

La raison principale de l'emploi de cette distance vient du comportement de l'oreille humaine, plus sensible aux variations logarithmiques.

➤ Le critère du Désajustement

Le critère du désajustement est un critère robuste dans l'évaluation des performances. Ce critère se calcule à partir de la distance euclidienne entre les coefficients des réponses impulsionnelles réelles et mesurées (estimées). Ce critère, noté SM pour "System Mismatch", est défini par l'expression suivante [36] :

$$(SM)_{dB} = 10 * \log_{10} \left( \frac{\|\hat{w}_{ij} - h_{ij}\|^2}{\|h_{ij}\|^2} \right) \quad \text{où } i \neq j \in \{1,2\} \quad (2.34)$$

où  $h_{ij}$  représente les coefficients du filtre réel ( $h_{12}$  et  $h_{21}$ ), et  $\hat{w}_{ij}$  représente la valeur moyenne des coefficients estimés (coefficients du filtre adaptatif  $w_{12}$  et  $w_{21}$ ).

#### 2.5.4.2 Critères subjectifs de qualité

Le jugement de la qualité vocale est avant tout une notion subjective liée à la perception humaine. La meilleure façon d'évaluer la qualité vocale est d'effectuer des tests dits « subjectifs » qui permettent de donner une évaluation qualitative. La méthode d'évaluation subjective de la qualité la plus couramment utilisée est le MOS (Mean Opinion Score).

##### ➤ MOS (Mean Opinion Score)

Cette méthode est recommandées par l'UIT-T P.800 [37], elle consiste à faire appel à des utilisateurs et de leur demander d'écouter un ensemble de fichier audio indépendamment, un à un, et de noter la qualité qu'ils perçoivent sur une échelle catégorielle à cinq niveaux, allant de 1 (mauvaise qualité) à 5 (excellente qualité), comme l'indique le tableau 2.1. Les notes des participants sont moyennées afin d'obtenir une note d'opinion moyenne ou MOS (Mean Opinion Score). Notons que la perception de la qualité diffère d'une personne à une autre. Elle dépend de plusieurs facteurs tels que l'état émotionnel de la personne, son âge, le contexte d'évaluation de la qualité.

Tableau 2.1 : Echelle de qualité utilisée lors du test MOS [38].

Qualité de la parole	Note
Excellente	5
Bonne	4
Passable	3
Médiocre	2
Mauvaise	1

#### 2.5.4.3 Analyse de la complexité

La complexité peut jouer un rôle déterminant dans la comparaison des méthodes de SAS. Une méthode très efficace mais qui demande beaucoup de calculs sera difficile à implémenter sur un composant électronique de type DSP ou FPGA. Il s'agit donc de faire attention à l'aspect de la complexité et la régularité de calcul.

Voici Deux points importants à prendre en considération lors d'une étude de la complexité d'une méthode de SAS

- La complexité algorithmique : plus la méthode demande de calcul mathématiques (nombre de multiplications et d'additions) lourds et complexes, plus difficile sera son implémentation sur circuit numérique en temps réel.
- Le nombre de paramètres fixes (constantes) à déterminer initialement, ainsi que des paramètres seuils que la méthode de SAS doit connaître à priori pour fonctionner. Plus le nombre de paramètres est grand et moins la méthode sera autonome et donc moins apte à s'adapter à des changements. Inversement, une méthode qui requiert peu de paramètres initiaux à déterminer de manière empirique pourra s'adapter plus facilement et conviendra mieux à une implémentation en temps réel.

## 2.6 Conclusion

Dans la première partie de ce chapitre, nous avons présenté une étude détaillée des deux structures de séparation de sources directe et récursive qui ont été utilisées pour le débruitage de la parole. Les avantages et les inconvénients de chaque structure ont été présentés. Nous avons remarqué qu'il est indispensable d'utiliser un système de DAVM dans le fonctionnement des structures de SAS classiques pour pouvoir contrôler l'adaptation de ces filtres, en discriminant les périodes où la parole est présente des périodes de bruit seul. La limitation de cette technique a été mentionnée.

Dans la seconde partie du chapitre, nous avons décrit les signaux de tests ainsi que le modèle de mélange convolutif utilisé qui prend en compte l'espacement entre les microphones utilisés dans la prise du son. Les critères utilisés pour l'étude comparative ont également été présentés à savoir les critères objectifs (le rapport signal à bruit en sortie de traitement, la distance cepstrale et le désajustement des coefficients des filtres adaptatifs), critère subjectif (MOS) et la complexité.

## CHAPITRE 3

### AMELIORATION DES TECHNIQUES DE SEPARATION AVEUGLE DE SOURCES (DIRECTE ET RECURSIVE)

#### 3.1 Introduction

Le contrôle d'adaptation des coefficients des deux filtres adaptatifs est une étape nécessaire pour un fonctionnement efficace de la structure de séparation de sources (directe et récursive). Pour la plupart des travaux de recherche issues de la littérature, l'adaptation est contrôlée par un système de détection d'activité vocale manuel (DAVM) de sorte à garantir la mise à jour des filtres adaptatifs aux moments opportuns, ce qui en pratique est irréalisable en raison d'absence d'informations a priori sur les signaux d'entrée. Donc, nous avons besoin de détecter les DAV automatiquement.

Au sein de ce chapitre, nous décrivons deux nouvelles versions améliorées pour les structures de séparations de sources (directe et récursive) avec l'utilisation d'un nouveau bloc qui réalise une technique de segmentation (DAV) automatique.

Une étude approfondie des résultats de simulations de la structure de séparation de sources améliorée de type directe est également présentée. Les travaux décrits dans ce chapitre ont été publiés dans [39].

#### 3.2 Structures améliorées

Dans ce chapitre, les deux méthodes proposées nommées (AFBSS et ABBSS) que nous allons décrire ci-après représentent une amélioration des deux structures de séparation de sources (directe et récursive), elles sont inspirées des méthodes d'Ikeda [40] et de Djendi [41] respectivement. Ces versions améliorées utilisent une nouvelle technique de DAV automatique (DAVA) dans leurs fonctionnements, en employant des pas d'adaptations variables en fonction du rapport signal à bruit (RSB) estimé sur chacune des voies d'observation afin de contrôler l'adaptation des deux filtres de séparation. La figure 3.1 représente un schéma bloc du principe des structures de séparation de sources améliorées.

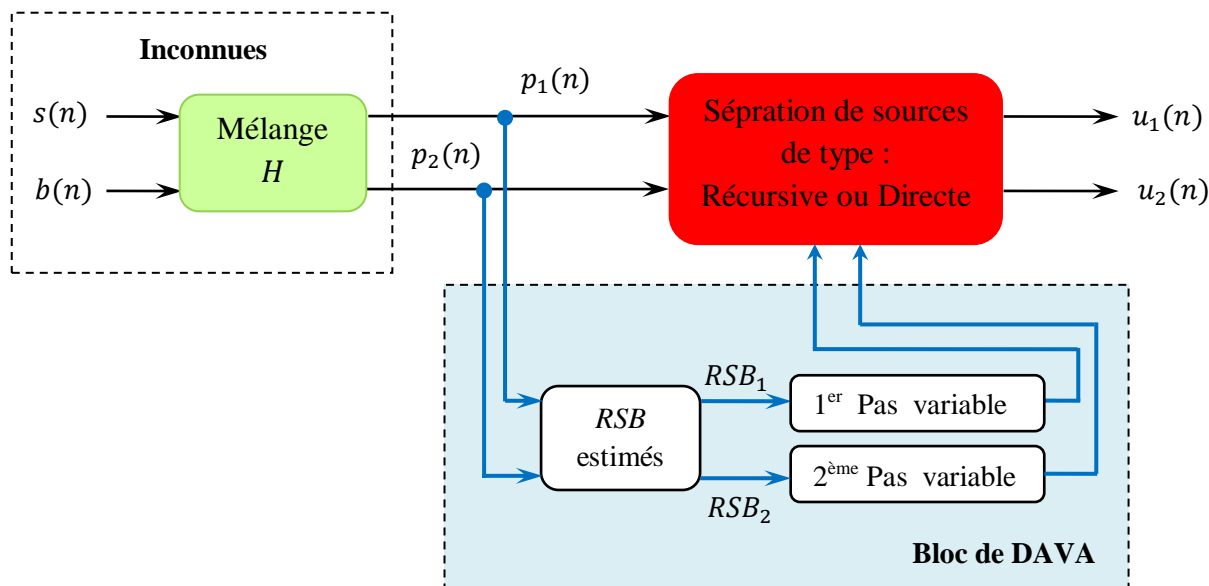


Figure 3.1 : Schéma bloc du principe des structures de séparation de sources améliorées.

### 3.2.1 Structure réursive améliorée

La figure (3.2) représente une amélioration de la structure de séparation de sources réursive appelée ABBSS ou Automatic Backward Blind Source Separation. Cette version améliorée correspond à une structure réursive contrôlée de manière adaptative par l'évolution du rapport signal à bruit (RSB) estimé sur les deux voies d'observations. Cette structure est basée sur l'utilisation de quatre filtres adaptatifs, à savoir, les filtres adaptatifs principaux ( $w_{21}(n)$ ,  $w_{12}(n)$ ) et les filtres adaptatifs secondaires ( $w_{cont1}(n)$ ,  $w_{cont2}(n)$ ). Le rôle principale de ces derniers ( $w_{cont1}(n)$ ,  $w_{cont2}(n)$ ) est de fournir une estimation des rapports signaux à bruit mesurés sur les deux signaux d'observations. Ces deux estimées sont ensuite utilisées pour contrôler les pas d'adaptations des coefficients des filtres adaptatifs principaux ( $w_{21}(n)$ ,  $w_{12}(n)$ ) [42].

### 3.2.2 Principe de fonctionnement des structures améliorées

Le principe général de ces structures est de faire varier les pas d'adaptations  $\mu_{21}(n)$  et  $\mu_{12}(n)$  des coefficients des filtres principaux ( $w_{21}(n)$ ,  $w_{12}(n)$ ) dans un interval  $[\mu_{min}, \mu_{max}]$  en fonction des valeurs des rapport signaux à bruit (RSBs) sur chacune des voies.

Pour le pas associé à  $w_{21}(n)$ , le comportement est le suivant :

- $\mu_{21}(n) \rightarrow \mu_{max}$  en période de bruit seul afin de faire converger  $w_{21}(n)$ ,
- $\mu_{21}(n) \rightarrow \mu_{min}$  en phase de parole utile seule,

- $\mu_{21}(n) \in [\mu_{min}, \mu_{max}]$  en phase de parole bruitée. Il faut que  $\mu_{21}(n) \rightarrow \mu_{min}$  si la parole utile est prépondérante devant le bruit et que  $\mu_{21}(n) \rightarrow \mu_{max}$  dans le cas contraire.

Le raisonnement inverse vaut pour le pas associé à  $w_{12}(n)$ .

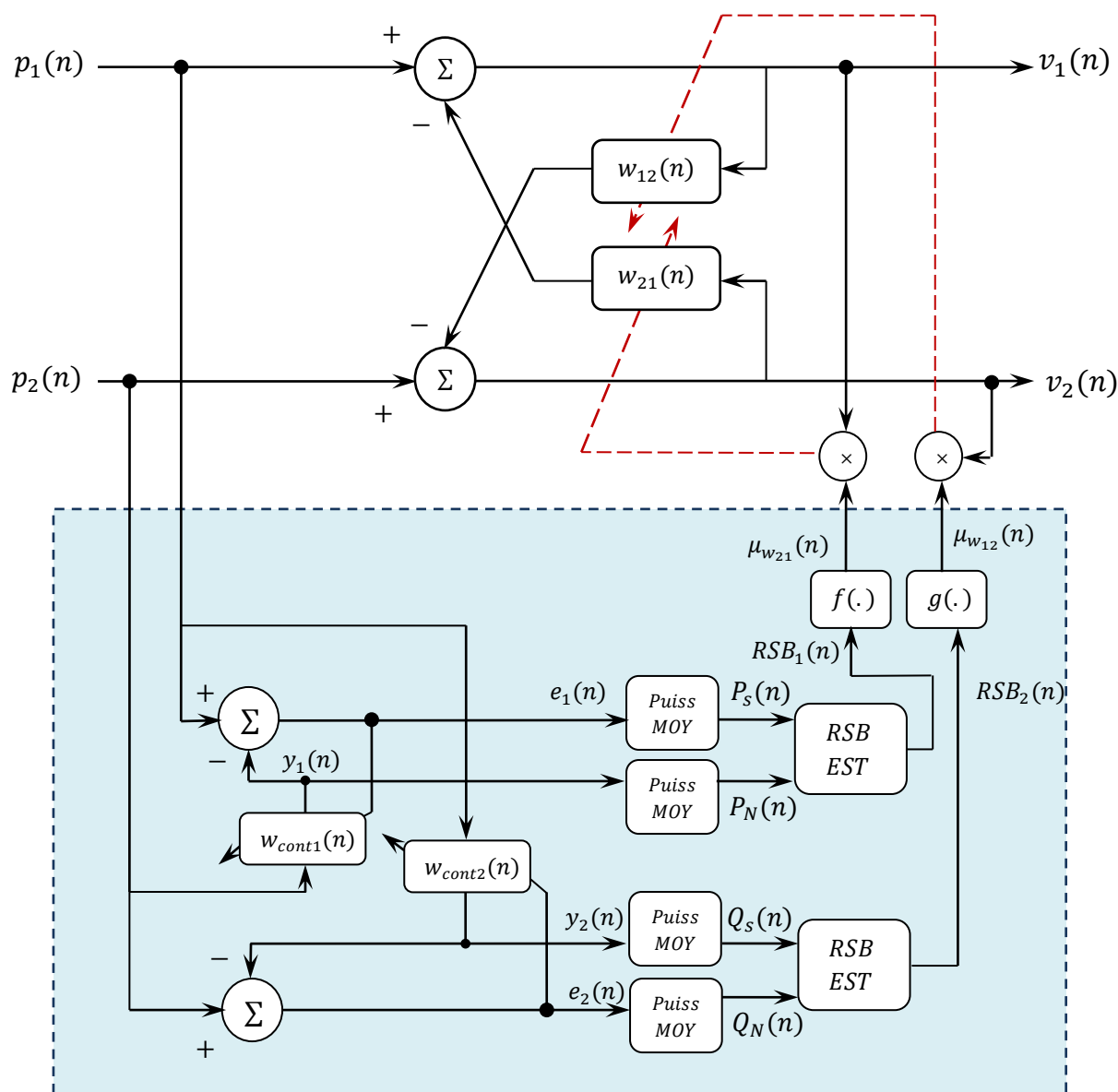


Figure 3.2 : Structure réursive améliorée (ABBSS).

### 3.2.2.1 Estimation des rapports signaux à bruit RSB

Pour contrôler les pas des filtres adaptatif de l'étage supérieur ( $w_{21}(n)$ ,  $w_{12}(n)$ ), il est nécessaire d'estimer les rapports signaux à bruit mesurés sur les deux signaux d'observations.



❖ Estimation du  $RSB_1$

Pour estimer le rapport signal à bruit  $RSB_1$  au niveau de la voie primaire, la puissance moyenne du bruit  $P_N(n)$  et celle de la parole  $P_S(n)$  sont calculées à partir des signaux disponibles en sortie du filtre secondaire  $w_{cont1}(n)$  suivant :

$$P_S(n) = \sum_{j=0}^{L-1} e_1^2(n-j) \quad (3.1)$$

$$P_N(n) = \sum_{j=0}^{L-1} y_1^2(n-j) \quad (3.2)$$

où  $y_1(n)$  et  $e_1(n)$  représentent respectivement la composante de bruit et l'estimée du signal utile mesurée sur la voie  $p_1(n)$ .

A partir de  $P_S(n)$  et  $P_N(n)$ , le  $RSB_1$  de la voie primaire est donné par l'expression suivante :

$$RSB_1(n) = 10 * \log_{10} \left\{ \frac{P_S(n)}{P_N(n)} \right\} \quad (3.3)$$

❖ Estimation du  $RSB_2$

La sortie  $y_2(n)$  du filtre secondaire  $w_{cont2}(n)$  et l'erreur  $e_2(n)$  sont utilisées pour estimer le rapport signal à bruit ( $RSB_2$ ) au niveau de la voie de référence, tel que le signal estimé  $y_2(n)$  et l'erreur  $e_2(n)$  correspondent respectivement à la composante du signal de parole et de bruit mesuré sur la voie  $p_2(n)$  à partir du filtre adaptatif  $w_{cont2}(n)$ .

Les puissances moyennes de la parole et du bruit ( $Q_S(n), Q_N(n)$ ) au niveau de la voie de référence sont données par les deux expressions suivantes :

$$Q_S(n) = \sum_{j=0}^{L-1} y_2^2(n-j) \quad (3.4)$$

$$Q_N(n) = \sum_{j=0}^{L-1} e_2^2(n-j) \quad (3.5)$$

A partir de  $Q_S(n)$  et  $Q_N(n)$ , le  $RSB_2$  de la voie de référence est donné par cette expression :

$$RSB_2(n) = 10 * \log_{10} \left\{ \frac{Q_S(n)}{Q_N(n)} \right\} \quad (3.6)$$

### 3.2.2.2 Contrôle des pas d'adaptations des filtres principaux

Le contrôle des pas d'adaptations variables  $\mu_{w_{21}}(n)$  et  $\mu_{w_{12}}(n)$  des coefficients des filtres principaux ( $w_{21}(n)$  et  $w_{12}(n)$ ) est basé sur les rapports signaux à bruit  $RSB_1(n)$  et  $RSB_2(n)$  estimés.

❖ Contrôle du pas d'adaptation  $\mu_{w_{21}}(n)$

Le pas d'adaptations  $\mu_{w_{21}}(n)$  est contrôlé à partir du  $RSB_1(n)$  estimé comme suit : Lorsque le  $RSB_1$  à l'entrée  $p_1(n)$  de la voie primaire est élevé, la valeur du pas d'adaptation  $\mu_{w_{21}}(n)$  du filtre  $w_{21}(n)$  doit être faible. Au contraire, pour un rapport signal à bruit faible sur l'entrée  $p_1(n)$ , le pas d'adaptation  $\mu_{w_{21}}(n)$  prend des valeurs importantes.

$\mu_{w_{21}}(n)$  est contrôlé, comme le montre l'équation suivante:

$$\mu_{w_{21}}(n) = \begin{cases} \mu_{1min} & RSB_1(n) > RSB_{1max} \\ \mu_{1max} & RSB_1(n) < RSB_{1min} \\ f(RSB_1(n)) & \text{sinon} \end{cases} \quad (3.7)$$

où les paramètres  $\mu_{1max}$ ,  $\mu_{1min}$  sont respectivement la valeur maximale et minimale du pas d'adaptation  $\mu_{w_{21}}(n)$ . La fonction  $f(.)$  est une fonction linéaire, continue et monotone décroissante en fonction du  $RSB_1(n)$ .

❖ Contrôle du pas d'adaptation  $\mu_{w_{12}}(n)$

Le pas d'adaptation  $\mu_{w_{12}}(n)$  est contrôlé à partir du  $RSB_2(n)$  estimé à l'entrée de la voie de référence  $p_2(n)$ . Dans le cas d'un rapport signal à bruit fort à l'entrée  $p_2(n)$ , la valeur du pas d'adaptation  $\mu_{w_{12}}(n)$  du filtre  $w_{12}(n)$  est grande. En revanche, lorsque le  $RSB_2$  est faible, la valeur du pas d'adaptation  $\mu_{w_{12}}(n)$  est petite.

L'équation suivante détermine les valeurs du  $\mu_{w_{12}}(n)$  basé sur  $RSB_2(n)$  :

$$\mu_{w_{12}}(n) = \begin{cases} \mu_{2min} & RSB_2(n) < RSB_{2min} \\ \mu_{2max} & RSB_2(n) > RSB_{2max} \\ g(RSB_2(n)) & \text{sinon} \end{cases} \quad (3.8)$$

où les paramètres  $\mu_{2max}$ ,  $\mu_{2min}$  sont respectivement la valeur maximale et minimale du pas d'adaptation  $\mu_{w_{12}}(n)$ . La fonction  $g(.)$  est une fonction linéaire, continue et monotone croissante en fonction du  $RSB_2(n)$ .

La figure (3.3) illustre la variation des pas d'adaptations par rapport aux  $RSB$  calculés au cours du temps ainsi que des seuils qui leur sont associés. Nous donnons la description algorithmique du contrôle des pas d'adaptations des structures améliorées, Algorithme 2, accompagnée du diagramme d'activités du calcul des pas (figure (3.4)).

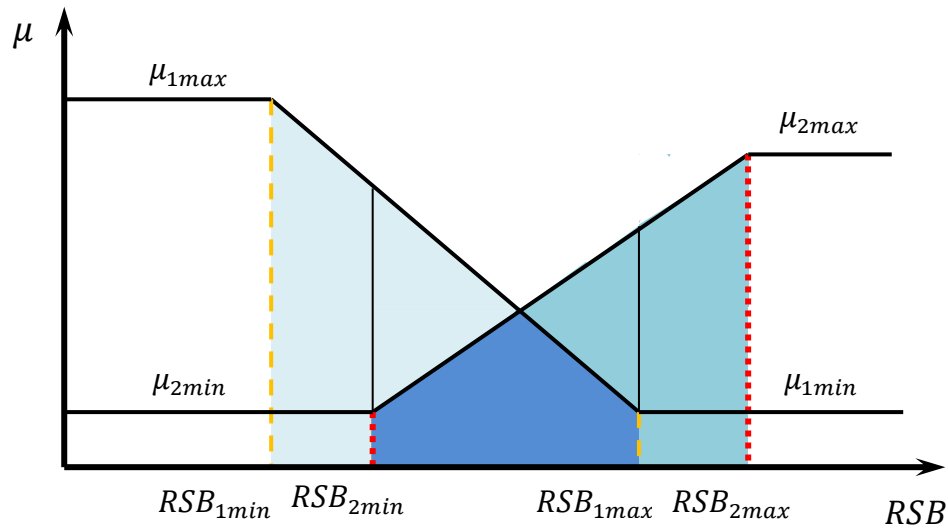


Figure 3.3 : Illustration de la variation des pas par rapport aux  $RSB$  en fonction de leurs seuils maximum et minimum sur chacune des voies.

---

**Algorithme 2** Algorithme du contrôle des pas d'adaptations

---

Début

- 1 : **si**  $RSB_1 > RSB_{1max}$  **alors**
- 2 :  $\mu_{w21} \leftarrow \mu_{w21_{min}}$
- 3 : **sinon si**  $RSB_1 < RSB_{1min}$  **alors**
- 4 :  $\mu_{w21} \leftarrow \mu_{w21_{max}}$
- 5 : **sinon**
- 6 :  $\mu_{w21} \leftarrow f(\mu_{w21_{min}}, \mu_{w21_{max}}, RSB_{1min}, RSB_{1max})$
- 7 : **fin si**
- 8 : **si**  $RSB_2 < RSB_{2min}$  **alors**
- 9 :  $\mu_{w12} \leftarrow \mu_{w12_{min}}$
- 10 : **sinon si**  $RSB_2 > RSB_{2max}$  **alors**
- 11 :  $\mu_{w12} \leftarrow \mu_{w12_{max}}$
- 12 : **sinon**
- 13 :  $\mu_{w12} \leftarrow g(\mu_{w12_{min}}, \mu_{w12_{max}}, RSB_{2min}, RSB_{2max})$
- 14 : **fin si**

Fin

---

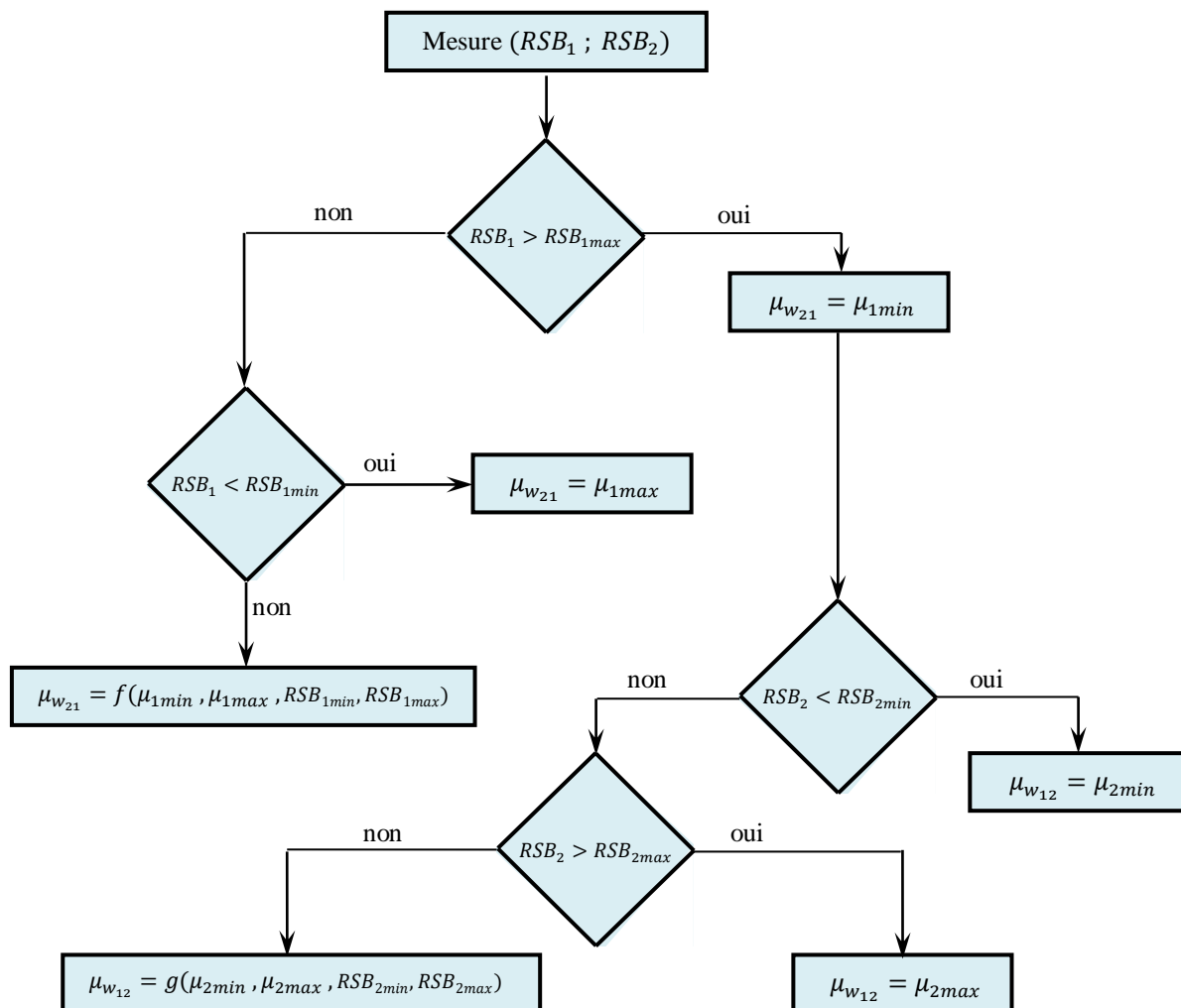


Figure 3.4 : Diagramme d'activités du calcul des pas d'adaptations.

Les paramètres utilisés dans la première structure améliorée (ABBSS) ont été choisis afin d'éviter la divergence à la sortie de cette structure.

### 3.2.3 Structure directe améliorée

La complexité de la première structure améliorée ABBSS réside dans le choix des paramètres seuils ainsi que le choix des pas d'adaptations, car l'étage supérieur de cette méthode correspond à une structure récursive, et celle-ci est très sensible, tel que un mauvais choix de ces paramètres peut perturber l'adaptation des filtres  $w_{21}(n)$  et  $w_{12}(n)$ , d'où les signaux en sortie divergent. Pour réduire la complexité concernant le choix difficile des paramètres seuils, nous avons proposé de remplacer la structure récursive dans le premier modèle par une structure directe car cette dernière est facile à contrôler (voir figure 3.5) [42].

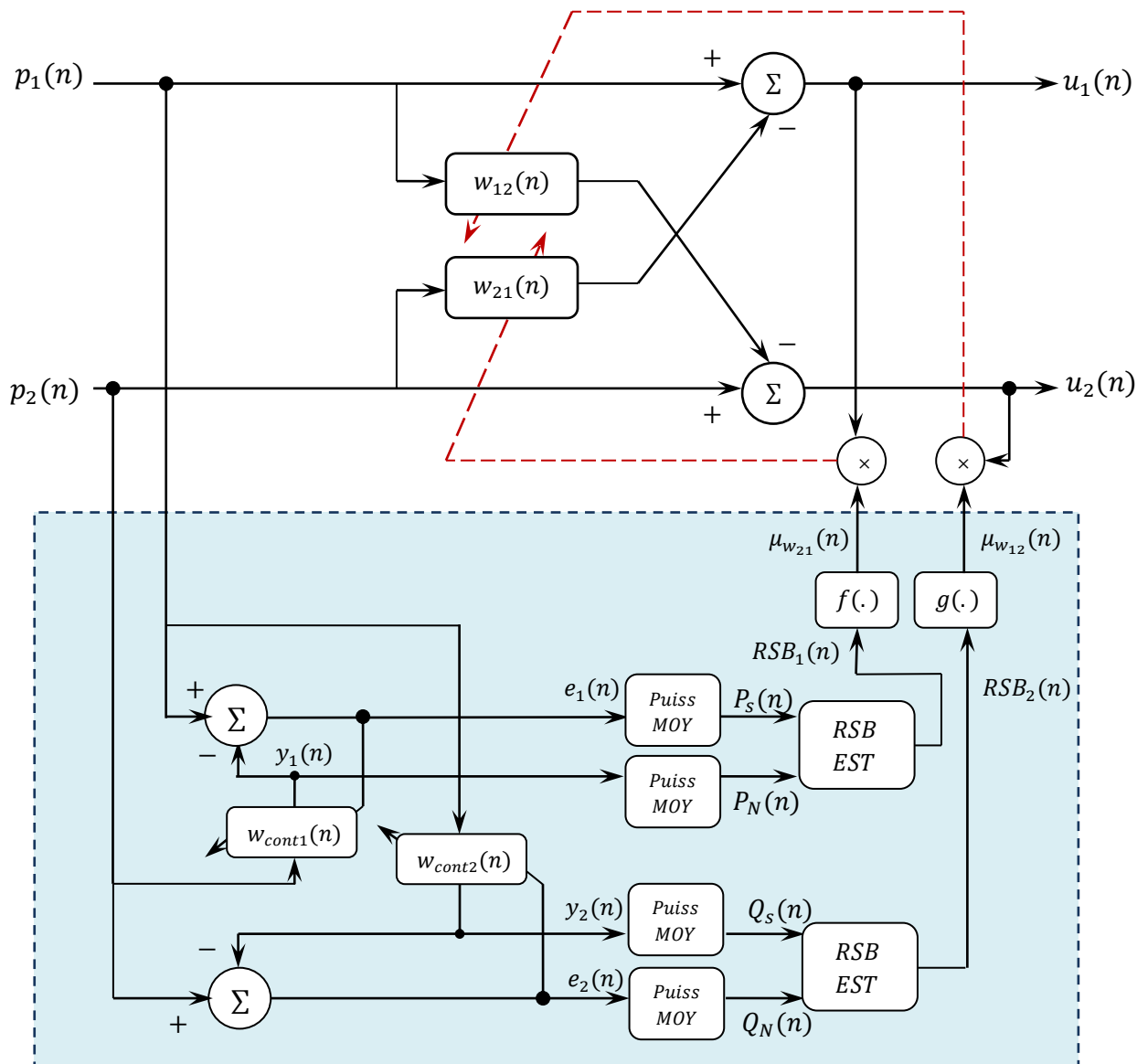


Figure 3.5 : Structure directe améliorée (AFBSS).

### 3.3 Résultats expérimentaux

Nous limitons notre étude expérimentale uniquement pour la structure directe améliorée AFBSS proposée dans ce chapitre. Dans cette étude, nous utilisons:

- le modèle du système de mélange convolutif montré dans la figure 2.6.
- les réponses impulsionnelles décrites à la section 2.5.2.
- les critères de comparaison qui ont été définis dans la section 2.5.4.

Les résultats complets des simulations du comportement de notre structure directe améliorée AFBSS seront présentés, ainsi qu'une évaluation des effets de certains paramètres décisifs ( $L$  : la taille du filtre,  $\mu$  : pas d'adaptation,  $RSB$  : rapport signal à bruit) sur celle-ci. Pour finir cette étude, une comparaison entre notre méthode améliorée et sa version classique FBSS sera détaillée.

### 3.3.1 Simulation de la structure directe améliorée (AFBSS)

Ce paragraphe présente les résultats obtenus à partir de la structure directe améliorée AFBSS décrite à la section 3.2.3 (voir figure 3.5), et qui correspond à une structure de séparation de sources directe contrôlée de manière adaptative par l'évolution du rapport signal à bruit (RSB) estimé sur les deux voies d'observations. Nous avons utilisé l'algorithme NLMS pour adapter chaque filtre de cette structure. Les paramètres de simulations pour le mélange convolutif, pour la structure directe classique (FBSS) ainsi que pour notre structure proposée (AFBSS) sont représentés dans le tableau 3.1.

Tableau 3.1 : Paramètres de simulations des structures (Mélange convolutif, FBSS et AFBSS)

Structures	Paramètres
Mélange convolutif	<p><b>Signal de parole originale :</b> signal de la figure (2.11)</p> <p><b>Un bruit Blanc :</b> signal de la figure (2.12)</p> <p><b>Longueur des filtres :</b> <math>L_{h12} = L_{h21} = 64</math></p> <p><b>RSBs d'entrées sur les deux voies :</b> <math>RSB_1 = 3 \text{ dB}; RSB_2 = 0 \text{ dB}</math></p>
FBSS	<p><b>Longueur des filtres :</b> <math>L_{w12} = L_{w21} = 64</math></p> <p><b>Pas d'adaptation des filtres :</b> <math>\mu_{w12} = 0.4; \mu_{w21} = 0.4</math></p>
AFBSS	<p><b>Longueur des filtres :</b> <math>L_{w12} = L_{w21} = L_{wcont1} = L_{wcont2} = 64</math></p> <p><b>Pas d'adaptation des filtres secondaires:</b> <math>\mu_{wcont1} = 0.4; \mu_{wcont2} = 0.001</math></p> <p><b>Pas d'adaptation des filtres principaux:</b> <math>\mu_{w12min} = 0; \mu_{w12max} = 0.2;</math> <math>\mu_{w21min} = 0; \mu_{w21max} = 0.2</math></p> <p><b>Valeurs seuils du RSB estimés :</b> <math>RSB_{1min} = -40 \text{ dB}; RSB_{1max} = -15 \text{ dB};</math> <math>RSB_{2min} = -5 \text{ dB}; RSB_{2max} = 2 \text{ dB}</math></p>

Les signaux sources (signal de parole et bruit Blanc), les signaux d'entrée (signaux de mélanges) et de sortie de la structure directe améliorée AFBSS sont représentés avec leurs spectrogrammes sur la figure (3.6).

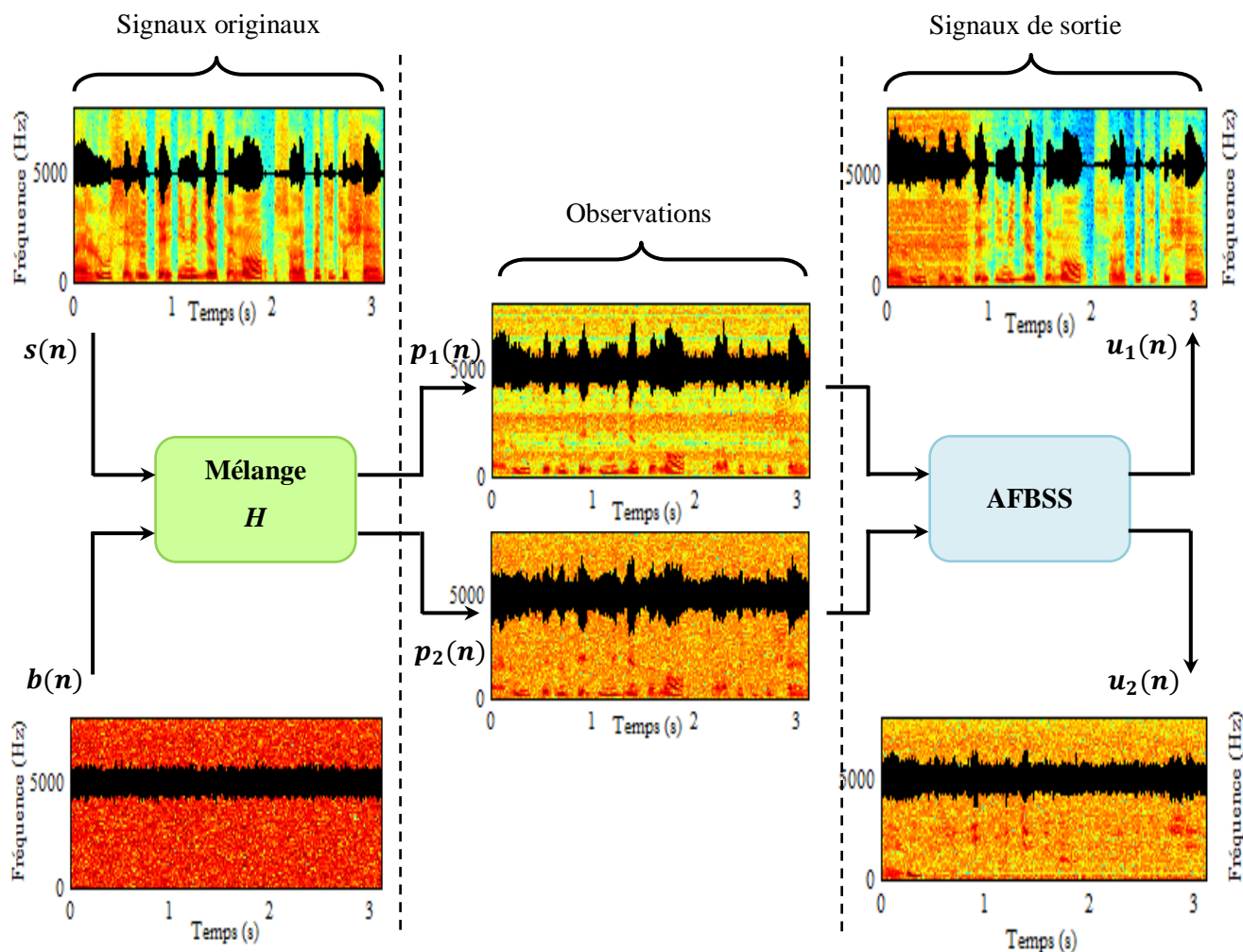


Figure 3.6 : Signaux sources (parole originale  $s(n)$ , bruit Blanc  $b(n)$ ), signaux de mélange ( $p_1(n)$ ,  $p_2(n)$ ) et signaux de sortie ( $u_1(n)$ ,  $u_2(n)$ ) de la structure AFBSS.

A partir de la figure (3.6), nous observons que le signal de parole en sortie  $u_1(n)$  est visuellement nettoyé du bruit et très proche du signal de parole originale. Ceci grâce au contrôle de l'adaptation du filtre  $w_{21}(n)$  par une technique automatique de détection d'activité vocale (DAV), ce filtre est adapté uniquement pendant les périodes de non activité vocale du mélange  $p_2(n)$ . Durant ces périodes d'inactivité vocale, le filtre adaptatif  $w_{21}(n)$  n'a à son entrée que du bruit, ce qui lui permet d'identifier la réponse de couplage  $h_{21}(n)$  puis supprimer le bruit au niveau du signal de sortie  $u_1(n)$ .

D'après cette même figure, nous remarquons que la parole est partiellement enlevée de la deuxième sortie  $u_2(n)$ , car si on veut supprimer complètement la parole on doit adapter le filtre  $w_{12}(n)$  pendant les périodes d'activité vocale. Malheureusement, durant ces périodes le bruit persiste et le filtre  $w_{12}(n)$  s'actualise avec un mélange à l'entrée  $p_1(n)$ .

Rappelons que la structure AFBSS utilise un système de détection d'activité vocale automatique (DAVA) pour contrôler l'adaptation des filtres principaux ( $w_{21}(n)$ ,  $w_{12}(n)$ ), cette technique de segmentation est basée sur la variation des pas d'adaptations ( $\mu_{w_{21}}(n)$ ,  $\mu_{w_{12}}(n)$ ) des coefficients des filtres adaptatifs de séparations ( $w_{21}(n)$ ,  $w_{12}(n)$ ) en fonction du rapport signal à bruit estimé sur chacune des voies d'observation.

Nous présentons sur la figure (3.7) le signal de parole originale, l'évolution du rapport signal à bruit  $RSB_1$  estimé au niveau de l'entrée primaire  $p_1(n)$  et le pas d'adaptation variable  $\mu_{w_{21}}(n)$  du filtre principal  $w_{21}(n)$ .

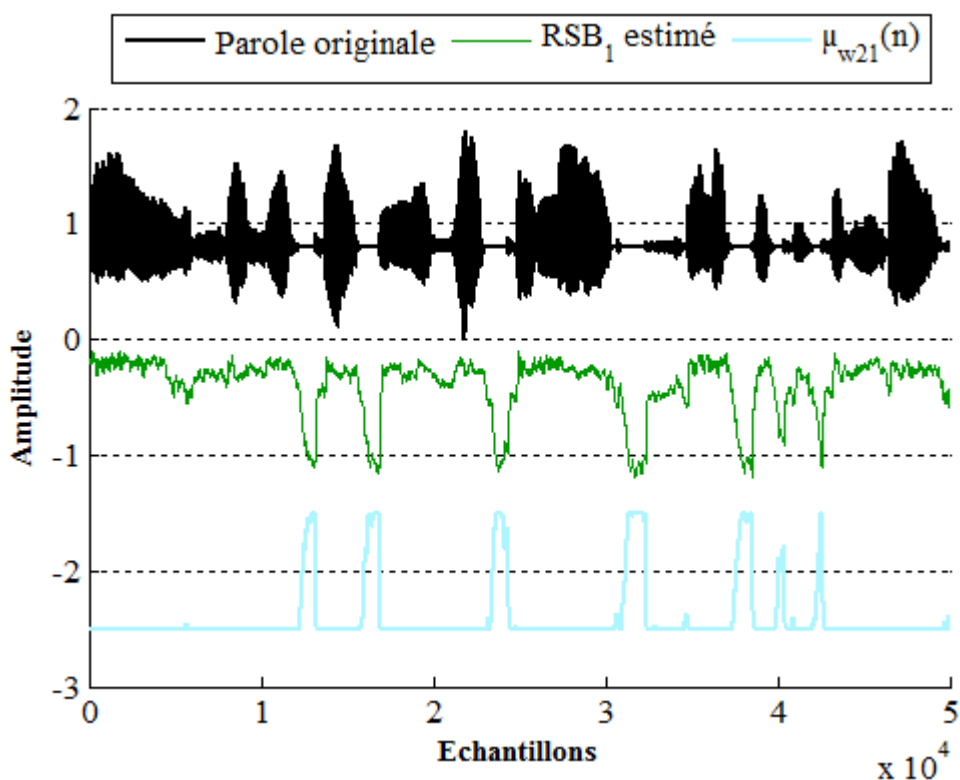


Figure 3.7 : Signal de parole original (en haut), le  $RSB_1$  estimé au niveau de l'entrée primaire de la structure AFBSS (au milieu) et le pas d'adaptation  $\mu_{w_{21}}(n)$  du filtre principal  $w_{21}(n)$  (en bas).

A partir de la figure 3.7, nous remarquons que lorsque le rapport signal à bruit  $RSB_1$  estimé au niveau de l'entrée primaire  $p_1(n)$  est élevé, la valeur du pas d'adaptation  $\mu_{w_{21}}(n)$  du



filtre  $w_{21}(n)$  est faible. Par contre, dans le cas d'un rapport signal à bruit  $RSB_1$  faible (lorsque le signal de parole est absent), le pas d'adaptation  $\mu_{w_{21}}(n)$  prend des valeurs importantes.

L'évolution du rapport signal à bruit  $RSB_2$  estimé au niveau de l'entrée de référence  $p_2(n)$  et le pas d'adaptation  $\mu_{w_{12}}(n)$  du filtre principal  $w_{12}(n)$  sont représentés avec le signal de parole originale sur la figure 3.8.

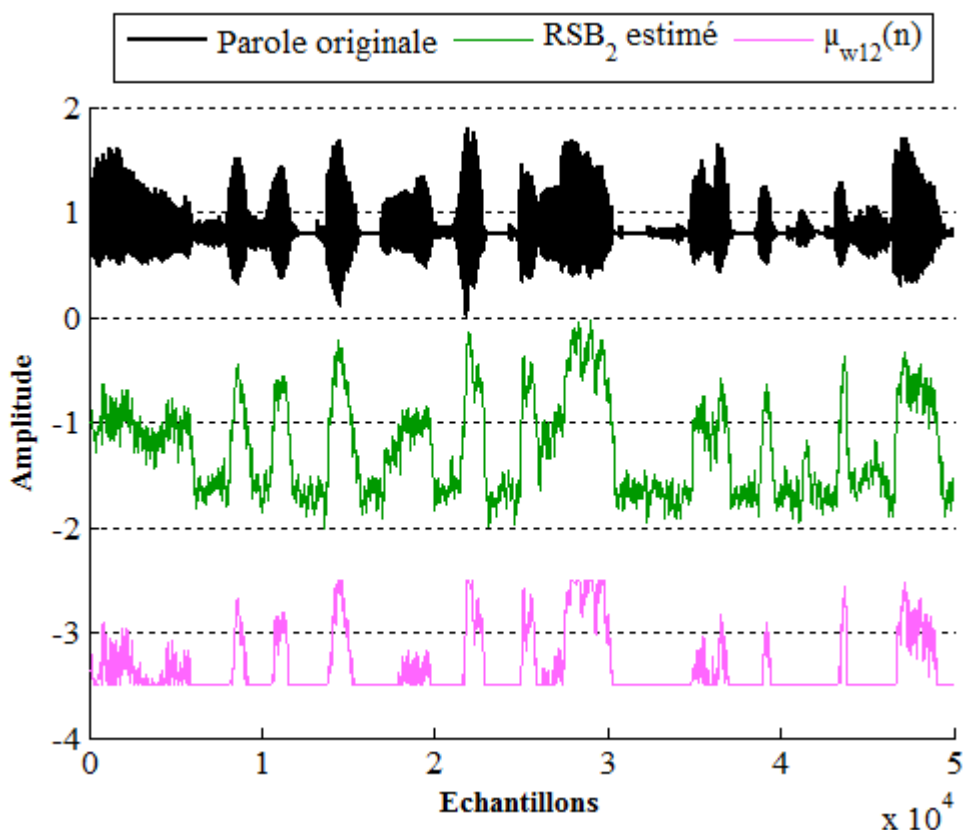


Figure 3.8 : Signal de parole original (en haut), le  $RSB_2$  estimé au niveau de l'entrée de référence de la structure AFBSS (au milieu) et le pas d'adaptation  $\mu_{w_{12}}(n)$  du filtre principal  $w_{12}(n)$  (en bas).

Selon cette figure, nous notons que le pas d'adaptation du filtre principal  $w_{12}(n)$  prend de grandes valeurs lorsque le rapport signal à bruit  $RSB_2$  est grand (ce qui signifie les sections où le signal de parole est présent). En revanche, le pas d'adaptation  $\mu_{w_{12}}(n)$  du filtre  $w_{12}(n)$  est faible lorsque le  $RSB_2$  est faible.

Nous présentons sur la même figure (figure 3.9) l'évolution des pas d'adaptations ( $\mu_{w_{12}}(n)$ ,  $\mu_{w_{21}}(n)$ ) des filtres principaux.

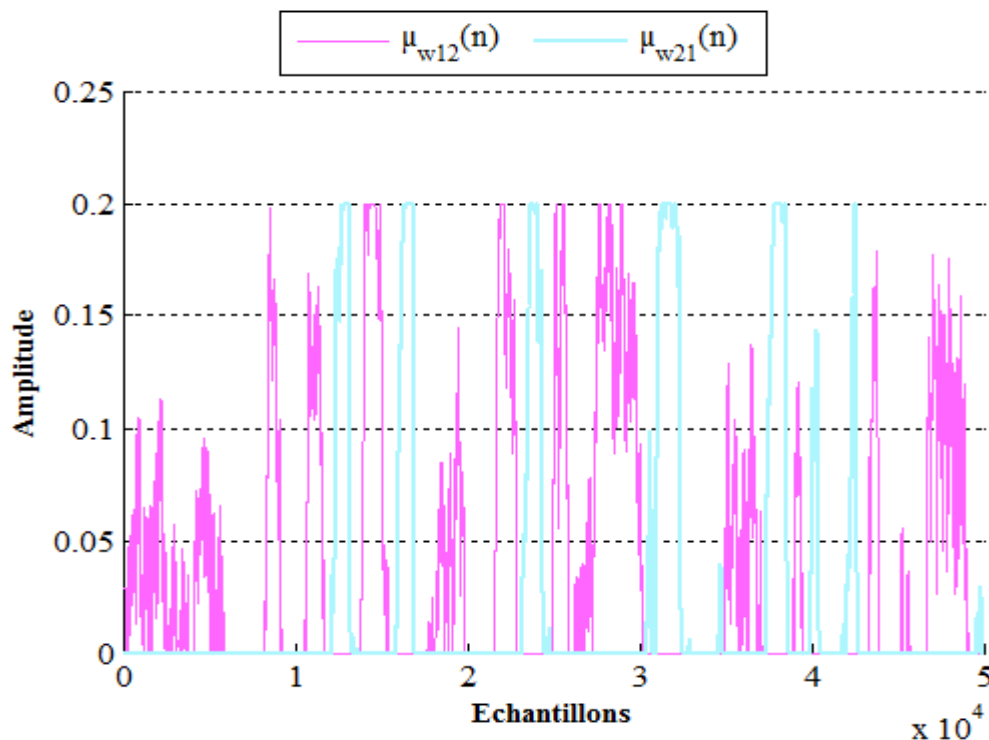


Figure 3.9 : Comportement des pas d'adaptations:  $\mu_{w_{12}}(n)$  (en magenta) et  $\mu_{w_{21}}(n)$  (en cyan).

A partir de cette figure, nous pouvons voir que l'ajustement du pas d'adaptation  $\mu_{w_{21}}(n)$  est désactivé pendant les périodes où l'ajustement du pas d'adaptation  $\mu_{w_{12}}(n)$  est permis, ces pas sont complémentaires. Par conséquent ces deux pas d'adaptation ( $\mu_{w_{21}}(n)$  et  $\mu_{w_{12}}(n)$ ) agissent comme un détecteur d'activité vocale (DAV) qui garantit que le filtre  $w_{21}(n)$  est uniquement adapté au cours des périodes de bruit, tandis que le filtre  $w_{12}(n)$  est adapté uniquement pendant les périodes d'activité vocale.

Dans les simulations ci-après, nous nous intéresserons uniquement à la voie restituant le signal de parole utile (une seule des deux sorties de traitement est exploitée) et la segmentation (DAV) utilisée correspond à l'évolution du pas  $\mu_{w_{21}}(n)$ .

Nous représentons sur la figure (3.10) les deux segmentations (DAV manuelle et automatique). D'après cette figure, nous remarquons que l'évolution du pas d'adaptation obtenu avec la segmentation automatique coïncide avec les valeurs du pas d'adaptation obtenus manuellement. Ceci confirme que les deux segmentations (manuelle et automatique) ont le même comportement.

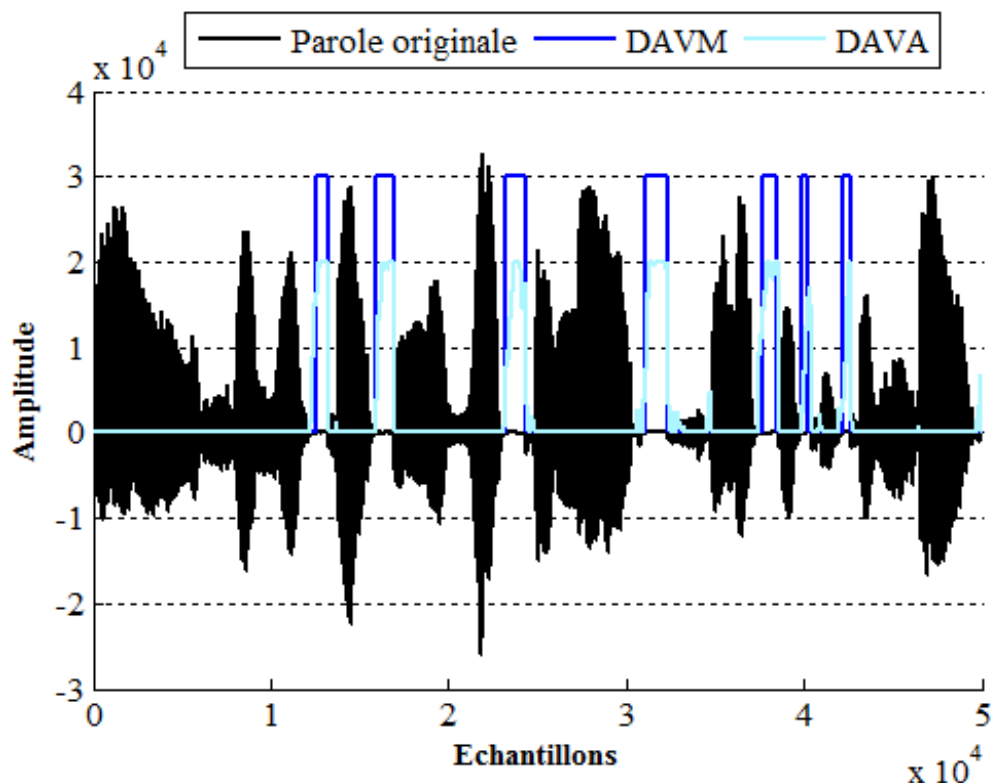


Figure 3.10 : Signal de parole utilisé avec les deux segmentations (manuelle et automatique (AFBSS)),  $F_e = 16$  KHz.

### 3.3.2 Etude des performances de la structure directe améliorée AFBSS

Pour améliorer le comportement de la structure (AFBSS) dans le domaine de débruitage du signal de parole. Nous avons réalisé plusieurs simulations pour voir l'influence des paramètres ( $L$ ,  $\mu$  et  $RSB$ ) sur les performances de cette structure. C'est à dire l'étude de performances de l'algorithme NLMS en termes de vitesse de convergence.

Ces paramètres sont :

$L$  : taille du filtre adaptatif,

$\mu$  : pas d'adaptation de l'algorithme NLMS,

$RSB_1$ ,  $RSB_2$  : rapport signal à bruit sur les deux voies d'entrées du mélange.

L'évaluation de l'effet de chaque paramètre cité précédemment sur les performances de notre structure, a donné les résultats suivants :

### 3.3.2.1 Effet de la taille du filtre adaptatif $L$

Pour voir l'effet de la taille du filtre  $L$  sur la structure directe améliorée AFBSS, nous avons varié celle-ci en prenant les valeurs suivantes :  $L = 32, 64$  et  $128$ . Les autres paramètres sont égales à :  $\mu = 0.4$ ,  $RSB_1 = 3$  dB et  $RSB_2 = 0$  dB.

Pour réaliser cette comparaison, nous avons tracé l'évolution temporelle du désajustement (System Mismatch) pour les trois tailles, et nous avons obtenu les résultats montrés sur la figure (3.11). D'après cette figure, nous remarquons que l'algorithme NLMS converge mieux lorsque la taille du filtre  $L$  est faible, ce qui montre que la vitesse de convergence de l'algorithme NLMS est inversement proportionnelle à la taille du filtre adaptatif.

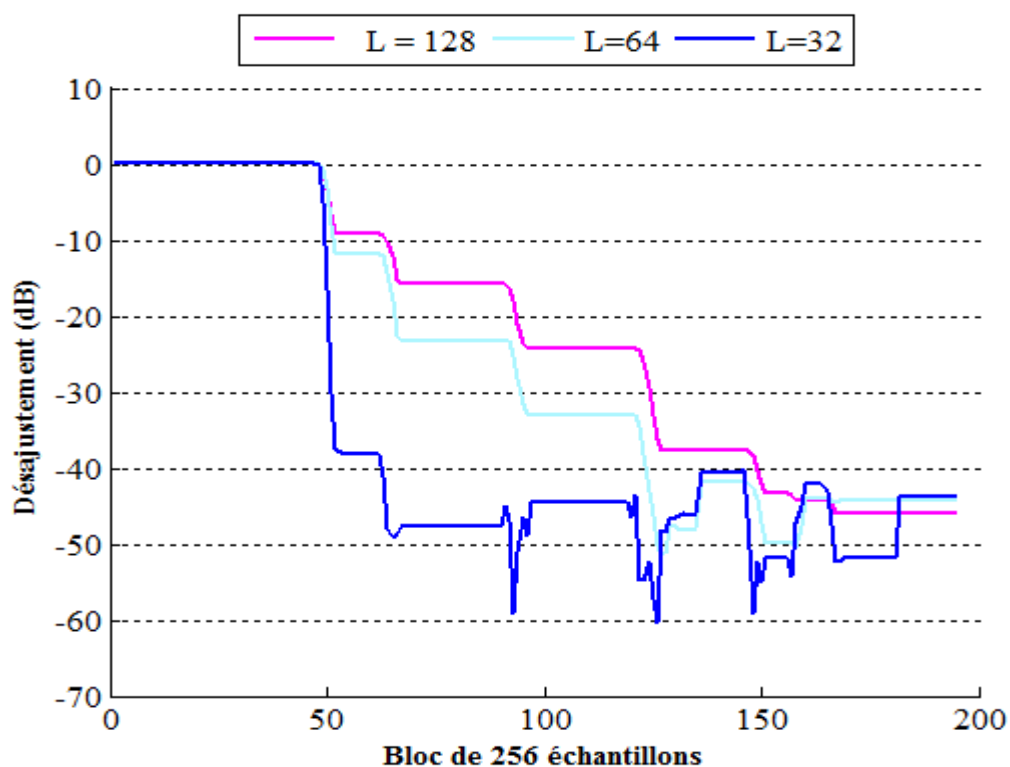


Figure 3.11 : Effet de la taille du filtre sur la structure directe améliorée AFBSS.

### 3.3.2.2 Effet du pas d'adaptation $\mu$

Dans cette simulation, nous avons varié le pas d'adaptation  $\mu$ , en prenant trois valeurs  $\mu = 0.2, 0.4$  et  $0.9$ , la taille du filtre  $L = 64$ , les rapports signal à bruit  $RSB_1 = 3$  dB et  $RSB_2 = 0$  dB.

Les résultats du désajustement de cette expérience sont représentés sur la figure (3.12). A partir de cette figure, nous remarquons que l'algorithme NLMS converge rapidement lorsque le pas d'adaptation  $\mu$  est grand, ce qui montre que la vitesse de convergence de l'algorithme NLMS est proportionnelle au pas d'adaptation.

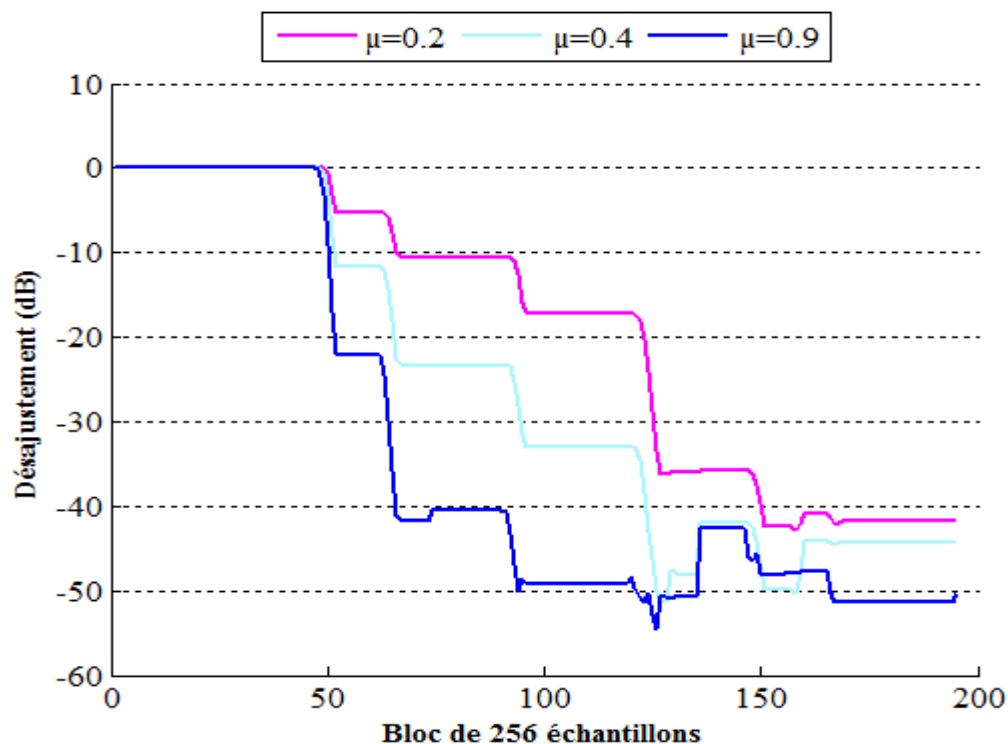


Figure 3.12 : Effet du pas d'adaptation sur la structure directe améliorée AFBSS.

### 3.3.2.3 Effet du *RSB* du mélange

Dans cette simulation, nous avons étudié l'influence du rapport signal à bruit au niveau du mélange  $p_2(n)$  sur les performances de la structure améliorée AFBSS. Les paramètres de cette simulation sont :

La taille du filtre :  $L=64$ , pas d'adaptation :  $\mu=0.4$ , nous avons fixé le rapport signal à bruit  $RSB_1 = 3\text{dB}$  et nous avons varié le  $RSB_2$ .

Les résultats de simulation du désajustement sont représentés sur la figure (3.13). D'après cette figure, nous avons remarqué que la vitesse de convergence de l'algorithme NLMS vers l'optimum est obtenue lorsque le rapport signal à bruit est faible.

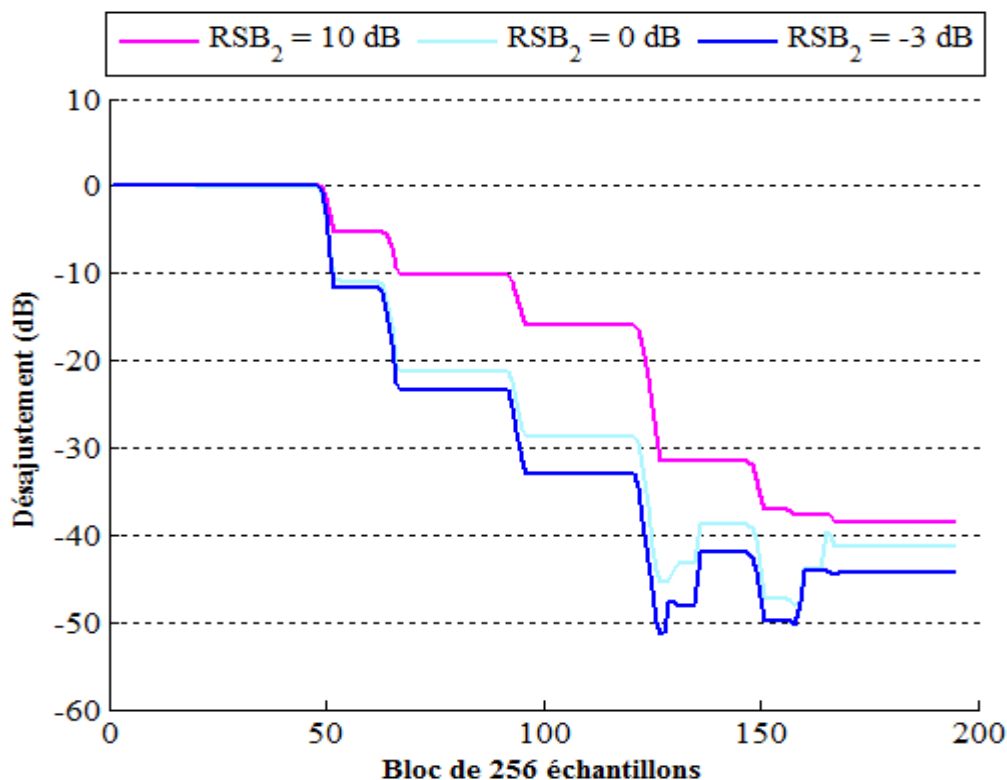


Figure 3.13 : Effet du *RSB* sur la structure directe améliorée AFBSS

### 3.3.3 Comparaison entre les structures AFBSS et FBSS

Pour évaluer les performances de la structure améliorée AFBSS qui utilise un système de détection d'activité vocale automatique (DAVA), nous avons réalisé une étude comparative de ses performances avec celles d'une structure de séparation de sources de type directe qui existe déjà dans la littérature. Cette structure directe considérée correspond à une structure classique (FBSS) qui utilise une segmentation manuelle (DAVM) dans son fonctionnement. Les paramètres de simulation des deux structures de séparation de sources directe (AFBSS et FBSS) ont été déjà illustrés dans le tableau 3.1.

Nous poursuivons cette étude par des étapes de comparaisons déjà illustrées dans l'organigramme de la figure 2.16.

#### 3.3.3.1 Représentation du signal rehaussé

Sur la figure (3.14), nous comparons les signaux de sorties obtenus avec les deux structures AFBSS et FBSS. Dans cet exemple, nous montrons l'évolution temporelle des différents signaux utilisés en simulation (signal de parole original, le signal de mélange ainsi que ceux de sortie de chaque structure). Pour chaque structure, nous avons également

représenté le spectrogramme du signal disponible en sortie de traitement. Nous pouvons conclure que les deux structures fonctionnent bien, c'est à dire que le signal de parole en sortie est presque totalement débruité et très proche du signal de parole original.

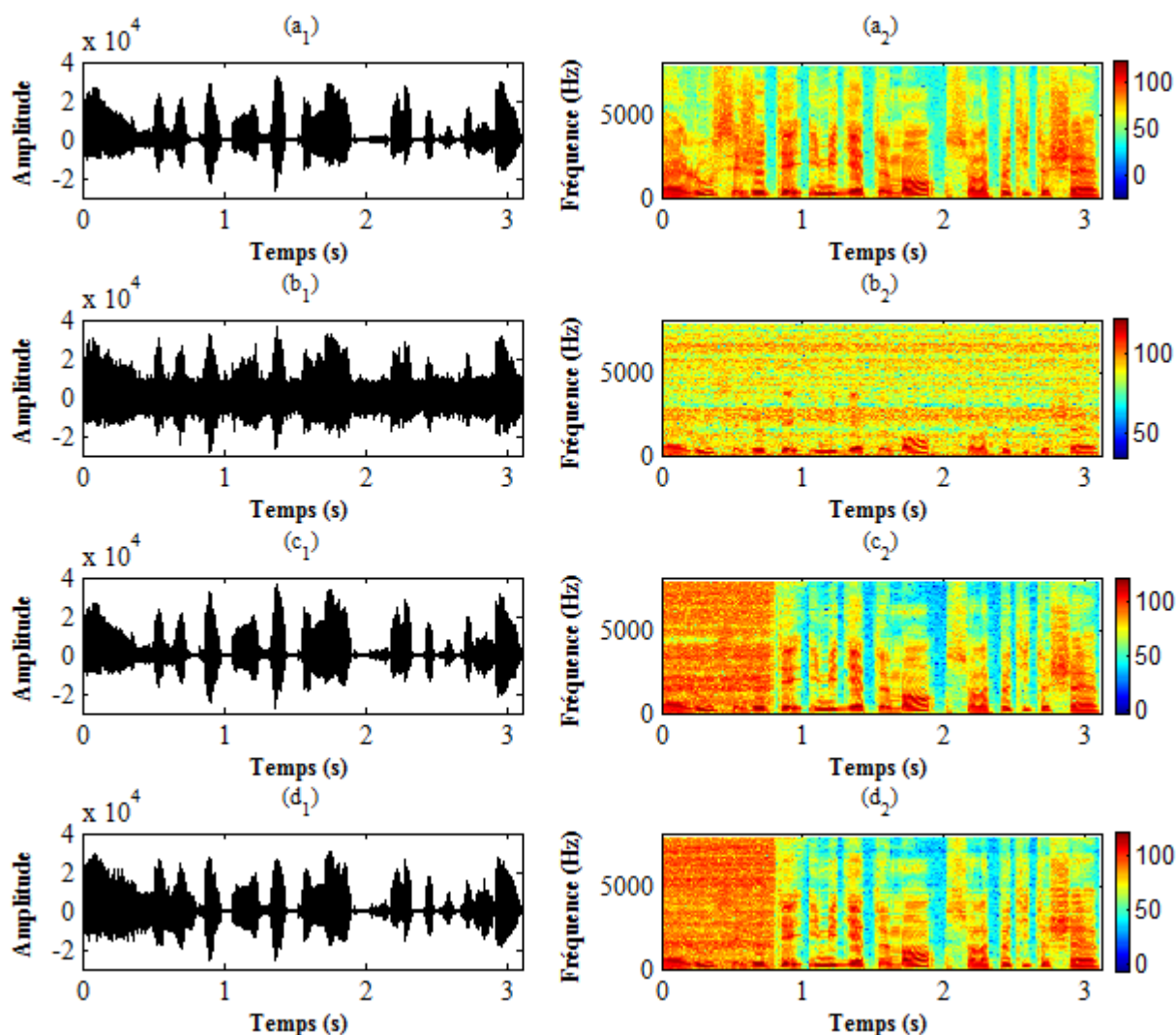


Figure 3.14 : (a<sub>1</sub>, a<sub>2</sub>) Signal de parole original  $s(n)$ , (b<sub>1</sub>, b<sub>2</sub>) signal de mélange  $p_1(n)$ , (c<sub>1</sub>, c<sub>2</sub>) signal de sortie  $u_1(n)$  de la structure (AFBSS), (d<sub>1</sub>, d<sub>2</sub>) signal de sortie  $u_1(n)$  de la structure (FBSS).

### 3.3.3.2 Évaluation de la distance cepstrale (DC)

Afin d'analyser les distorsions introduites par chacune des deux structures, nous avons représenté à la figure (3.15) la distance cepstrale entre le signal de parole original et le signal disponible en sortie de traitement pour chaque structure. Sur cette figure, chaque point correspond à une valeur moyenne obtenue sur 256 échantillons consécutifs.

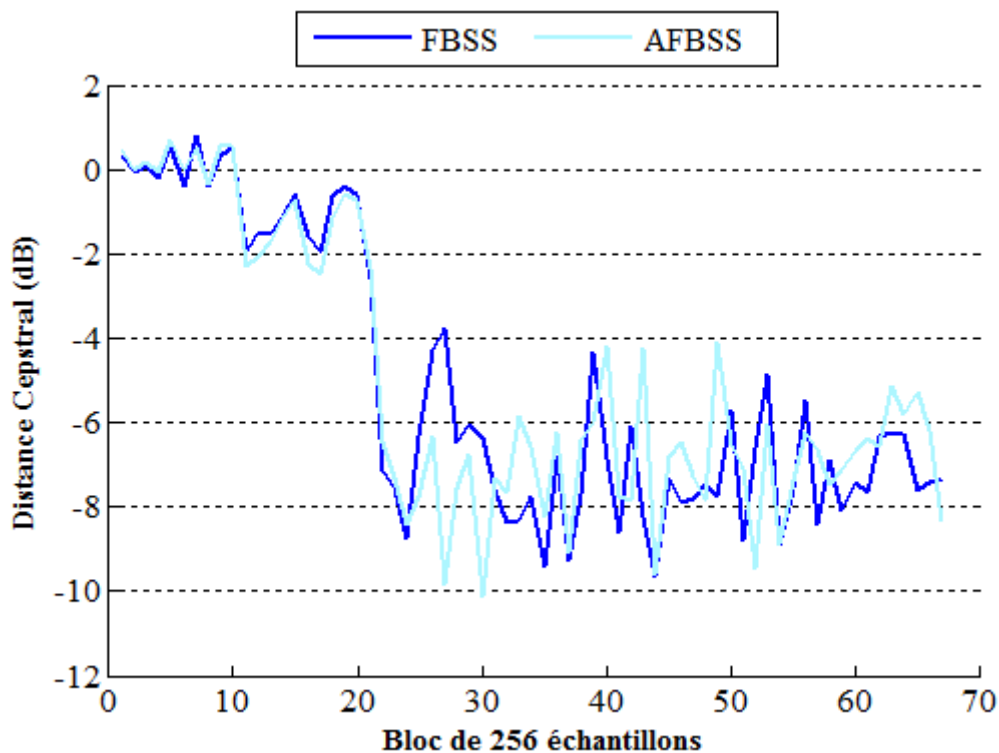


Figure 3.15 : Evolution du critère de la distance cepstrale des deux structures AFBSS (en cyan) et FBSS (en bleu).

D'après cette figure, nous remarquons que les valeurs moyennes de la distance cepstrale sont de -7.86 dB et -7.80 dB respectivement pour les structures AFBSS et FBSS. Ce qui signifie que la distance cepstrale obtenue à partir de la structure améliorée AFBSS est très proche de celle obtenue avec la structure classique FBSS. Ces résultats montrent également le bon comportement de la structure directe améliorée (AFBSS).

### 3.3.3.2 Évaluation du rapport signal à bruit (RSB)

La figure (3.16) représente l'évolution temporelle du rapport signal à bruit pour les deux structures. Chaque point de cette figure correspond à une valeur moyenne obtenue sur une trame de 1024 échantillons. Nous notons que la valeur moyenne du rapport signal à bruit est de 50.32 dB et de 50.21 dB, respectivement pour la structure AFBSS et la structure FBSS. Il existe donc un écart très faible entre les deux structures. Cela montre que la structure améliorée (AFBSS) fournit presque les mêmes performances que l'autre structure (FBSS).



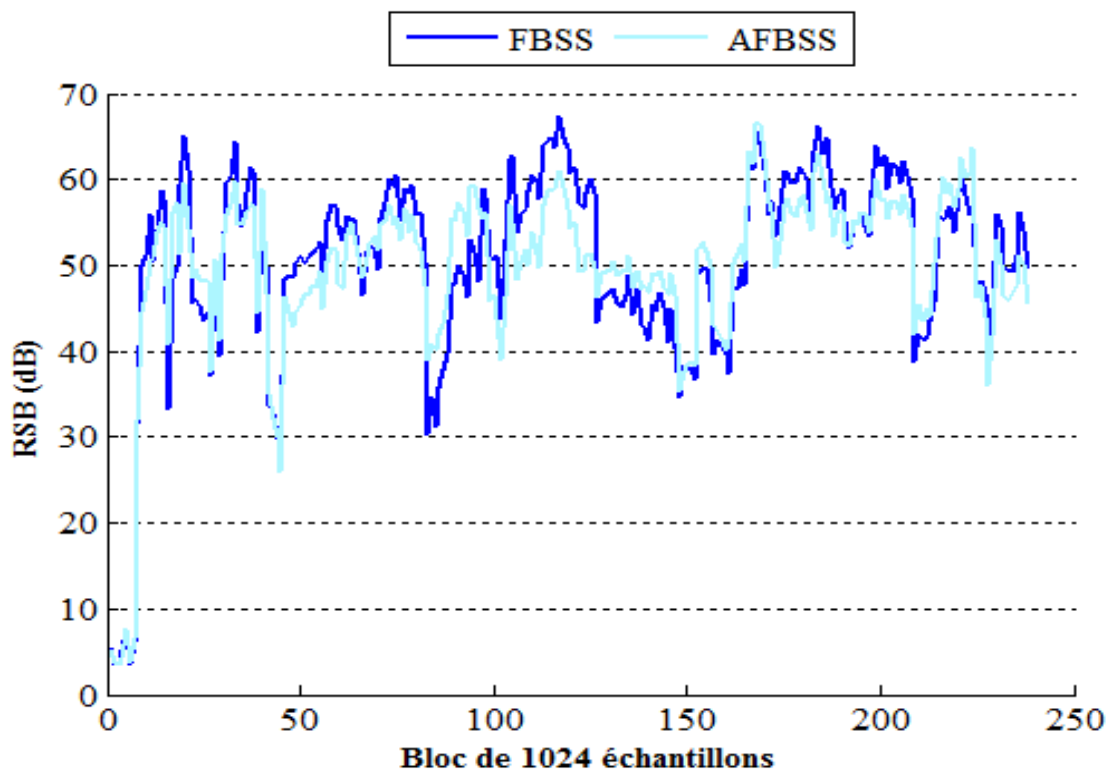


Figure 3.16 : Evolution du critère du rapport signal à bruit (RSB) pour les deux structures AFBSS (en cyan) et FBSS (en bleu).

### 3.3.3.3 Évaluation du système mismatch (SM)

De manière à compléter l'analyse de comportement de la structure améliorée AFBSS, nous présentons sur la figure (3.17) l'évolution temporelle du désajustement mesuré sur les coefficients du filtre adaptatif  $w_{21}(n)$  pour les deux structures.

À partir du résultat de la figure (3.17), nous notons que la vitesse de convergence est presque la même pour les deux structures. Ce qui démontre encore une fois le bon comportement de la structure améliorée AFBSS.

À partir des différents résultats présentés (signaux de sorties, distance cepstrale, rapport signal à bruit et le désajustement) nous pouvons conclure que la structure de débruitage directe améliorée (AFBSS) donne de bon résultats.

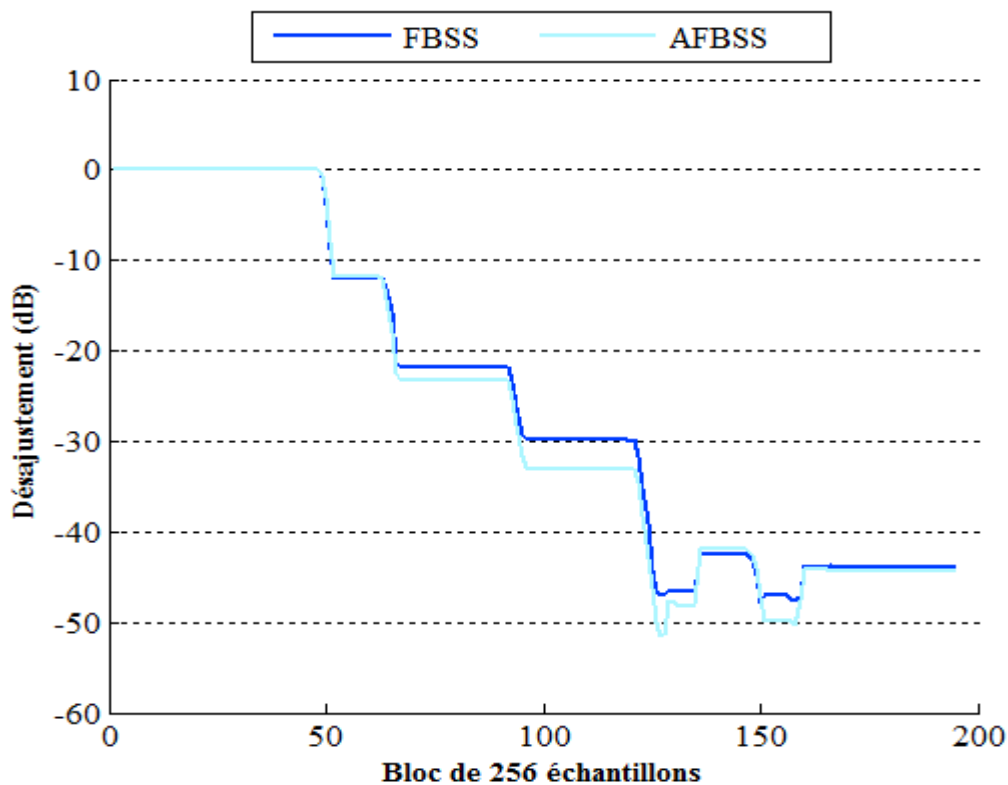


Figure 3.17 : Comparaison du désajustement obtenu avec les deux structures AFBSS (en cyan) et FBSS (en bleu).

### 3.3.3.4 Evaluation de la complexité

Le tableau 3.2 donne la complexité des structures de séparation de sources améliorée AFBSS et classique FBSS en termes du : nombre d'opérations (additions et multiplications) nécessaire au calcul de chaque structure et du nombre de paramètres seuils à fixer initialement.

Tableau 3.2 : Comparaison de la complexité pour les structures AFBSS et FBSS.

Structures	Nombre d'opérations			Nombre de paramètres seuils
	Additions	Multiplications	Total	
FBSS	$6L+8$	$6L+4$	$12L+12$	0
AFBSS	$12L+6$	$12L+8$	$24L+14$	6

D'après le tableau 3.2, nous pouvons voir que notre structures améliorée AFBSS possède une

complexité plus importante car elle nécessite environ 12 fois plus d'opérations et six valeurs de paramètres seuils plus que la structure classique décrite au chapitre précédent.

### 3.4 Conclusion

Dans ce chapitre, nous avons présenté une technique automatique de segmentation du signal de parole utilisée pour les deux structures de séparation de sources directe et récursive. Cette technique est basée sur l'utilisation des pas d'adaptation variables en fonction du rapport signal à bruit afin de contrôler l'adaptation des filtres  $w_{21}(n)$  et  $w_{12}(n)$  automatiquement.

Nous avons présenté les résultats de simulations obtenus à partir de la structure de séparation de sources directe améliorée AFBSS, ces résultats montrent bien que le débruitage du signal de parole avec cette structure a donné de bons résultats lorsque le signal de parole est segmenté automatiquement. Nous avons ainsi montré que la vitesse de convergence de l'algorithme NLMS appliqué à cette structure est directement proportionnelle au pas d'adaptation et inversement proportionnelle à la taille du filtre et le rapport signal à bruit sur la voie de référence.

Nous avons montré aussi le comportement presque similaire de cette structure AFBSS en comparaison avec une structure de la littérature FBSS. Malgré le bon comportement des structures améliorées, il convient de noter une complexité algorithmique assez importante de ces méthodes par rapport aux structures classiques. D'autre part, elles nécessitent aussi plusieurs paramètres ajustés expérimentalement. Ces limitations nous a mené à réfléchir sur des nouvelles idées concernant le contrôle automatique de la SAS qui soient plus performantes et moins complexes. En ce sens, le chapitre qui suit va se concentrer sur l'obtention des nouvelles solutions pour le contrôle automatique des deux structures de séparation de source résolvant le problème de la complexité.

## CHAPITRE 4

### SOLUTIONS STRUCTURELLES ET ALGORITHMIQUES POUR LA SEPARATION AVEUGLE DE SOURCES

#### 4.1 Introduction

L'utilisation d'une DAVM pour la mise à jour aux instants adéquats de chacun des filtres de séparation est une solution inévitable pour permettre un fonctionnement efficace de la structure de séparation de sources (directe et récursive), en particulier dans le traitement de la parole. Ce type de DAV donne une segmentation parfaite ce qui n'est pas le cas en pratique, car on n'a pas d'information a priori sur les signaux d'entrée (sources). Les méthodes exposées précédemment sont donc un moyen de résoudre ce problème mais souffrent, d'une part de la complexité qu'elles introduisent et d'autre part d'un libre choix difficile sur la valeur des seuils. Or, ceux-ci ne peuvent être choisis au hasard. Cependant nous allons présenter dans ce chapitre deux grandes catégories de solutions : structurelle et algorithmique afin de résoudre le problème de complexité. Seuls les résultats de simulation des versions de la structure de SAS de type directe seront présentés dans ce chapitre.

#### 4.2 Solutions structurelles proposées

Pour réduire la complexité de nos méthodes améliorées [39] [42], nous nous sommes concentrés sur la simplification de leur bloc de calcul des pas d'adaptation (bloc de DAVA). Nous avons vu au chapitre précédent que ces pas ( $\mu_{21}(n)$  et  $\mu_{12}(n)$ ) sont complémentaires (c-à-d ils fonctionnent d'une manière inverse), pour cela nous allons éliminer l'étape de calcul du 2<sup>ème</sup> pas et nous allons nous appuyer uniquement sur le comportement du 1<sup>er</sup> pas qui sera basé sur l'énergie estimée  $E_e$  afin de contrôler l'adaptation des deux filtres de séparation automatiquement. L'utilisation de ce nouveau bloc simplifié avec les deux structures de séparation de sources (directe et récursive) représente les deux solutions structurelles que nous allons décrire ci-après. Le principe général des solutions structurelles est donné par la figure 4.1.

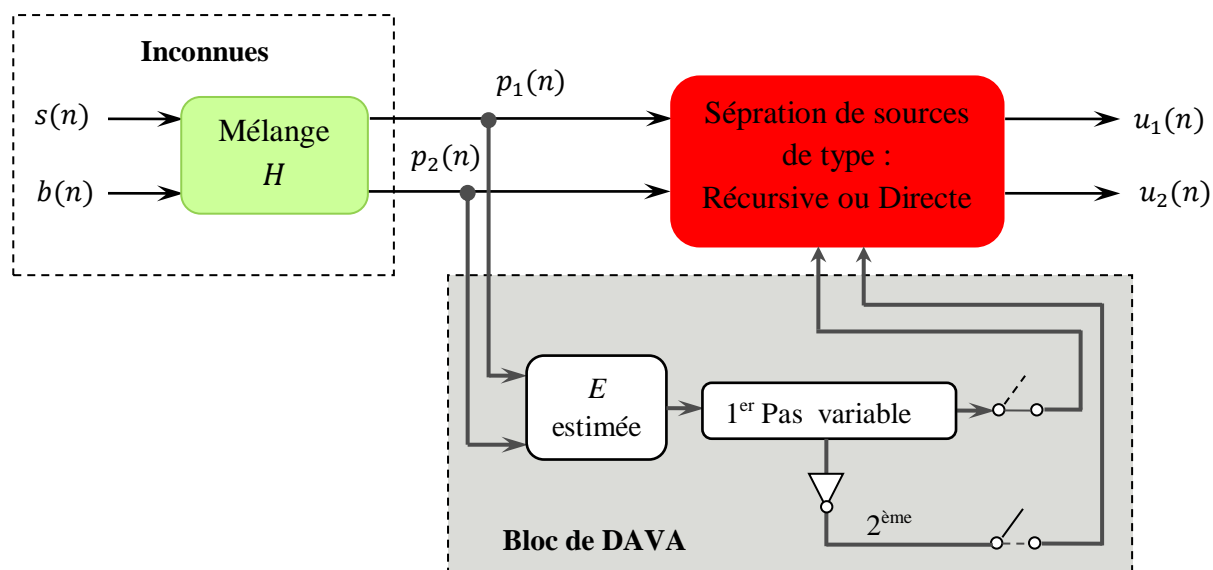


Figure 4.1 : Schéma bloc du principe des solutions structurales.

#### 4.2.1 Solution 1 : Structure automatique de séparation de sources directe à complexité réduite (LC-AFBSS)

La technique LC-AFBSS [43], pour Low Complexity Automatic Forward Blind Source Separation, est une solution que nous proposons pour la structure de séparation de source directe afin de garantir l'adaptation automatique de ces deux filtres de séparation  $w_{12}(n)$  et  $w_{21}(n)$  aux instants opportuns avec moins de complexité. L'idée principale consiste à stopper l'adaptation du filtre  $w_{21}(n)$  lorsque le signal du locuteur est présent et l'autoriser en pause silencieuse. Contrairement pour le filtre  $w_{12}(n)$ , l'adaptation est faite dans les zones où le signal de parole est présent et le filtre sera gelé dans le cas échéant. Pour conserver ce comportement, nous proposons donc de gérer les pas d'adaptations de ces deux filtres adaptatifs à partir d'un nouveau bloc de contrôle de manière inverse puisque pour l'un il s'agit d'apprendre le bruit alors que pour l'autre c'est le signal de parole. Le schéma général de cette nouvelle structure LC-AFBSS est présenté à la figure 4.2. Dans le système de séparation de la structure proposée LC-AFBSS, on identifie deux étages de traitement, principal et secondaire.

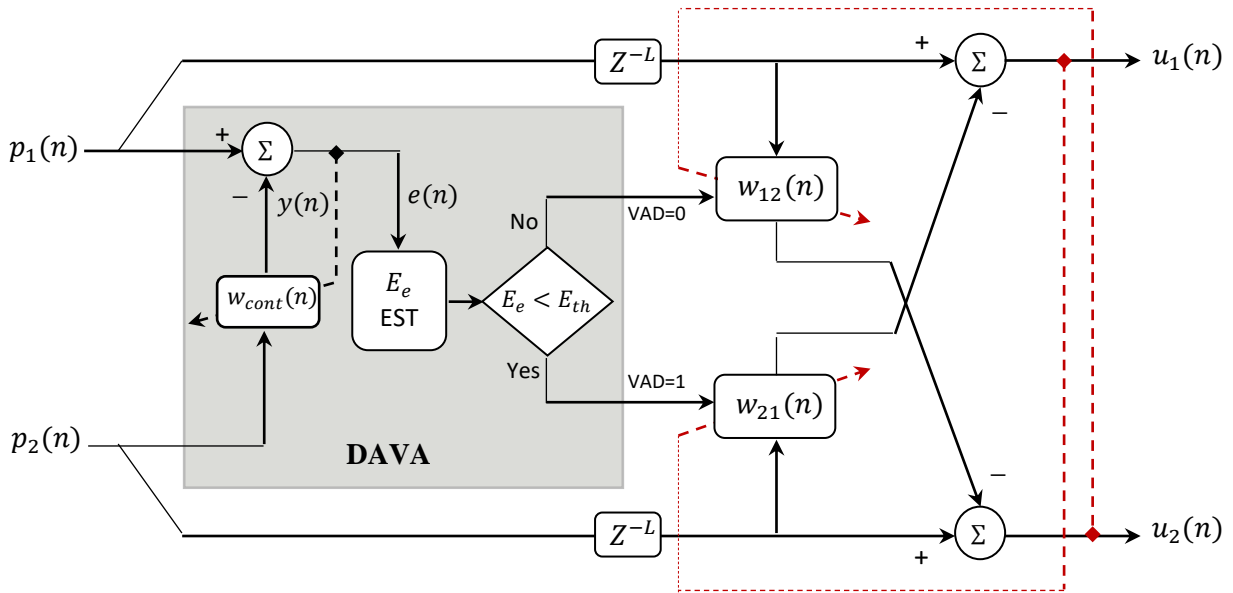


Figure 4.2 : La première solution structurelle proposée LC-AFBSS.

#### 4.2.1.1 Etage 1 : SAS

Le premier étage représente une structure de séparation de type Forward à deux voies de filtrage adaptatif ( $w_{12}(n)$  et  $w_{21}(n)$ ), ces filtres sont appliqués aux deux signaux d'observations ( $p_1(n)$  et  $p_2(n)$ ), afin de délivrer deux signaux filtrés ( $u_1(n)$  et  $u_2(n)$ ), dont l'un  $u_1(n)$ , est utilisé pour restaurer le signal de parole estimé. Les deux sorties  $u_1(n)$  et  $u_2(n)$  du premier étage s'écrivent :

$$u_1(n) = p_1(n - L) - \sum_{k=0}^{L-1} w_{21}(k) p_2(n - L - k) \quad (4.1)$$

$$u_2(n) = p_2(n - L) - \sum_{k=0}^{L-1} w_{12}(k) p_1(n - L - k) \quad (4.2)$$

où les signaux d'observations  $p_1(n)$  et  $p_2(n)$  ont été donnés à la section 2.2.1.2 du chapitre 2. Nous avons utilisé l'algorithme NLMS pour l'adaptation des filtres principaux  $w_{12}(n)$  et  $w_{21}(n)$  de la structure proposée LC-AFBSS. Les formules de mise à jour de ces filtres adaptatifs sont données par les équations suivantes:

$$\mathbf{w}_{21}(n) = \mathbf{w}_{21}(n - 1) + 2 \mu_{21} \frac{u_1(n) \mathbf{p}_2(n-L)}{\mathbf{p}_2^T(n-L) \mathbf{p}_2(n-L) + \xi_1} \quad (4.3)$$

$$\mathbf{w}_{12}(n) = \mathbf{w}_{12}(n - 1) + 2 \mu_{12} \frac{u_2(n) \mathbf{p}_1(n-L)}{\mathbf{p}_1^T(n-L) \mathbf{p}_1(n-L) + \xi_2} \quad (4.4)$$

où les scalaires  $\xi_1$  et  $\xi_2$  sont introduits dans les formules des filtres  $w_{21}(n)$  et  $w_{12}(n)$  respectivement, pour éviter la division par zéros lorsque le signal de parole est absent. Les équations (4.3) et (4.4) peuvent être réécrites sous la forme suivante :

$$\mathbf{w}_{21}(n) = \mathbf{w}_{21}(n-1) + 2 \mu_{21} \frac{u_1(n) \mathbf{p}_2(n-L)}{E_{p_2}(n) + \xi_1} \quad (4.5)$$

$$\mathbf{w}_{12}(n) = \mathbf{w}_{12}(n-1) + 2 \mu_{12} \frac{u_2(n) \mathbf{p}_1(n-L)}{E_{p_1}(n) + \xi_2} \quad (4.6)$$

les termes  $E_{p_1}(n)$  et  $E_{p_2}(n)$  représentent les énergies des signaux de mélanges  $p_1(n)$  et  $p_2(n)$  respectivement. Afin de réduire la complexité des équations de mise à jour des deux filtres adaptatifs  $w_{21}(n)$  et  $w_{12}(n)$ , les scalaires  $E_{p_1}(n)$  et  $E_{p_2}(n)$  sont estimés de manière récursive selon les formules suivantes:

$$E_{p_1}(n) = \theta_1 E_{p_1}(n-1) + (1 - \theta_1) |p_1(n-L)|^2 \quad (4.7)$$

$$E_{p_2}(n) = \theta_2 E_{p_2}(n-1) + (1 - \theta_2) |p_2(n-L)|^2 \quad (4.8)$$

où  $\theta_1$  et  $\theta_2$  sont des facteurs d'oubli qui permettent d'oublier le passé lointains des valeurs de  $p_1(n-L)$  et  $p_2(n-L)$ , respectivement. Ils sont généralement sélectionnés par  $\theta_1 = \theta_2 = 1 - \frac{1}{L}$  où  $L$  est la longueur des filtres adaptatifs  $w_{21}(n)$  et  $w_{12}(n)$ .

#### 4.2.1.2 Étage 2 : Bloc de contrôle automatique de la DAV (DAVA)

Le second étage de traitement représente une nouvelle forme du bloc qui réalise l'opération de DAVA, dont la sortie de ce bloc commande l'adaptation des coefficients des filtres adaptatifs du premier étage, automatiquement. Ce system de DAVA est basé sur l'estimation d'énergie calculée sur les deux voies d'observations dont le but est de piloter un système de décision permettant de modifier, via une règle, la valeur optimale des pas d'adaptations selon le contexte souhaité. Le nouveau bloc de DAVA proposé repose sur la mise en œuvre des fonctionnalités suivantes :

##### ❖ Filtre secondaire $w_{cont}(n)$

C'est un filtre qui permet de réaliser une opération de filtrage adaptatif dans le but est d'identifier le canal acoustique  $h_{21}(n)$  et de bien suivre ses variations dans le temps afin de fournir une estimée du bruit ( $y(n)$ ). Ce filtre est calculé, à chaque itération, à partir de l'estimé du signal de parole  $e(n)$  donné par :

$$e(n) = p_1(n) - \sum_{k=0}^{L-1} w_{cont}(k) p_2(n-k) \quad (4.9)$$

En ce qui concerne l'adaptation du filtre adaptatif de contrôle  $w_{cont}(n)$ , nous avons utilisé l'algorithme NLMS, la formule de mise à jour de ce filtre est donnée par :

$$\mathbf{w}_{cont}(n) = \mathbf{w}_{cont}(n-1) + 2 \mu_{cont} \frac{e(n) \mathbf{p}_2(n)}{E_{p_2}(n) + \xi_3} \quad (4.10)$$

où  $\xi_3$  est une faible constante permettant de ne pas diverger lorsque  $\mathbf{p}_2^T(n) \mathbf{p}_2(n) \sim 0$  ( $E_{p_2}(n) \sim 0$ ) et le paramètre  $\mu_{cont}$  est un pas d'adaptation, choisi entre 0 et 2, il contrôle la vitesse de convergence du filtre  $w_{cont}(n)$ .

#### ❖ Estimation d'énergie $E_e(n)$

Une fois l'estimé du signal de parole  $e(n)$  est obtenu, le calcul de son énergie est basé sur l'équation réursive suivante :

$$E_e(n) = \theta_3 E_e(n-1) + (1 - \theta_3) |e(n)|^2 \quad (4.11)$$

où  $\theta_3$  est un facteur d'oubli, il est sélectionné de la même façon comme  $\theta_1$  et  $\theta_2$  (voir équations (4.7) et (4.8), respectivement).

#### ❖ Prise de décision

Dans cette étape, nous définissons la règle de décision qui indique la situation de la DAV afin de piloter les instants d'adaptation des filtres du premier étage de la structure SAS. Cette décision est prise en fonction de l'énergie estimée  $E_e$  comparée à un seuil  $E_{th}$ . A partir de cette comparaison, nous décidons si la sortie du bloc de DAV est égale à 0 ou 1, indiquant respectivement, pour l'étage suivant, la présence ou l'absence du signal de parole. Le résultat de détection est ensuite utilisé pour gérer les pas d'adaptations  $\mu_{12}$  et  $\mu_{21}$  des filtres adaptatifs  $w_{12}(n)$  et  $w_{21}(n)$  respectivement. Les valeurs de ces pas sont modifiées en procédant comme suit :

- Si l'énergie estimée est inférieure au seuil  $E_{th}$  (zones de silence), alors la valeur du pas  $\mu_{21}$  associée au filtre  $w_{21}(n)$  est différente de sa valeur minimale et nous affectons le pas  $\mu_{12}$  associé au filtre  $w_{12}(n)$  à zéro. Ainsi, en garantissant que les filtres ne s'adaptent pas conjointement mais intermittemment cela veut dire que lorsque la décision d'adapter le filtre  $w_{21}(n)$  est prise alors l'adaptation de  $w_{12}(n)$  est gelée et vis versa. Ce principe de la prise de décision suit la relation suivante :



$$\begin{cases} \text{si } E_e < E_{th}, \text{ DAV} = 1 \rightarrow \begin{cases} \mu_{21} \neq 0, w_{21}(n) \text{ est adapté} \\ \mu_{12} = 0, w_{12}(n) \text{ n'est pas adapté} \end{cases} \\ \text{sinon, DAV} = 0 \rightarrow \begin{cases} \mu_{21} = 0, w_{21}(n) \text{ n'est pas adapté} \\ \mu_{12} \neq 0, w_{12}(n) \text{ est adapté} \end{cases} \end{cases} \quad (4.12)$$

#### 4.2.2 Solution 2 : Structure automatique de séparation de sources récursive à complexité réduite (LC-ABBSS)

Dans ce paragraphe, nous proposons une nouvelle structure automatique de séparation de source récursive à complexité réduite (LC-ABBSS) [44], pour Low Complexity Automatic Backward Blind Source Séparation. Cette nouvelle solution emploie également le même bloc de contrôle automatique que celle initialement proposée (LC-AFBSS), afin de contrôler la structure de l'étage principale qui correspond à une structure de SAS de type backward (voir figure 4.3). Les équations du premier étage de cette solution se résument dans le tableau 4.1. Pour le fonctionnement du deuxième étage, Un diagramme d'activités de la gestion des pas d'adaptations ainsi qu'une description algorithmique sont représentés à la figure (4.4) et à l'Algorithme 3 respectivement. Les résultats de simulations de cette solution ne sont pas présentés dans ce chapitre.

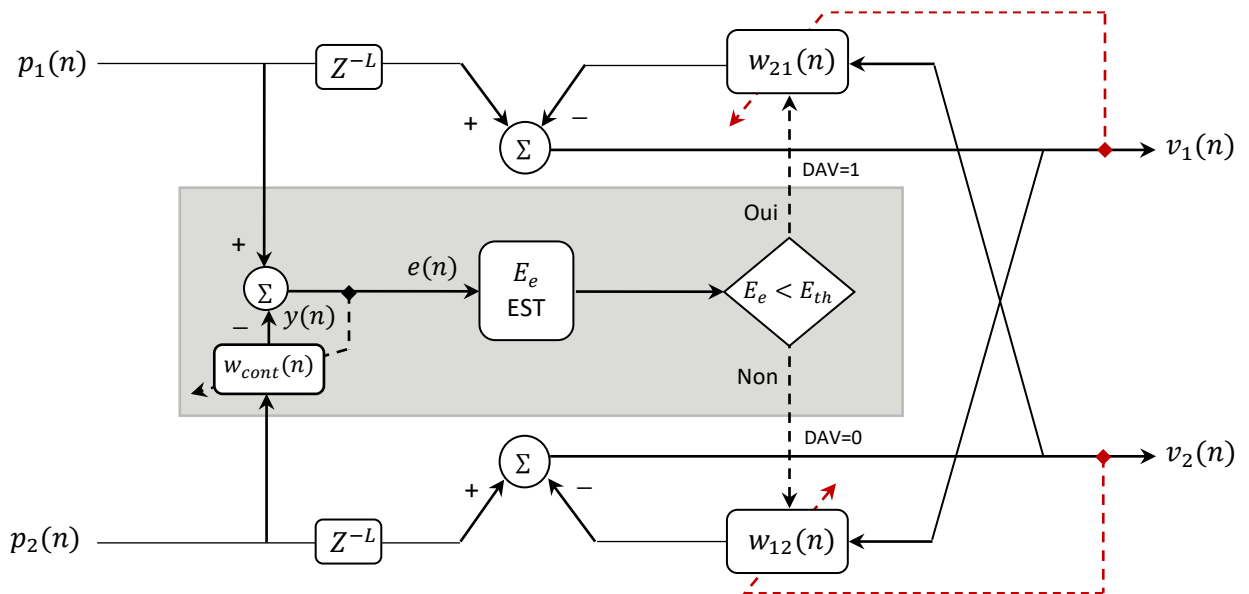
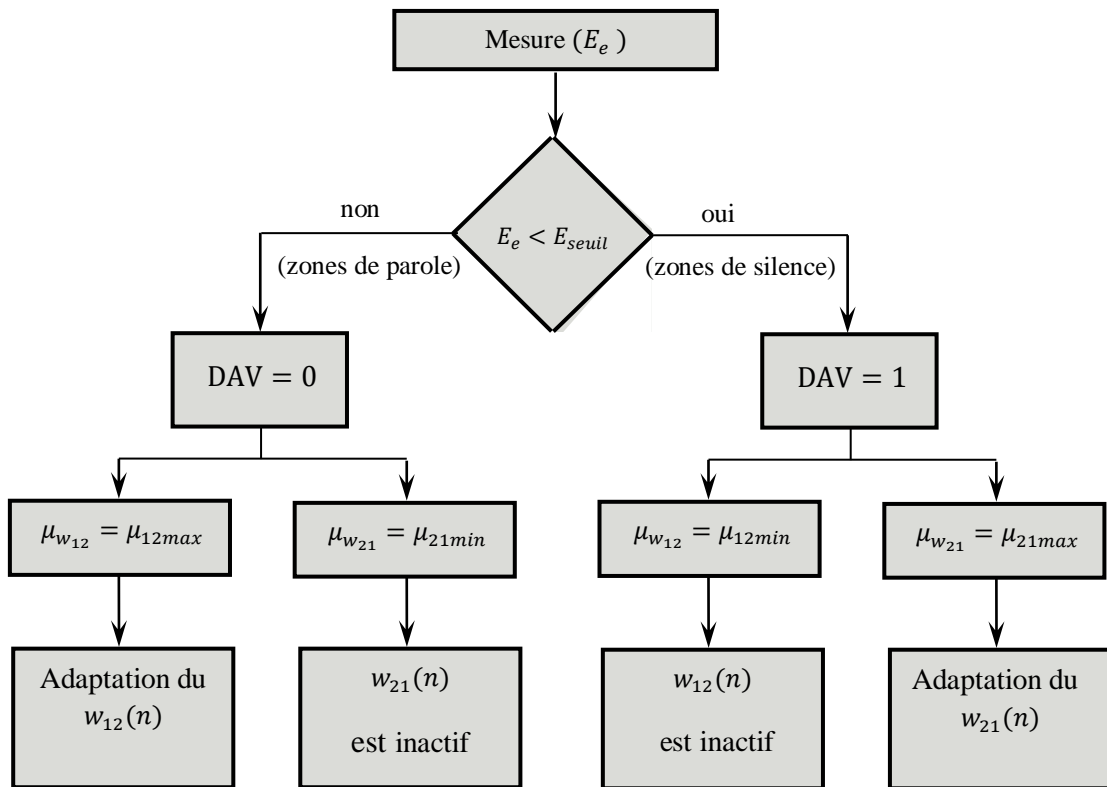


Figure 4.3: La deuxième solution structurelle proposée LC-ABBSS.

Tableau 4.1 : Les équations du premier étage de la solution structurelle LC-ABBSS.

<b>1<sup>er</sup> Etage : Structure de SAS récurrente</b>	
Signaux d'entrée :	Signaux de sortie :
$p_1(n) = s(n) + h_{21}(n) * b(n)$	$v_1(n) = p_1(n - L) - \sum_{k=0}^{L-1} w_{21}(k) v_2(n - L - k)$
$p_2(n) = b(n) + h_{12}(n) * s(n)$	$v_2(n) = p_2(n - L) - \sum_{k=0}^{L-1} w_{12}(k) v_1(n - L - k)$
Adaptation des filtres :	
$w_{21}(n) = w_{21}(n - 1) + 2 \mu_{21} \frac{v_1(n) k_2(n-L)}{E_{k_2}(n) + \xi_1}; E_{k_2}(n) = \theta_1 E_{k_2}(n - 1) + (1 - \theta_2)  k_2(n - L) ^2$	
$w_{12}(n) = w_{12}(n - 1) + 2 \mu_{12} \frac{v_2(n) k_1(n-L)}{E_{k_1}(n) + \xi_2}; E_{k_1}(n) = \theta_2 E_{k_1}(n - 1) + (1 - \theta_2)  k_1(n - L) ^2$	
$\theta_1 = \theta_2 = 1 - \frac{1}{L}$	

Figure 4.4 : Diagramme d'activités de la gestion des pas d'adaptations du 2<sup>ème</sup> étage de la structure LC-ABBSS.

---

**Algorithme 3** Algorithme de la gestion des pas d'adaptations du 2<sup>ème</sup> étage de la structure LC-ABBSS.

---

Début

1 : **si**  $E_e < E_{th}$  **alors**

2 : **DAV** = 1 // absence du signal de parole (zones de silence)

$\mu_{w_{21}} \leftarrow \mu_{w_{21_{max}}}$  //  $w_{21}(n)$  est actif

$\mu_{w_{12}} \leftarrow \mu_{w_{12_{min}}}$  //  $w_{12}(n)$  est inactif

3 : **sinon** // **DAV** = 0 présence du signal de parole

$\mu_{w_{12}} \leftarrow \mu_{w_{12_{max}}}$  //  $w_{12}(n)$  est actif

$\mu_{w_{21}} \leftarrow \mu_{w_{21_{min}}}$  //  $w_{21}(n)$  est inactif

4 : **fin si**

Fin

---

Pour l'évaluation des solutions structurelle et algorithmique, nous reprenons les mêmes démarches de simulation que celles présentées dans le chapitre précédent.

#### 4.3 Evaluation de la première solution structurelle LC-AFBSS

Dans ce paragraphe, nous étudions la première solution structurelle proposée LC-AFBSS qui correspond à une structure de séparation de sources directe intégrant une nouvelle forme simplifiée du bloc qui réalise la DAV automatique, dont la description a été donnée à la section 4.2.1 (voir figure 4.2). Les signaux et les réponses impulsionnelles utilisés pour générer les signaux de mélange correspondent à ceux présentés au chapitre 2. Pour mettre à jour les coefficients des filtres adaptatifs  $w_{21}(n)$  et  $w_{12}(n)$  nous utilisons l'algorithme NLMS. Les valeurs des paramètres utilisées pour cette solution structurelle sont résumées dans le tableau 4.2. Le signal de parole originale, les signaux de mélange ( $p_1(n), p_2(n)$ ) et les signaux de sortie  $u_1(n), u_2(n)$  de la première solution structurelle LC-AFBSS sont représentés avec leurs spectrogrammes sur la figure (4.5). Nous pouvons également observer à partir de cette figure, que le signal disponible sur la première sortie de traitement  $u_1(n)$  est presque complètement débruité et très proche du signal de parole originale. Nous notons que ce résultat est obtenu grâce à l'utilisation d'un bloc de DAVA comme indiqué auparavant. Contrairement pour le signal de la deuxième sortie  $u_2(n)$  qui est partiellement nettoyé du signal de parole. Ceci est dû à la mauvaise identification du filtre adaptatif  $w_{12}(n)$  qui s'adapte durant les périodes de présence d'un mélange du signal de parole et du bruit et

ceci malgré l'utilisation d'un système de détection des périodes d'activité vocale (DAV).

Tableau 4.2 : Les valeurs optimales des paramètres utilisées dans la solution structurelle LC-AFBSS.

Structures	Paramètres
LC-AFBSS	<p><b>Longueur des filtres :</b>  <math>L_{w12} = L_{w21} = L_{wcont} = 64</math></p> <p><b>Pas d'adaptation du filtre de l'étage de contrôle :</b>  <math>\mu_{wcont} = 0.4 ;</math></p> <p><b>Valeur seuil de l'énergie estimée :</b>  <math>E_e = 0.5</math></p> <p><b>Pas d'adaptation des filtres du 1<sup>er</sup> étage:</b>  <math>\mu_{w12min} = 0 ; \mu_{w12max} = 0.2 ;</math>  <math>\mu_{w21min} = 0 ; \mu_{w21max} = 0.2</math></p>

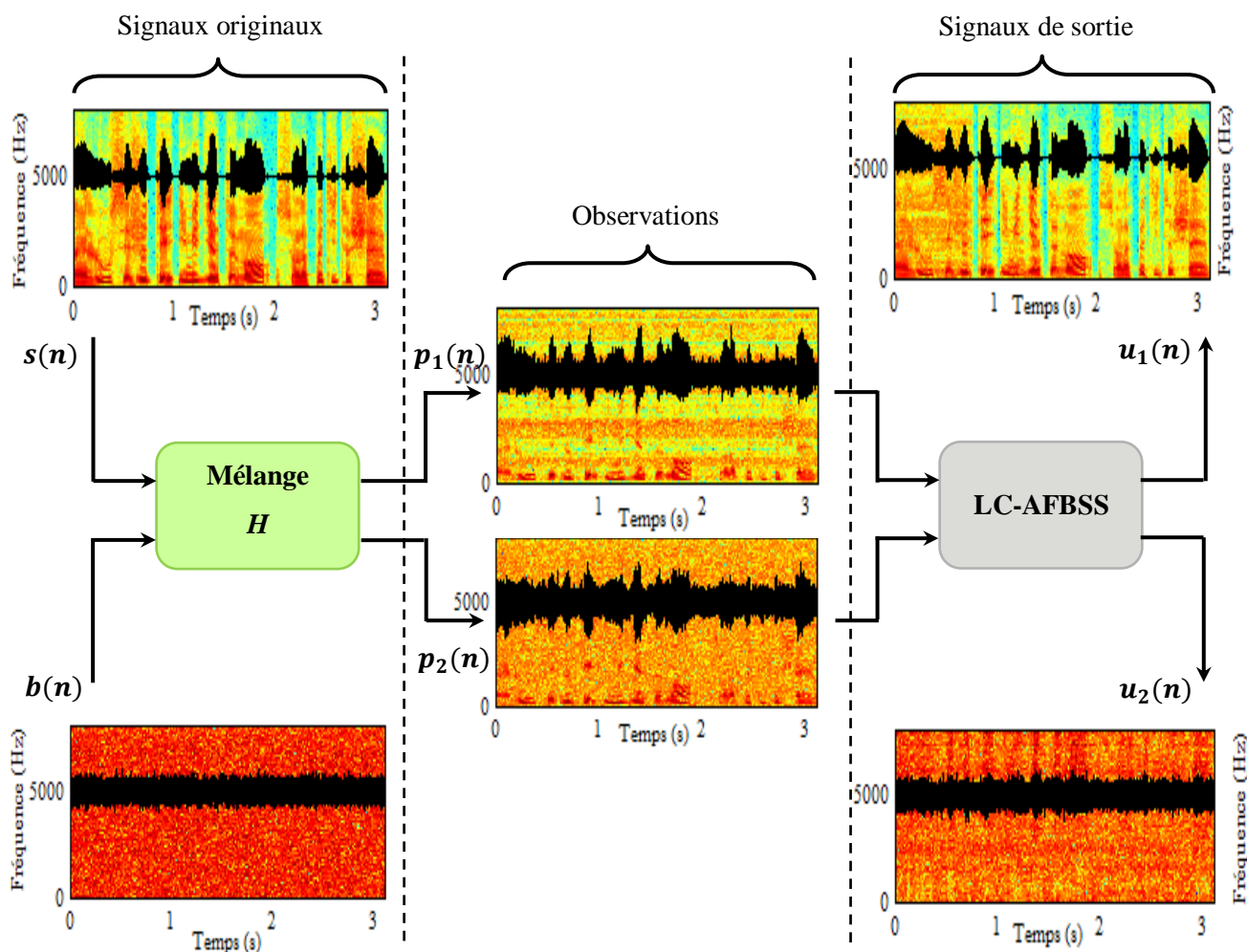


Figure 4.5: Signaux sources (parole originale  $s(n)$ , bruit Blanc  $b(n)$ ), signaux de mélange ( $p_1(n)$ ,  $p_2(n)$ ) et signaux de sortie ( $u_1(n)$ ,  $u_2(n)$ ) de la structure LC-AFBSS.

Nous notons qu'un bon fonctionnement de notre nouvelle structure LC-AFBSS est directement lié à une bonne détection des instants de non-activité vocale, cette dernière étant basée sur la bonne estimation de l'énergie  $E_e(n)$  du signal utile.

L'évolution de l'énergie estimée  $E_e$  au niveau de l'entrée primaire  $p_1(n)$  et le pas d'adaptation  $\mu_{w_{21}}(n)$  du filtre principal  $w_{21}(n)$  sont représentés avec le signal de parole originale sur la figure 4.6.

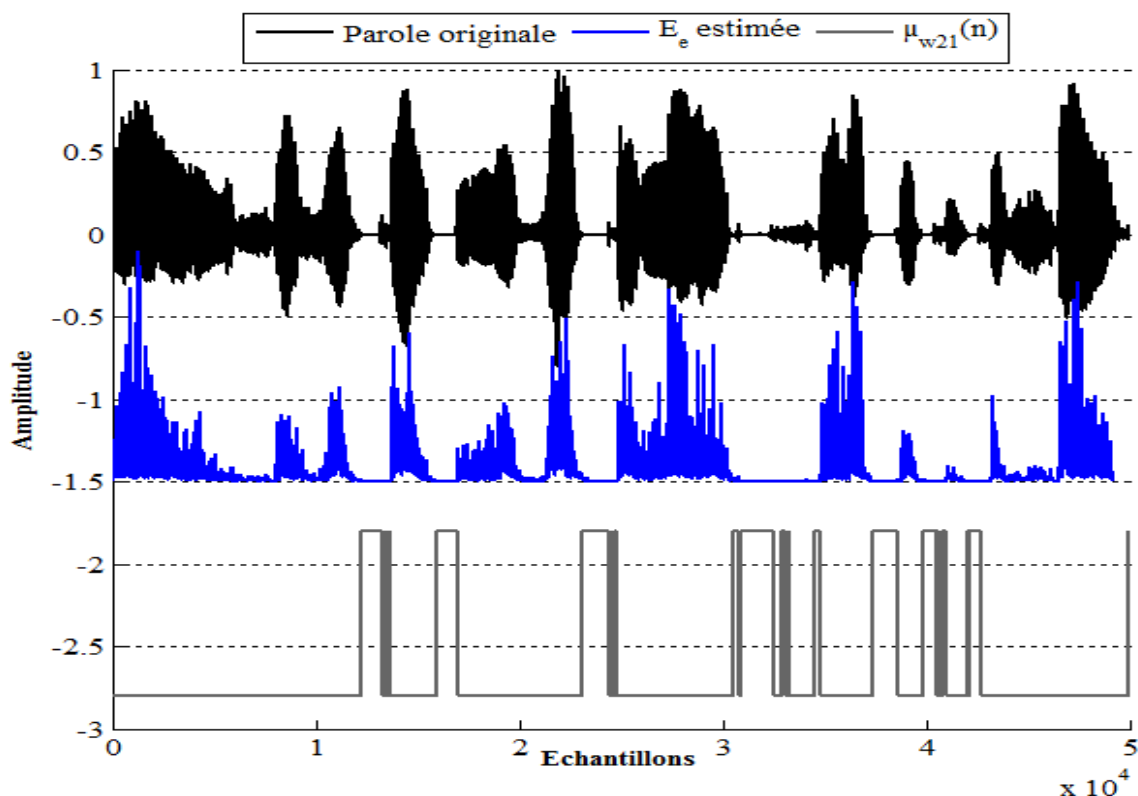


Figure 4.6 : Signal de parole originale (en haut), l'énergie estimée  $E_e$  au niveau de l'entrée primaire (LC-AFBSS) (au milieu) et le pas d'adaptation  $\mu_{w_{21}}(n)$  du filtre principal  $w_{21}(n)$  (en bas).

Selon la figure 4.6, nous remarquons que le pas d'adaptation  $\mu_{w_{21}}(n)$  du filtre principal  $w_{21}(n)$  est grand lorsque l'énergie estimée  $E_e$  est très faible (cela signifie que le pas d'adaptation  $\mu_{w_{21}}(n)$  prend des valeurs importantes lorsque le signal de parole est absent car les zones de faible énergie correspondent aux instants de silence de la parole).

#### 4.4 Comparaison entre les structures LC-AFBSS et AFBSS

Afin de valider l'efficacité de notre nouvelle solution structurelle (LC-AFBSS) proposée qui utilise un nouveau bloc de DAV automatique simplifié, nous avons comparé ces

performances à celles de la méthode améliorée (AFBSS) qui a été bien étudiée et discutée dans le chapitre 3. Rappelons que la méthode améliorée (AFBSS) utilise aussi une technique de segmentation automatique mais qui exige beaucoup de calcul. Les paramètres de simulations pour la solution structurelle LC-AFBSS et la structure améliorée AFBSS ont été représentés dans les tableaux (4.1) et (3.2) respectivement. Dans ces simulations, nous nous intéressons uniquement à la voie restituant le signal de parole utile (seule la première sortie de traitement  $u_1(n)$  est exploitée). Nous commençons cette étude par une comparaison entre les signaux de sortie de chaque structure, ces résultats sont représentés sur la figure 4.7.

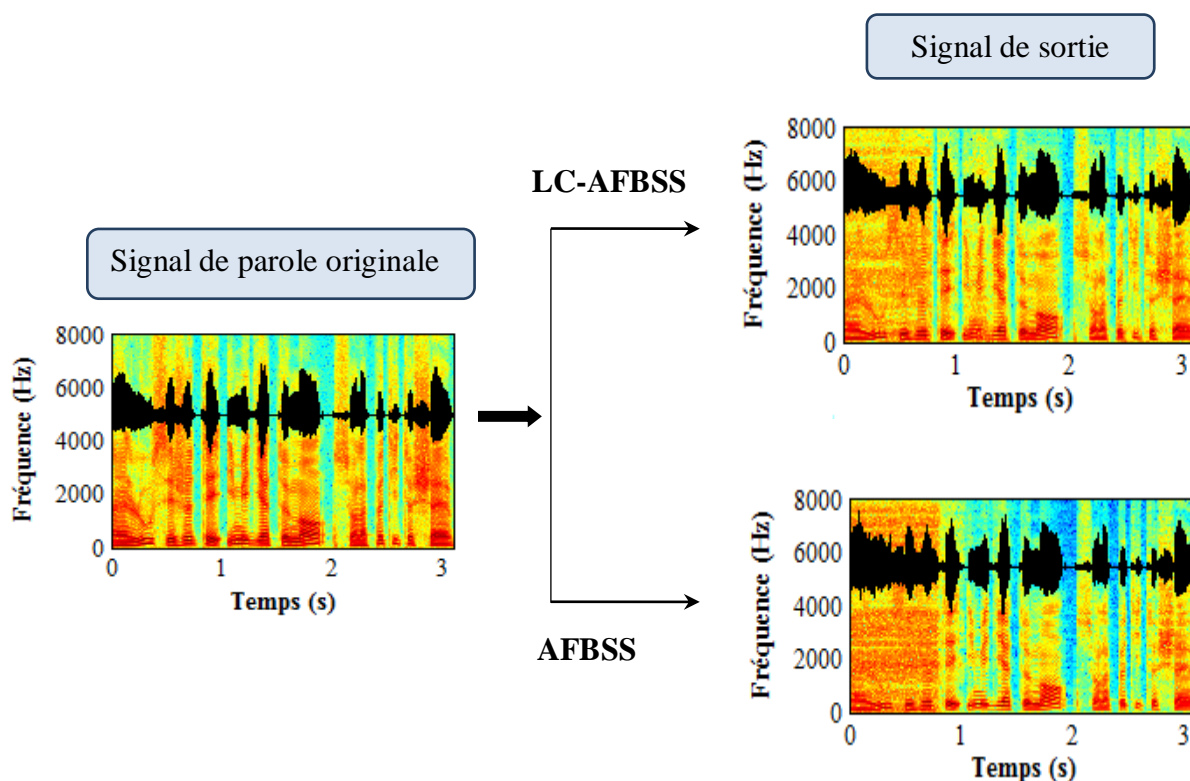


Figure 4.7 : Signal de parole originale  $s(n)$  (à gauche), signal de sortie  $u_1(n)$  de la structure LC-AFBSS (en haut), signal de sortie  $u_1(n)$  de la structure AFBSS (en bas).

Nous pouvons observer à partir de la figure (4.7) que les signaux disponibles en sortie de traitement à partir de ces deux structures (LC-AFBSS et AFBSS) sont visuellement nettoyé du bruit et sont proches du signal de parole originale. Pour confirmer ces observations visuelles, nous mesurons la distance cepstrale entre le signal de parole original et celui obtenu en sortie de traitement de chaque structure (LC-AFBSS et AFBSS). Nous observons que les deux structures ont le même comportement en termes de distance cepstrale (DC).

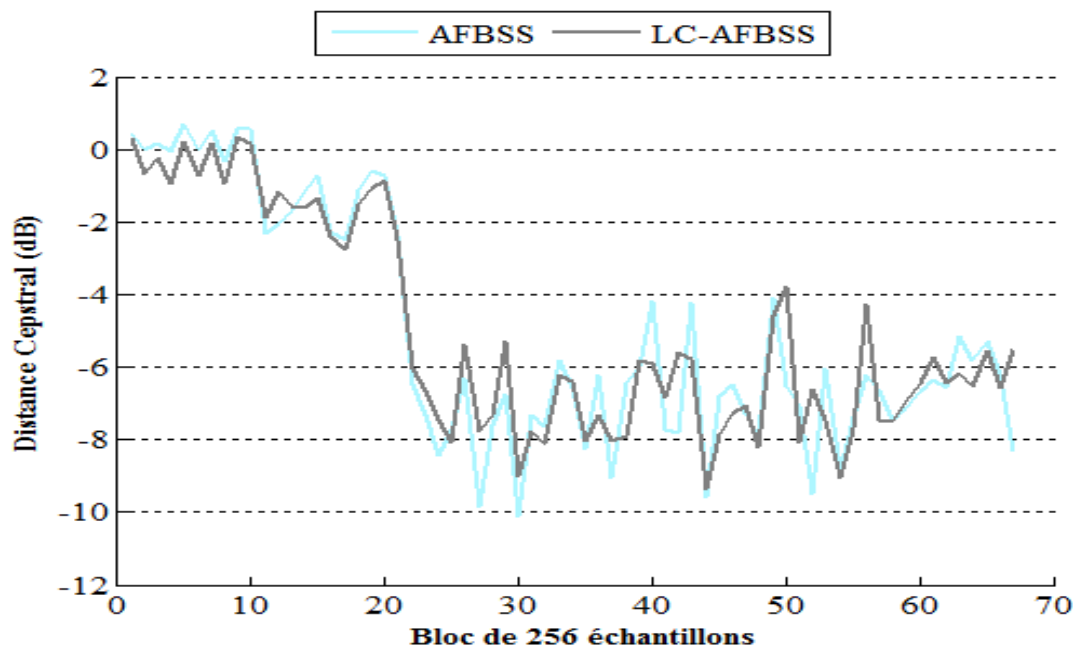


Figure 4.8 : Evolution du critère de la distance cepstrale des deux structures LC-AFBSS (en gris) et AFBSS (en cyan).

Sur la figure (4.9) suivante, nous montrons l'évolution du critère du rapport signal à bruit (RSB) et ceci pour les deux structures. Nous remarquons bien que le comportement de la solution structurelle proposée (LC-AFBSS) est presque similaire à celui de l'autre structure (AFBSS). Ceci confirme le bon comportement de notre première solution structurelle proposée (LC-AFBSS). Nous présentons sur la figure (4.10) l'évolution temporelle du désajustement mesuré sur les coefficients du filtre adaptatif  $w_{21}(n)$  pour les deux structures (LC-AFBSS et AFBSS).

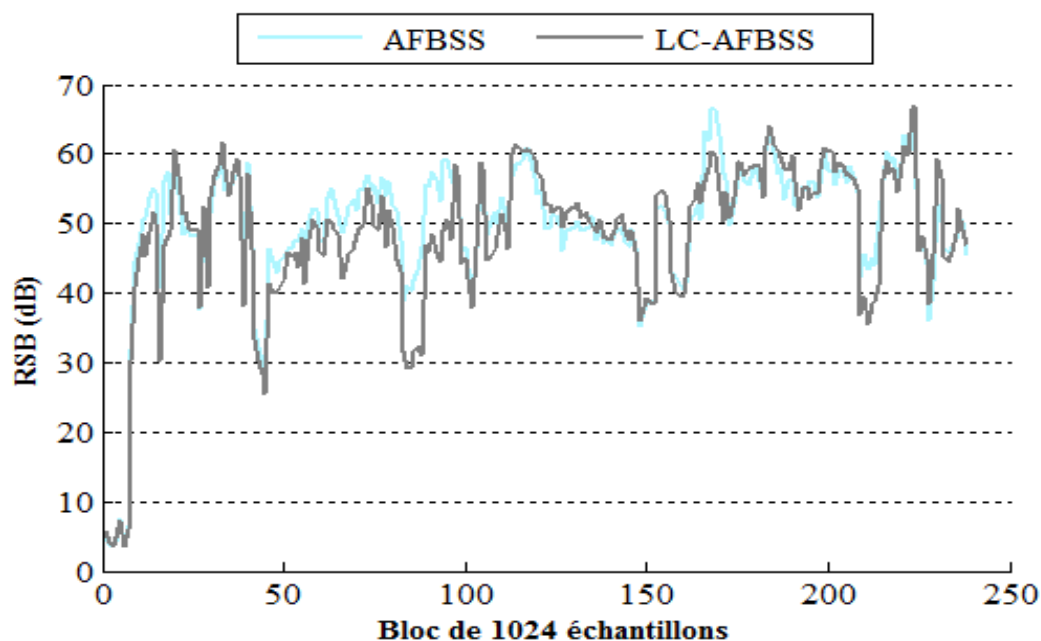


Figure 4.9 : Evolution du critère du rapport signal à bruit (RSB) pour les deux structures LC-AFBSS (en gris) et AFBSS (en cyan).

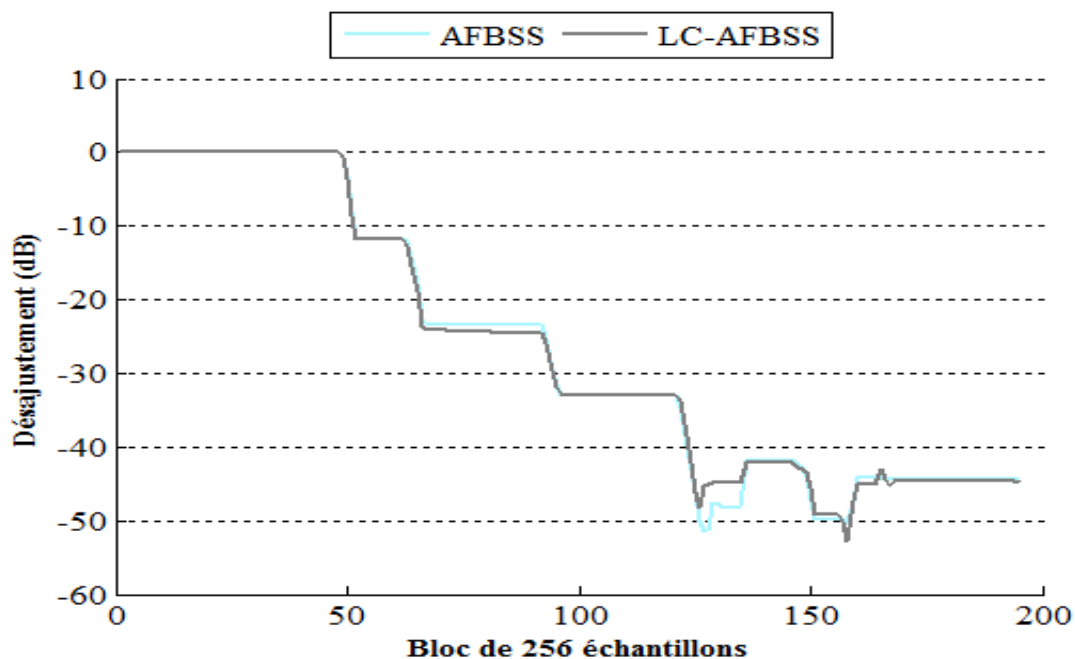


Figure 4.10 : Comparaison du désajustement obtenu avec les deux structures LC-AFBSS (en gris) et AFBSS (en cyan).

Cette figure montre bien la superposition presque totale des tracés de l'évolution temporelle du désajustement de chaque structure, ce qui démontre encore une fois les performances excellentes de la nouvelle solution structurelle LC-AFBSS.

Afin de compléter cette étude comparative, la complexité algorithmique et le nombre de paramètres seuils utilisé pour la solution structurelle LC-AFBSS sont données dans le tableau (4.3) suivant. En comparant avec les résultats précédents associés à la structure améliorée AFBSS (voir tableau 3.2), nous pouvons voir que la méthode étudiée LC-AFBSS fait l'économie environ de  $6L$  multiplications et  $6L$  additions et demande moins de paramètres seuils que la méthode AFBSS.

Tableau 4.3 : La complexité de la solution structurelle LC-AFBSS.

Structures	Nombre d'opérations			Nombre de paramètres seuils
	Additions	Multiplications	Total	
LC-AFBSS	$6L+17$	$6L+8$	$12L+25$	3



Les résultats obtenus avec la solution structurelle (LC-AFBSS) se sont avérés très proches de ceux obtenus avec la structure directe améliorée AFBSS. Le seul avantage de cette solution et sa faible complexité par rapport à la structure améliorée. Ces résultats sont très satisfaisants et confirment le bon comportement de la première solution structurelle proposée dans cette thèse.

#### 4.5 Solution algorithmique proposée

Les solutions structurelles [43-44] permettent de réduire la complexité en termes du coût de calcul par rapport aux méthodes améliorées, tout en exigeant la fixation empirique d'un certain paramètre seuil ce qui dégrade en quelque sorte le potentiel de ces solutions proposées. Cette limitation nous a conduits à proposer d'autres solutions performantes, résolvant cet inconvénient. Nous avons vu, pour les méthodes précédentes (méthodes améliorées et solutions structurelles), qu'il était important de commander l'adaptation des filtres de séparations à partir d'un module complémentaire de variation des pas qui garantit l'adaptation des filtres  $w_{21}(n)$  et  $w_{12}(n)$  pendant les périodes d'absence/présence du signal de parole respectivement. Nous allons conserver ce comportement dans cette partie pour les solutions algorithmiques proposées ci-après ; il nous faut donc une méthode de variation des pas fondée sur des approches algorithmiques pures qui soit à l'image du comportement souhaité. Dans ce qui suit, nous allons présenter une solution algorithmique qui permet la réduction du bruit acoustique d'une manière automatique, sans avoir besoin d'un bloc de DAV pour contrôler les filtres adaptatifs de séparation aveugles des signaux.

##### 4.5.1 Solution 3 : Algorithme régularisé pour la FBSS (RFBSS)

Nous proposons dans cette section, un algorithme nommé RFBSS pour Regularized Forward Blind Source Separation [45] combiné avec la FBSS de la figure 2.8 et dédié pour la réduction du bruit acoustique et le rehaussement du signal de parole. Ce nouvel algorithme robuste est automatique et basé sur l'utilisation d'un pas variable assurant en même temps les tâches suivantes :

- La mise à jour automatique des coefficients des filtres de séparation ( $w_{12}(n)$ ,  $w_{21}(n)$ ).
- Le contrôle automatique d'adaptation de ces filtres adaptatifs se fait selon la présence ou l'absence des périodes d'activités du signal de parole.

Dans cette section, nous donnons la formulation exacte du nouvel algorithme RFBSS proposé. Nous introduisons d'abord la méthode de Newton que nous utilisons pour dériver la formulation de cet algorithme proposé.

#### 4.5.1.1 Méthode de Newton

Considérons une fonction  $f(x)$  continue sur  $\mathcal{R}^n$ , et on cherche une solution  $x$  de cette fonction tel que  $f(x) = 0$ . Supposons que la matrice Jacobienne  $Df(x) \neq 0$  est non singulière au voisinage de  $x$ . En commençant par une valeur initiale  $x_0$  qui est assez proche de  $x$  en extrapolant la tangente en  $x_0$  jusqu'à son intersection avec l'axe  $OX$ , on continue jusqu'à obtenir une valeur de  $x$  tel que :  $f(x) = 0$ . La figure suivante nous donne une description graphique de la méthode de Newton [46].

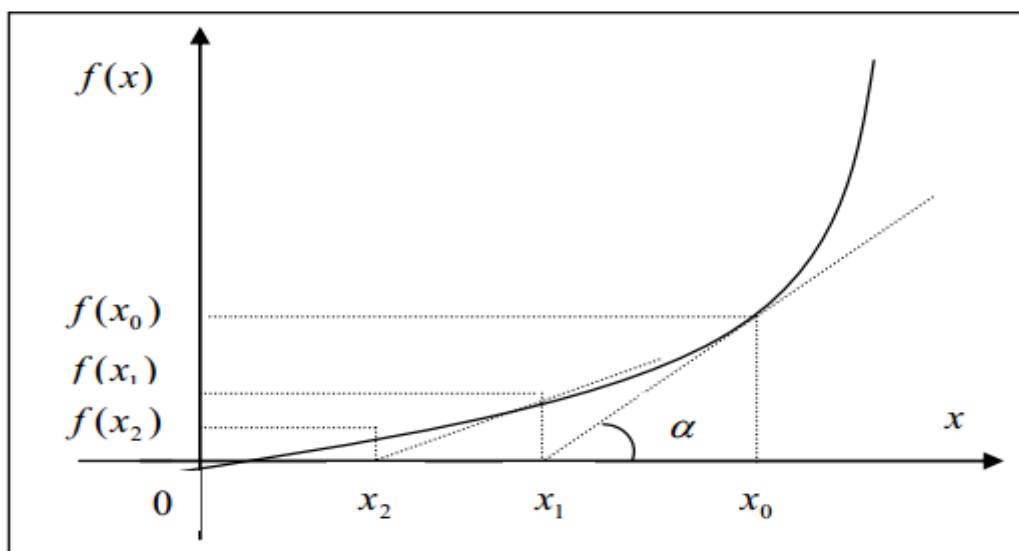


Figure 4.11 : Représentation graphique de la méthode de Newton.

D'après cette figure, nous écrivons :

$$\operatorname{tg}(\alpha) = f'(x_0) = \frac{f(x_0)}{x_0 - x_1} \quad (4.13)$$

A partir de cette dernière équation, nous pouvons écrire les termes suivants :

$$x_1 = x_0 - \frac{f(x_0)}{f'(x_0)}, \quad x_2 = x_1 - \frac{f(x_1)}{f'(x_1)}, \quad x_3 = x_2 - \frac{f(x_2)}{f'(x_2)}, \quad \dots \quad x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)} \quad (4.14)$$

D'où la formule générale s'écrira :

$$x(t) = x(t-1) - \frac{f[x(t-1)]}{f'[x(t-1)]} \quad (4.15)$$

Dans le cas discret, nous écrivons l'équation (4.15) comme suit :

$$x(n+1) = x(n) - Df[x(n)]^{-1}f[x(n)] \quad (4.16)$$

Cependant, cette méthode de Newton produit une séquence  $\{x(n)\}_{n=0}^{\infty}$  qui converge de façon quadratique vers la solution  $x$  [46].

#### 4.5.1.2 Dérivation de l'algorithme RFBSS proposé

Nous présentons dans cette partie, la dérivation de la formulation complète du nouvel algorithme proposé RFBSS. La construction de ce dernier se base principalement sur le critère de minimisation de l'erreur quadratique moyenne défini par :

$$J = E[e(n)^2] = E[(d(n) - \mathbf{w}^T(n)\mathbf{x}(n))^2] \quad (4.17)$$

où  $d(n)$  est le signal primaire;  $\mathbf{x}(n)$  le vecteur contenant les  $N+1$  dernières valeurs du signal de référence,  $\mathbf{w}(n)$  le vecteur des coefficients du filtre (voir figure 1.3) et " $T$ " représente l'opérateur de transposition.

Rappelons les signaux de sortie de la structure directe qui seront notés dorénavant signaux d'erreurs par leurs relations suivantes :

$$u_1(n) = p_1(n) - \mathbf{w}_{21}^T(n) \mathbf{p}_2(n) \quad (4.18)$$

$$u_2(n) = p_2(n) - \mathbf{w}_{12}^T(n) \mathbf{p}_1(n) \quad (4.19)$$

où

$$\mathbf{w}_{21}^T(n) = [w_{21}^0(n) w_{21}^1(n) w_{21}^2(n) \cdots w_{21}^N(n)] \quad (4.20)$$

$$\mathbf{w}_{12}^T(n) = [w_{12}^0(n) w_{12}^1(n) w_{12}^2(n) \cdots w_{12}^N(n)] \quad (4.21)$$

sont les coefficients variables des filtres adaptatifs  $\mathbf{w}_{21}(n)$  et  $\mathbf{w}_{12}(n)$  d'ordre  $N+1$ , et

$$\mathbf{p}_1^T(n) = [p_1(n) p_1(n-1) \cdots p_1(n-N)] \quad (4.22)$$

$$\mathbf{p}_2^T(n) = [p_2(n) p_2(n-1) \cdots p_2(n-N)] \quad (4.23)$$

Appliquons le critère de l'EQM minimale à la structure directe de la figure 2.7, nous obtenons :

$$J_1 = E[u_1(n)^2] = E[(p_1(n) - \mathbf{w}_{21}^T(n)\mathbf{p}_2(n))^2] \quad (4.24)$$

$$J_2 = E[u_2(n)^2] = E[(p_2(n) - \mathbf{w}_{12}^T(n)\mathbf{p}_1(n))^2] \quad (4.25)$$

où  $J_1$  et  $J_2$  représentent les EQMs sur les voies primaire et secondaire respectivement.

La structure directe est symétrique, pour cela nous nous intéresserons aux formules de la première branche qui permet d'obtenir le signal de parole estimé  $u_1(n)$ . Puis, nous allons adapter les différentes formules obtenues sur la deuxième branche de cette structure.

L'équation de la mise à jour du filtre adaptatif  $\mathbf{w}_{21}(n)$  est obtenue en minimisant l'EQM  $J_1$  selon l'approche des moindres carrés. Nous annulons donc le gradient de  $J_1$  par rapport à  $\mathbf{w}_{21}(n)$ , en utilisant la méthode de Newton, alors nous écrivons :

$$\text{Min } J_1 \Rightarrow \left. \frac{\partial J_1}{\partial \mathbf{w}_{21}} \right|_{\mathbf{w}_{21}=\mathbf{w}_{21}^{opt}} = 0 \quad (4.26)$$

On pose :

$$f(x) = \frac{\partial J_1}{\partial \mathbf{w}_{21}} = -2[\mathbf{P}_{p_1 p_2} - \mathbf{R}_{p_2 p_2} \mathbf{w}_{21}] \quad (4.27)$$

$$Df(x) = \frac{\partial^2 J_1}{\partial^2 \mathbf{w}_{21}} = 2 \mathbf{R}_{p_2 p_2} \quad (4.28)$$

En remplaçant les équations (4.27) et (4.28) dans l'équation (4.16), nous obtenons :

$$\mathbf{w}_{21}(n+1) = \mathbf{w}_{21}(n) + \mu_{21} [\mathbf{R}_{p_2}]^{-1} [\mathbf{P}_{p_1 p_2} - \mathbf{R}_{p_2} \mathbf{w}_{21}] \quad (4.29)$$

où

$\mu_{21}$  : est un pas d'adaptation.

$\mathbf{P}_{p_1 p_2}$  : représente le vecteur d'inter-corrélation entre les signaux de mélange bruités ( $p_1(n)$ ,  $p_2(n)$ ), donné par :

$$\mathbf{P}_{p_1 p_2} = E[p_1(n)p_2(n)] \quad (4.30)$$

$\mathbf{R}_{p_2}$  : est la matrice d'auto-corrélation du vecteur  $p_2(n)$ , exprimée par :

$$\mathbf{R}_{p_2} = E[p_2(n)p_2^T(n)] \quad (4.31)$$

Des modifications seront apportées à l'équation de mise à jour du filtre adaptatif  $\mathbf{w}_{21}(n)$  (4.29), l'ensemble de ces dernières (modifications) décrivent le nouvel algorithme nommé Regularized Forward Blind Source Separation (RFBSS). Ces modifications sont illustrées comme suit :

- ✚ En premier lieu, nous avons introduit le paramètre  $\varepsilon(n) \mathbf{I}$  dans l'équation (4.29) afin d'éviter la division par zéro, où  $\varepsilon > 0$  est un petit paramètre de régularisation et  $\mathbf{I}$  est une matrice d'identité  $M \times M$ . Nous obtenons alors l'équation (4.32) suivante :

$$\mathbf{w}_{21}(n+1) = \mathbf{w}_{21}(n) + \mu_{21} [\varepsilon(n) \mathbf{I} + \mathbf{R}_{p_2}]^{-1} [\mathbf{P}_{p_1 p_2} - \mathbf{R}_{p_2} \mathbf{w}_{21}] \quad (4.32)$$

✚ En second lieu, nous remplaçons  $\mathbf{P}_{p_1 p_2}$  et  $\mathbf{R}_{p_2}$  par leurs valeurs instantanées, car ce nouvel algorithme est une variante de l'algorithme LMS, donc:

$$\mathbf{P}_{p_1 p_2} = [p_1(n) \mathbf{p}_2(n)] \quad (4.33)$$

$$\mathbf{R}_{p_2} = [\mathbf{p}_2(n) \mathbf{p}_2^T(n)] \quad (4.34)$$

Cela conduit à l'expression suivante :

$$\mathbf{w}_{21}(n+1) = \mathbf{w}_{21}(n) + \mu_{21} [\varepsilon(n) \mathbf{I} + \mathbf{p}_2(n) \mathbf{p}_2^T(n)]^{-1} [p_1(n) \mathbf{p}_2(n) - \mathbf{p}_2(n) \mathbf{p}_2^T(n) \mathbf{w}_{21}]^T \quad (4.35)$$

✚ Par la suite, nous proposons de remplacer le petit paramètre de régularisation  $\varepsilon(n)$  dans l'équation précédente par la norme quadratique du vecteur d'erreur  $u_1(n)$ ,  $\|\mathbf{u}_{1L}(n)\|^2$ , estimée sur les  $L$  dernières valeurs de ce dernier et qui est calculée comme suit :

$$\|\mathbf{u}_{1L}(n)\|^2 = \sum_{i=0}^{L-1} |u_1(n-i)|^2 \quad (4.36)$$

La nouvelle relation de la mise à jour du filtre adaptatif  $\mathbf{w}_{21}(n)$  est donnée par :

$$\mathbf{w}_{21}(n+1) = \mathbf{w}_{21}(n) + \mu_{21} [\alpha \|\mathbf{u}_{1L}(n)\|^2 \mathbf{I} + \gamma \mathbf{p}_2(n) \mathbf{p}_2^T(n)]^{-1} \mathbf{p}_2(n) u_1(n) \quad (4.37)$$

Sachant que nous avons introduit deux constantes positives  $\alpha$  et  $\gamma$  dans le dénominateur de l'équation du filtre  $\mathbf{w}_{21}(n)$ .

✚ Finalement, pour simplifier l'écriture de la relation (4.37) nous avons appliqué au dénominateur de celle-ci, le lemme d'inversion matricielle suivant :

$$[A + BCD]^{-1} = A^{-1} - A^{-1}B[C^{-1} + DA^{-1}B]^{-1}DA^{-1} \quad (4.38)$$

en posant  $A = \alpha \|\mathbf{u}_{1L}(n)\|^2 \mathbf{I}$ ,  $B = \mathbf{p}_2(n)$ ,  $C = \gamma$ , et  $D = \mathbf{p}_2^T(n)$ , il vient immédiatement :

$$[\alpha \|\mathbf{u}_{1L}(n)\|^2 \mathbf{I} + \gamma \mathbf{p}_2(n) \mathbf{p}_2^T(n)]^{-1} = \alpha^{-1} \|\mathbf{u}_{1L}(n)\|^{-2} \mathbf{I} - \alpha^{-1} \|\mathbf{u}_{1L}(n)\|^{-2} \mathbf{I} \mathbf{p}_2(n) \frac{\mathbf{p}_2^T(n) \alpha^{-1} \|\mathbf{u}_{1L}(n)\|^{-2}}{\gamma^{-1} + \mathbf{p}_2^T(n) \alpha^{-1} \|\mathbf{u}_{1L}(n)\|^{-2} \mathbf{p}_2(n)} \quad (4.39)$$

Après multiplication de l'équation (4.39) par le vecteur  $\mathbf{p}_2(n)$  et en faisant quelques réarrangements, cette équation peut être réécrite comme suit:

$$[\alpha \|\mathbf{u}_{1L}(n)\|^2 \mathbf{I} + \gamma \mathbf{p}_2(n) \mathbf{p}_2^T(n)]^{-1} = \frac{1}{\alpha \|\mathbf{u}_{1L}(n)\|^2 + \gamma \|\mathbf{p}_2(n)\|^2} \quad (4.40)$$

Nous substituons cette relation dans (4.37) et nous remplaçons  $\gamma$  par  $1 - \alpha$  alors nous obtenons la relation de mise à jour finale du filtre adaptatif  $\mathbf{w}_{21}(n)$  suivante :

$$\mathbf{w}_{21}(n+1) = \mathbf{w}_{21}(n) + \frac{\mu_{21}}{\alpha \|\mathbf{u}_{1L}(n)\|^2 + (1-\alpha) \|\mathbf{p}_2(n)\|^2} u_1(n) \mathbf{p}_2(n) \quad (4.41)$$

La quantité fractionnaire dans (4.41) représente le pas variable,  $\mu_{21}(n)$ , du nouvel algorithme RFBSS donnée par :

$$\mu_{21}(n) = \frac{\mu_{21}}{\alpha \|\mathbf{u}_{1L}(n)\|^2 + (1-\alpha) \|\mathbf{p}_2(n)\|^2} \quad (4.42)$$

Notons que ce pas d'adaptation variable est contrôlé par la normalisation de l'erreur et le vecteur d'entrée du filtre adaptatif  $\mathbf{w}_{21}(n)$ . L'algorithme RFBSS proposé diffère de l'algorithme NLMS par le terme ajouté  $\|\mathbf{u}_{1L}(n)\|^2$  avec une constante proportionnelle. Ce terme représente la variable qui contrôle l'adaptation du nouvel algorithme RFBSS. A chaque instant où la parole est présente, le terme ajouté  $\|\mathbf{u}_{1L}(n)\|^2$  augmentera le dénominateur de l'équation (4.42), dans ce cas le pas d'adaptation  $\mu_{21}(n) \sim 0$  et donc l'adaptation du filtre  $\mathbf{w}_{21}(n)$  sera bloquée et l'on a simplement :

$$\mathbf{w}_{21}(n+1) = \mathbf{w}_{21}(n) \quad (4.43)$$

Lorsque la parole n'est pas présente, le pas  $\mu_{21}(n)$  augmente, alors le filtre  $\mathbf{w}_{21}(n)$  est mise à jour selon l'équation (4.41).

La structure FBSS étant symétrique et utilise deux filtres adaptatifs  $\mathbf{w}_{12}(n)$  et  $\mathbf{w}_{21}(n)$  afin d'annuler les composantes du bruit acoustique à partir des signaux d'observations bruitées. Par analogie, nous déduisons la formulation du deuxième filtre adaptatif  $\mathbf{w}_{12}(n)$  à partir de l'analyse des équations (4.25) - (4.43) comme suit :

$$\mathbf{w}_{12}(n+1) = \mathbf{w}_{12}(n) + \mu_{12} u_2(n) \mathbf{p}_1(n) [\alpha \|\mathbf{u}_{2L}(n)\|^2 + (1-\alpha) \|\mathbf{p}_1(n)\|^2]^{-1} \quad (4.44)$$

#### 4.6 Evaluation de la solution algorithmique RFBSS

Ce paragraphe présente les résultats obtenus à partir de la nouvelle solution algorithmique

proposée décrite à la section 4.5.1 sous l'acronyme RFBSS. Comme nous l'avons mentionné dans les résultats de simulations ci-avant, nous avons utilisé les signaux sources de la base de données AURORA [33], ainsi que, l'utilisation des réponses impulsionnelles simulées selon le modèle de mélange convolutif [34] pour générer les signaux de mélanges  $(p_1(n), p_2(n))$ . Les paramètres utilisés pour le réglage de cette solution algorithmique RFBSS proposée sont les suivants :

- Taille des filtres :  $L = 64$ ,
- Pas d'adaptions :  $\mu_{w_{12}} = 0.01$  ;  $\mu_{w_{21}} = 0.1$  ,
- Paramètres de contrôle :  $\alpha = 0.98$  ;  $\gamma = 1 - \alpha$  .

Les signaux disponibles aux sorties de la solution algorithmique sont représentés sur la figure (4.12).

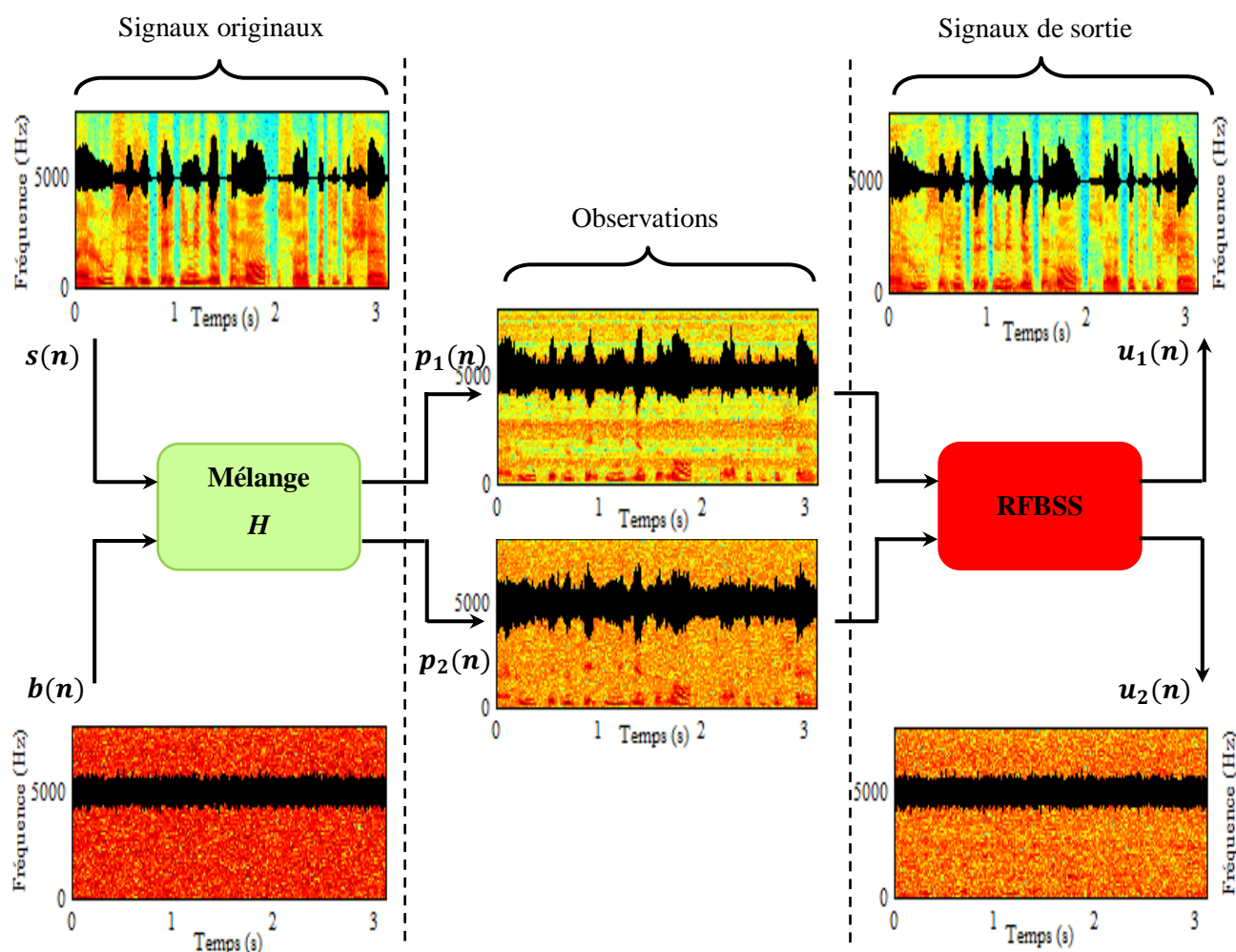


Figure 4.12 : Signaux sources (parole originale  $s(n)$ , bruit Blanc  $b(n)$ ), signaux de mélange  $(p_1(n), p_2(n))$  et signaux de sortie  $(u_1(n), u_2(n))$  de la structure RFBSS.

Nous pouvons également observer à partir de la figure 4.12 que le signal de la première sortie est complètement débruité et très proche du signal de parole original, ce bon résultat est obtenu grâce à l'utilisation du nouvel algorithme RFBSS qui adapte et contrôle en même temps les coefficients des filtres de séparation d'une manière automatique sans besoin d'un module de DAV. Pour le signal de la deuxième sortie, nous remarquons que la parole n'est pas complètement enlevée de cette sortie, ce résultat est dû en grande partie à la mauvaise identification du filtre  $w_{12}(n)$ .

Remarque 4 :

Notons que le pas d'adaptation variable  $\mu_{w_{21}}(n)$  associé à cette solution algorithmique RFBSS, se comporte quasiment comme un détecteur de bruit (soit l'inverse d'une DAV) qui garantit l'adaptation automatique du filtre  $w_{21}(n)$  sur les instants propices aux zones de silence dans le signal de parole. Donc c'est la valeur du pas variable  $\mu_{w_{21}}(n)$  qui déterminera l'efficacité de la méthode RFBSS. Une erreur dans cette détection entraîne forcément une erreur dans l'identification du filtre  $w_{21}(n)$ . Le comportement de ce pas variable est illustré sur la figure 4.13.

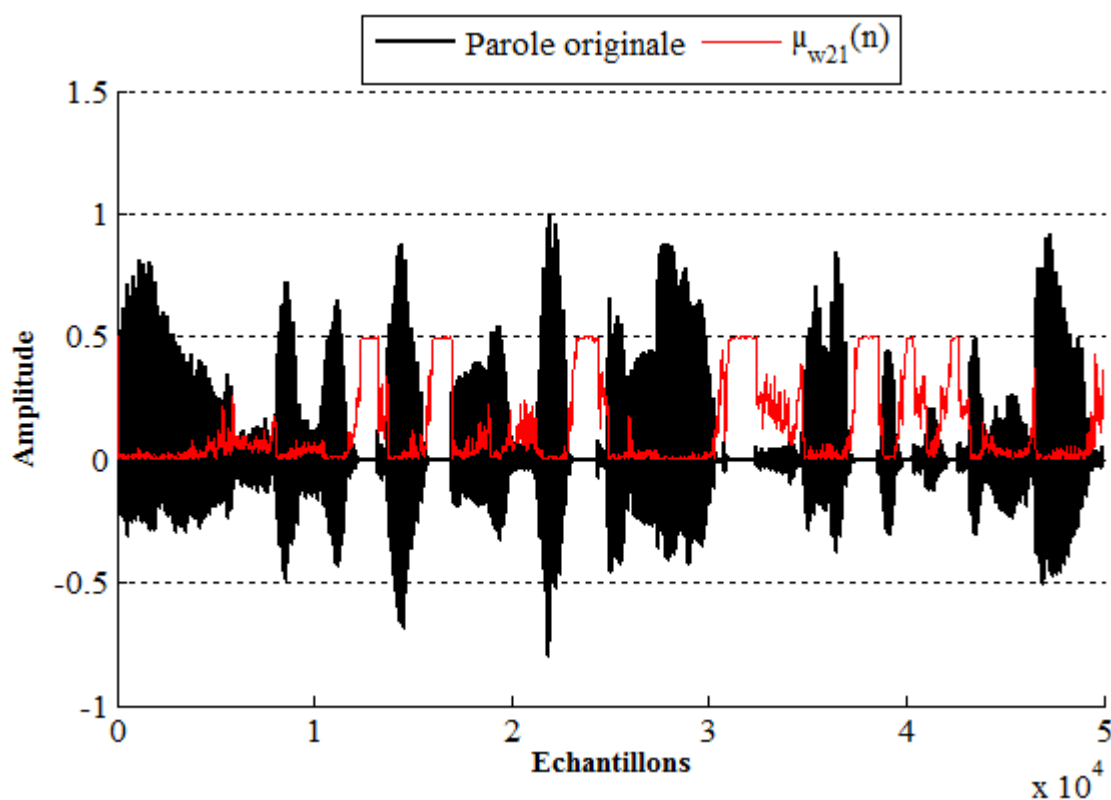


Figure 4.13 : L'évolution du pas d'adaptation  $\mu_{w_{21}}(n)$  du filtre principal  $w_{21}(n)$  pour la solution algorithmique RFBSS.



#### 4.7 Comparaison entre les structures RFBSS et LC-AFBSS

Pour voir en claire les avantages de notre solution algorithmique RFBSS par rapport à notre solution structurelle LC-AFBSS décrite précédemment, nous allons comparer ces deux solutions en reprenant les mêmes démarches de comparaison que dans la section 4.4. Les paramètres de simulations pour la solution structurelle LC-AFBSS sont les mêmes cités dans le tableau 4.2. Afin de comparer les performances de ces deux structures (RFBSS et LC-AFBSS), nous présentons sur les figures suivantes : les signaux disponibles en sortie de traitement, l'évolution temporelle du rapport signal à bruit, la distance cepstrale ainsi que les évolutions temporelles du désajustement mesuré sur les coefficients du filtre adaptatif  $w_{21}(n)$  pour chaque structure.

Les résultats après réduction de bruit pour les deux solutions algorithmique et structurelle sont également présentés dans la figure (4,14), ces signaux sont comparés avec le signal de parole original. Nous pouvons également observer que les signaux disponibles en sortie de traitement pour chaque solution sont débruités.

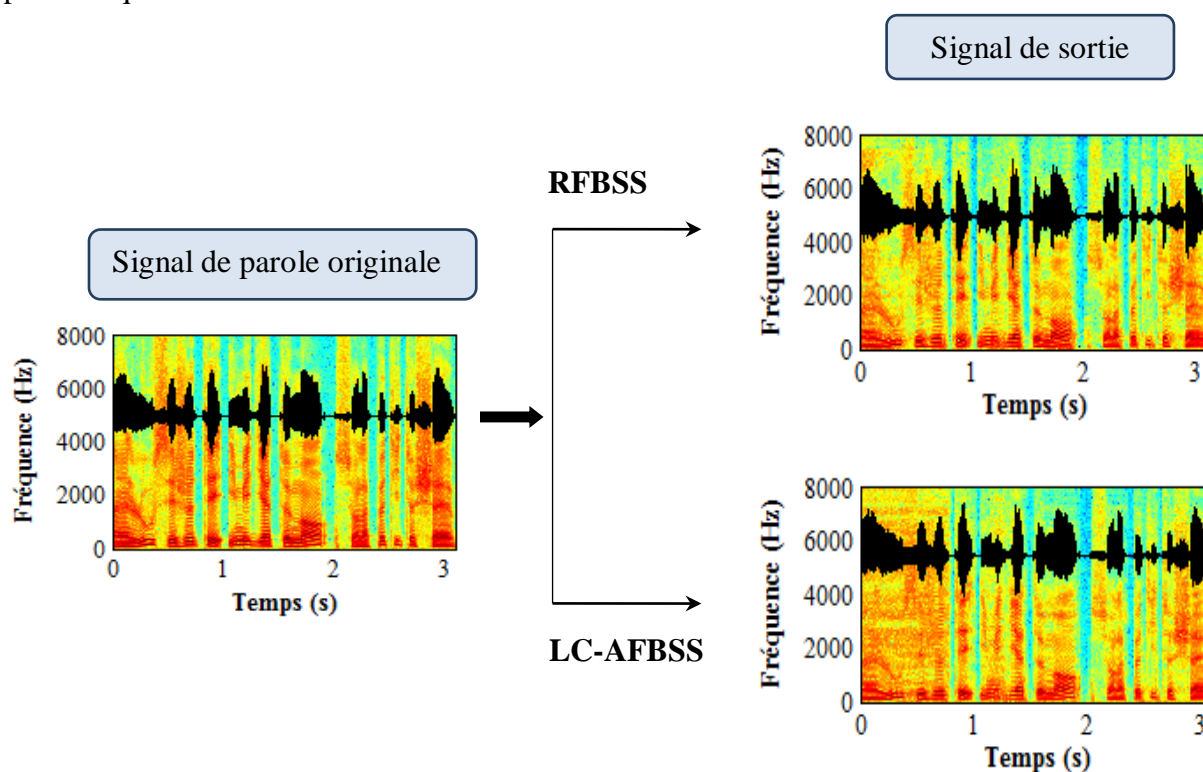


Figure 4.14 : Signal de parole originale  $s(n)$  (à gauche), signal de sortie  $u_1(n)$  de la structure RFBSS (en haut), signal de sortie  $u_1(n)$  de la structure LC-AFBSS (en bas).

En nous basant sur les spectrogrammes de la figure 4.14, nous pouvons remarquer que le

signal de sortie de la méthode RFBSS est plus proche du signal de parole original (en contenu spectral) en comparaison avec l'autre résultat (LC-AFBSS). De ces résultats, il devient évident que la solution algorithmique RFBSS présente une meilleure réduction de bruit comparée à la solution structurelle LC-AFBSS. Cela est justifié par le bon comportement du pas  $\mu_{w21}(n)$  associé à la méthode RFBSS comparé à celui de la solution structurelle LC-AFBSS (voir figure 4.15). D'après cette figure nous observons très nettement la bonne détection des zones de silence avec le pas variable de la solution algorithmique RFBSS.

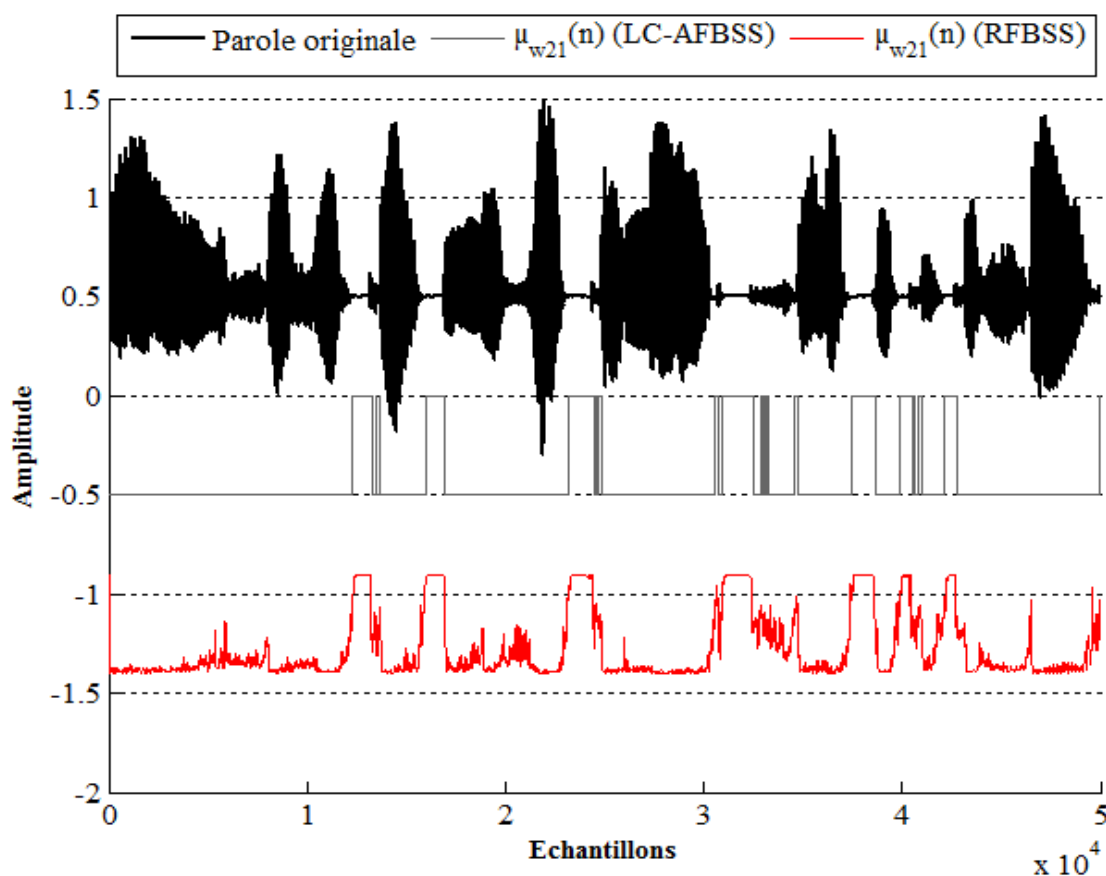


Figure 4.15 : Comparaison des évolutions des pas d'adaptations  $\mu_{w21}(n)$  pour les deux méthodes LC-FBSS et RFBSS.

Nous avons évalué la distance cepstrale mesurée entre le signal de parole original et la composante du signal de parole présenté au sein du signal de sortie pour les deux solutions. Les résultats correspondants sont présentés sur la figure 4.16. Cette figure met en évidence la supériorité de la solution algorithmique RFBSS sur la solution structurelle LC-AFBSS en terme de distance cepstrale (DC). Nous observons un écart de performance entre les deux solutions, cet écart est de l'ordre de 3 dB en faveur de la méthode RFBSS.

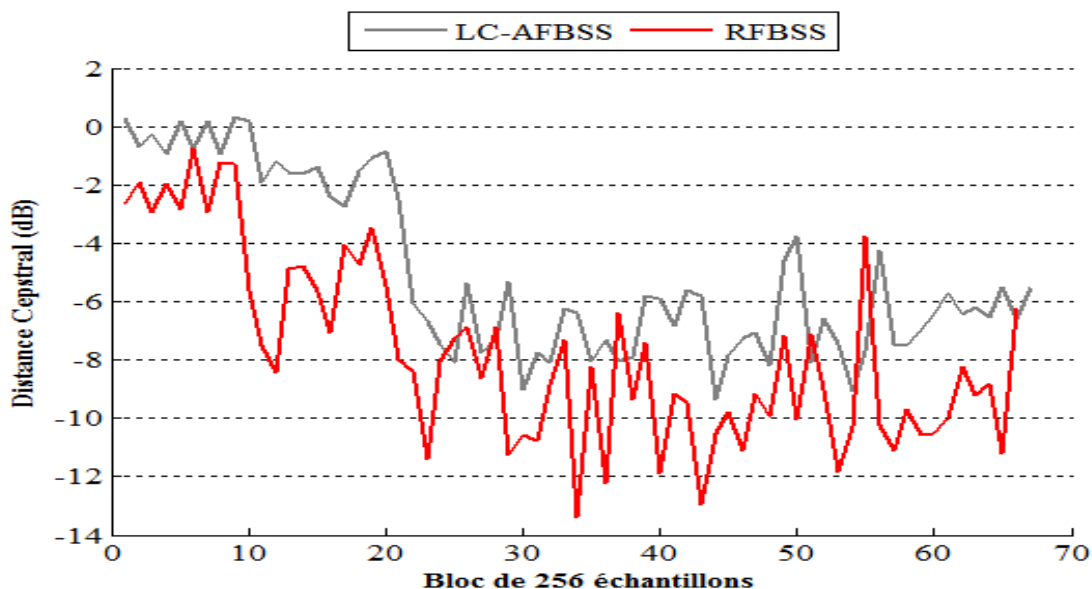


Figure 4.16: Evolution du critère de la distance cepstrale des deux structures RFBSS (en rouge) et LC-AFBSS (en gris).

L'évolution du rapport signal à bruit entre les deux méthodes précédentes est représentée à la figure 4.17. Nous pouvons constater le bon comportement de la méthode RFBSS proposée. Les valeurs moyennes des rapports signal à bruit sont d'environ 55.75 dB et 50.31 dB pour, respectivement, la méthode RFBSS et la méthode LC-AFBSS. Il existe donc un gain de 4 dB pour la méthode RFBSS sur la méthode LC-AFBSS. Ceci confirme la supériorité de la solution algorithmique RFBSS proposée sur la solution structurelle LC-AFBSS.

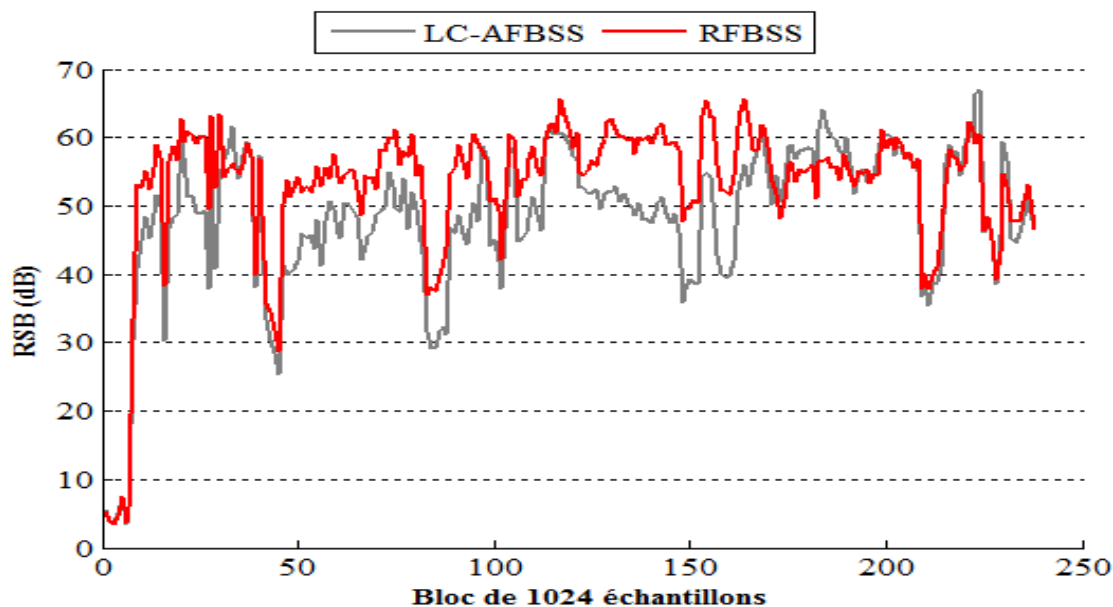


Figure 4.17 : Evolution du critère du rapport signal à bruit (RSB) pour les deux structures RFBSS (en rouge) et LC-AFBSS (en gris).

La figure 4.18 montre une comparaison entre le désajustement obtenu par les deux solutions : algorithmique et structurelle. D'après ces résultats, on peut constater que la solution algorithmique RFBSS présente une convergence rapide par rapport à la solution structurelle LC-AFBSS.

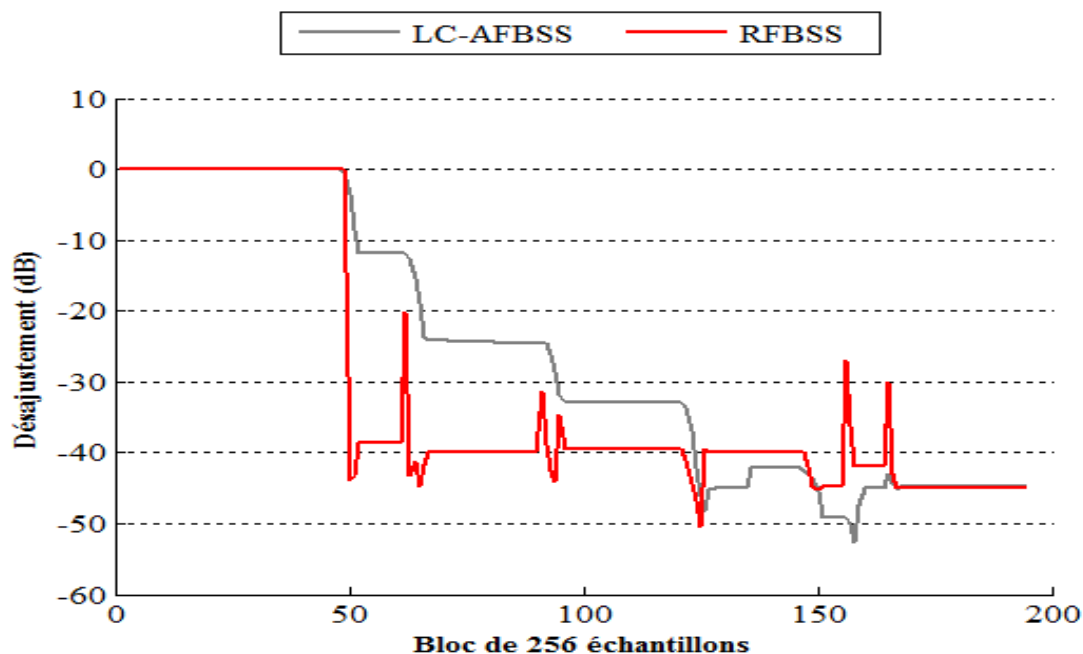


Figure 4.18 : Comparaison du désajustement obtenu avec les deux structures RFBSS (en rouge) et LC-AFBSS (en gris).

Comme nous l'avons fait pour la solution structurelle LC-AFBSS, le tableau 4.4 fournit la complexité de la solution algorithmique RFBSS en termes de coût de calcul et du nombre de paramètres seuils utilisé. Par comparaison avec les résultats associés au LC-AFBSS, nous constatons que notre solution algorithmique RFBSS est moins complexe, et n'utilise aucun paramètre seuil dans son fonctionnement.

Tableau 4.4 : La complexité de la solution algorithmique RFBSS.

Structures	Nombre d'opérations			Nombre de paramètres seuils
	Additions	Multiplications	Total	
RFBSS	$6L+4$	$6L+6$	$12L+10$	0

Afin de conclure sur la comparaison des performances expérimentales de toutes les méthodes de séparation de source de type directe utilisée (FBSS) et proposée dans ce travail de thèse (AFBSS, LC-AFBSS et RFBSS), Nous avons utilisé le critère subjectif (MOS) pour évaluer la qualité de la parole en sortie de chaque version de la SAS, les résultats obtenus ont été résumés dans le tableau 4.5 suivant. Dans ce tableau nous avons interrogés 5 sujets âgés entre 20 et 30 et qui sont d'une bonne qualité d'audition.

Tableau 4.5 : Comparaison subjective des différentes méthodes de la SAS.

RSB d'entrée (dB)	Méthodes			
	FBSS	AFBSS	LC-AFBSS	RFBSS
-12	3.17	3.13	3.74	<b>4.03</b>
-6	3.82	3.75	3.93	<b>4.15</b>
0	3.98	3.86	4.25	<b>4.67</b>
6	4.08	4.2	4.56	<b>4.95</b>
12	4.29	4.34	4.84	<b>5.00</b>

Les meilleurs résultats de la qualité vocale du signal de parole ont été obtenus par la solution algorithmique proposée RFBSS. Ces tests ont confirmé les bonnes performances et la supériorité de notre solution algorithmique en comparaison avec les autres méthodes directes (FBSS, AFBSS et LC-AFBSS).

#### 4.8 Conclusion

Dans ce chapitre, nous nous sommes focalisés sur le contrôle automatique des structures de SAS tout en tenant compte de l'aspect de la complexité. Les méthodes que nous avons proposé se répartissent en deux classes selon la manière du pilotage des pas d'adaptations afin de contrôler la mise à jour des filtres adaptatif de la structure de SAS automatiquement et à moindre coût.

La première classe correspond aux solutions structurelles qui reposent sur la variation des pas d'adaptations des filtres de séparation à partir d'un module de DAVA simplifié basé sur l'estimation d'énergie. Ces solutions ont pour but de commander l'adaptation automatiquement et offrent l'avantage de réduire la complexité arithmétique par rapport aux

méthodes améliorées. Les résultats expérimentaux obtenus à partir de la solution structurelle de type directe LC-AFBSS montrent la validité de cette technique et son moindre coût calculatoire par rapport à la structure améliorée AFBSS.

La deuxième classe s'appuie sur des approches algorithmes, d'où vient son nom : solution algorithmique. Le but de cette solution est de guider l'évolution des pas d'adaptation de chacun des filtres de séparation pour rendre le contrôle d'adaptation de ces filtres automatique. La validation expérimentale de la solution algorithmique pour la structure de SAS de type directe (RFBSS) a été réalisée. Ces résultats ont confirmé les bonnes performances de cette solution.

Une étude comparative a été réalisée et à montrer la supériorité de la solution algorithmique de type directe RFBSS en terme de (mesures objectives, testes subjectifs et la complexité) par rapport à la solution structurelle LC-AFBSS.

L'ensemble des résultats obtenus dans cette thèse rendent la solution algorithmique RFBSS très compétitif dans les applications de réduction de bruit et le rehaussement du signal de parole. Nous recommandons l'utilisation cette méthode en pratique

## CONCLUSION

Dans les terminaux mains-libres, le signal de parole reçu par le capteur est dégradé par la présence du bruit de l'environnement, donc un module de réduction du bruit est indispensable pour ce type de système afin d'éliminer le bruit indésirable et améliorer la qualité de la parole transmise vers les réseaux de communication.

Le but de ce travail de thèse consistait à proposer une méthode de réduction de bruit adaptatif fiable, nous nous sommes orientés vers la structure de séparation aveugle de sources SAS (directe et récursive) qui permet de séparer le signal et le bruit à partir des observations bruitées dans le cas d'une prise de son bi-capteurs sans aucune connaissance a priori sur les signaux sources.

Dans un premier temps, nous nous sommes consacrés, au chapitre 2, à l'étude des structures de SAS (directe et récursive) en donnant leurs avantages et inconvénients, ainsi que les étapes nécessaires à leur bon fonctionnement. Parmi celles-ci nous avons choisi de cibler notre étude sur l'étape de contrôle de la mise à jour des filtres adaptatifs. Ce contrôle est effectué dans les structures de SAS déjà existantes dans la littérature (FBSS et BBSS) par un système de DAVM. Le manque d'information sur les signaux d'entrées empêche ce système de DAVM d'être réalisable en pratique.

Dans ce contexte, nous avons proposé dans le chapitre 3, deux nouvelles méthodes appelées AFBSS et ABBSS qui correspondent à des versions améliorées des deux structures classiques de séparation aveugle de sources (FBSS et BBSS). Ces nouvelles méthodes utilisent un module de DAVA dans leur fonctionnement. Ce dernier est basé sur l'estimation des rapports signal à bruit (RSB) sur les deux voies de la structure de séparation.

Nous avons testé et comparé la structure améliorée directe AFBSS à une structure de l'état de l'art (FBSS), à partir des résultats obtenus nous avons remarqué que la structure proposée AFBSS a permis d'obtenir des performances plus proches que celles de la structure directe classique (FBSS) en termes de distance cepstrale (DC), le rapport signal à bruit (RSB) en sortie de traitement et le désajustement des coefficients des filtres adaptatifs (SM).

Nous avons évalué l'effet du pas d'adaptation, la taille des filtres et le rapport signal à bruit de la voie de référence ( $RSB_2$ ) sur les performances de la structure AFBSS.

Nous avons remarqué que cette structure est sensible aux valeurs faibles du pas d'adaptation, aux tailles importantes des filtres de séparation et aux valeurs importantes du  $RSB_2$ .

La complexité (en terme de calcul et du nombre de paramètres seuils) induite par l'utilisation du bloc de DAVA avec les deux structures de SAS n'encourage pas au développement de ces méthodes améliorées dans la pratique, ce qui nous a conduit à proposer deux catégories de solutions. La première catégorie (solutions structurelles : LC-AFBSS et LC-BBSS), décrite à la première partie du chapitre 4, consiste à simplifier l'architecture du bloc de DAVA des méthodes améliorées. Ce nouveau système de DAVA simplifié qui est basé sur l'estimation de l'énergie ( $E_e$ ) permet de contrôler l'adaptation des deux filtres de séparation automatiquement et à faible complexité algorithmique. Les résultats de simulation de la solution structurelle de type directe LC-AFBSS montrent à la fois des performances presque identiques à celles de la structure AFBSS, et une complexité assez faible devant celle-ci.

La deuxième catégorie que nous avons proposé (solution algorithmique : RFBSS), décrite à la deuxième partie du chapitre 4, est fondée sur un nouvel algorithme assurant en même temps l'adaptation des coefficients des filtres de séparation et le contrôle automatique de cette adaptation. Dans l'ensemble des résultats, il a été observé que la solution algorithmique RFBSS performe beaucoup mieux que la solution structurelle LC-AFBSS. En ce qui concerne la complexité, cette solution algorithmique est moins complexe par rapport à la solution structurelle et n'utilise aucun paramètre seuil dans son fonctionnement.

Un teste subjectif (MOS) démontre la supériorité de la solution algorithmique RFBSS par rapport aux autres structure de SAS citées auparavant.



## APPENDICE A

### LISTE DES ABBREVIATIONS

ABBSS	Automatic Backward Blind Source Separation
AEC	Acoustic Echo Cancelation
AFBSS	Automatic Forward Blind Source Separation
ANC	Adaptive Noise Canceller
AVAD	Automatic Voice Activity Detection
BSS	Blind Source Separation
BBSS	Backward Blind Source Separation
DAV	Detection d'Activité Vocale.
DAVA	Detection d'Activité Vocale Automatique.
DAVM	Detection d'Activité Vocale Manuelle
dB	décibel
DC	Distance Cepstrale
DR	Déréverbération
DSP	Densité Spectrale de Puissance
EQM	Erreur Quadratique Moyenne
EQMM	Erreur Quadratique Moyenne Minimum
FBSS	Forward Blind Source Separation
FPGA	Field Programmable Gate Arrays
LMS	Least-Mean Square.
LC-ABBSS	Low Complexity Automatic Backward Blind Source Separation.
LC-AFBSS :	Low Complexity Automatic Forward Blind Source Separation.
MVAD	Manuel Voice Activity Detection
MMSE	Minimum Mean Square Erreur
MOS	Mean Opinion Score
NLMS	Normalized Least-Mean Square
NR	Noise Reduction
RFBSS	Regulized Forward Blind Sorce Separation
RIF	Réponse Impulsionnelle Finie
RII	Réponse Impulsionnelle Infinie

RSB	Rapport Signal à Bruit
SAS	Séparation Aveugle de Sources
SM	System Mismatch
SSA	Soustraction Spectrale d'Amplitude
SSP	Soustraction Spectrale de Puissance
TFCT	Transformée de Fourier à Court Terme
TFD	Transformée de Fourier Discrète
UIT	Union Internationale des Télécommunications
VAD	Voice Activity detection

## APPENDICE B

### LISTE DES SYMBOLES

$f$	Indice de fréquence
$L$	Taille des filtres d'analyse et de synthèse
$M$	Taille du filtre adaptatif
$n$	Indice de temps
$k$	Indice de trame
$B$	Petite constante positive
$E[ \ ]$	Espérance mathématique
$J$	Fonction coût
$\mathbf{R}_{xx}$	Matrice d'autocorrélation du signal $x(n)$
$\mathbf{P}_{xd}$	Vecteur d'intercorrélation entre $x(n)$ et $d(n)$
$\mathbf{w}^{opt}$	Coefficients du filtre optimal de Wiener
$\mathbf{w}(n)$	Coefficients du filtre adaptatif
$\mu$	Pas d'adaptation fixe
$\lambda_{max}$	Valeur maximale propre de la matrice d'autocorrélation
$\sigma_x^2$	Variance de $x$
$( )^T$	Transposé d'une matrice ou d'un vecteur
$\gamma_x(f)$	Densités spectrales de puissances (DSP) de $x$
$\left. \begin{array}{l} h_{11}(n), \\ h_{22}(n) \end{array} \right\}$	Les réponses impulsionnelles des canaux directs de couplage
$\left. \begin{array}{l} h_{21}(n), \\ h_{12}(n) \end{array} \right\}$	Les effets de couplage croisé
$\mu_{w12}$	Pas d'adaptation fixe du filtre adaptatif $w_{12}(n)$
$\mu_{w21}$	Pas d'adaptation fixe du filtre adaptatif $w_{21}(n)$
$\mu_{w12}(n)$	Pas d'adaptation variable du filtre adaptatif $w_{12}(n)$
$\mu_{w21}(n)$	Pas d'adaptation variable du filtre adaptatif $w_{21}(n)$
$\delta(n)$	Impulsion de Dirac
$\mu_{w12min}$	Valeur minimale du pas d'adaptation variable $\mu_{w12}(n)$
$\mu_{w12max}$	Valeur maximale du pas d'adaptation variable $\mu_{w12}(n)$

$\mu_{w21min}$	Valeur minimale du pas d'adaptation variable $\mu_{w21}(n)$
$\mu_{w21max}$	Valeur maximale du pas d'adaptation variable $\mu_{w21}(n)$
$\mu_{wcont}$	Pas d'adaptation fixe du filtre adaptatif de contrôle $w_{cont}(n)$
$\theta$	Facteur d'oubli
$\varepsilon$	Petit paramètre de régularisation
$\alpha, \gamma$	Paramètres de contrôle

## REFERENCE

- [1] M. Djendi, A. Gilloire and P. Scalart, « Noise cancellation using two closely spaced microphones: experimental study with a specific model and two adaptive algorithms », IEEE International Conference on Acoustics Speech and Signal Processing Proceedings, vol. 3, pp. 744–747, 2006.
- [2] M. Djendi, A. Gilloire and P. Scalart, « New Frequency Domain Post-Filters for Noise Cancellation Using Two Closely Spaced Microphones», 15th European Signal Processing Conference (EUSIPCO 2007), Poznan, Poland, September 3-7, 2007.
- [3] M. Djendi, P. Scalart and A. Gilloire « Comparative study of new blind source separation structures for two-cannel noise cancellation », 17th European Signal Processing Conference (EUSIPCO), Glasgow, UK, 23–29, vol. 1, pp. 165–169, 2009
- [4] L. Manh Tuan, «Analyse acoustique de sons bien identifiés par un système de reconnaissance automatique de la parole », Mémoire de fin d'étude, Institut de la Francophonie pour l'Informatique INRIA de Lorraine – LORIA, 2007.
- [5] R. Boite, H. Boulard, T. Dutoit, J. Hancq et H. Leich « Traitement de la Parole », Presses Polytechniques Universitaires Romandes, Lausanne, 2000.
- [6] A. Amehraye, « Débruitage perceptuel de la parole », Thèse de Doctorat, École Nationale Supérieure des Télécommunications de Bretagne, 2009.
- [7] F. Ykhlef, « Etude de la stabilité des algorithmes moindres carrés transversaux rapides avec des signaux parfaitement prédictibles application a la parole », Thèse de Magister, Université Saad Dahlab de Blida, 2002.

- [8] C. Plapous, « Traitements pour la réduction de bruit. Application à la communication parlée », Thèse de Doctorat, Université de Rennes 1, 2005.
- [9] K. Meghriche, « Filtrage Adaptatif Utilisant les Statistiques d'Ordre Supérieur », Thèse de Doctorat d'Etat, École Nationale Polytechnique, 2006.
- [10] H. Alaeddine, « Application de la transformée en nombres entiers à la conception d'algorithmes de faible complexité pour l'annulation d'échos acoustiques », Thèse de Doctorat, Université de Bretagne Occidentale, 2010.
- [11] J.-F. Bercher et P. Jardin, « Introduction au filtrage adaptatif », ESIEE-Paris, I4-TTS, 2003.
- [12] B. Widrow, J.R. Glover, J. McCool, J. Kaunitz, C. Williams, R. Hearn, J. Zeidler, E. Dong, and R. Goodlin, « Adaptive noise cancelling: Principles and Applications », Proc. of the IEEE, vol. 63, no. 12, pp.1692–1716, 1975.
- [13] G. Bellanger, « Adaptive Digital Filters », 2<sup>ème</sup> édition, Marcel Dekker, Inc., New York, 2001.
- [14] J. El Mhamdi, F. Regragui et M. Harnafi, « Traitement adaptatif appliqué au signal sismique », Bulletin de l'Institut Scientifique, Rabat, section Sciences de la Terre, no.30, 2008.
- [15] F. Ykhlef, « Réduction de bruit et contrôle s'écho pour les applications radio mobile et audioconférence » Thèse de Doctorat, École Nationale Polytechnique, 2008.
- [16] M. Talbi et C. Adnen « Débruitage de la parole par paquet d'ondelettes » The 4<sup>th</sup> International Conference: Sciences of Electronic Technologies of Information and Telecommunication (SETIT 2007), March 25-29, 2007, TUNISIA.

- [17] Steven F. Boll, « Suppression of acoustic noise in speech using spectral subtraction », IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. 27, no.2, pp. 113-120, 1979.
- [18] H. Glotin, « Elaboration et comparaison de systèmes adaptatifs multi-flux de reconnaissance robuste de la parole : incorporation des indices de voisement et de localisation », Thèse de Doctorat, Institut Nationale Polytechnique de Grenoble, 2001.
- [19] P. Howells, « Intermediate frequency side lobe canceller », US patent 3202990, Aug.24, 1965.
- [20] J. Kaunitz, « Adaptive filtering of broadband signals as applied to noise cancelling », Stanford Electronics Lab., Stanford Univ., Stanford Calif., Rep. SU-SEL-72-038, Aug. 1972 (Ph.D. Dissertation).
- [21] M. Sondhi, « An Adaptive echo canceller », Bell Syst. Tech. J., vol. 46, pp. 497-511, Mar. 1967.
- [22] J. Rosenberger and E. Thomas, « Performance of an adaptive echo canceller operating in a noisy, linear, time-invariant environment », Bell Syst. Tech. J., vol.50, pp.785-813, Mar. 1971.
- [23] B. Widrow and M. E. Hoff Jr., « Adaptive switching circuits », in IRE WESCON Conv. Rec. , pt. 4, pp. 96-104, 1960.
- [24] N. Charkani, « Séparation auto-adaptative de sources pour des mélanges convolutifs. Application à la téléphonie mains-libres dans les voitures », Thèse de Doctorat, Institut Nationale Polytechnique de Grenoble, 1996.
- [25] B. Rivet, « La bimodalité de la parole au secours de la séparation de sources », Thèse de Doctorat, Institut National Polytechnique de Grenoble, 2006.

- [26] M. Akil, « Séparation de sources convolutives », Thèse de Doctorat, Université Joseph Fourier – Grenoble 1, 2008.
- [27] F. Abrard, « Méthodes de séparation aveugle de sources et applications : Des statistiques d'ordre supérieur à l'analyse temps-fréquence », Thèse de Doctorat, Université Paul Sabatier – Toulouse III, 2003.
- [28] H.L. Nguyen Thi, J. Caelen et Ch. Jutten, « Rehaussement de la parole par la séparation de sources dans un mélange convolutif », Colloque C5, supplément au Journal de Physique III, vol.4, 1994.
- [29] Al-Kindi, M. J., & Dunlop, J. (1989), « Improved adaptive noise cancellation in the presence of signal leakage on the noise reference channel », *Signal Processing*, 17(3), 241–250, 1989.
- [30] Van Gerven, S., & Van Compernelle, D, « Signal separation by symmetric adaptive decorrelation: Stability, convergence, and uniqueness », *IEEE Transactions on Signal Processing*, 74(3), 1602–1612, (1995).
- [31] M. Djendi, P. Scalart and A. Gilloire, « Comparative study of new blind source separation structures for two-channel acoustic noise cancellation », 17th European Signal Processing Conference (EUSIPCO 2009), Glasgow, Scotland, August 24-28, 2009.
- [32] M. Zoulikha, K. Mokhnache, « Rehaussement du signal de parole par les méthodes adaptatives de réduction de bruit », Mémoire d'Ingéniorat, Université de Blida1, 2010.
- [33] Combescure, P. (1981). 20 listes de dix phrases phonétiquement équilibrées. *Revue d'Acoustique*, 56, 34–38.



- [34] M. Djendi, A. Gilloire and P. Scalart, « Noise cancellation using two closely spaced microphones: experimental study with a specific model and two adaptive algorithms», IEEE Int. Conf. ASSP, vol.3, pp. 744-747, 2006.
- [35] L. Rabiner, B.-H. Juang, « Fundamentals of Speech Recognition », Prentice-Hall, Englewood Cliffs, NJ, 1993.
- [36] M. Djendi, « Advanced techniques for two-microphone noise reduction in mobile communications », Ph.D. dissertation (in French), University of Rennes 1, France, 2010.
- [37] ITU-T Recommendation P.800, « Methods for Subjective Determination of Transmission Quality », International Telecommunication Union, Geneva, 1996.
- [38] Philipos C. Loizou , « Speech Quality Assessment », Multimedia Analysis, Processing and Communications. Studies in Computational Intelligence, vol 346. Springer, Berlin, Heidelberg, 2011.
- [39] M. Djendi, M. Zoulikha, « New automatic forward and backward blind sources separation algorithms for noise reduction and speech enhancement », Journal of Computers and Electrical Engineering 40 (2014) 2072–2088.
- [40] S. Ikeda and A. Sugiyama, « An adaptive noise canceller with low signal- distortion in the presence of crosstalk », IEICE Trans. Fundamentals, vol. E82-A, no.8, 1999.
- [41] M. Djendi and P. Scalart, « Double pseudo affine projection algorithm for speech enhancement and acoustic noise reduction », 20th European Signal Processing Conference (EUSIPCO 2012), Bucharest, Romania, August 27-31, 2012.
- [42] M. Zoulikha, « Méthodes améliorées pour la réduction du bruit acoustique et le rehaussement du signal de parole », Mémoire de Magister, Université de Blida1, 2014.

- [43] M. Zoulikha and M.Djendi « A new robust forward BSS adaptive algorithm based on automatic voice activity detector for speech quality enhancement », *International Journal of Speech Technology* 21,1007–1020 (2018).
- [44] M. Zoulikha, M.Djendi and A. Guessoum, « An Efficient Automatic Backward Blind Source Separation Method for Hands-Free System», *The 20<sup>th</sup> International Conference Certificate on Information, Communications and Signal Processing (ICICSP 2018)*, Istanbul, 23-24 July 2018, Turkey.
- [45] M. Zoulikha and M.Djendi, « A new regularized forward blind source separation algorithm for automatic speech quality enhancement », *Journal of Applied Acoustics* 112 (2016) 192–200.
- [46] A. Maddi, « algorithmes implementables d'identification paramétrique : application au signal de parole bruitée», *Thèse de Doctorat, Ecole Nationale Polytechnique*, 2008.

## PUBLICATIONS ET CONFERENCES

### Publications :

1. M. Djendi and M. Zoulikha, “ New automatic forward and backward blind sources separation algorithms for noise reduction and speech enhancement”. *Journal of Computers and Electrical Engineering* 40 (2014) 2072–2088.
2. M. Zoulikha and M. Djendi, “A new regularized forward blind source separation algorithm for automatic speech quality enhancement. *Journal of Applied Acoustics* 112 (2016) 192–200.
3. M. Zoulikha and M. Djendi, “A new robust forward BSS adaptive algorithm based on automatic voice activity detector for speech quality enhancement. *International Journal of Speech Technology* 21,1007–1020(2018).
4. M. Djendi and M. Zoulikha, “A new efficient backward BSS crosstalk-resistant algorithm for automatic blind speech quality enhancement. *International Journal of Speech Technology* 21, 809–823(2018).

### Communications :

1. Meriem Zoulikha, Mohamed Djendi and Abderezzak Guessoum (2017) “A Variable Step Size-Forward Blind Source Separation Algorithm for Speech Enhancement”, The 3<sup>rd</sup> International Conference on Electrical Engineering and Control Applications (ICEECA2017), Constantine, 21-22 November 2017, Algeria.
2. Mohamed Djendi and Meriem Zoulikha (2017) “A New Robust Forward Blind Source Separation Algorithm for Speech Enhancement”, The 3<sup>rd</sup> International Conference on

Electrical Engineering and Control Applications (ICEECA2017), Constantine, 21-22 November 2017, Algeria.

3. Meriem Zoulikha, Mohamed Djendi and Abderezzak Guessoum (2017) “An Efficient Automatic Backward Blind Source Separation Method for Hands-Free System ”, The 20<sup>th</sup> International Conference Certificate on Information, Communications and Signal Processing (ICICSP 2018), Istanbul, 23-24 July 2018, Turkey.