

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

UNIVERSITE SAAD DAHLEB DE BLIDA

Faculté de Technologie
Département d'Electronique

MEMOIRE DE MAGISTER

Option : Signaux et Systèmes

METHODES AMELIOREES POUR LA REDUCTION DU BRUIT ACOUSTIQUE ET
LE REHAUSSEMENT DU SIGNAL DE PAROLE

Par

Meriem ZOULIKHA

Devant le jury composé de :

Mr A.GUESSOUM	Professeur, (USDB)	Président
Mr D.BERKANI	Professeur, (ENSP)	Examineur
Mr Z.BENSELAMA	Maître de conférences A, (USDB)	Examineur
Mr M.DJENDI	Maître de conférences A, (USDB)	Rapporteur
Mr F.YKHLEF	Maître de conférences B, (USDB)	Invité
Mme N.AMIROUCHE	Chargée de cours, (USDB)	Invitée

Blida, Avril 2014

Remerciements

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

J'adresse toute ma reconnaissance à mon encadreur, Monsieur Mohamed Djendi, Professeur à l'université de Blida, pour m'avoir fait confiance et m'avoir dirigé pendant cette année. Je tiens tout particulièrement à le remercier pour sa patience tout au long de ce travail, ainsi que pour la richesse de ses idées. Espérant d'avoir l'occasion de travailler avec lui dans le futur, qu'il trouve ici l'expression de ma respectueuse gratitude.

Que soient également remerciés tous les membres du jury, pour l'honneur et l'intérêt qu'ils m'ont accordé en acceptant d'examiner et d'évaluer mon mémoire.

Mes pensées vont vers ma famille, et surtout mes parents, qui m'ont permis de poursuivre mes études jusqu'à aujourd'hui.

Mes remerciements les plus sincères à toutes les personnes qui ont contribué de près ou de loin à l'élaboration de ce mémoire.

ملخص

في هذه المذكرة، قمنا بمعالجة مشكلة الحد من الضوضاء وتعزيز إشارة الكلام. لقد اهتمنا بأساليب إلغاء الضوضاء ثنائي الالتقاط وعلى وجه الخصوص بالهيكلين (BSS) المباشر و الرجعي لفصل المصادر. العائق الرئيسي لهذه الهياكل هو استخدام نظام يدوي للكشف عن فترات وجود الإشارة الصوتية (VAD). في هذه المذكرة، نقترح هيكلين جديدين اللذان يوافقان إصدارات محسنة للهياكل التقليدية لفصل المصادر (المباشر و الرجعي). هذه الهياكل الجديدة تستخدم نظام أوتوماتيكي للكشف عن فترات وجود الإشارة الصوتية. ويستند هذا النظام المقترح على تقييم نسبة الإشارة إلى الضوضاء (SNR) على قناتي هيكل الانفصال. وقد أظهرت هذه الهياكل الجديدة الكثير من الايجابية في سلوكها.

Résumé

Dans ce mémoire, nous avons traité le problème de réduction du bruit et le rehaussement du signal de parole. Nous nous sommes intéressés aux méthodes d'annulation de bruit bi-capteurs et en particulier les deux structures de séparation de sources (BSS) de type Directe et Récursive. L'inconvénient majeur de ces structures est l'utilisation d'un système de détection d'activité vocal manuel (DAV). Dans ce mémoire, nous proposons deux nouvelles structures qui correspondent à des versions améliorées des structures classiques de séparation de sources (Directe et Récursive). Ces nouvelles structures utilisent un système de détection d'activité vocale (DAV) automatique dans leur fonctionnement. Ce système proposé est basé sur l'estimation des rapports signal à bruit (RSB) sur les deux voies de la structure de séparation. Ces nouvelles structures ont montré un comportement très efficace.

Abstract

In this thesis, we treated the problem of acoustic noise reduction and the speech enhancement. We are interested by the two-sensors noise cancellation methods and in particularly the Forward and Backward Blind source separation structures (BSS). The major drawback of these structures is the use of a manual Voice Activity Detector (VAD) system. In this thesis, we propose two new structures that correspond to improved versions of the classical blind source separation structures (Farward and Backward). These new structures use an automatic Voice Activity Detector (VAD) system. The later is based on the estimated signal-to-noise ratios (SNR) on the two channels of the separation structure. These new structures showed a very efficient behavior.

Table des matières

Introduction générale.....01

Chapitre I **Généralités sur le traitement du signal**

I.1 Introduction.....03

I.2 Généralités sur le traitement du signal.....03

 I.2.1 Notion de signal.....03

 I.2.2 Classification des signaux.....03

 I.2.2.1 Classification phénoménologique.....03

 I.2.2.2 Classification énergétique.....04

 I.2.2.3 Classification morphologique.....05

 I.2.3 Numérisation.....06

 I.2.3.1 Echantillonnage.....06

 I.2.3.2 Quantification.....07

 I.2.3.3 Codage.....07

 I.2.4 Notion de système.....08

 I.2.5 Convolution.....08

 I.2.6 Corrélation.....08

 I.2.7 Filtrage numérique.....09

 I.2.7.1 Définition.....09

 I.2.7.2 Transformée en z10

 I.2.7.3 Causalité et stabilité.....12

 I.2.7.4 Classification des filtres numériques.....13

 I.2.8 Filtrage de Wiener.....14

 I.2.9 Algorithmes du gradient stochastique LMS.....18

I.2.10	Algorithmes du gradient stochastique normalisé NLMS.....	20
I.2.11	Filtrage adaptatif.....	21
I.3	Généralités sur la parole et le bruit.....	22
I.3.1	Qu'est ce que c'est la parole ?.....	22
I.3.2	Production du signal de parole.....	22
I.3.3	Classification des sons de la parole.....	24
I.3.3.1	Les sons voisés.....	24
I.3.3.2	Les sons non-voisés.....	25
I.3.3.3	Les silences.....	26
I.3.2	Nature et caractéristiques du bruit.....	26
I.4	Conclusion.....	27

Chapitre II

Méthodes mono/bi/multi-voies pour la réduction du bruit et le rehaussement de la parole

II.1	Introduction.....	28
II.2	Méthodes de débruitage mono-voie.....	28
II.2.1	Soustraction spectrale.....	31
II.2.2	Filtrage de Wiener.....	33
II.3	Méthodes de débruitage bi-voies.....	34
II.3.1	Structure d'annulation adaptative de bruit (ANC).....	34
II.3.1.1	Principe de base de l'annulation adaptative de bruit.....	35
II.3.1.2	Solution optimale de Wiener au problème d'annulation du bruit.....	36
II.3.2	Séparation aveugle de sources.....	38
II.3.2.1	Modèle de mélange convolutif d'une prise de son.....	43
II.3.2.2	Solution de la séparation de sources.....	45
II.4	Méthodes de débruitage multi-voies.....	46
II.4.1	Réduction de bruit par le réseau d'antennes de Frost.....	46
II.4.2	Structure de Griffiths et Jim.....	49

II.5	Conclusion.....	50
------	-----------------	----

Chapitre III

Amélioration des techniques de séparation aveugle de sources (Directe et Réursive)

III.1	Introduction.....	52
III.2	La structure Réursive.....	52
III.2.1	Limitation de la structure Réursive.....	55
III.3	La structure Directe.....	55
III.3.1	Avantage de la structure Directe.....	57
III.3.2	Inconvénient de la structure Directe	57
III.4	Contrôle des deux structures Directe et Réursive par un système de DAV.....	57
III.5	Amélioration des deux structures de séparation de sources (Directe et Réursive).....	58
III.5.1	Structure Réursive améliorée.....	58
III.5.1.1	Structure développée par Ikeda.....	59
III.5.1.2	Proposition d'une nouvelle structure (Propos1).....	66
III.5.2	Structure Directe améliorée.....	68
III.5.2.1	Structure développée par Djendi.....	68
III.5.2.2	La deuxième structure proposée (Propos2).....	70
III.6	Conclusion.....	72

Chapitre IV

Résultat des simulations

IV.1	Introduction.....	73
IV.2	Signaux utilisés en simulation.....	73
IV.3	Simulation de réponses impulsionnelles.....	75
IV.4	Détection d'activité vocale (DAV).....	78
IV.5	Critères objectifs de qualité.....	79
IV.5.1	Le critère du Rapport Signal à Bruit (RSB).....	79
IV.5.2	Le critère de la Distance Cepstrale (CD).....	79

IV.5.3	Le critère du Désajustement.....	80
IV.6	Simulation de la première structure proposée (Propos1).....	80
IV.7	Etude des performances de la première structure proposée (Propos1)	89
IV.7.1	Effet de la taille du filtre adaptatif	90
IV.7.2	Effet du pas d'adaptation.....	90
IV.7.3	Effet du RSB du mélange.....	91
IV.8	Comparaison de la structure Propos1 avec deux structures de séparation de sources Réursive.....	92
IV.9	Simulation de la deuxième structure proposée (Propos2).....	98
IV.10	Etude des performances de la deuxième structure proposée (Propos2).....	101
IV.10.1	Effet de la taille du filtre adaptatif	101
IV.10.2	Effet du pas d'adaptation.....	102
IV.10.3	Effet du RSB du mélange.....	103
IV.11	Comparaison de la structure Propos2 avec deux structures de séparation de sources Directe.....	104
IV.12	Conclusion.....	109
	Conclusion générale	110
	Bibliographie	112

Liste des tableaux

Tableau I.1	Les différentes grandeurs intervenant dans le filtre de Wiener	15
Tableau I.2	Complexité algorithmique de l'algorithme LMS	19
Tableau I.3	Complexité algorithmique de l'algorithme NLMS.....	21
Tableau I.4	Différentes classes du bruit.....	26
Tableau III.1	Paramètres utilisés pour la méthode d'Ikeda (Method1).....	66
Tableau III.2	Paramètres utilisés pour la structure Propos1.....	68
Tableau III.3	Paramètres utilisés pour la méthode de Djendi (Method2).....	70
Tableau III.4	Paramètres utilisés pour la structure Propos2.....	72
Tableau IV.1	Paramètres de simulation de la structure Propos1.....	81
Tableau IV.2	Paramètres de simulations des structures BBSS et Method1.....	93
Tableau IV.3	Paramètres de simulation de la structure Propos2.....	98
Tableau IV.4	Paramètres de simulations des structures FBSS et Method2.....	105

Liste des figures

Figure I.1	Les quatre types de signaux classés suivant leur morphologie (continue ou discrète).....	05
Figure I.2	Signal réel et signal échantillonné.....	06
Figure I.3	La représentation d'un système.....	08
Figure I.4	Filtre numérique.....	09
Figure I.5	Représentation schématique d'un filtre de Wiener.....	14
Figure I.6	Schéma général d'un système de filtrage adaptatif.....	21
Figure I.7	Appareil phonatoire.....	22
Figure I.8	Modèle simple de production de la parole.....	24
Figure I.9	Forme d'onde d'une trame de signal vocal voisé et son spectre d'amplitude.....	25
Figure I.10	Forme d'onde d'une trame de signal vocal non-voisé et son spectre d'amplitude.....	25
Figure II.1	Modèle de débruitage utilisé.....	29
Figure II.2	Structure d'annulation adaptative de bruit (ANC).....	35
Figure II.3	Schéma du Filtre de Wiener (ANC).....	36
Figure II.4	Structure générale du problème de la séparation aveugle de sources.....	39
Figure II.5	Schéma général de la relation sources-observations en présence de bruit.....	40
Figure II.6	Exemple d'un mélange instantané de sources.....	41
Figure II.7	Illustration d'un mélange convolutif acoustique à deux sources et deux capteurs.....	42
Figure II.8	Modèle général du mélange convolutif.....	44
Figure II.9	Modèle simplifier du mélange convolutif.....	44
Figure II.10	Structure symétrique Directe "Farward" de débruitage.....	45
Figure II.11	Structure symétrique Récursive "Backward" de débruitage.....	46
Figure II.12	Structure de FROST.....	47
Figure II.13	Structure de Griffiths et Lim.....	49

Figure III.1	Structure de séparation de sources Récursive (BBSS)	52
Figure III.2	Structure de séparation de sources Directe (FBSS)	55
Figure III.3	Schéma bloc du principe de la structure de séparation de sources classique.....	57
Figure III.4	Schéma bloc du principe de la structure de séparation de sources améliorée.....	58
Figure III.5	Structure développée par Ikeda (Method1).....	59
Figure III.6	Illustration de la variation des pas par rapport aux RSB en fonction de leurs seuils maximum et minimum sur chacune des voies.....	63
Figure III.7	Diagramme d'activités de la gestion des pas d'adaptations.....	65
Figure III.8	La première structure proposée (Propos1).....	67
Figure III.9	Structure développée par Djendi (Method2)	69
Figure III.10	La deuxième structure proposée (Propos2).....	71
Figure IV.1	Signal de parole utilisé, $F_e = 16$ KHz	74
Figure IV.2	Signal du bruit Blanc utilisé, $F_e = 16$ KHz	74
Figure IV.3	Modèle de mélange	75
Figure IV.4	Exemples d'une réponse impulsionnelle simulée ($h_{12}(n)$), cas des capteurs espacés.....	77
Figure IV.5	Exemples d'une réponse impulsionnelle simulée ($h_{21}(n)$), cas des capteurs espacés.....	77
Figure IV.6	Signal de parole utilisé avec une segmentation (DAV) manuelle.....	78
Figure IV.7	Signal original de parole $s(n)$ en haut avec son spectrogramme en bas.....	82
Figure IV.8	Signal de bruit Blanc $b(n)$ en haut avec son spectrogramme en bas.....	82
Figure IV.9	Signal de mélange $p_1(n)$ en haut avec son spectrogramme en bas.....	83
Figure IV.10	Signal de mélange $p_2(n)$ en haut avec son spectrogramme en bas.....	83
Figure IV.11	Signal de sortie $v_1(n)$ en haut avec son spectrogramme en bas de sources.....	84
Figure IV.12	Signal de sortie $v_2(n)$ en haut avec son spectrogramme en bas bruit.....	84
Figure IV.13	(a) Le RSB_1 estimé au niveau de l'entrée primaire (Propos1) et (b) le pas d'adaptation $\mu_{w_{21}}(n)$ du filtre principal $w_{21}(n)$	86
Figure IV.14	(c) Le RSB_2 estimé au niveau de l'entrée de référence (Propos1) et (d) le pas d'adaptation $\mu_{w_{12}}(n)$ du filtre principal $w_{12}(n)$	87
Figure IV.15	Comportement des pas d'adaptations: $\mu_{w_{12}}(n)$ (en bleu) et $\mu_{w_{21}}(n)$ (en rouge).....	88

Figure IV.16	Signal de parole utilisé avec les deux segmentations (manuelle et automatique (Propos1) $F_e = 16$ KHz.....	89
Figure IV.17	Effet de la taille du filtre sur la première structure proposée (Propos1) pour $\mu=0.4$, $RSB_1=3$ dB et $RSB_2=0$ dB.....	90
Figure IV.18	Effet du pas d'adaptation sur la première structure proposée (Propos1) pour $L=64$, $RSB_1=3$ dB et $RSB_2=0$ dB.....	91
Figure IV.19	Effet du RSB sur la structure proposée (Propos1) pour $L=64$, $\mu=0.4$ et $RSB_1=3$ dB.....	92
Figure IV.20	(a_1, a_2) Signal de parole originale $s(n)$, (b_1, b_2) signal de mélange $p_1(n)$, (c_1, c_2) signal de sortie $v_1(n)$ de la structure (Propos1), (d_1, d_2) signal de sortie de la structure (BBSS), (e_1, e_2) signal de sortie de la structure (Method1).....	94
Figure IV.21	Evolution du critère de la distance cepstrale des trois structures Propos1 (en rouge), Method1 (en bleu) et BBSS (en vert).....	95
Figure IV.22	Evolution du critère du rapport signal à bruit (RSB) pour les trois structures Propos1 (en rouge), Method1 (en bleu) et BBSS (en vert).....	96
Figure IV.23	Comparaison du désajustement obtenu avec les trois structures Propos1 (en rouge), Method1 (en bleu) et BBSS (en vert).....	97
Figure IV.24	(a_1, a_2) Signal de parole originale $s(n)$, (b_1, b_2) signal de mélange $p_1(n)$, (c_1, c_2) signal de sortie $u_1(n)$ de la structure (Propos2).....	99
Figure IV.25	(a) Le RSB_1 estimé au niveau de l'entrée primaire (Propos2) et (b) le pas d'adaptation $\mu_{w_{21}}(n)$ du filtre principal $w_{21}(n)$	100
Figure IV.26	Signal de parole utilisé avec les deux segmentations (manuelle et automatique (Propos2)), $F_e = 16$ KHz.....	101
Figure IV.27	Effet de la taille du filtre sur la structure proposée (Propos2) pour $\mu=0.4$, $RSB_1=3$ dB et $RSB_2=0$ dB.....	102
Figure IV.28	Effet du pas d'adaptation sur la structure proposée (Propos2) pour $L=64$, $RSB_1=3$ dB et $RSB_2=0$ dB.....	103
Figure IV.29	Effet du RSB sur la structure proposée (Propos2) pour $L=64$, $\mu=0.4$ et $RSB_1=3$ dB.....	104
Figure IV.30	(a_1, a_2) Signal de parole originale $s(n)$, (b_1, b_2) signal de mélange $p_1(n)$, (c_1, c_2) signal de sortie $u_1(n)$ de la structure Propos2, (d_1, d_2) signal de sortie de la structure FBSS, (e_1, e_2) signal de sortie de la structure Method2.....	106

Figure IV.31 Evolution du critère de la distance cepstrale des trois structures Propos2 (en rouge), Method2 (en bleu) et FBSS (en vert).....	107
Figure IV.32 Evolution du critère du rapport signal à bruit (RSB) pour les trois structures Propos2 (en rouge), Method2 (en bleu) et FBSS (en vert).....	108
Figure IV.33 Comparaison du désajustement obtenu avec les trois structures Propos2 (en rouge), Method2 (en bleu) et FBSS (en vert).....	108

Liste d'abréviation

- ANC : Adaptive Noise Canceller.
- AR : Auto-Régressif.
- ARMA : Auto-Régressif, Moyenne Ajustée.
- BSS : Blind Source Separation.
- BBSS : Backward Blind Source Separation.
- DAV: Detection d'Activité Vocale.
- DC : Distance Cepstrale.
- DNLMs : Double Normalized Least Mean Square.
- DSP : Densité Spectrale de Puissance.
- EQM : Erreur Quadratique Moyenne.
- EQMM : Erreur Quadratique Moyenne Minimum.
- FBSS : Farward Blind Source Separation.
- LMS : Least-Mean Square.
- MA : Moyenne Ajustée.
- MMSE : Minimum Mean Square Erreur.
- NLMS : Normalized Least-Mean Square.
- RIF : Réponse Impulsionnelle Finie.
- RII : Réponse Impulsionnelle Infinie.
- RSB : Rapport Signal à Bruit.
- RSB_{post} : rapport signal à bruit a posteriori.
- RSB_{prio} : rapport signal à bruit a priori.
- 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.

Introduction générale

Les évolutions technologiques dans le monde ne cessent de s'accroître à haute cadence, notamment pour les systèmes de télécommunications mobiles.

A ses premiers temps, la téléphonie mobile a dû son essor extraordinaire au seul fait d'apporter une liberté nouvelle. La possibilité de communiquer à distance, avec qui l'on veut et quel que soit l'endroit où l'on se trouve a rendu ce système un objet indispensable dans notre vie quotidienne, 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. L'utilisation du téléphone mobile rend certes le téléphone « classique » fixe obsolète mais fait apparaître de nouveaux problèmes notamment dans l'environnement véhicule. En effet, outre le fait d'immobiliser une main du conducteur, réduit dangereusement la concentration du conducteur ce qui altère fortement la sécurité de la conduite.

Plusieurs pays dans le monde ont alors émis des recommandations prohibant l'utilisation d'un combiné téléphonique dans le véhicule en déplacement : « *Téléphoner ou conduire, il faut choisir* ». A partir de ce constat, le mains-libres s'impose de lui-même.

L'intérêt majeur du téléphone mains-libres à bord des véhicules est d'éviter la distraction du conducteur lors de la conduite et d'assurer sa sécurité. C'est aussi une façon d'offrir plus de confort au conducteur.

De nouvelles complications ont surgi dans le cas d'utilisation du système mains-libres. Ces systèmes sont perturbés par la présence du bruit ambiant à l'intérieur du véhicule (moteur, roues, radio, conversations...).

La gêne occasionnée par la perturbation qualifiée de bruit est généralement source d'inconfort et de fatigue pour le locuteur distant. De plus, la volonté de dématérialiser la prise de son (système mains-libres) va favoriser l'émergence du bruit du fait de l'augmentation de la distance entre la bouche et le microphone. Dans ces conditions, la conversation téléphonique peut s'avérer rapidement fastidieuse ce qui justifie le besoin d'un traitement à même de réduire la gêne des utilisateurs.

La téléphonie mains-libres doit donc inclure une étape de réduction de bruit pour assurer une communication de bonne qualité. Ce sujet 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. Dans le cadre de ce mémoire, nous portant notre effort principal sur l'amélioration des deux structures de séparation de sources classiques.

Nous présentons ci-après un aperçu des différents chapitres de ce mémoire :

- ❖ **Le chapitre I** présente en premier lieu, les principaux outils mathématiques du traitement de signal. En deuxième lieu, nous présentons des généralités sur le signal de parole et du bruit.
- ❖ **Le chapitre II** présente un état de l'art des techniques mono-capteur, bi-capteurs et 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 chapitre III** présente, en premier lieu, les structures de séparation de source classiques Directe et Récursive ainsi qu'une analyse de leurs solutions. En deuxième lieu, nous présentons deux nouvelles structures que nous proposons dans le cadre de ce mémoire. Ces structures correspondent à des versions améliorées des structures classiques.
- ❖ **Le chapitre IV** est réservé pour la validation expérimentale des différents résultats théoriques du chapitre III.

Nous terminerons notre document de mémoire par une conclusion générale.

I.1 Introduction

Le traitement du signal est une discipline indispensable de nos jours, il a pour objet l'élaboration ou l'interprétation des signaux porteurs d'informations. Son but est donc de réussir à extraire un maximum d'information utile d'un signal perturbé par du bruit en s'appuyant sur les ressources de l'électronique, de l'informatique et de la physique appliquée. Son champ d'application se situe donc dans tous les domaines concernés par la perception, la transmission ou l'exploitation des informations véhiculées par ces signaux.

Pour ce chapitre, nous allons présenter dans un premier temps, quelques définitions de base ainsi que les outils les plus utilisés dans le domaine du traitement de signal.

Dans un deuxième temps, nous présenterons les principales caractéristiques de la parole et du bruit.

I.2 Généralités sur le traitement du signal

I.2.1 Notion de signal

Un signal est une représentation physique d'une information, qui sera transporté d'une source à une destination. Il s'agit donc d'une notion tout à fait générale que l'on peut rencontrer dans des domaines aussi variés que l'électricité, l'électronique, l'acoustique, l'optique, la mécanique, la biologie, l'économie, ...etc [1].

I.2.2 Classification des signaux

On peut envisager plusieurs modes de classification pour les signaux suivant leurs propriétés [1].

I.2.2.1 Classification phénoménologique

On considère la nature de l'évolution du signal en fonction du temps. Il apparaît deux types de signaux :

- **Les signaux déterministes** : ou signaux certains, leur évolution en fonction du temps peut être parfaitement modélisée par une fonction mathématique. On retrouve dans

cette classe les signaux périodiques, les signaux transitoires, les signaux pseudo-aléatoires, etc....

- **Les signaux aléatoires** : leur comportement temporel est imprévisible. Il faut faire appel à leurs propriétés statistiques pour les décrire. Si leurs propriétés statistiques sont invariantes dans le temps, on dit qu'ils sont stationnaires.

I.2.2.2 Classification énergétique

Les signaux peuvent être à énergie finie ou à puissance moyenne finie :

- **Les signaux à énergie finie** : Un signal $x(t)$ est à énergie finie si l'intégrale suivante existe,

$$E_x = \int_{-\infty}^{+\infty} |x(t)|^2 dt \quad (\text{I.1})$$

c'est-à-dire si :

$$\int_{-\infty}^{+\infty} |x(t)|^2 dt < +\infty \quad (\text{I.2})$$

Les signaux à énergie finie sont aussi appelés signaux de carré sommable ou de carré intégrable.

- **Les signaux à puissance moyenne finie** : un signal $x(t)$, définie sur \mathcal{R} , est à puissance moyenne finie sur cet intervalle si :

$$0 < P_x = \lim_{T \rightarrow \infty} \left[\frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} |x(t)|^2 dt \right] < +\infty \quad (\text{I.3})$$

La définition exclut le cas de signaux à puissance moyenne nulle, qui correspond à des signaux à énergie finie.

I.2.2.3 Classification morphologique

On distingue les signaux à variable continue des signaux à variable discrète ainsi que ceux dont l'amplitude est discrète ou continue (voir figure I.1).

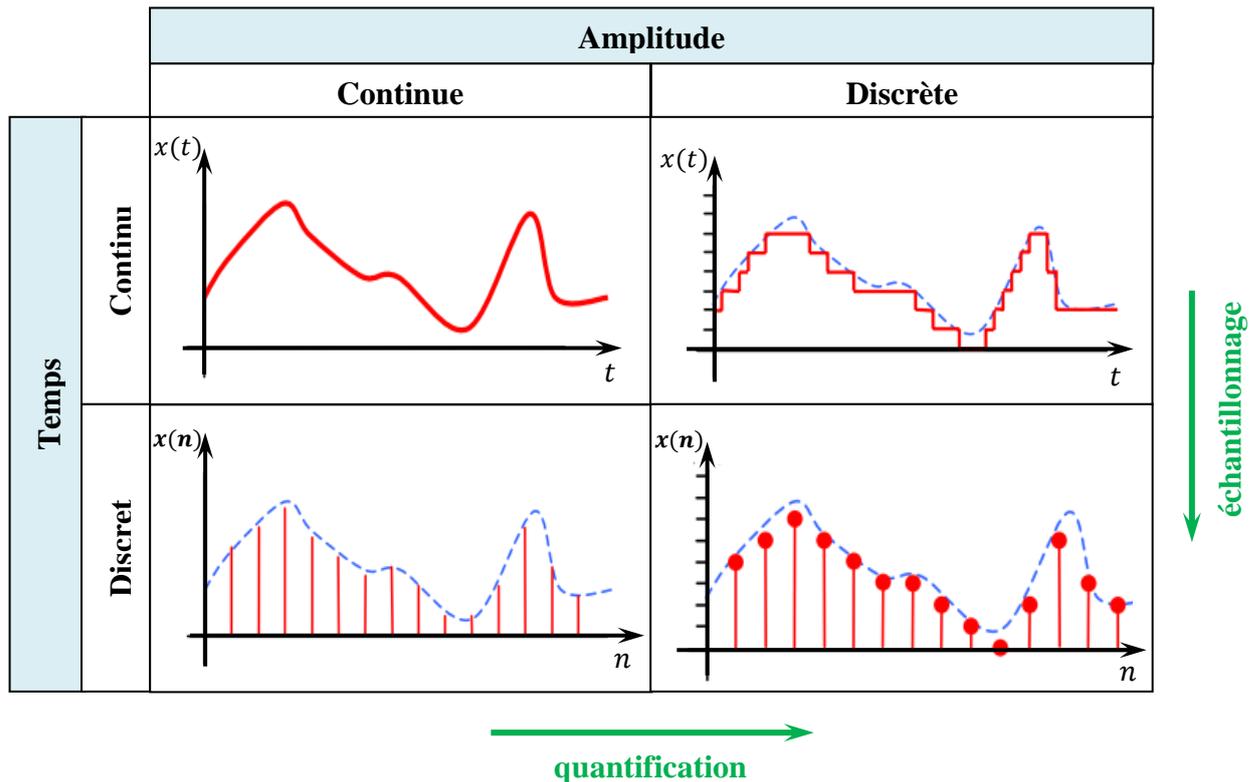


Figure I.1 : Les quatre types de signaux classés suivant leur morphologie (continue ou discrète).

On obtient donc 4 classes de signaux :

- Les **signaux analogiques** dont l'amplitude et le temps sont continus.
- Les **signaux quantifiés** dont l'amplitude est discrète et le temps continu.
- Les **signaux échantillonnés** dont l'amplitude est continue et le temps discret.
- Les **signaux numériques** dont l'amplitude et le temps sont discrets.

I.2.3 Numérisation

L'importance des systèmes numériques de traitement de l'information ne cesse de croître (radio, télévision, téléphone ...). Ce choix est souvent justifié par des avantages techniques tels que la grande stabilité des paramètres, une excellente reproductibilité des résultats et des fonctionnalités accrues. Le monde extérieur étant par nature "analogique", une opération préliminaire de conversion analogique numérique est nécessaire. La conversion analogique numérique est la succession de trois effets sur le signal analogique de départ [2]:

- l'**échantillonnage** pour rendre le signal discret.
- la **quantification** pour associer à chaque échantillon une valeur.
- le **codage** pour associer un code à chaque valeur.

I.2.3.1 Echantillonnage

Les signaux primaires porteurs d'information sont pratiquement toujours de type analogique (amplitude et temps continus). Un système d'acquisition ou tout autre système numérique traite des données, c'est-à-dire des suites de nombres. Il faut donc représenter le signal par une suite de valeurs ponctuelles : c'est l'échantillonnage [2].

L'échantillonnage consiste à prélever à des instants précis, le plus souvent équidistants, les valeurs instantanées d'un signal. Le signal analogique $x(t)$, continu dans le temps, est alors représenté par un ensemble de valeur discrètes : $x_e(t) = x(n T_e)$. Cette opération est réalisée par un échantillonneur souvent symbolisé par un interrupteur [3].

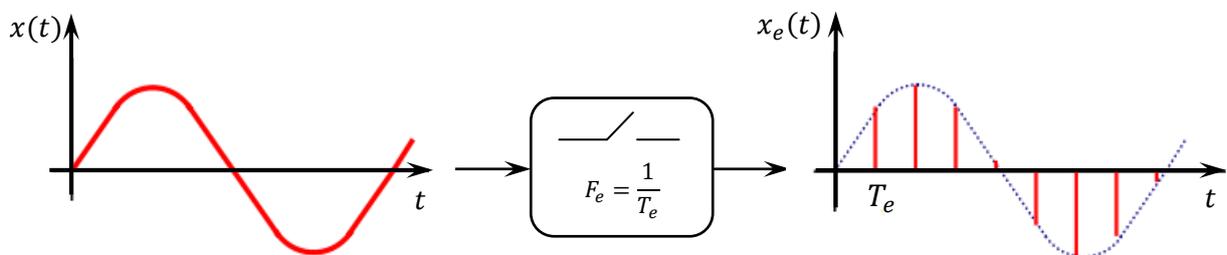


Figure I.2 : Signal réel et signal échantillonné.

L'intervalle T_e entre deux échantillons successifs est appelé pas d'échantillonnage et $F_e = \frac{1}{T_e}$ fréquence d'échantillonnage.

L'échantillonnage idéal est modélisé par la multiplication du signal continu $x(t)$ par peigne de Dirac de période T_e . Le signal échantillonné correspond à :

$$x_e(t) = x(t) \cdot \delta_{T_e}(t) = x(t) \cdot \sum_{n=-\infty}^{\infty} \delta(t - nT_e) \quad (\text{I.4})$$

En utilisant la propriété $x(t) \cdot \delta(t - t_0) = x(t_0) \cdot \delta(t - t_0)$, on obtient :

$$x_e(t) = \sum_{n=-\infty}^{\infty} x(nT_e) \cdot \delta(t - nT_e) \quad (\text{I.5})$$

La condition nécessaire et suffisante pour échantillonner un signal sans perte d'information est que la fréquence d'échantillonnage F_e soit supérieure ou égale au double de la fréquence maximale du signal F_{max} c-à-d $F_e \geq 2F_{max}$. La fréquence $\frac{F_e}{2}$ est appelée **fréquence de Nyquist** et la condition d'échantillonnage est appelée **critère de Shannon**.

I.2.3.2 Quantification

En traitement des signaux, la quantification est le procédé qui permet d'approcher un signal continu (ou à valeurs dans un ensemble discret de grande taille) par les valeurs d'un ensemble discret d'assez petite taille [3].

I.2.3.3 Codage

Le codage consiste à associer à un ensemble de valeurs discrètes un code composé d'éléments binaires. Les codes les plus connus : code binaire naturel, code binaire décalé, code complément à 2, code DCB, code Gray.

I.2.4 Notion de système

Un système est un ensemble d'éléments fonctionnels interagissant entre eux et qui établit un lien de cause à effet entre ses signaux d'entrée et ses signaux de sortie. La représentation graphique d'un tel système est donnée par la figure suivante :

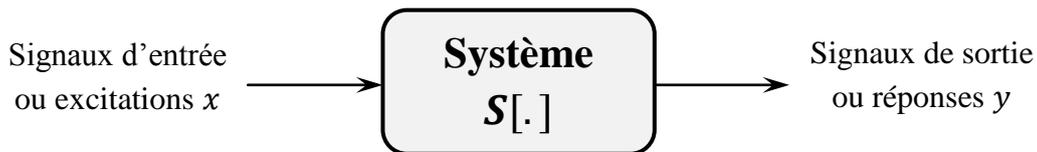


Figure I.3 : La représentation d'un système.

Par exemple l'oreille humaine est un système qui transforme une variation de pression (acoustique) en un signal électrique (sur le nerf auditif). Un microphone est un système analogue à l'oreille, de même que les cordes vocales et les enceintes acoustique.

I.2.5 Convolution

La convolution est une méthode pour combiner deux signaux et en produire un troisième, c'est la technique la plus importante en traitement de signaux. Le produit de convolution $z = x * y$ de deux signaux discrets $x(n)$ et $y(n)$ est défini par :

$$z(n) = \sum_{k=-\infty}^{+\infty} x(k).y(n - k) \quad (\text{I.6})$$

I.2.6 Corrélation

L'opération de corrélation permet d'exprimer la ressemblance entre deux signaux $x(n)$ et $y(n)$ au niveau de la forme et de la position en fonction d'un paramètre de translation. La fonction de corrélation entre deux signaux $x(n)$ et $y(n)$ différents, notée $C_{xy}(n)$, est

appelée fonction d'**intercorrélacion** (ou corrélation croisée ou corrélation mutuelle) et elle est définie par [4] :

$$C_{xy}(k) = \sum_{k=-\infty}^{+\infty} x(n) \cdot y^*(n - k) \quad (\text{I.7})$$

On appelle **autocorrélacion** d'un signal, son intercorrélacion avec lui-même [4] :

$$C_{xx}(k) = \sum_{k=-\infty}^{+\infty} x(n) \cdot x^*(n - k) \quad (\text{I.8})$$

I.2.7 Filtrage numérique

I.2.7.1 Définition

On appelle filtre numérique un système utilisé pour modifier la distribution fréquentielle d'un signal numérique d'entrée en le transformant en un signal numérique désiré en sortie. Avec le progrès en informatique les filtres numériques sont caractérisés par leurs: précision, fiabilité, stabilité, adaptabilité et facilité de commande [1].

La figure (I.4) est une représentation d'un filtre numérique à une entrée et une sortie.

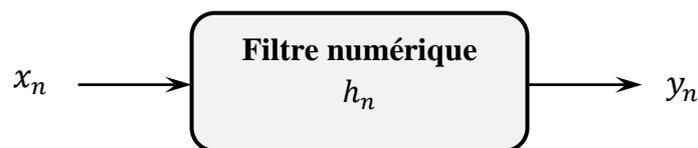


Figure I.4 : Filtre numérique.

Ce filtre est un système discret linéaire invariant dans le temps, il relie la sortie $y(nT_e) = y_n$ à l'entrée $x(nT_e) = x_n$ chaque instant nT_e , T_e étant la période d'échantillonnage.

➤ **Linéarité**

Soit deux suites $x_1(n)$ et $x_2(n)$ avec les sorties correspondantes $y_1(n)$ et $y_2(n)$.

Dire que le système est linéaire signifie que :

$$\begin{cases} \forall \lambda_1 \in \mathcal{R} \\ \forall \lambda_2 \in \mathcal{R} \end{cases} \Rightarrow \lambda_1 x_1(n) + \lambda_2 x_2(n) \rightarrow \lambda_1 y_1(n) + \lambda_2 y_2(n) \quad (\text{I.9})$$

➤ **Invariance dans le temps**

Soit la suite $x(n)$ et la sortie correspondante $y(n)$, dire que le système est invariant dans le temps signifie qu'à la suite $x(n - n_0)$ correspond la sortie $y(n - n_0)$, est ceci quelque soit n_0 .

La sortie d'un filtre numérique est le produit de convolution de l'entrée par sa réponse impulsionnelle h_n .

$$y_n = h_n * x_n$$

$$y_n = \sum_{k=-\infty}^{+\infty} h_k \cdot x_{(n-k)} = \sum_{k=-\infty}^{+\infty} x_k \cdot h_{(n-k)} \quad (\text{I.10})$$

I.2.7.2 Transformée en z

L'utilisation de la transformée en z permet de caractériser un filtre numérique par sa fonction de transfert qui est, par définition, la transformée en z de sa réponse impulsionnelle. Elle a pour les systèmes échantillonnés ou numériques le même rôle que la fonction de transfert (transformée de Laplace) pour les systèmes continus et permet une analyse temporelle et/ou fréquentielle des filtres numériques [5].

La transformée en z d'une suite numérique x_n est définie par :

$$X(z) = \sum_{n=-\infty}^{+\infty} x_n \cdot z^{-n}, \quad z \in \mathbb{C} \quad (\text{I.11})$$

Pour un signal donné, l'ensemble des valeurs de z pour lesquelles la série converge est la *région de convergence* (RDC). Pour la déterminer, on peut utiliser le critère de Cauchy qui permet d'étudier l'existence de la transformée en z [6] :

$$|\sum_{n=0}^{+\infty} x_n| < \sum_{n=0}^{+\infty} |x_n| < +\infty \quad \text{si} \quad \lim_{n \rightarrow \infty} |x_n|^{\frac{1}{n}} < 1 \quad (\text{I.12})$$

Séquences causales :

Une séquence x_n est dite causale si elle est nulle pour $n < 0$. Pour une séquence causale, la transformée en z est monolatérale :

$$X(z) = \sum_{n=0}^{+\infty} x_n \cdot z^{-n} \quad (\text{I.13})$$

$X(z)$ existe si :

$$\lim_{n \rightarrow \infty} |x_n \cdot z^{-n}|^{\frac{1}{n}} < 1 \quad \text{c-à-d} \quad \lim_{n \rightarrow \infty} |x_n|^{\frac{1}{n}} |z|^{-1} < 1 \quad \text{soit} \quad |z| > R_+ = \lim_{n \rightarrow \infty} |x_n|^{\frac{1}{n}}$$

La transformée en z d'une suite causale est donc définie à l'extérieur d'un cercle de rayon R_+ .

Séquences anticausales :

Une séquence x_n est dite anticausale si elle est nulle pour $n \geq 0$. Pour une séquence anticausale, la transformée en z s'écrit:

$$X(z) = \sum_{n=-\infty}^{-1} x_n \cdot z^{-n} = \sum_{n=1}^{+\infty} x_{-n} \cdot z^n \quad (\text{I.14})$$

$X(z)$ existe si :

$$\lim_{n \rightarrow \infty} |x_{-n} \cdot z^n|^{\frac{1}{n}} < 1 \quad \text{c-à-d} \quad \lim_{n \rightarrow \infty} |x_{-n}|^{\frac{1}{n}} |z| < 1 \quad \text{soit} \quad |z| < R_- = \lim_{n \rightarrow \infty} |x_{-n}|^{-\frac{1}{n}}$$

La transformée en z d'une suite anticausale est donc définie à l'intérieur d'un cercle de rayon R_- .

Suite quelconque :

une suite quelconque x_n est la somme d'une suite causale x_n^+ et d'une suite anticausale x_n^- :

$$x_n = x_n^+ + x_n^- \quad (\text{I.15})$$

avec :

$$\begin{array}{l} x_n^+ = x_n \text{ si } n \geq 0 \\ x_n^+ = 0 \text{ si } n < 0 \end{array} \quad \text{et} \quad \begin{array}{l} x_n^- = x_n \text{ si } n < 0 \\ x_n^- = 0 \text{ si } n \geq 0 \end{array}$$

La transformée en z d'une suite quelconque est donc définie dans une couronne : $R_+ < |z| < R_-$, quand cette couronne existe.

I.2.7.3 Causalité et stabilité

❖ **Causalité**

Un filtre numérique est dit causal, si sa réponse impulsionnelle h_n est nulle pour $n < 0$. Sa transformé en z converge alors à l'extérieur d'un cercle.

Un filtre numérique est dit anticausal, si sa réponse impulsionnelle h_n est nulle pour $n \geq 0$. Sa transformé en z converge alors à l'intérieur d'un cercle [6].

❖ **Stabilité**

On dit qu'un filtre numérique est stable, si à toute entrée bornée x_n correspond une sortie y_n bornée. Pour toute suite (x_n) bornée [6].

$$\exists M \quad \forall n |x_n| < M \quad \rightarrow \quad \exists M' \quad \forall n |y_n| < M' \quad (\text{I.16})$$

Pour un filtre numérique linéaire et invariant, cela se traduit sur sa réponse impulsionnelle par:

$$\sum_{n=-\infty}^{+\infty} |h_n| < +\infty \quad (\text{I.17})$$

I.2.7.4 Classification des filtres numériques

Les filtres numériques se divisent en deux classes selon la durée de la réponse impulsionnelle.

❖ Filtres à réponse impulsionnelle finie (RIF) [MA]

Ces filtres sont caractérisés par des réponses impulsionnelles de durée finie. Ils sont appelés aussi filtres à moyenne ajustée (MA). La réponse y_n du filtre s'écrit [7] :

$$y_n = \sum_{i=0}^{N-1} b_i x_{n-i} = \sum_{i=0}^{N-1} h_i x_{n-i} \quad (\text{I.18})$$

Le filtre RIF est toujours stable. Sa fonction de transfert s'écrit :

$$H(z) = \sum_{i=0}^{N-1} b_i z^{-i} \quad (\text{I.19})$$

➤ Principales caractéristiques des filtres RIF :

- Une bande de transition plus large.
- Des méthodes de synthèse efficaces.
- Une stabilité inhérente ($\sum_{n=0}^{N-1} |h_n| < \infty$).
- Une meilleure stabilité numérique que les RII.
- Une phase qui peut être exactement linéaire.

❖ Filtres à réponse impulsionnelle infinie (RII) [ARMA]

Dans ce cas la réponse impulsionnelle est illimitée et la réponse $y(n)$ s'écrit de manière générale :

$$y_n = \sum_{i=0}^N b_i x_{n-i} - \sum_{i=1}^L a_i y_{n-i} \quad (\text{I.20})$$

Sa fonction de transfert s'écrit :

$$H(z) = \frac{N(z)}{D(z)} = \frac{\sum_{i=0}^N b_i z^{-i}}{1 + \sum_{i=1}^L a_i z^{-i}} \quad (\text{I.21})$$

- Principales caractéristiques des filtres RII :
 - Une bande de transition qui peut être étroite ;
 - Une instabilité potentielle due à des pôles situés en dehors du cercle unité (i.e. $\exists i, |p_i| \geq 1$).
 - Une instabilité numérique (i.e. après quantification des coefficients et du signal).
 - Si $\begin{cases} b_i = 0 & i \neq 0 \\ b_0 = 1 \end{cases} \Rightarrow$ ce modèle est appelé filtre auto-régressif (AR).

I.2.8 Filtrage de Wiener

La figure (I.5) 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 M étant l'ordre 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) [8].

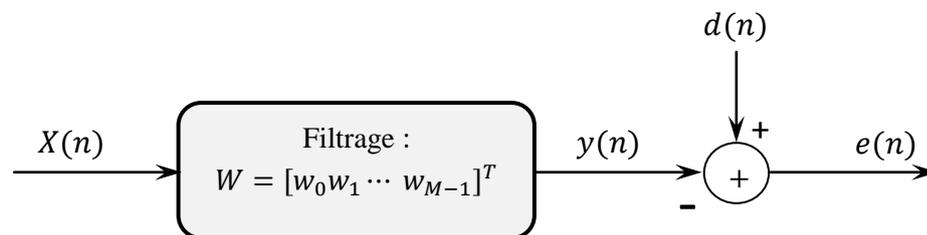


Figure I.5 : Représentation schématique d'un filtre de Wiener.

Les différentes grandeurs utilisées dans le filtre de Wiener sont représentées sur le tableau I.1.

Tableau I.1 : Les différentes grandeurs intervenant dans le filtre de Wiener.

Grandeur	Filtre de Wiener
Entrée du filtre	$X(n) = \begin{bmatrix} x(n) \\ x(n-1) \\ \vdots \\ x(n-M+1) \end{bmatrix}$
Coefficients du filtre	$W = \begin{bmatrix} w_0 \\ w_1 \\ \vdots \\ w_{M-1} \end{bmatrix}$
Sortie du filtre	$y(n) = \sum_{k=0}^{M-1} w_k x(n-k) = W^T X(n)$
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.

$$J = E[e(n)^2] \tag{I.22}$$

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

$$e(n) = d(n) - y(n) = d(n) - W^T X(n) = d(n) - X(n)^T W \tag{I.23}$$

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

En faisant l'hypothèse que le signal $\{x\}$ est stationnaire, et si on introduit l'équation (I.23) dans l'équation(I.22), on peut écrire la fonction coût de la manière suivante [9] :

$$\begin{aligned}
 J &= E[(d(n) - W^T X(n))^2] \\
 J &= E[d(n)^2 - 2 W^T X(n)d(n) + W^T X(n)X(n)^T W] \\
 J &= E[d(n)^2] - 2 W^T E[X(n)d(n)] + W^T E[X(n)X(n)^T] W \\
 J &= E[d(n)^2] - 2 W^T P_{xd} + W^T R_{xx} W
 \end{aligned} \tag{I.24}$$

où R est une matrice d'autocorrélation du signal d'entrée $X(n)$ de taille $(M \times M)$, elle est symétrique et définie positive.

$$R_{xx} = E[X(n)X(n)^T] \tag{I.25}$$

et P est un vecteur d'intercorrélation entre le signal d'entrée $X(n)$ et le signal désiré $d(n)$ de taille M définie par :

$$P_{xd} = E[X(n)d(n)] \tag{I.26}$$

Les coefficients du filtre optimal (W^{opt}) sont obtenus en annulant le gradient de (I.22) par rapport aux coefficients du filtre [8].

$$\begin{aligned}
 \nabla J &= \frac{\partial J}{\partial W} \\
 \nabla J &= 2E[e(n) \frac{\partial e(n)}{\partial W}] \\
 \nabla J &= -2E[e(n)X(n)]
 \end{aligned} \tag{I.27}$$

où $E[.]$ et ∇ 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

$$E[e_{min}(n)X(n)] = 0 \quad (I.28)$$

C'est le *principe d'orthogonalité* signifiant que toutes les entrées $X(n)$ sont décorréliées de l'erreur $e_{min}(n)$ [10].

En développant cette dernière équation on obtient:

$$E[X(n)X(n)^T]W^{opt} = E[X(n) d(n)] \quad (I.29)$$

qu'on peut réécrire sous la forme :

$$R_{xx}W^{opt} = P_{xd} \quad (I.30)$$

L'expression (I.30) permet d'aboutir au filtre optimal de Wiener W^{opt} donné par (I.31) plus connue sous le nom représentation matricielle de l'équation de Wiener-Hopf [11],

$$W^{opt} = R_{xx}^{-1}P_{xd} \quad (I.31)$$

avec W^{opt} étant le vecteur optimum des coefficients.

Une première approche pour résoudre l'équation de Wiener-Hopf (I.31) 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 W [8].

La méthode du gradient consiste à [10] :

- Choisir un vecteur initial de coefficients $W(0)$.

- Obtenir, à partir d'un vecteur $W(n)$ donné, un vecteur $W(n + 1)$ par incrémentation de $W(n)$ dans la direction opposée du gradient de la fonction coût J .

$$W(n + 1) = W(n) - \frac{1}{2}\mu \nabla J|_{W=W(n)} \quad (\text{I.32})$$

où μ 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 :

$$\nabla J|_{W=W(n)} = 2R_{xx} W(n) - 2P_{xd} \quad (\text{I.33})$$

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

$$W(n + 1) = W(n) + \mu (P_{xd} - R_{xx} W(n)) \quad (\text{I.34})$$

I.2.9 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 , $\nabla J|_{W=W(n)} = 2R_{xx} W(n) - 2P_{xd}$, obtenue en remplaçant les estimations des coefficients de R_{xx} et P_{xd} fait à l'aide de moyennes empiriques par une estimation instantanée [12] :

$$\hat{R}_{xx} = X(n)X(n)^T \quad (\text{I.35})$$

$$\hat{P}_{xd} = X(n) d(n) \quad (\text{I.36})$$

Remplaçant (I.35) et (I.36) dans l'expression du gradient déterministe qui est donnée par (I.34), on obtient l'algorithme du gradient stochastique :

$$W(n + 1) = W(n) + \mu e(n)X(n) \quad (\text{I.37})$$

où $e(n)$ est donnée par la relation (I.23).

D'après [12], La condition nécessaire et suffisante de convergence en moyenne de l'algorithme (LMS) est :

$$0 < \mu < \frac{2}{\lambda_{max}} \quad (\text{I.38})$$

où λ_{max} est la plus grande valeur propre de la matrice d'autocorrélation R_{xx} donnée par (I.25).

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

$$0 < \mu \leq \frac{1}{\text{Trace}(R_{xx})} = \frac{1}{M\sigma_x^2} \quad (\text{I.39})$$

où $\text{Trace}(R_{xx})$ représente la somme des éléments de la diagonale de la matrice R_{xx} et σ_x^2 désigne l'énergie du signal d'entrée $X(n)$.

La complexité algorithmique de l'algorithme LMS est connue : chaque itération n comporte $(2M + 1)$ multiplications et $(2M)$ additions. Le tableau (I.2) résume les opérations nécessaires pour l'algorithme LMS à chaque itération n [9].

Tableau I.2 : Complexité algorithmique de l'algorithme LMS.

Etape de l'algorithme	Nombre d'additions	Nombre de multiplications
$e(n) = d(n) - X(n)^T W$	M	M
$W(n + 1) = W(n) + \mu e(n)X(n)$	M	$M + 1$
Total par itération	$2M$	$2M + 1$

I.2.10 Algorithme du gradient stochastique normalisé NLMS

L'algorithme NLMS (Normalized Least Mean Square) consiste à normaliser le pas d'adaptation μ 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 (I.37), le pas d'adaptation μ est alors remplacé par un pas d'adaptation μ_n défini à chaque itération par [9]:

$$\mu_n = \frac{\mu}{X(n)X(n)^T} \quad (\text{I.40})$$

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

$$W(n+1) = W(n) + \frac{\mu}{X(n)X(n)^T + \beta} e(n)X(n) \quad (\text{I.41})$$

où β 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 (I.3) résume le nombre d'opérations nécessaires dans l'algorithme NLMS à chaque itération n [9].

Tableau I.3 : Complexité algorithmique de l'algorithme NLMS.

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

I.2.11 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 [13]. Le rôle primordial d'un filtre adaptatif est d'ajuster un vecteur de paramètre W pour un objectif bien défini (minimisation de l'EQM : erreur quadratique moyenne). La figure (I.6) schématise le principe d'un filtre adaptatif bouclé par un algorithme d'adaptation.

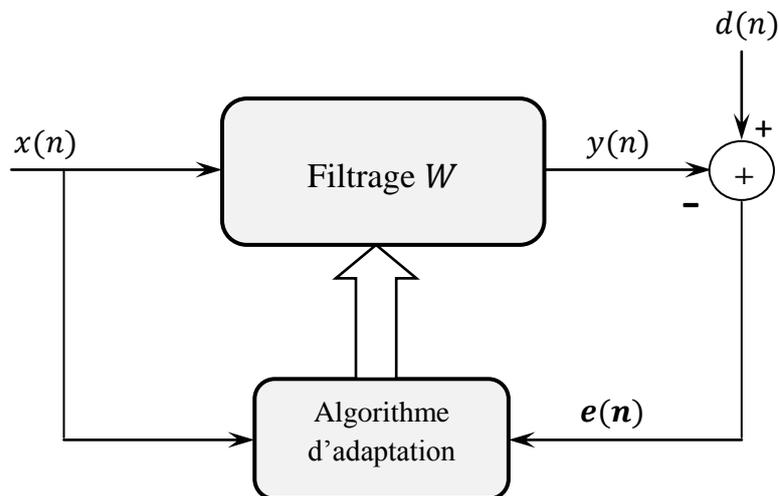


Figure I.6 : Schéma général d'un système de filtrage adaptatif.

La technique de filtrage adaptatif se décompose classiquement en deux étapes [9] :

1. Une étape de filtrage qui permet d'obtenir une estimation du signal inconnu en convoluant le signal d'entrée $x(n)$ avec les coefficients du filtre adaptatif 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 W suivant un algorithme donné.

I.3 Généralités sur la parole et le bruit

I.3.1 Qu'est ce que c'est 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 [14].

I.3.2 Production du signal de parole

La parole est produite par le système articulatoire, présenté à la figure (I.7) :

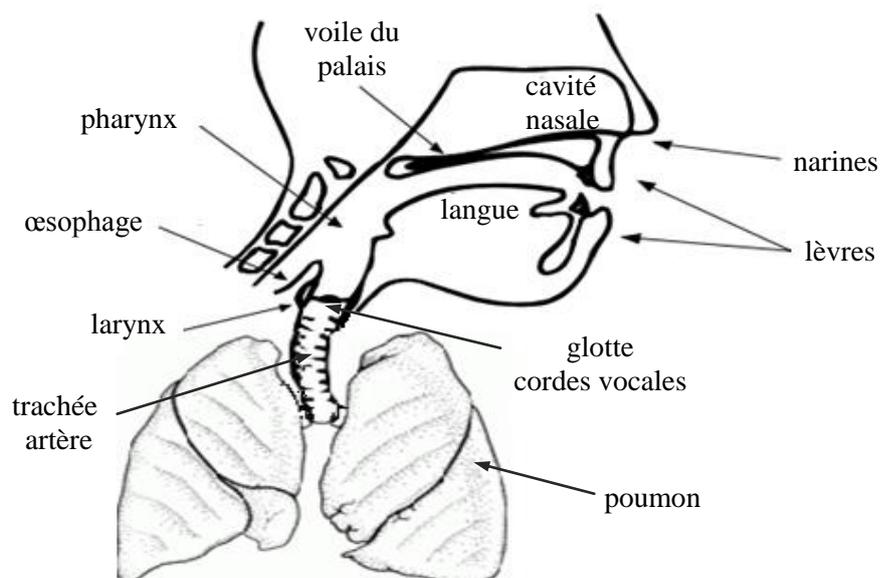


Figure I.7 : Appareil phonatoire.

La parole peut être décrite comme le résultat de l'action volontaire et coordonnée des appareils respiratoire et masticatoire. Cette action se déroule sous le contrôle du système nerveux central qui reçoit en permanence des informations par rétroaction auditive et par les sensations cénesthésiques [14].

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.

L'intensité du son émis est liée à la pression de l'air en amont du larynx, sa hauteur est fixée par la fréquence de vibration des cordes vocales, appelée *fréquence fondamentale* F_0 (ou *pitch*). 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 450 Hz pour une voix féminine,
- de 200 à 600 Hz pour une voix d'enfant.

Le signal de parole est 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 I.8). 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 (du moins pas dans tout le conduit vocal) sans provoquer de vibration des cordes vocales [15].

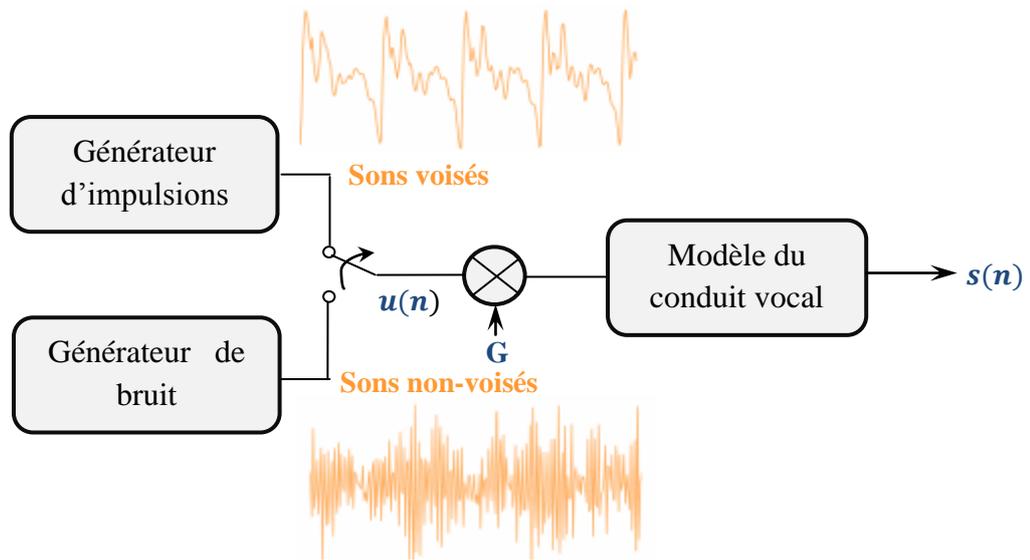


Figure I.8 : Modèle simple de production de la parole.

I.3.3 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 [16] 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 :

I.3.3.1 Les sons voisés

Un son voisé est par définition un signal quasi-périodique qui possède un spectre fréquentiel très caractéristique, comme on peut le voir sur l'exemple de la figure (I.9). La première raie de ce spectre correspond au fondamental (F_0) et les raies suivantes à ses harmoniques (multiples de F_0). L'enveloppe de ces raies possède des maxima locaux appelés formants (localisés en F_i avec $i = 1, \dots, 4$) [17].

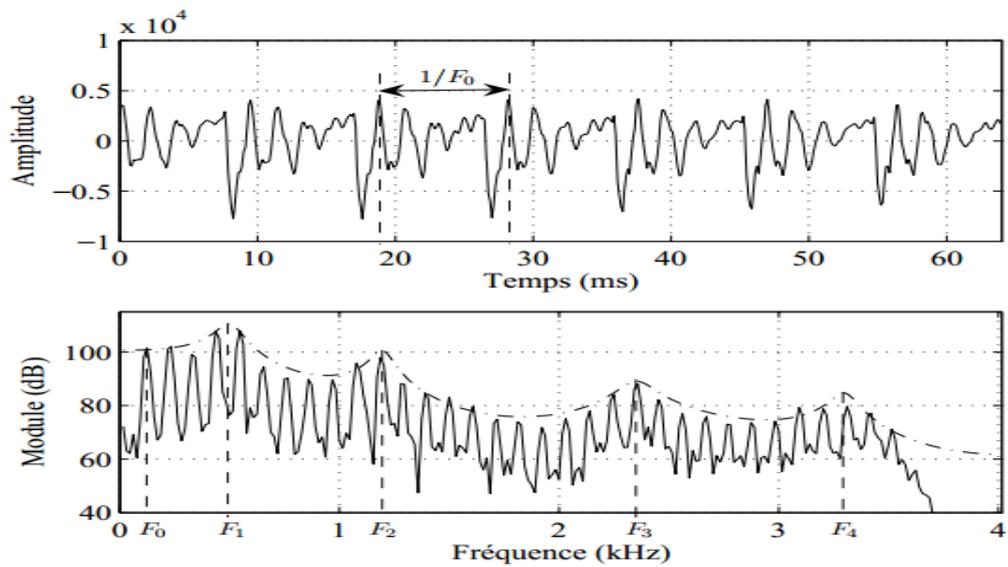


Figure I.9 : Forme d'onde d'une trame de signal vocal voisé et son spectre d'amplitude.

I.3.3.2 Les sons non-voisés

Un son non voisé ne présente pas de structure périodique, 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 comme l'illustre la figure (I.10) [17].

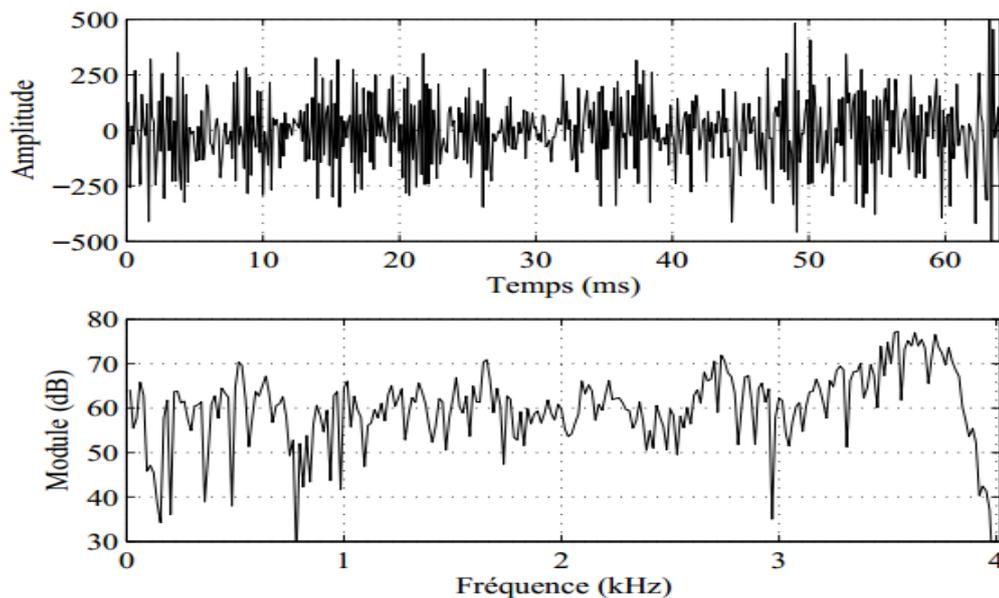


Figure I.10 : Forme d'onde d'une trame de signal vocal non voisé et son spectre d'amplitude.

I.3.3.3 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.

I.3.4 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 I.4 représente les différentes classes auxquelles un bruit peut appartenir [15].

Tableau I.4 : Différentes propriétés du bruit acoustique.

Propriétés du bruit	
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.

- **Bruit blanc**

Un bruit blanc est un signal aléatoire stationnaire de variance infinie dont la fonction d'auto-corrélation est proportionnelle à un Dirac (c'est-à-dire un spectre complexe constant sur toute la plage des fréquences). Cela se traduit par le fait que les valeurs du signal pris à deux instants, même très proches, ne sont pas du tout corrélées [18].

- **Bruit coloré**

Un bruit coloré est un bruit blanc filtré, ainsi la fonction d'autocorrélation d'un tel bruit ne sera pas une impulsion de Dirac, mais plutôt une courbe étroite [19]. Un bruit coloré de basse fréquence est parfois appelé bruit rose car il ne conserve que les grandes longueurs d'onde, ce qui dans le spectre du visible correspond aux teintes rouges.

- **Bruit musical**

Le bruit musical est un bruit résiduel perceptuellement gênant qui apparaît suite au débruitage de la parole par des algorithmes d'atténuation spectrale à court terme tels que la soustraction spectrale ou le filtrage de Wiener. Le spectre du bruit musical est particulièrement tonal, d'où le caractère musical [15].

I.4 Conclusion

Dans ce chapitre, nous avons présenté, dans une première partie, quelques notions de base du traitement du signal, le principe du filtrage adaptatif et enfin les algorithmes du gradient stochastique LMS et NLMS.

Dans la deuxième partie de ce chapitre, les propriétés et les caractéristiques de la parole et du bruit ont été rappelées afin de mieux les traiter dans le cadre de ce mémoire.

Dans le prochain chapitre, nous allons présenter les techniques mono/bi/multi voies pour la réduction du bruit et le rehaussement du signal de parole.

II.1 Introduction

Le domaine de la réduction du 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. Nous présentons dans ce chapitre quelques méthodes de débruitage de la parole. La classification que nous adoptons pour ces méthodes repose sur le nombre d'observations disponibles (relié directement au nombre de microphones utilisés).

Nous traitons en premier le cas des méthodes mono-voie où un seul capteur est utilisé, le principe général de ces méthodes est exposé dans la partie (II.2). Nous rappelons dans la partie (II.2.1) et (II.2.2) l'essentiel des techniques classiques de réduction de bruit mono-capteur qui a un lien direct avec le sujet de ce mémoire.

Nous abordons ensuite le cas où l'on dispose de deux observations, pour cela nous présentons les méthodes bi-voies les plus utilisées dans le domaine de réduction de bruit et de rehaussement du signal de parole. Dans la partie (II.3.1), Nous étudions la méthode la plus connue et la plus ancienne qui est l'annulation adaptative de bruit (ANC), et dans la partie (II.3.2), nous formulons le problème de la séparation de sources et nous présentons les définitions et les notations associées.

Finalement, nous présentons les méthodes multi-voies basées sur la combinaison des filtrages spatial et temporel par les approches de traitement d'antennes linéaires.

II.2 Méthodes de débruitage mono-voie

Le domaine du rehaussement de la parole dans le cas où un seul capteur est utilisé est riche en solutions et fait appel à de nombreux concepts de traitement du signal. Nous présentons dans cette partie, les méthodes qui sont à base du traitement mono-voie (soustraction spectrale, filtre de Wiener) dont le modèle de débruitage utilisé pour ces méthodes est illustré sur la figure (II.1).

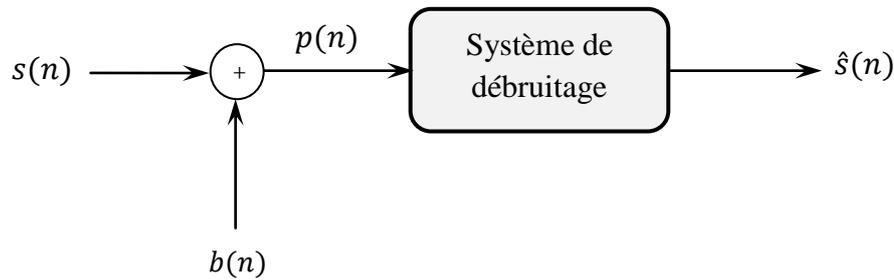


Figure II.1 : Modèle de débruitage utilisé.

L'objectif, dans le cas mono-voie, consiste à estimer un signal $s(n)$ noyé dans un bruit additif $b(n)$ supposé indépendant, à partir de la seule observation $p(n)$ qui est la somme de ses deux composantes pour chaque indice temporel discret n :

$$p(n) = s(n) + b(n) \quad (\text{II.1})$$

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 (II.1) s'écrit [20] :

$$\gamma_p(f) = \gamma_s(f) + \gamma_b(f) \quad (\text{II.2})$$

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 transformé 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 [17].

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

$$P(f, k) = S(f, k) + B(f, k) = |P(f, k)|e^{j\Phi_p(f, k)} \quad (\text{II.3})$$

❖ Définitions des Rapports signal à bruit (RSB)

Il n'existe pas de solution simple à l'estimation spectrale de $S(f, k)$. Celle-ci est généralement obtenue par l'application, sur le spectre du signal bruité $P(f, k)$, d'un gain spectral $G(f, k)$ dépendant du RSB :

$$\hat{S}(f, k) = G(f, k) P(f, k) \quad (\text{II.4})$$

Le gain spectral $G(f, k)$ possède toujours le comportement asymptotique suivant :

- Un RSB élevé correspond très certainement à une forte composante de la parole, et le gain $G(f, k)$ doit donc être proche de 1 pour préserver cette composante.
- Un RSB faible doit mener à une atténuation importante de $P(f, k)$ ($G(f, k) \ll 1$), cette observation ne contenant que du bruit.

Tout le problème consiste donc à estimer ce RSB. Selon les hypothèses choisies pour exprimer le gain spectral, deux types de RSB sont utilisés, le RSB *a priori* et le RSB *a posteriori* :

$$RSB_{prio}(f) = \frac{\gamma_s(f)}{\gamma_b(f)} = \frac{E[|S(f)|^2]}{E[|B(f)|^2]} \quad (\text{II.5})$$

$$RSB_{post}(f, k) = \frac{|P(f, k)|^2}{\gamma_b(f)} = \frac{|P(f, k)|^2}{E[|B(f)|^2]} \quad (\text{II.6})$$

La quantité $RSB_{prio}(f)$ exprime le RSB long terme en supposant les caractéristiques statistiques du signal utile connues *a priori*, d'où son nom. La grandeur $RSB_{post}(f, k)$ exprime le RSB de la trame courante en considérant le spectre instantané du signal bruité. Il faut noter que le $RSB_{prio}(f)$ ne dépend pas du temps puisqu'il s'agit d'une estimation sur une période importante. En pratique, la DSP du signal de parole est bien évidemment amenée à évoluer au cours du temps, on considérera par la suite la notation $RSB_{prio}(f, k)$ avec un indice de trame, tout en gardant en mémoire que ses variations sont plus lentes que celles de $RSB_{post}(f, k)$ [20].

II.2.1 Soustraction spectrale

Dans le domaine de la réduction du bruit, la soustraction spectrale initialement proposée par Boll s'octroie une place de choix [21] : son caractère intuitif et sa simplicité en ont fait un algorithme particulièrement répandu. Elle est basée sur la connaissance *a priori* de la densité spectrale de puissance du bruit supposé stationnaire, qui peut être apprise lors des phases de silence. Historiquement, Boll a proposé d'estimer le module de la TFCT du signal de parole par :

$$|\hat{S}(f, k)| = |P(f, k)| - \sqrt{E[|B(f, k)|^2]} \quad (\text{II.7})$$

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 [17]:

$$|\hat{S}(f, k)| = \begin{cases} |P(f, k)| - \sqrt{E [|B(f, k)|^2]} & \text{si } |P(f, k)| \geq \sqrt{E [|B(f, k)|^2]} , \\ 0 & \text{sinon.} \end{cases} \quad (\text{II.8})$$

Toutefois dans un souci d'homogénéisation des écritures avec les autres techniques présentées, la SSA peut se formuler sous forme d'un gain dépendant du RSB *a posteriori* (II.6). Ainsi, le spectre estimé du signal utile peut être obtenu par la relation suivante :

$$\hat{S}(f, k) = G_{SSA}(f, k)P(f, k) \quad (\text{II.9})$$

avec

$$G_{SSA}(f, k) = \begin{cases} 1 - \sqrt{\frac{1}{RSB_{post}(f, k)}} & \text{si } RSB_{post}(f, k) \geq 1, \\ 0 & \text{sinon.} \end{cases} \quad (\text{II.10})$$

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) [17] :

$$|\hat{S}(f, k)|^2 = |P(f, k)|^2 - E [|B(f, k)|^2] \quad (\text{II.11})$$

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 :

$$|\hat{S}(f, k)|^2 = \begin{cases} |P(f, k)|^2 - E [|B(f, k)|^2] & \text{si } |P(f, k)|^2 \geq E [|B(f, k)|^2] , \\ 0 & \text{sinon.} \end{cases} \quad (\text{II.12})$$

McAulay a montré que sous hypothèse d'indépendance et de distribution gaussienne des composantes spectrales du bruit et de la parole, l'équation (II.12) correspond à l'estimateur au sens du maximum de vraisemblance de la densité spectrale de puissance du signal utile [22], ce qui assure théoriquement une distorsion nulle de son spectre [23].

La SSP peut aussi s'exprimer en fonction du RSB *a posteriori* :

$$G_{SSP}(f, k) = \begin{cases} \sqrt{1 - \frac{1}{RSB_{post}(f, k)}} & \text{si } RSB_{post}(f, k) \geq 1, \\ 0 & \text{sinon.} \end{cases} \quad (\text{II.13})$$

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) [17].

II.2.2 Filtrage de Wiener

C'est le filtre linéaire optimal au sens du Minimum de l'Erreur Quadratique Moyenne (MEQM), c'est-à-dire qu'il minimise la fonction d'erreur suivante [20] :

$$E[|S(f, k) - G_W(f, k)P(f, k)|^2] \quad (\text{II.14})$$

On montre que ce filtre, noté $G_W(f, k)$, s'écrit :

$$G_W(f, k) = \frac{\gamma_s(f, k)}{\gamma_s(f, k) + \gamma_b(f, k)} \quad (\text{II.15})$$

En utilisant la définition du RSB *a priori*, le filtre de Wiener s'exprime sous la forme :

$$G_W(f, k) = \frac{RSB_{prio}(f, k)}{1 + RSB_{prio}(f, k)} \quad (\text{II.16})$$

En pratique, il est possible de remplacer la DSP du signal utile $E[|S(f, k)|^2]$ par une estimée obtenue directement à partir du spectre bruité instantané selon le principe de la SSP: $|P(f, k)|^2 - E[|B(f, k)|^2]$, bien que cela n'ait pas de fondement théorique. On s'éloigne alors des hypothèses de stationnarité nécessaires au filtre de Wiener, C'est pourquoi on parle de filtre pseudo-Wiener selon l'appellation proposée dans [23]. Il est alors possible d'exprimer le filtre pseudo-Wiener en fonction du RSB *a posteriori* :

$$G_W(f, k) = 1 - \frac{1}{RSB_{post}(f, k)} \quad (\text{II.17})$$

On peut d'ailleurs remarquer que dans ce cas $G_W(f, k) = G_{SSP}^2(f, k)$ attestant du fait que ce filtre fait partie de la famille des méthodes de soustraction spectrale.

II.3 Méthodes de débruitage bi-voies

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-voies 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 du bruit et le rehaussement du signal de parole.

II.3.1 Structure d'annulation adaptative de bruit (ANC)

Le principe de l'annulation adaptative de bruit remonte aux contributions de Howells et Applebaum [24] vers la fin des années 50. Le développement de l'algorithme LMS (Least Mean Square) par Widrow et Hoff [25] a permis l'élaboration de cette technique sous sa

forme actuelle. Nous présentons ici le principe de base de cette structure.

II.3.1.1 Principe de base de l'annulation adaptative de bruit

Considérons une observation bruitée $d(n)$ d'un signal $s(n)$ c.à.d $d(n) = s(n) + b_0(n)$ où $b_0(n)$ est une contribution décorrélée du signal utile $s(n)$. $d(n)$ 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 $b_0(n) = h(n) * b(n)$ 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 [26]. Une illustration graphique de ce principe est présentée en figure II.2.

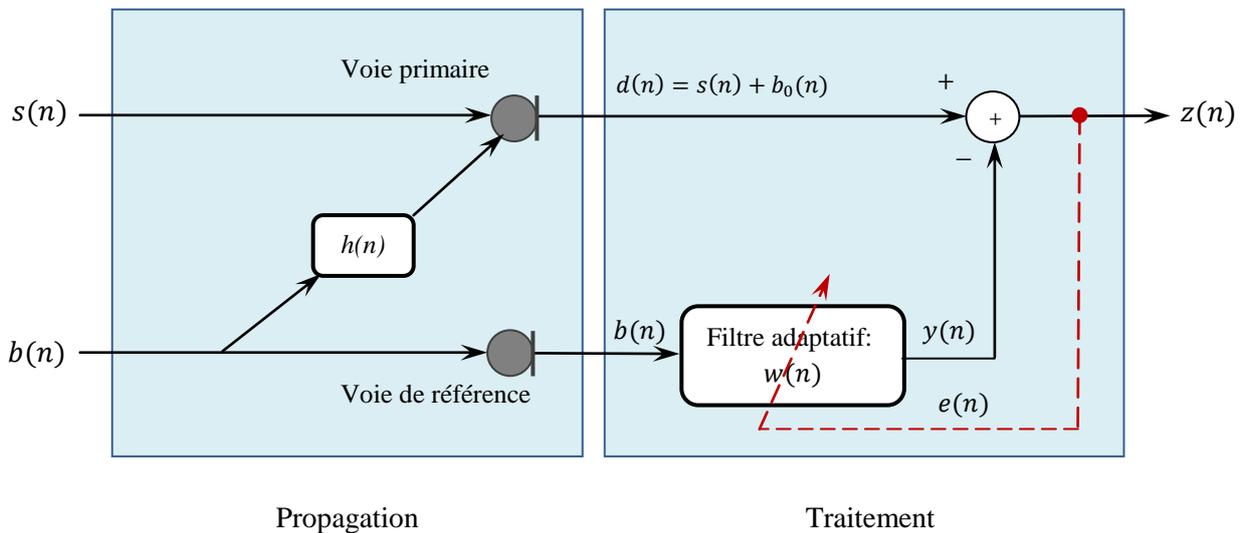


Figure II.2 : Structure d'annulation adaptative de bruit (ANC).

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

$$z(n) = s(n) + (h(n) - w(n)) * b(n) \quad (\text{II.18})$$

La solution d'un débruitage idéal $z(n) = s(n)$ 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 $z(n)$. En effet, compte tenu de la décorrélation entre $b_0(n)$ et $s(n)$, on a d'après (II.18) :

$$E [z^2(n)] = E [s^2(n)] + E [(h(n) - w(n)) * b(n)]^2 \quad (\text{II.19})$$

La valeur minimale de $E [z^2(n)]$ 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.

II.3.1.2 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 donné par la figure suivante :

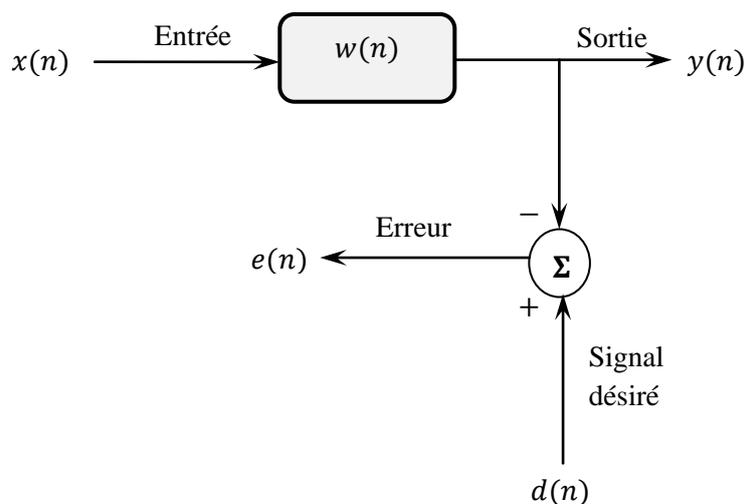


Figure II.3 : Schéma du Filtre de Wiener (ANC).

Sur cette figure, nous utilisons les notations suivantes : le signal d'entrée est $x(n)$, le signal en sortie de traitement est $y(n)$ et le signal désiré est $d(n)$. Nous supposons en plus que les signaux d'entrée/sortie sont stationnaires. 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 transfert du filtre de Wiener est donnée par :

$$W^{opt}(f) = \frac{\phi_{dx}(f)}{\phi_{xx}(f)} \quad (\text{II.20})$$

avec

$$\phi_{dx}(f) = TF \{ r_{dx}(k) \} \quad (\text{II.21})$$

$$\phi_{xx}(f) = TF \{ r_{xx}(k) \} \quad (\text{II.22})$$

où $\phi_{dx}(f)$ représente la densité inter-spectrale de puissance entre le signal d'entrée et le signal désiré et $\phi_{xx}(f)$ 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 :

$$r_{dx}(k) = E [d(n)x^*(n - k)] \quad (\text{II.23})$$

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

$$r_{xx}(k) = E [x(n)x^*(n - k)] \quad (\text{II.24})$$

Appliquons la solution de Wiener donnée par l'équation (II.20) au cas de l'annulation adaptative du bruit suivant le schéma détaillé de la figure II.2. Le signal désiré $d(n)$ correspond au signal observé sur la voie primaire (c.à.d. $d(n) = s(n) + b_0(n)$) 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 :

$$W^{opt}(f) = \frac{\phi_{ab}(f)}{\phi_{bb}(f)} \quad (\text{II.25})$$

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

$$\phi_{ab}(f) = \phi_{bb}(f) H(f) \quad (\text{II.26})$$

En remplaçant l'expression (II.26) dans l'expression du filtre de Wiener (II.25), on obtient :

$$W^{opt}(f) = H(f) \quad (\text{II.27})$$

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

II.3.2 Séparation aveugle de sources

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 [27].

La SAS se rencontre dans plusieurs applications tels que: radiocommunication, traitement des images, rehaussement de la parole, débruitage, etc...

❖ **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 II.4) [28]:

- **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.

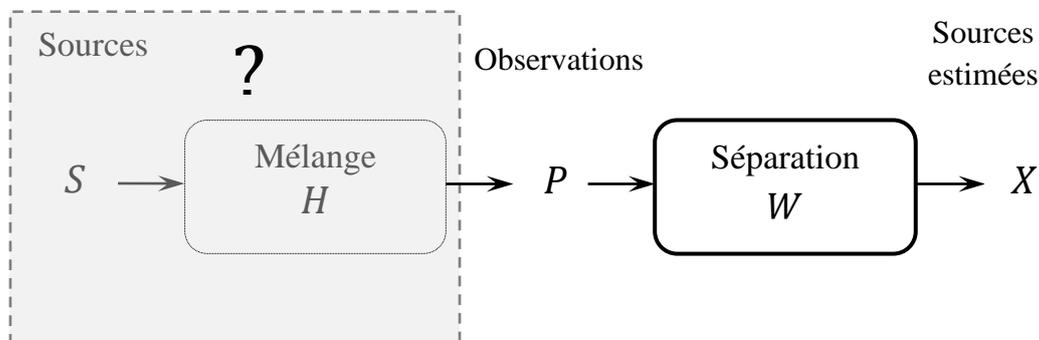


Figure II.4 : Structure générale du problème de la séparation aveugle de sources.

➤ **Modélisation du mélange**

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

$$P(t) = H [S(t)] + B (t) \tag{II.28}$$

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 qui modélise les erreurs de mesure des capteurs.
- H est l'opérateur de mélange.

La figure (II.5) symbolise la relation sources-observation.

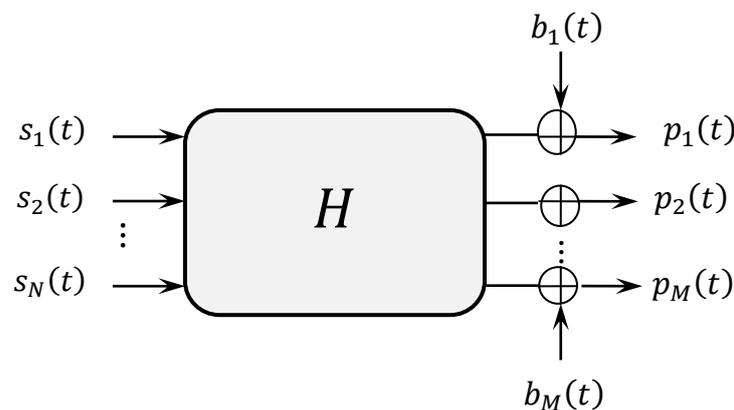


Figure II.5 : Schéma général de la relation sources-observations en présence de bruit.

▪ **Les différents types de mélanges**

Nous pouvons classer les mélanges en deux grandes catégories selon la nature de l'opérateur H [29] :

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

$$P(t) = H(t) * S(t) + B(t) \tag{II.29}$$

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 (II.29) devient alors :

$$P(n) = H(n) * S(n) + B(n) \quad (\text{II.30})$$

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 (II.30) devient :

$$P(n) = H S(n) + B(n) \quad (\text{II.31})$$

Un exemple du mélange instantané de sources est illustré sur la figure (II.6).

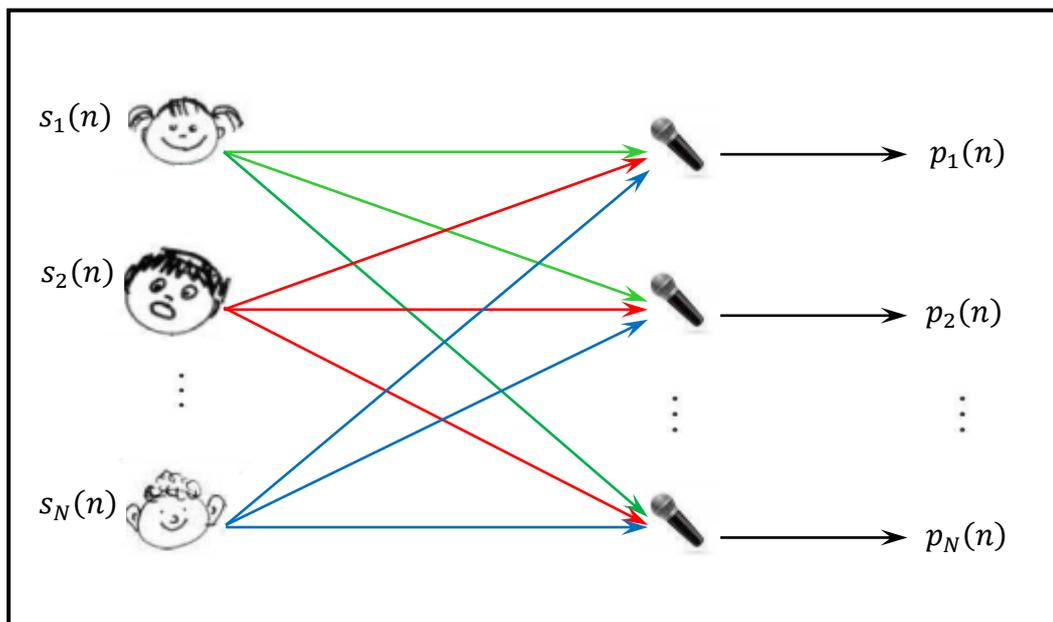


Figure II.6 : Exemple d'un mélange instantané de sources.

- ❖ **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 (II.7) présente un exemple du mélange convolutif de deux sources et deux observations.

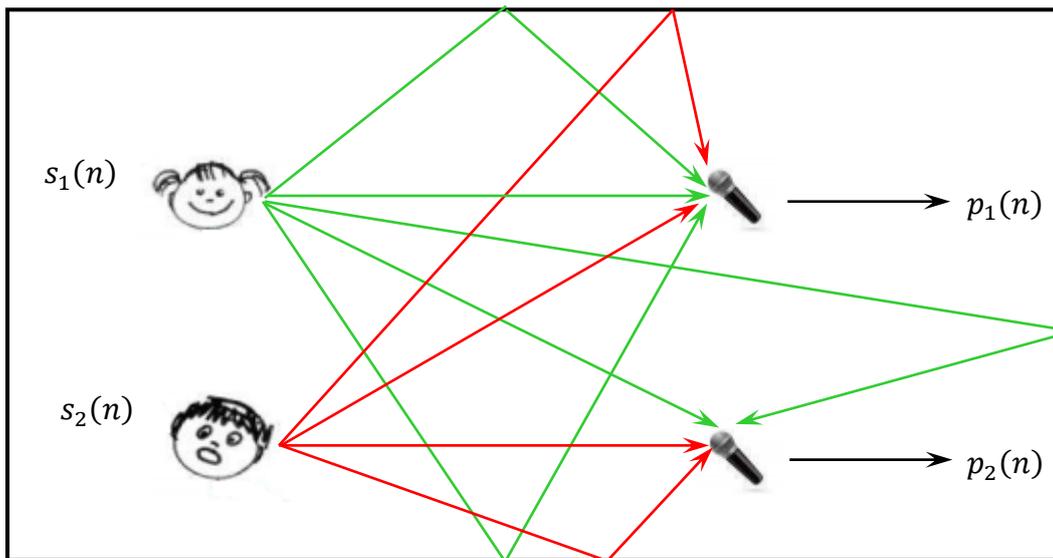


Figure II.7 : Illustration d'un mélange convolutif acoustique à deux sources et deux capteurs.

- 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 [29].

Dans ce mémoire, 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.

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 ce mémoire, nous nous intéressons uniquement aux mélanges convolutifs déterminés d'ordre 2 ($N = M = 2$).

II.3.2.1 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^{er} microphone, l'autre source de perturbation (un bruit ou un autre signal de la parole) est près du 2^{ème} microphone. A la sortie des microphones, on observe une superposition des signaux primitifs inconnus selon un mélange inconnu (figure II.8). 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 [30] :

$$p_1(n) = h_{11}(n) * s_1(n) + h_{21}(n) * s_2(n) + b_1(n) \quad (\text{II.32})$$

$$p_2(n) = h_{22}(n) * s_2(n) + h_{12}(n) * s_1(n) + b_2(n) \quad (\text{II.33})$$

où : $s_1(n)$ et $s_2(n)$ 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.

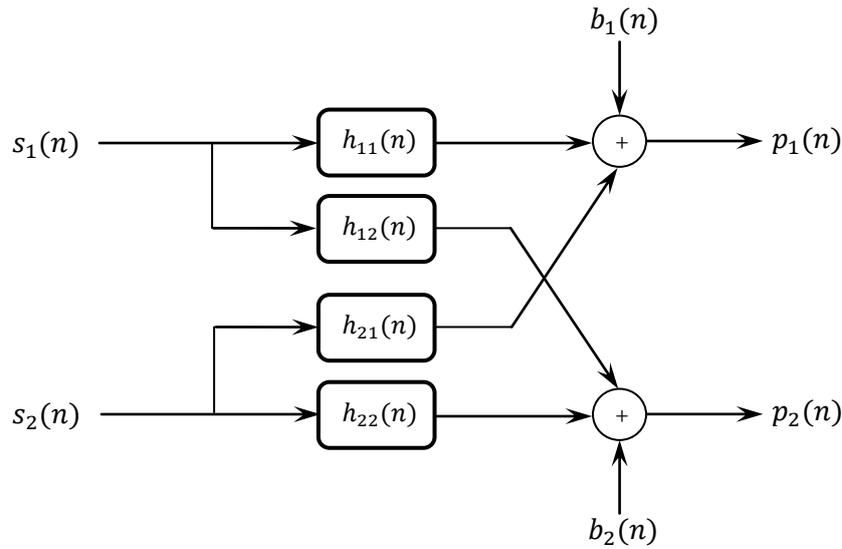


Figure II.8 : Modèle général du mélange convolutif.

- **Modèle convolutif simplifié**

Pour simplifier le modèle de mélange convolutif qui est représenté dans la figure (II.8), 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 canaux 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 (II.9).

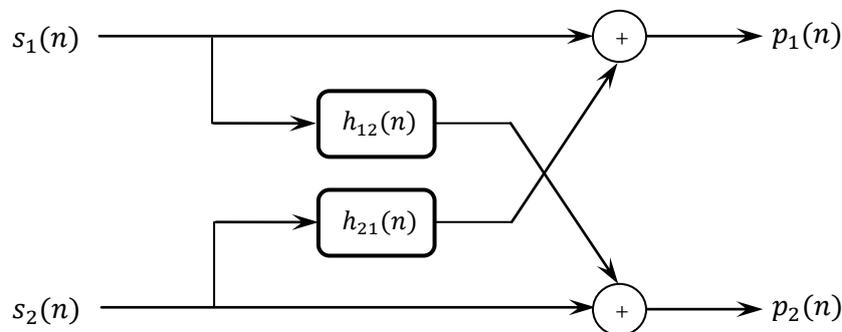


Figure II.9 : Modèle simplifier du mélange convolutif.

Les équations des signaux du mélange deviennent :

$$p_1(n) = s_1(n) + h_{21}(n) * s_2(n) \quad (\text{II.34})$$

$$p_2(n) = s_2(n) + h_{12}(n) * s_1(n) \quad (\text{II.35})$$

où les sources $s_1(n)$ et $s_2(n)$ et les filtres $h_{12}(n)$ et $h_{21}(n)$ sont tous inconnus.

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

II.3.2.2 Solution de la séparation de sources

Pour résoudre ce problème, deux structures de séparation de sources bien connues sont classiquement utilisées :

- La structure Directe

Connue sous le terme anglo-saxon feedforward. Dans cette configuration, chaque sortie du système est une combinaison linéaire des observations, le schéma de la séparation que l'on qualifie de directe, est illustré par la figure (II.10).

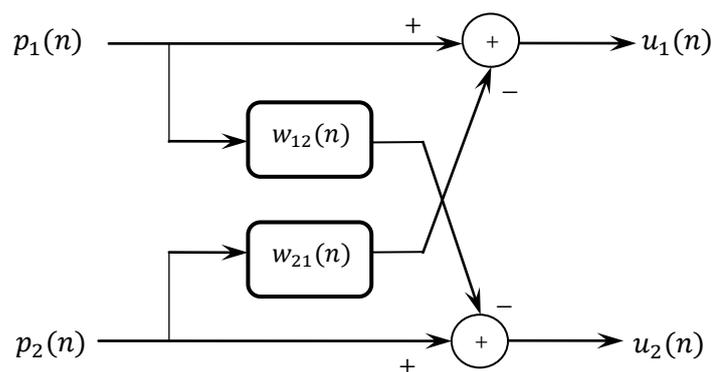


Figure II.10 : Structure symétrique Directe “Farward” de débruitage.

- La structure Réursive ou Bachward

Dans cette configuration, chaque sortie du système de séparation est une combinaison linéaire d'une observation $p_i(n)$ et de l'autre sortie $v_j(n)$, cette structure est représentée par la figure (II.11).

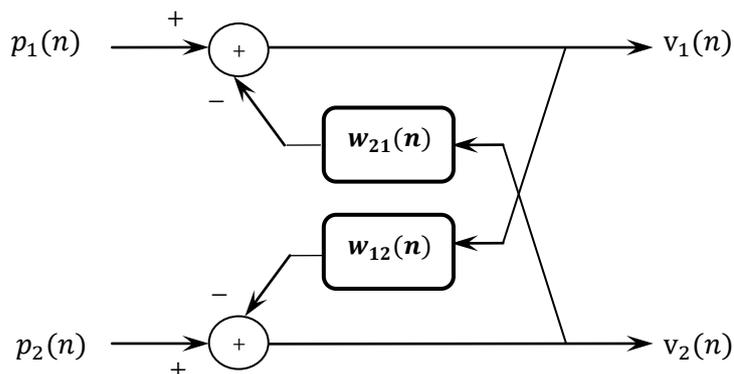


Figure II.11 : Structure symétrique Réursive “Backward” de débruitage.

II.4 Méthodes de débruitage multi-voies

Dans cette partie, nous supposons que l'on dispose d'un nombre élevé de microphones. Cette hypothèse autorise bien sûr la duplication des techniques mono-voie et bi-voies précédemment présentées mais elle permet surtout d'envisager la combinaison des filtrages temporel et spatial par les méthodes de traitement d'antenne linéaire.

II.4.1 Réduction de bruit par le réseau d'antennes de Frost

Considérons un réseau de N microphones identiques, omnidirectionnels, linéairement disposés et uniformément séparés par une distance d (voir figure II.12). Soit un signal $x(n)$ à large bande représentant une onde plane émanant de la direction θ et soient b_1, \dots, b_N des bruits mesurés au niveau des différents microphones.

D'une manière analogue au filtrage temporel, on peut augmenter le RSB au niveau de la réception du signal utile par filtrage spatial. Ceci peut être réalisé par une redirection mécanique des capteurs ou en les associant en un formateur de voie. Le but consiste à

privilégier la direction d'arrivée du signal de parole utile en réduisant le bruit issu des autres directions [26].

Nous présentons en figure (II.12) le formateur de voie proposé par O.L.Frost [31]. Son principe repose sur l'utilisation d'un système à base de filtres RIF (de même ordre) dont les coefficients sont ajustés pour réduire la puissance en sortie de l'antenne tout en maintenant une réponse prédéfinie pour le signal utile. Les hypothèses utilisées sont :

- ❖ Les signaux utiles reçus par les microphones sont identiques à un retard près.
- ❖ Les bruits reçus par les différents capteurs sont corrélés.

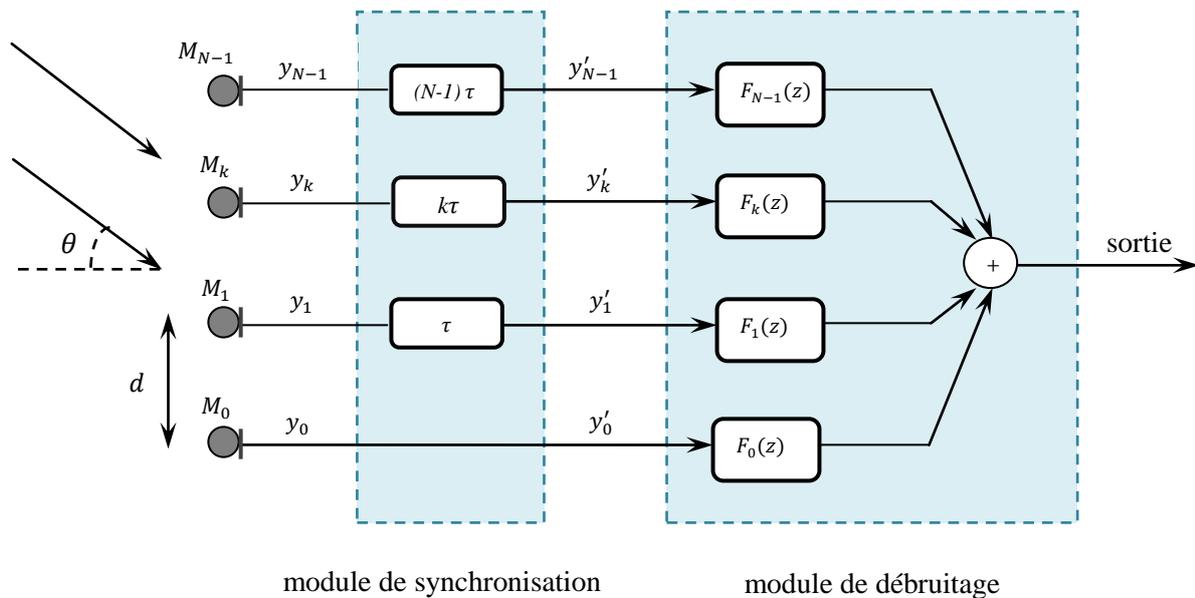


Figure II.12 : Structure de FROST [31].

La structure de Frost est une mise en cascade de deux étages :

- ❖ **Le module de synchronisation** dont le but est de compenser le retard de propagation du signal utile en introduisant des retards égaux aux temps de propagation entre capteurs. Pour un signal utile émanant de la direction θ , une distance inter-microphone

d et une vitesse du son c , le retard élémentaire entre deux microphones adjacents correspond à :

$$\tau = \frac{d}{c} \sin(\theta) \quad (\text{II.36})$$

Ainsi, si l'on désigne respectivement par $x_k(n)$ et $b_k(n)$ le signal utile et le bruit mesurés au niveau du microphone M_k , on aura la relation :

$$y_k(n) = x_k(n) + b_k(n) = x_0(n + k\tau) + b_k(n) \quad (\text{II.37})$$

où $x_0(n)$ représente le signal utile au niveau du microphone M_0 . En appliquant un retard $k\tau$ à $y_k(n)$, on obtient la nouvelle observation qui contient le même signal utile que le microphone M_0 :

$$y'_k(n) = y_k(n - k\tau) = x_0(n) + b_k(n - k\tau) \quad (\text{II.38})$$

- ❖ **Le module de débruitage** qui optimise le choix des filtres F_k pour réduire le bruit en sortie tout en maintenant une réponse prédéfinie dans la direction du signal utile. On note que les filtres $F_k(z)$ sont à réponse impulsionnelle finie et de même longueur M , soit :

$$F_k(z) = \sum_{p=0}^{M-1} f_k(p) z^{-p}, \quad \forall k \in [1, N] \quad (\text{II.39})$$

Pour un signal émanant de la direction de vue (direction d'incidence du signal utile), le récepteur sera équivalent au filtre $F_u(z)$ défini par :

$$F_u(z) = \sum_{k=1}^N F_k(z) = \sum_{p=0}^{M-1} (\sum_{k=1}^N f_k(p)) z^{-p} \quad (\text{II.40})$$

En fixant les valeurs des expressions $(\sum_{k=1}^N f_k(p))$ pour $p \in [1, M]$, Frost fixe la réponse fréquentielle associée au signal utile. Les $(M - 1)N$ coefficients restants sont ajustés pour réduire la puissance du bruit en sortie. La mise en œuvre pratique de cette technique utilise un algorithme LMS avec contraintes linéaires.

II.4.2 Structure de Griffiths et Jim

La méthode proposée par Griffiths et Jim sous le sigle GSC (Generalized Sidelobe Canceler) est une combinaison d'un formateur de voie classique et d'une structure d'annulation adaptative de bruit.

Cette structure, représentée à la figure (II.13), repose sur un principe général identique à celui de l'algorithme de Frost à ceci près que la contrainte n'est plus imposée au niveau de l'algorithme d'optimisation mais au niveau de la structure elle-même ce qui conduit à une réalisation algorithmique plus simple. Ce dispositif consiste en deux sous-structures :

- Un étage de synchronisation associé un ensemble de coefficients de pondération fixes H_1, \dots, H_N qui réalise un formateur de voie classique non adaptatif. Les retards τ_i ($i = 1, \dots, N$) permettent de sélectionner une direction de vue tandis que les coefficients permettent de contrôler la largeur de bande, de réaliser un gain de lobe secondaire donné ou de réduire les bruits décorrélés. La sortie de cet étage passe à travers un filtre RIF formant le filtre de contrainte $F_u(z)$.

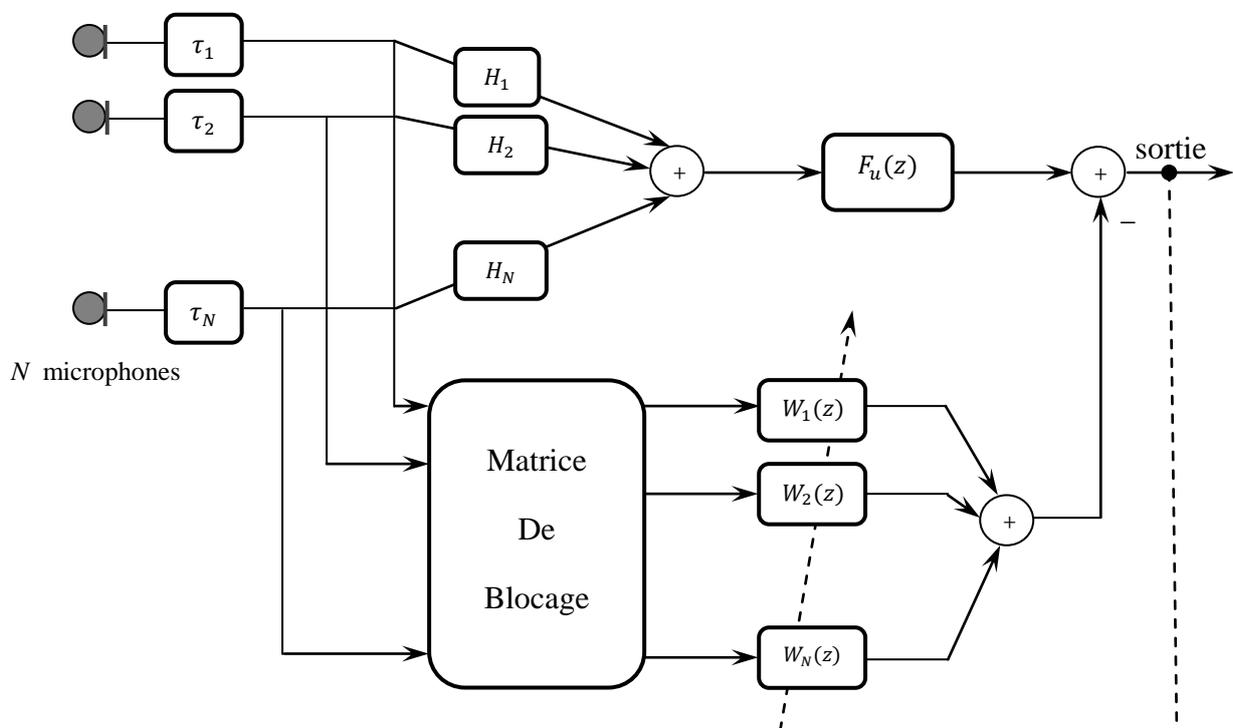


Figure II.13 : Structure de Griffiths et Lim.

- Un étage de réduction de bruit qui consiste en une matrice de blocage qui élimine le signal utile (la somme des éléments de chaque ligne est nulle) suivie de filtres W_1, \dots, W_N qui sont adaptativement estimés pour réduire la puissance du bruit en sortie.

En fait, ce module correspond à un schéma d'ANC avec plusieurs références de bruit.

II.5 Conclusion

Au cours de ce chapitre, nous avons présenté un ensemble de méthodes de rehaussement de la parole. La classification adoptée repose sur le nombre de microphones disponibles.

Nous avons vu que les méthodes classiques mono-voie 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-voies décrites dans ce chapitre, qui sont basées sur la structure d'annulation adaptative de bruit (ANC). Le grand intérêt de l'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 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) résoud théoriquement ce problème, c'est pourquoi cette structure fera l'objet d'une étude complète qui sera présentée au sein du chapitre III.

Pour ce qui concerne les techniques multi-voies, deux méthodes de traitement ont été présentées : La structure de Frost et la structure GSC de Griffith.

Ces deux méthodes supposent que la position de la source est connue et fixe et que les signaux sources après compensation du retard sont identiques sur les différentes voies (propagation en ondes planes). Ces deux suppositions ne sont pas remplies dans les conditions plus réalistes (téléphonie mains-libres par exemple) où la position de la source est susceptible de varier au cours du temps, et où la distance réduite entre la source et les capteurs favorise l'hypothèse d'une propagation en onde sphérique [26].

III.1 Introduction

Le problème de la séparation aveugle de sources consiste à retrouver, avec le moins de connaissance a priori, des signaux utiles qui ont été mélangés. Pour résoudre ce problème deux structures appelées Directe et Réursive, sont fréquemment utilisées pour extraire les signaux sources à partir de la seule connaissance des signaux de mélange.

Ce chapitre est organisé en deux parties. Nous présentons dans la première partie une étude détaillée du fonctionnement des deux structures classique de séparation de sources Directe et Réursive ainsi que de leur avantage et inconvénient.

Dans la deuxième partie, nous décrivons deux versions améliorées pour chaque structure classique (Directe et Réursive).

III.2 La structure Réursive

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 (III.1).

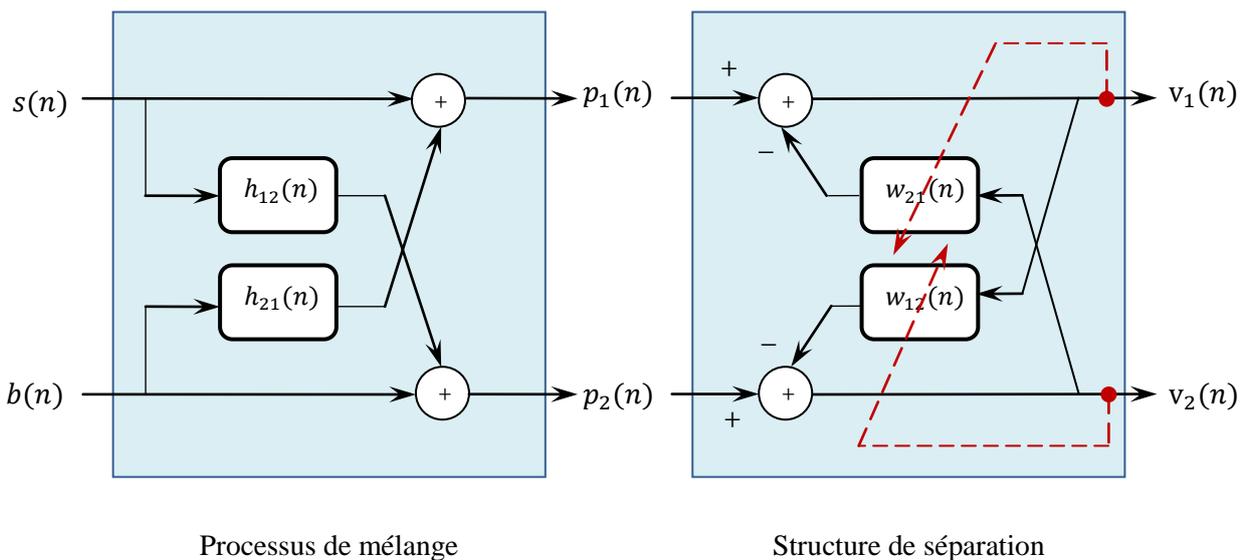


Figure III.1 : Structure de séparation de sources Réursive (BBSS).

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 (III.1), sont donnés par :

$$p_1(n) = s(n) + h_{21}(n) * b(n) \quad (\text{III.1})$$

$$p_2(n) = b(n) + h_{12}(n) * s(n) \quad (\text{III.2})$$

La structure Réursive conduit aux relations observations/sorties suivantes :

$$v_1(n) = p_1(n) - v_2(n) * w_{21}(n) \quad (\text{III.3})$$

$$v_2(n) = p_2(n) - v_1(n) * w_{12}(n) \quad (\text{III.4})$$

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

$$v_1(n) = s(n) + h_{21}(n) * b(n) - [p_2(n) - v_1(n) * w_{12}(n)] * w_{21}(n) \quad (\text{III.5})$$

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

$$v_1(n) = \frac{1}{\delta(n) - w_{12}(n) * w_{21}(n)} (s(n) * (\delta(n) - h_{12}(n) * w_{21}(n)) + b(n) * (h_{21}(n) - w_{21}(n))) \quad (\text{III.6})$$

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 :

$$v_2(n) = \frac{1}{\delta(n) - w_{12}(n) * w_{21}(n)} (b(n) * (\delta(n) - h_{21}(n) * w_{12}(n)) + s(n) * (h_{12}(n) - w_{12}(n))) \quad (\text{III.7})$$

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:

$$v_1(n) = s(n) \quad (\text{III.8})$$

$$v_2(n) = b(n) \quad (\text{III.9})$$

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 I, 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 :

$$w_{12}(n) = w_{12}(n-1) + \mu_{12} \frac{v_2(n) k_1(n)}{k_1^T(n) k_1(n)} \quad (\text{III.10})$$

$$w_{21}(n) = w_{21}(n-1) + \mu_{21} \frac{v_1(n) k_2(n)}{k_2^T(n) k_2(n)} \quad (\text{III.11})$$

où

$$k_1(n) = [v_1(n), v_1(n-1), \dots, v_1(n-L+1)]^T$$

$$k_2(n) = [v_2(n), v_2(n-1), \dots, v_2(n-L+1)]^T$$

μ_{12} et μ_{21} sont respectivement les pas d'adaptations des deux filtres adaptatifs $w_{12}(n)$ et $w_{21}(n)$.

III.2.1 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 [32] :

- 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.

III.3 La structure Directe

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 (III.2).

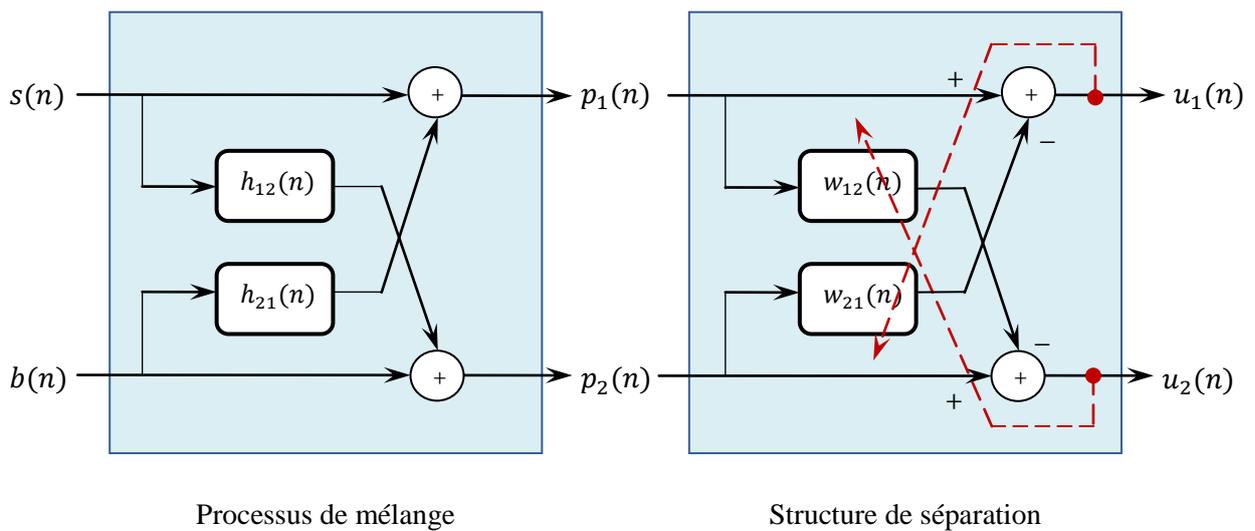


Figure III.2 : Structure de séparation de sources Directe (FBSS).

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

$$u_1(n) = p_1(n) - p_2(n) * w_{21}(n) \quad (\text{III.12})$$

$$u_2(n) = p_2(n) - p_1(n) * w_{12}(n) \quad (\text{III.13})$$

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

$$u_1(n) = b(n) * [h_{21}(n) - w_{21}(n)] + s(n) * [\delta(n) - h_{12}(n) * w_{21}(n)] \quad (\text{III.14})$$

$$u_2(n) = s(n) * [h_{12}(n) - w_{12}(n)] + b(n) * [\delta(n) - h_{21}(n) * w_{12}(n)] \quad (\text{III.15})$$

si

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

alors

$$u_1(n) = s(n) * [\delta(n) - h_{12}(n) * h_{21}(n)] \quad (\text{III.16})$$

$$u_2(n) = b(n) * [\delta(n) - h_{21}(n) * h_{12}(n)] \quad (\text{III.17})$$

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

$$w_{12}(n) = w_{12}(n-1) + \mu_{12} \frac{u_2(n) m_1(n)}{m_1^T(n) m_1(n)} \quad (\text{III.18})$$

$$w_{21}(n) = w_{21}(n-1) + \mu_{21} \frac{u_1(n) m_2(n)}{m_2^T(n) m_2(n)} \quad (\text{III.19})$$

où

$$m_1(n) = [p_1(n), p_1(n-1), \dots, p_1(n-L+1)]^T$$

$$m_2(n) = [p_2(n), p_2(n-1), \dots, p_2(n-L+1)]^T$$

III.3.1 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.

III.3.2 Inconvénient de la structure Directe

A partir des équations (III.16) et (III.17), 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 [33], consiste l'utilisation des post-filtres placés en sortie de la structure Directe.

III.4 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 (c'est-à-dire une DAV manuelle), 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 et le filtre $w_{12}(n)$ est mis à jour uniquement pendant les périodes d'activité vocale. Le principe de la structure de séparation de sources classique (Directe ou Récursive) est montré par la figure III.3.

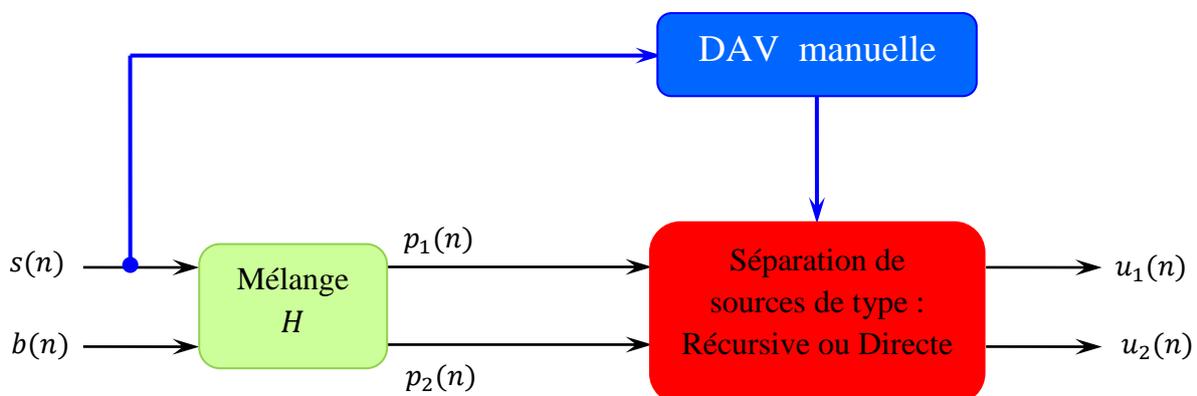


Figure III.3 : Schéma bloc du principe de la structure de séparation de sources classique.

III.5 Amélioration des deux structures de séparation de sources (Directe et Réursive)

L'utilisation d'une DAV manuelle dans le fonctionnement des deux structures classique de séparation de sources (Directe et Réursive) donne une segmentation parfaite ce qui n'est pas le cas en pratique, car on n'a pas d'informations a priori sur les signaux d'entrées. Pour cela nous avons besoin de détecter les DAV automatiquement, c'est la raison principale pour laquelle nous proposons, dans ce mémoire, deux nouvelles structures appelées : Propos1 et Propos2 qui utilisent une technique de DAV automatique basée sur l'estimation des rapports signaux à bruit (RSB) sur les deux voies de la structure de séparation. La figure III.4 représente un schéma bloc du principe de la structure de séparation de sources améliorée.

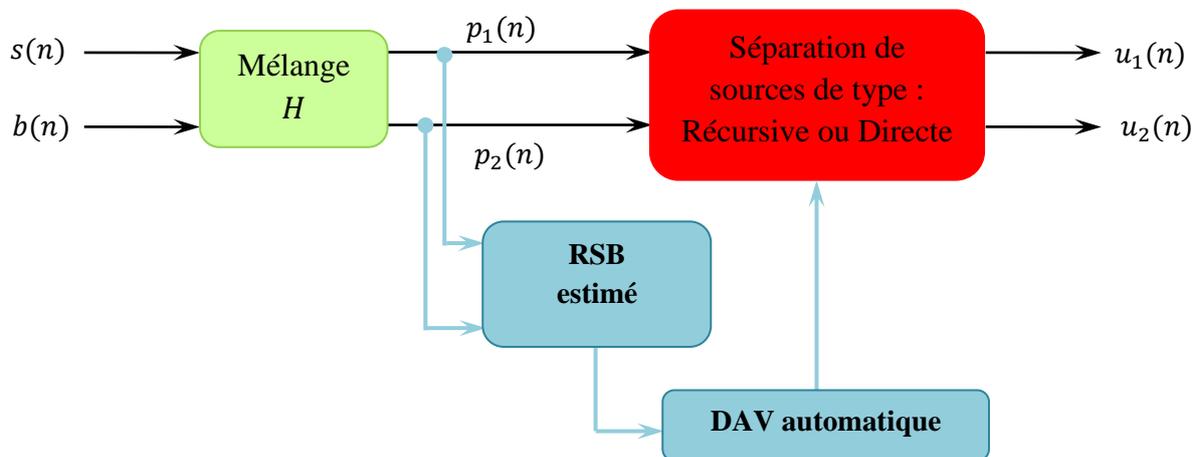


Figure III.4 : Schéma bloc du principe de la structure de séparation de sources améliorée.

A titre d'information, deux structures ont été déjà proposées concernant l'utilisation d'une DAV automatique.

La première structure [34] correspond à une structure Réursive améliorée, la seconde structure [35] correspond à une structure Directe améliorée. Ces deux structures seront détaillées dans la suite de ce document.

III.5.1 Structure Réursive améliorée

Dans cette partie, nous présentons deux versions améliorées de la structure Réursive, celle développée par Ikeda dans [34] et celle que nous allons proposer (Propos1) dans ce travail.

III.5.1.1 Structure développée par Ikeda [34]

La figure (III.5) représente une amélioration de la structure Récursive (développée par Ikeda) sous l'acronyme Method1. Cette version améliorée correspond à une structure Récursive contrôlée de manière adaptative par l'évolution du rapport signal à bruit (RSB) estimé sur les deux voies d'observations, elle 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 ($w_{cont1}(n)$, $w_{cont2}(n)$).

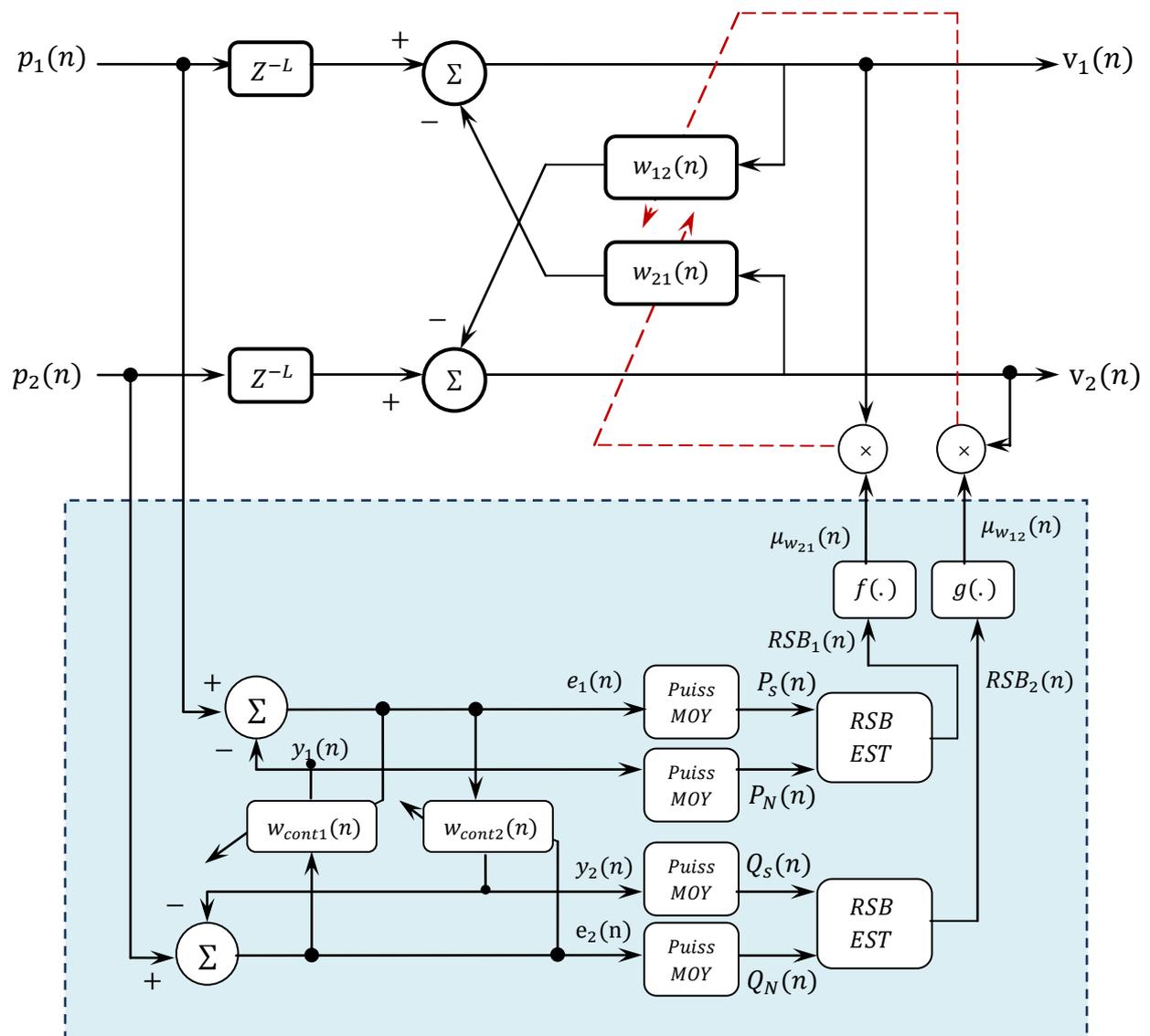


Figure III.5 : Structure développée par Ikeda (Method1) [34].

a) Principe du contrôle des structures de séparations de sources 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 (RSB) 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 $w_{12}(n)$.

b) 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 :

$$P_S(n) = \sum_{j=0}^{M-1} e_1^2(n-j) \tag{III.20}$$

$$P_N(n) = \sum_{j=0}^{M-1} y_1^2(n-j) \tag{III.21}$$

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 :

$$RSB_1(n) = 10 * \log_{10} \left\{ \frac{P_S(n)}{P_N(n)} \right\} \quad (\text{III.22})$$

❖ Estimation du RSB_2

La sortie $y_2(n)$ du filtre secondaire $w_{cont\ 2}(n)$ 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 $w_{cont\ 2}(n)$.

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 :

$$Q_S(n) = \sum_{j=0}^{M-1} y_2^2(n-j) \quad (\text{III. 23})$$

$$Q_N(n) = \sum_{j=0}^{M-1} e_2^2(n-j) \quad (\text{III.24})$$

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

$$RSB_2(n) = 10 * \log_{10} \left\{ \frac{Q_S(n)}{Q_N(n)} \right\} \quad (\text{III.25})$$

c) 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:

$$\mu_{w_{21}}(n) = \begin{cases} \mu_{1min} & RSB_1(n) > RSB_{1max} \\ \mu_{1max} & RSB_1(n) < RSB_{1min} \\ f(RSB_1(n)) & \text{sinon} \end{cases} \quad (\text{III.26})$$

où les paramètres μ_{1max} , μ_{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 $RSB_2(n)$:

$$\mu_{w_{12}}(n) = \begin{cases} \mu_{2min} & RSB_2(n) < RSB_{2min} \\ \mu_{2max} & RSB_2(n) > RSB_{2max} \\ g(RSB_2(n)) & \text{sinon} \end{cases} \quad (\text{III.27})$$

où les paramètres μ_{2max} , μ_{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 (III.6) illustre la variation des pas d'adaptations par rapport aux RSB calculés au cours du temps ainsi que des seuils qui leur sont associés.

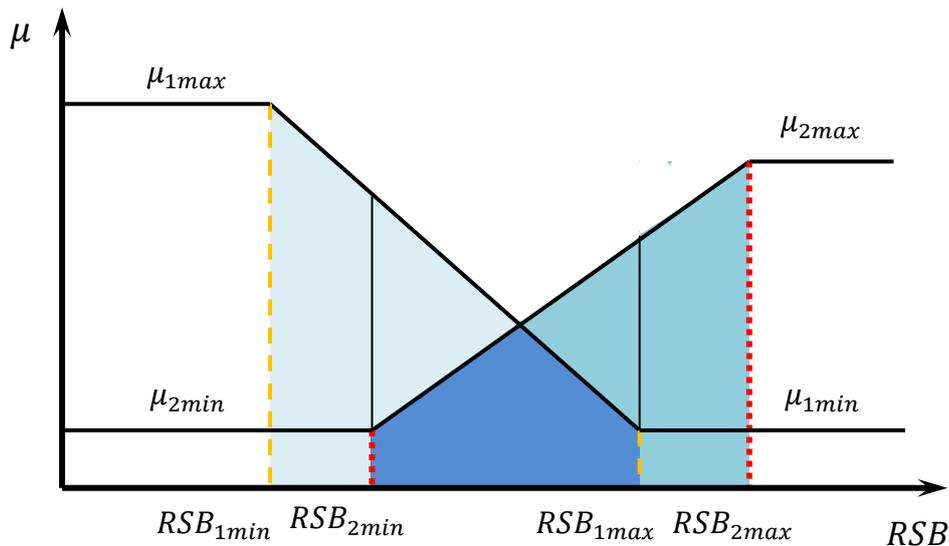


Figure III.6 : Illustration de la variation des pas par rapport aux RSB en fonction de leurs seuils maximum et minimum sur chacune des voies.

Nous donnons la description algorithmique du contrôle des pas d'adaptations des structures améliorées, Algorithme1, accompagnée du diagramme d'activités de la gestion des pas, figure (III.7).

Algorithme 1 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

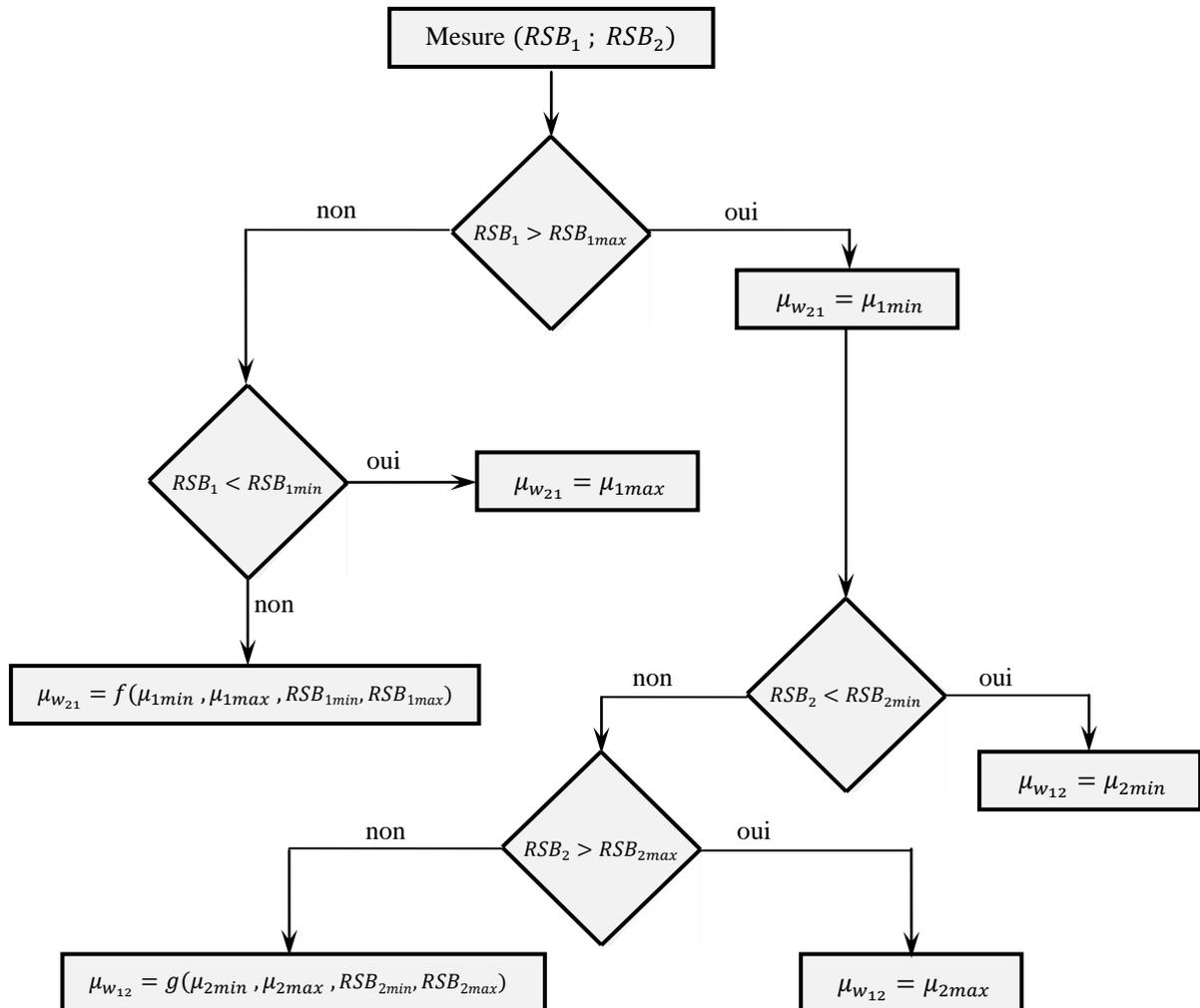


Figure III.7 : Diagramme d'activités de la gestion des pas d'adaptations.

Les paramètres utilisés dans la structure développée par Ikeda (Method1) ont été choisis afin d'éviter la divergence à la sortie de cette structure, ils sont représentés dans le tableau suivant :

Tableau III.1 : Paramètres utilisés pour la méthode d'Ikeda (Method1).

Paramètres	Valeurs
μ_{wcont1}	0.4
μ_{wcont2}	0.001
μ_{1min}	0
μ_{1max}	0.2
RSB_{1min}	-45 dB
RSB_{1max}	-15 dB
μ_{2min}	0
μ_{2max}	0.02
RSB_{2min}	-13 dB
RSB_{2max}	-1 dB

III.5.1.2 Proposition d'une nouvelle structure (Propos1)

Dans la méthode d'Ikeda, La structure Récursive a été contrôlée par un bloc qui réalise une DAV automatique basée sur l'estimation du RSB. Ce bloc correspond à une structure de type backward.

Nous proposons dans ce travail, une nouvelle structure appelée Propos1 qui est donnée par figure (III.8), tel que le bloc utilisé pour le contrôle de la structure Récursive correspond à une structure de type forward.

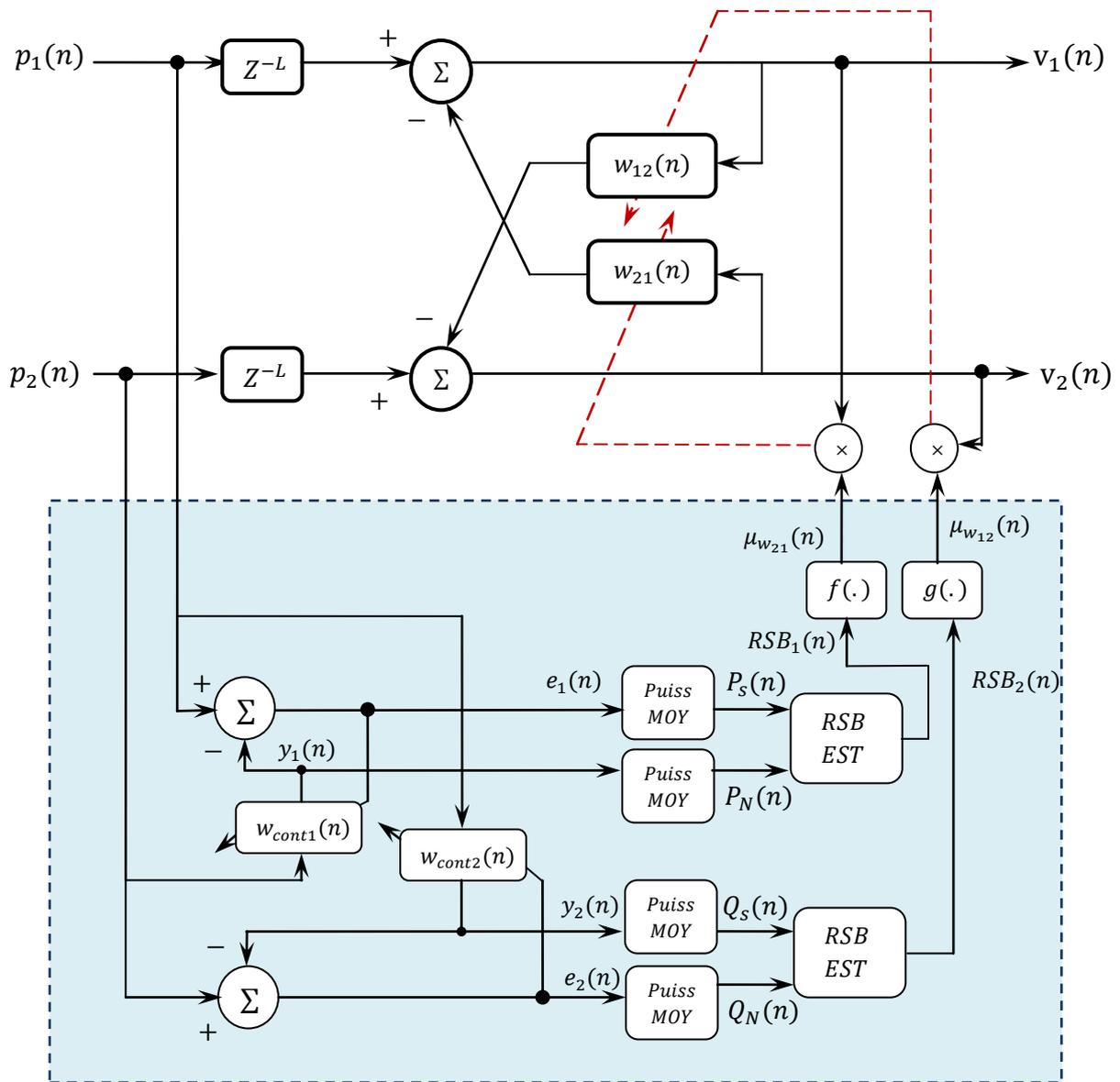


Figure III.8 : La première structure proposée (Propos1).

Le tableau III.2, représente les valeurs des paramètres utilisés dans cette structure.

Tableau III.2 : Paramètres utilisés pour la structure Propos1.

Paramètres	Valeurs
μ_{wcont1}	0.4
μ_{wcont2}	0.001
μ_{1min}	0
μ_{1max}	0.2
RSB_{1min}	-40 dB
RSB_{1max}	-15 dB
μ_{2min}	0
μ_{2max}	0.02
RSB_{2min}	-1 dB
RSB_{2max}	6 dB

III.5.2 Structure Directe améliorée

Deux versions améliorées de la structure Directe sont présentées dans cette partie, celle développée par Djendi dans [35] et l'autre que nous proposons dans ce travail, nommée Propos2.

III.5.2.1 Structure développée par Djendi [35]

La complexité de la méthode d'Ikeda 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 de la structure mise en œuvre par Ikeda concernant le choix difficile des paramètres seuils, Djendi a proposé de remplacer la structure Récursive dans le

modèle d'Ikeda par une structure Directe car cette dernière est facile à contrôler (voir figure III.9).

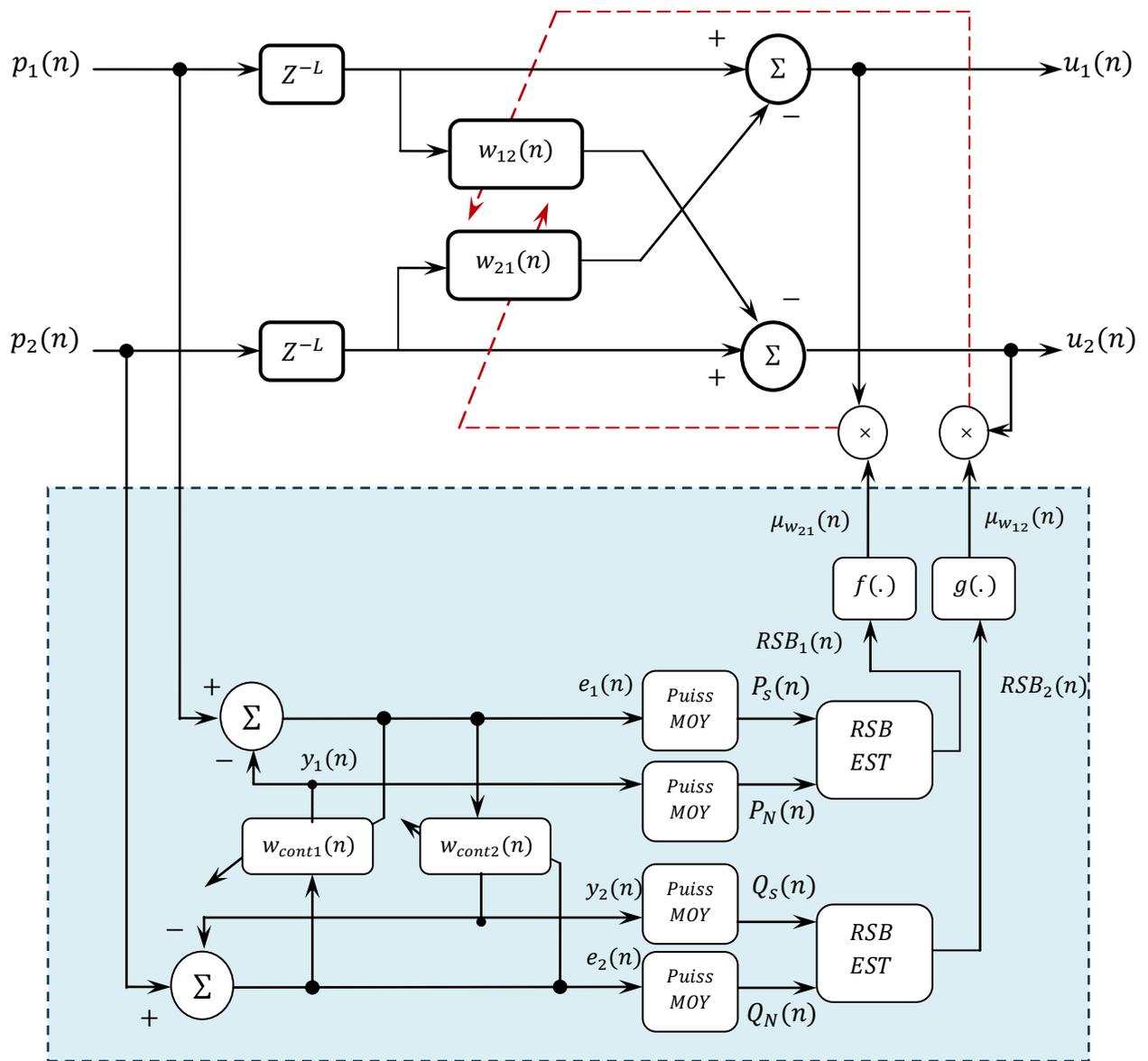


Figure III.9 : Structure développée par Djendi (Method2) [35].

Le tableau III.3, représente les valeurs des paramètres utilisés dans cette structure.

Tableau III.3 : Paramètres utilisés pour la méthode de Djendi (Method2).

Paramètres	Valeurs
μ_{wcont1}	0.4
μ_{wcont2}	0.001
μ_{1min}	0
μ_{1max}	0.2
RSB_{1min}	-35 dB
RSB_{1max}	-15 dB
μ_{2min}	0
μ_{2max}	0.02
RSB_{2min}	-10 dB
RSB_{2max}	0 dB

III.5.2.2 La deuxième structure proposée (Propos2)

Dans [35] la structure directe a été contrôlée par une technique de détection d'activité vocale basée sur les rapports signaux à bruit estimés à partir d'une instruction backward, dans cette version, nous proposons de contrôler la structure Directe avec une structure de type forward (voir figure III.10).

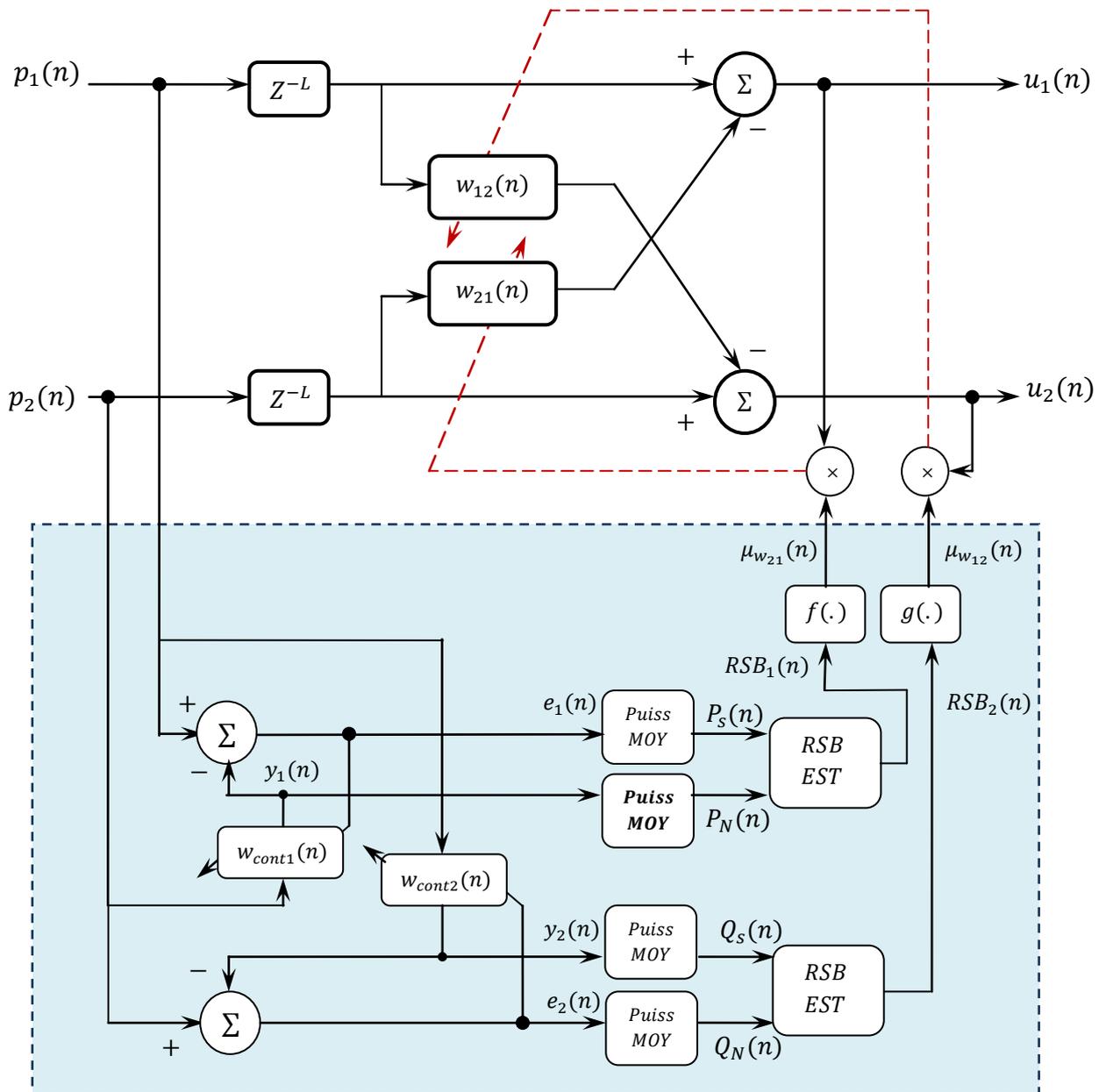


Figure III.10 : la deuxième structure proposée (Propos2).

Les valeurs des paramètres seuils utilisées dans cette structure sont représentées dans le tableau III.4.

Tableau III.4: Paramètres utilisés pour la structure Propos2.

Paramètres	Valeurs
μ_{wcont1}	0.4
μ_{wcont2}	0.001
μ_{1min}	0
μ_{1max}	0.2
RSB_{1min}	-40 dB
RSB_{1max}	-15 dB
μ_{2min}	0
μ_{2max}	0.02
RSB_{2min}	-5 dB
RSB_{2max}	2 dB

III.6 Conclusion

Dans ce chapitre, nous avons présenté deux techniques de segmentation du signal de parole utilisées pour les deux structures de séparation de sources Directe et Réursive. La première technique repose sur un mécanisme de détection d'activité vocale (DAV) manuel pour contrôler l'adaptation des filtres $w_{21}(n)$ et $w_{12}(n)$, la deuxième technique est une segmentation automatique qui utilise des pas d'adaptation variables en fonction du rapport signal à bruit.

Afin de confronter cette étude théorique avec les résultats expérimentaux, nous présentons dans le chapitre suivant les résultats de simulations obtenus à l'aide des structures de séparation de sources améliorées (Directe et Réursive).

IV.1 Introduction

Nous présentons dans ce chapitre une validation expérimentale des deux nouvelles structures (Propos1 et Propos2). Ces deux dernières ont été présentées au sein du chapitre précédent.

Les sorties de ces structures sont obtenues grâce à une nouvelle technique proposée de segmentation (DAV) automatique.

L'organisation de ce chapitre est la suivante :

Nous définissons, en premier lieu, le cadre expérimental de nos simulations à travers notamment l'utilisation d'un modèle de mélange qui a été proposé dans [36]. Les différents signaux sources (parole, bruit) utilisés seront également précisés.

Nous présentons dans un second temps, les résultats complets de simulations du comportement des deux structures proposées Propos1 et Propos2 avec une étude comparative de chaque structure avec d'autres méthodes qui existent dans la littérature, et ceci dans la situation de capteurs espacés.

Une évaluation des effets de certains paramètres (L : la taille du filtre, μ : pas d'adaptation, RSB : rapport signal à bruit) sur les deux structures proposées seront détaillées.

IV.2 Signaux utilisés en simulation

Dans l'ensemble des simulations présentées dans ce chapitre, une fréquence d'échantillonnage de 16 KHz a été utilisée. Un signal de parole échantillonné à 16 KHz d'environ 4 secondes prononcé par un locuteur masculin suivant « un loup s'est jeté immédiatement sur la petite chèvre » utilisé pour obtenir les résultats dans ce chapitre, il est présenté dans la figure ci-dessous.

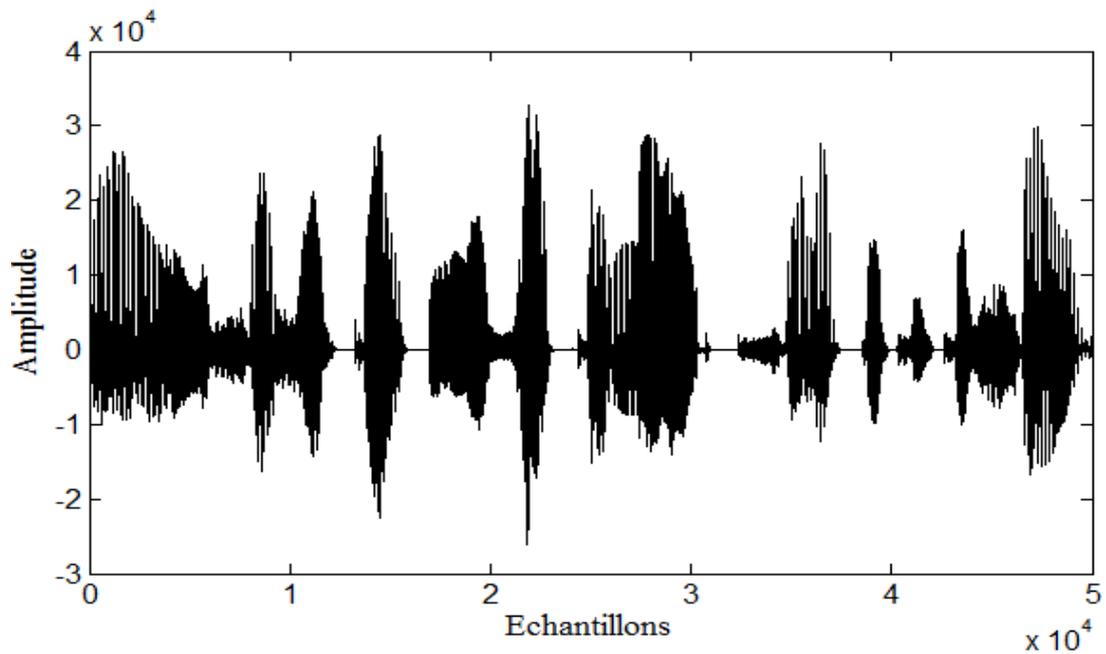


Figure IV.1 : Signal de parole utilisé, $F_e = 16$ KHz.

En ce qui concerne la seconde source (bruit stationnaire), celle-ci correspond à un bruit Blanc qui 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 (IV.2).

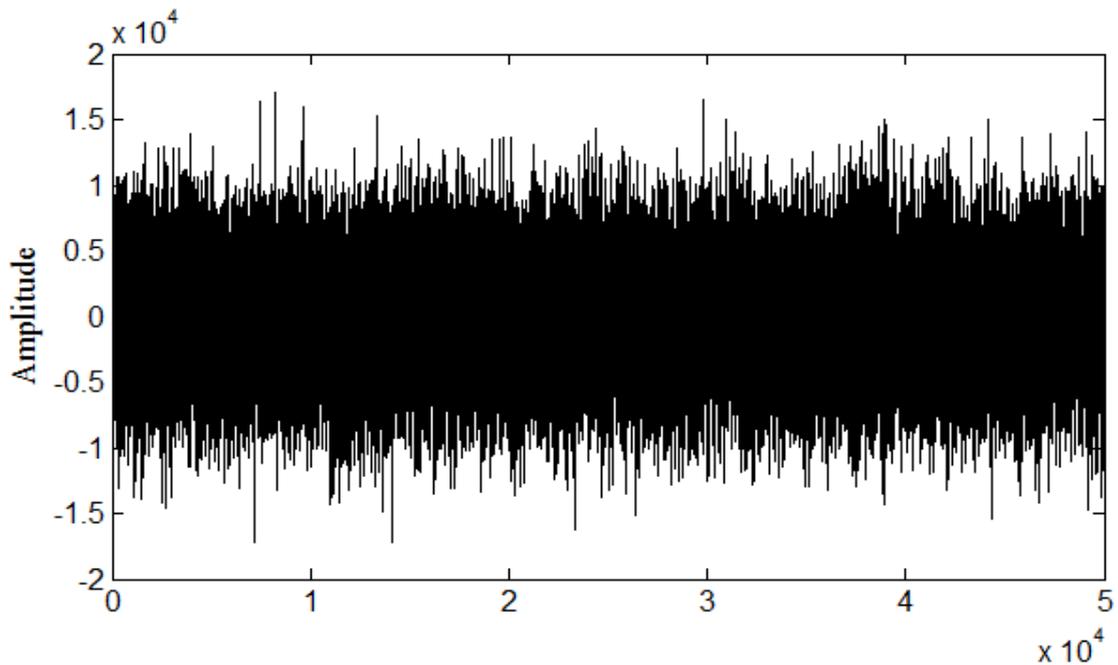


Figure IV.2 : Signal du bruit Blanc utilisé, $F_e = 16$ KHz.

IV.3 Simulation de réponses impulsionnelles

Nous utilisons dans cette partie une implémentation spécifique du modèle de mélange de la figure (III.1) du chapitre précédent, ce modèle étant rappelé sur la figure ci-dessous. 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.

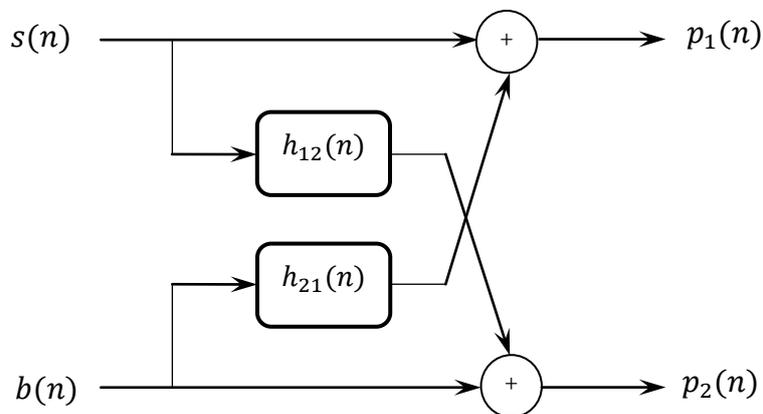


Figure IV.3 : Modèle de mélange.

Les réponses impulsionnelles du couplage croisé $h_{12}(n)$ et $h_{21}(n)$ simulées sont donc construites selon [36] comme suit :

$$h_{12}(n) = \delta(n) + h'_{12}(n) \quad (\text{IV.1})$$

$$h_{21}(n) = \delta(n) + h'_{21}(n) \quad (\text{IV.2})$$

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 :

$$f(n) = Ae^{-Bn} \quad (\text{IV.3})$$

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 :

$$B = 3 \ln(10) / n_r = 6.9078 / n_r \quad (\text{IV.4})$$

Les amplitudes des réponses impulsionnelles $h'_{12}(n)$ et $h'_{21}(n)$ sont ajustées à partir de la variance d'un bruit aléatoire qui est lié physiquement à l'espacement entre les deux microphones.

Les figures (IV.4) et (IV.5) 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 [36] ce qui correspond à une configuration de microphones relativement espacés. Le facteur d'atténuation est pris égal à $B = 0.028$ [36], avec une période d'échantillonnage $T_e = 625 \mu s$. Le temps de réverbération correspondant est de 30.8 ms [36]. 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.

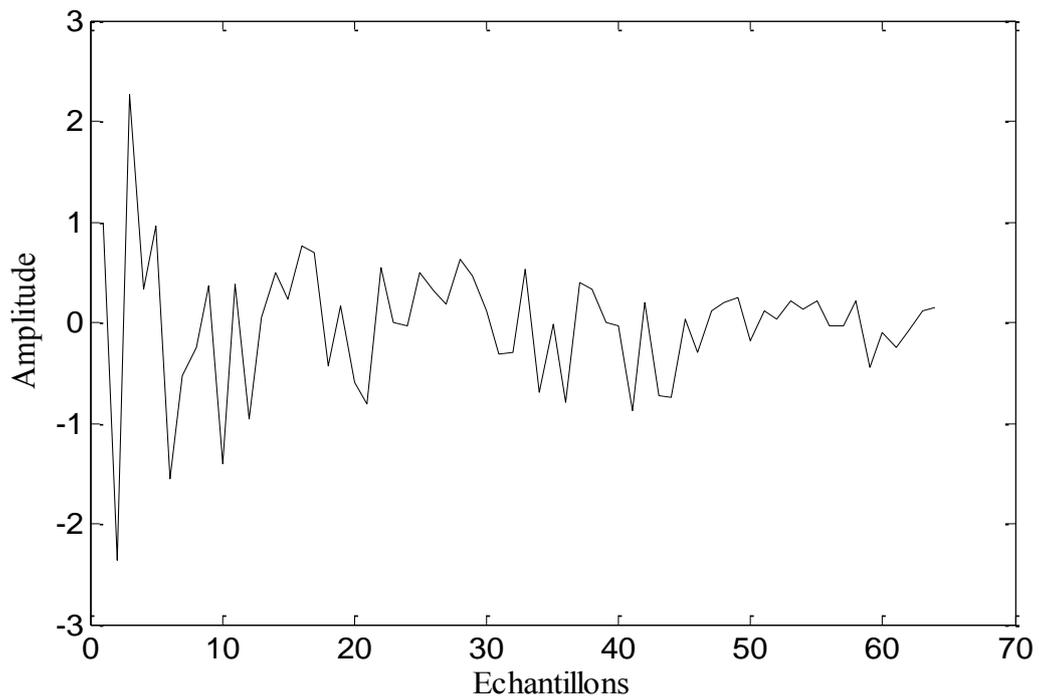


Figure IV.4 : Exemples d'une réponse impulsionnelle simulée ($h_{12}(n)$), cas des capteurs espacés.

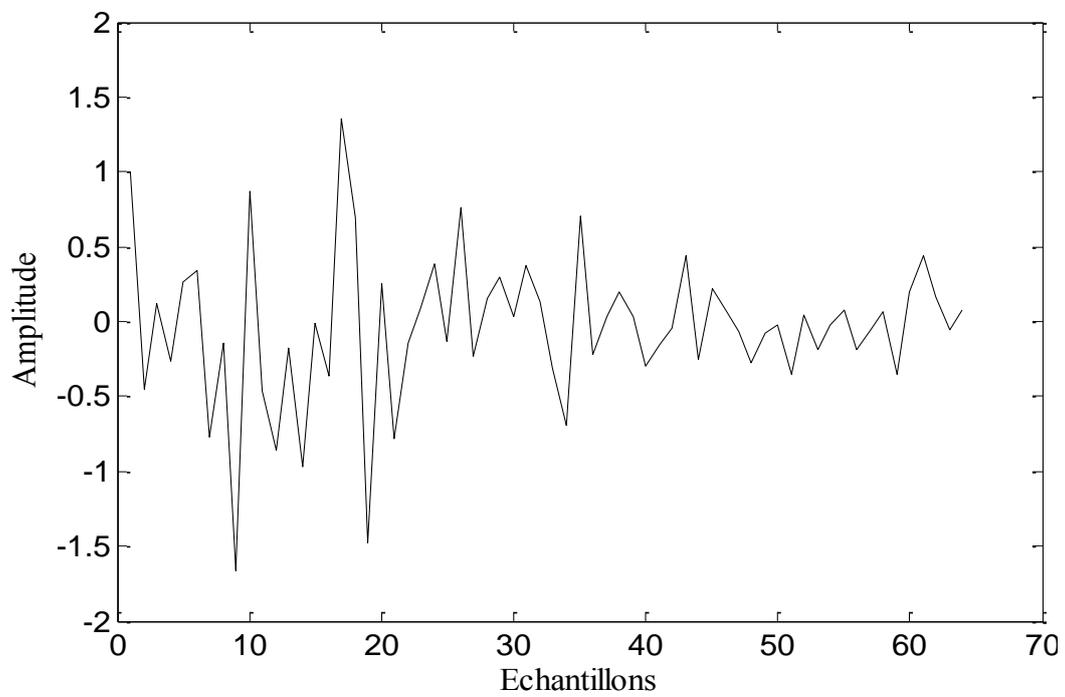


Figure IV.5 : Exemples d'une réponse impulsionnelle simulée ($h_{21}(n)$), cas des capteurs espacés.

IV.4 Détection d'activité vocale (DAV)

Il a déjà été noté que les signaux en sortie de la structure classique de séparation symétrique de sources de type Directe (FBSS) ou Récursive (BBSS) sont obtenus à partir d'un système d'aiguillage (c'est-à-dire que, selon la configuration souhaitée par l'utilisateur, il est possible de maîtriser le fait que l'estimée d'un signal source soit disponible sur la sortie primaire ou secondaire de cette structure).

Les deux structures classiques de séparation de sources utilisent un mécanisme de Détection d'Activité Vocale (DAV) manuel pour contrôler l'adaptation des filtres $w_{21}(n)$ et $w_{12}(n)$. Cette adaptation est contrôlée de la manière suivante : le filtre $w_{21}(n)$ est adapté uniquement en période de bruit seul, alors que le filtre $w_{12}(n)$ est adapté uniquement pendant les périodes d'activité vocale (période de présence du signal de parole utile). La figure (IV.6) présente un exemple de DAV manuelle sur le signal de parole originale.

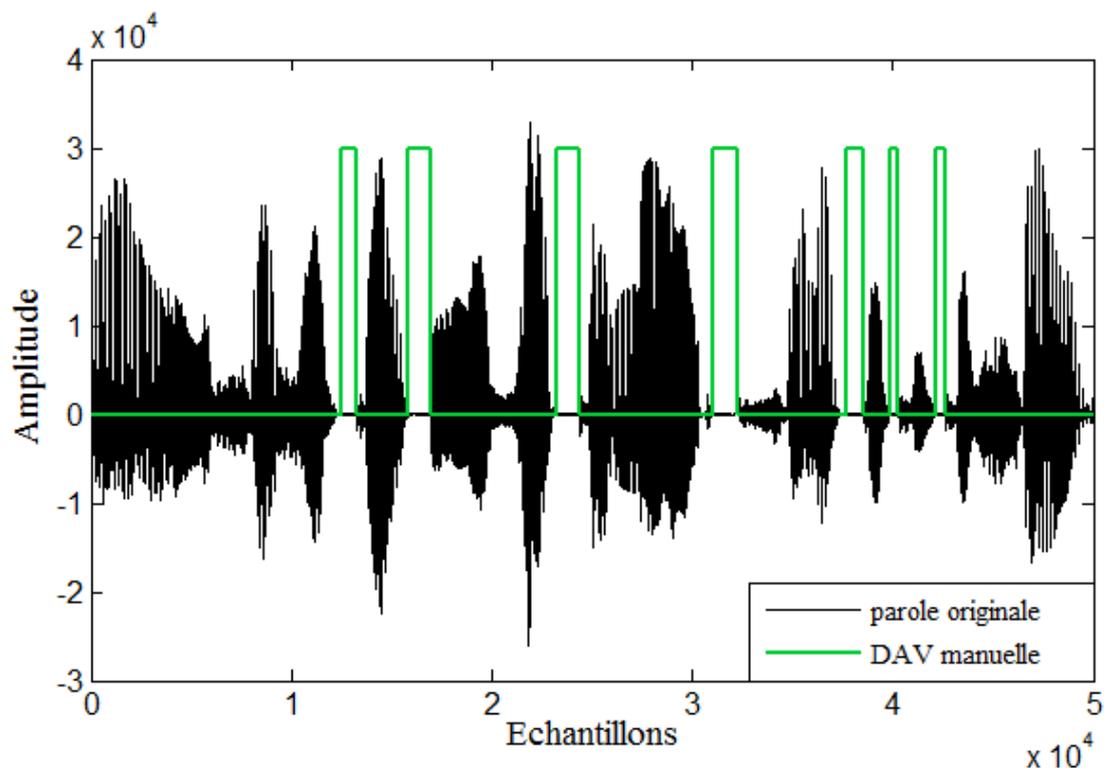


Figure IV.6 : Signal de parole utilisé avec une segmentation (DAV) manuelle.

IV.5 Critères objectifs de qualité

L'évaluation des performances des techniques de réduction de bruit représente une véritable problématique. En effet, ce sont des êtres humains qui sont amenés à juger de la qualité des traitements, jugement qui est donc par essence subjectif. Malheureusement l'évaluation de la qualité subjective est coûteuse en termes de temps et de ressources, il est par contre possible d'utiliser des mesures objectives qui sont beaucoup plus pratiques et moins coûteuses.

Les mesures objectives de qualité des signaux vocaux utilisées dans ce mémoire sont :

- Critère du rapport signal à bruit.
- Critère de la distance cepstrale.
- Critère du désajustement des coefficients.

IV.5.1 Le critère du Rapport Signal à Bruit (RSB)

Le critère du rapport signal à bruit donne une information globale (pleine bande) du rapport entre la puissance du signal de parole rehaussé et celle du bruit, suivant :

$$\text{RSB} = 10 \log_{10} \left\{ \frac{P_S(n)}{P_N(n)} \right\} \quad (\text{IV.1})$$

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.

IV.5.2 Le critère de la Distance Cepstrale (CD)

La distance cepstrale 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 :

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

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.

IV.5.3 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:

$$(SM)_{dB} = 10 * \log_{10} \left(\frac{\|\hat{w}_{ij} - h_{ij}\|^2}{\|h_{ij}\|^2} \right) \quad \text{où } i \neq j \in \{1,2\} \quad (IV.3)$$

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}).

IV.6 Simulation de la première structure proposée (Propos1)

Ce paragraphe présente les résultats obtenus à partir de la structure proposée Propos1 décrite au chapitre III (voir figure III.8), et qui correspond à une structure de séparation de sources Réursive 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 Double, noté DNLMS pour adapter chaque filtre de cette structure.

Les paramètres de simulations pour la structure (Propos1) dans le cas de microphones espacés sont représentés dans le tableau IV.1.

Tableau IV.1 : Paramètres de simulation de la structure Propos1.

Structure	Paramètres
Propos1	<p>Longueur des filtres : $L_{w12} = L_{w21} = L_{wcont1} = L_{wcont2} = 64$</p> <p>Pas d'adaptation des filtres secondaires: $\mu_{wcont1} = 0.4 ; \mu_{wcont2} = 0.001$</p> <p>Pas d'adaptation des filtres principaux: $\mu_{w12min} = 0 ; \mu_{w12max} = 0.02 ;$ $\mu_{w21min} = 0 ; \mu_{w21max} = 0.2$</p> <p>Valeurs seuils du RSB estimés : $RSB_{1min} = -40 \text{ dB} ; RSB_{1max} = -15 \text{ dB} ;$ $RSB_{2min} = -1 \text{ dB} ; RSB_{2max} = 6 \text{ dB}$</p>

Nous notons que, le contexte de simulation des signaux de mélange est le suivant : le rapport de la puissance moyenne du signal à celle du bruit est fixé à 3 dB (sur la voie 1) et à 0 dB (sur la voie 2).

Le signal de parole originale, le signal de bruit Blanc et les signaux de mélanges sont représentés avec leurs spectrogrammes sur les figures VI. (7, 8, 9, 10), les signaux de sorties de la première structure proposée Propos1 sont représentés avec leurs spectrogrammes sur les figures (IV.11) et (IV.12).

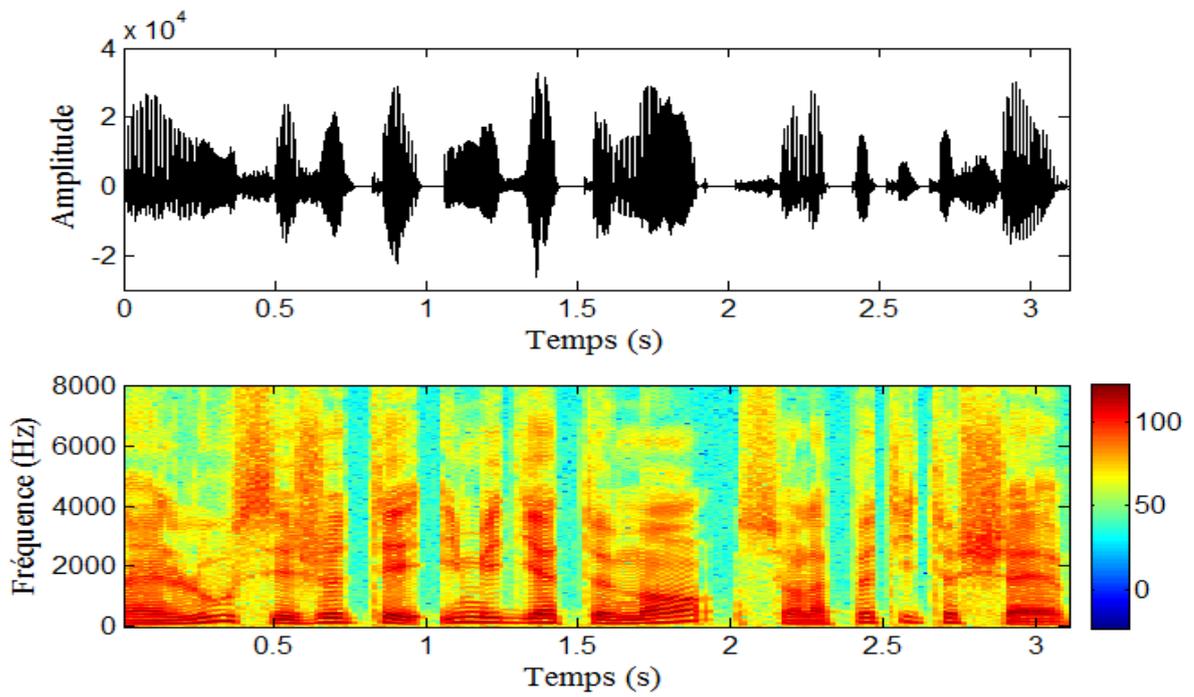


Figure IV.7 : Signal original de parole $s(n)$ en haut avec son spectrogramme en bas.

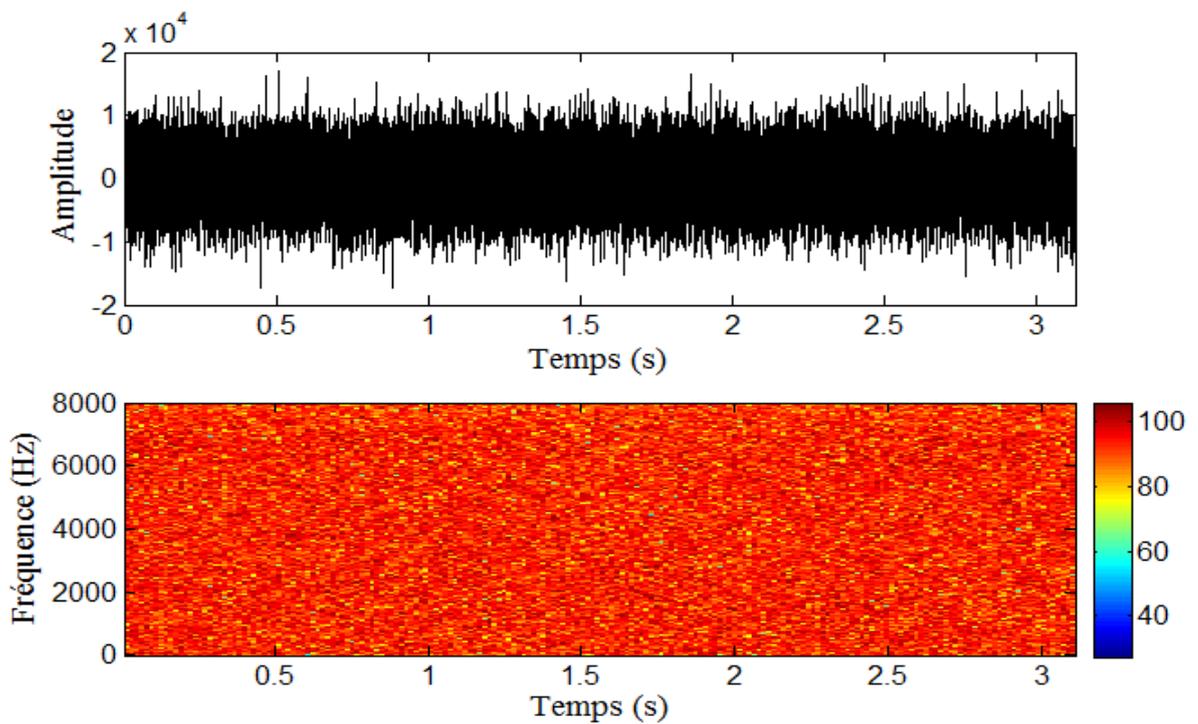


Figure IV.8 : Signal de bruit Blanc $b(n)$ en haut avec son spectrogramme en bas.

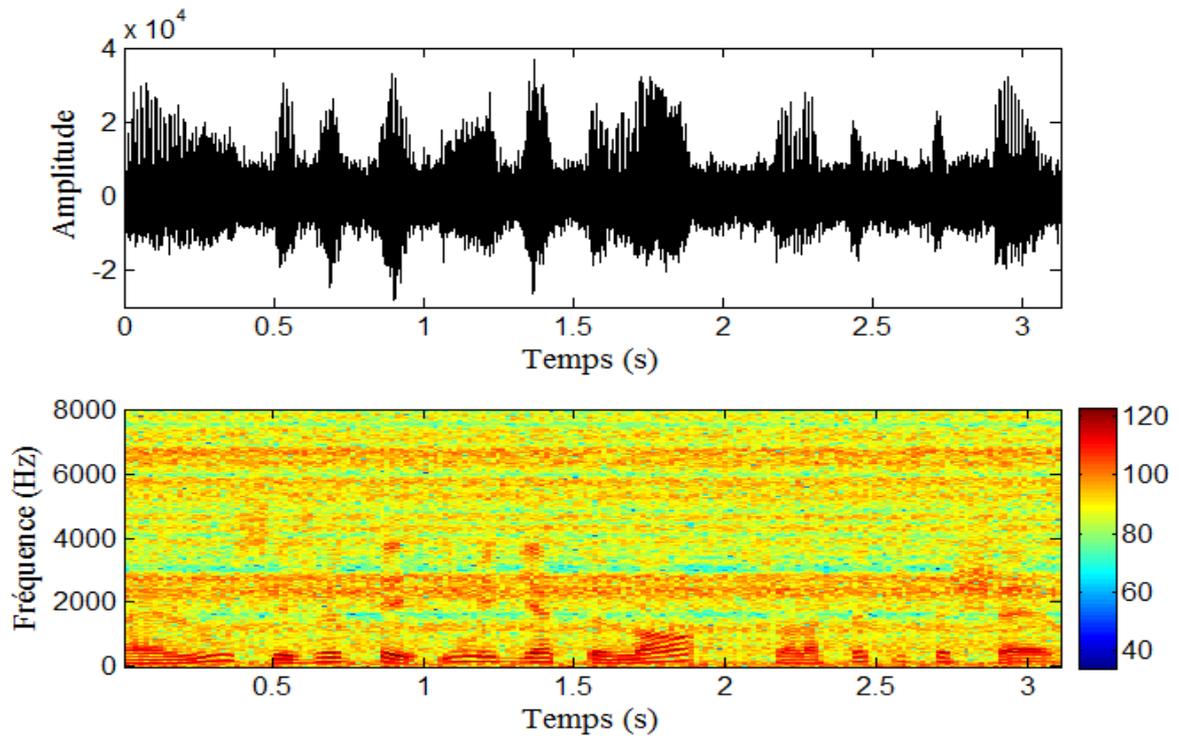


Figure IV.9 : Signal de mélange $p_1(n)$ en haut avec son spectrogramme en bas.

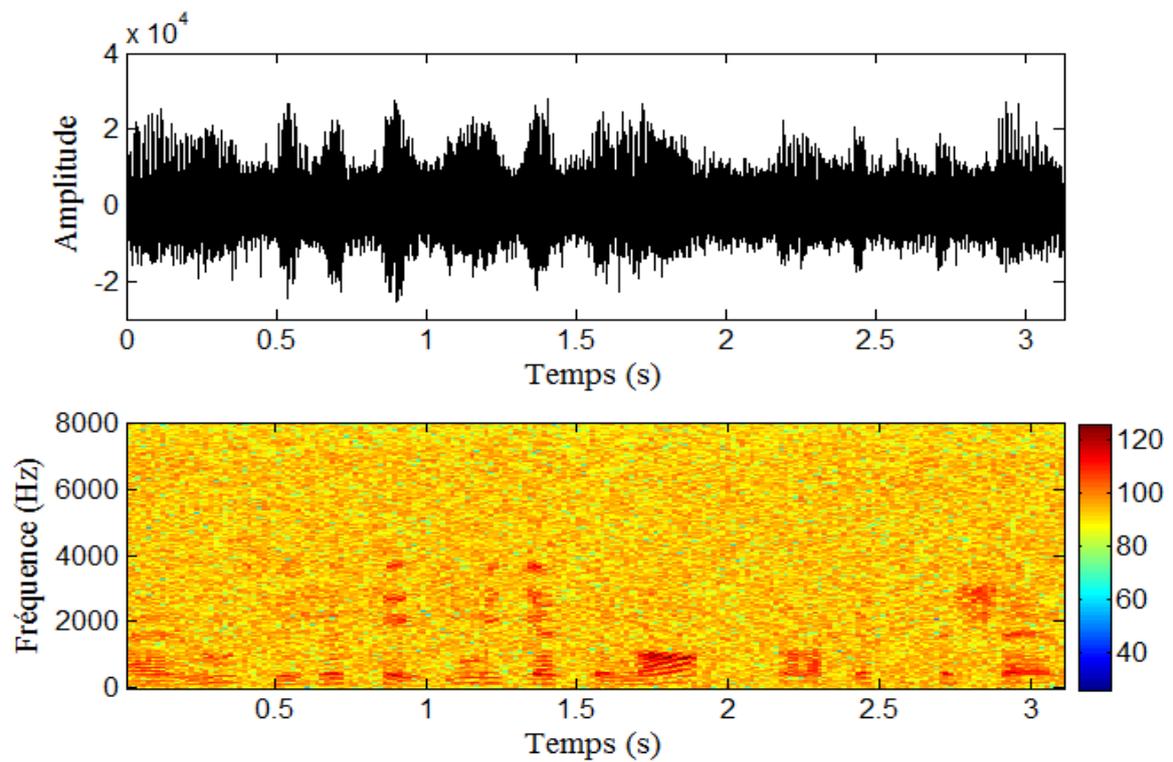


Figure IV.10 : Signal de mélange $p_2(n)$ en haut avec son spectrogramme en bas.

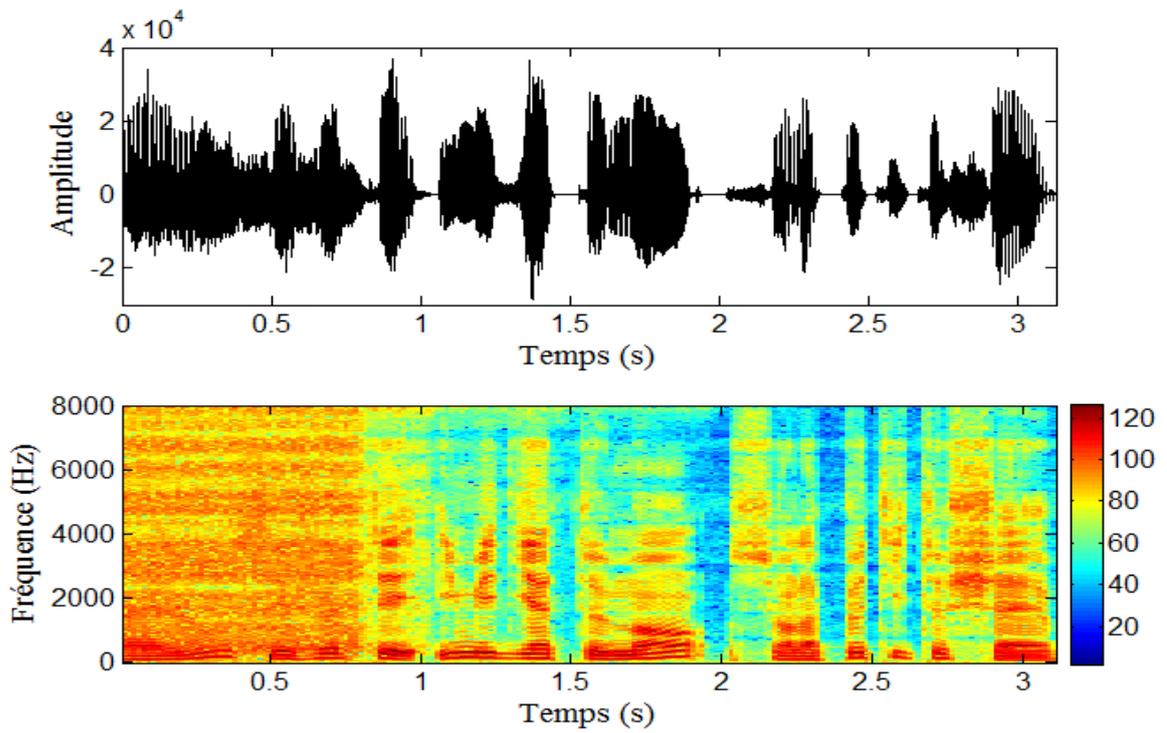


Figure IV.11 : Signal de sortie $v_1(n)$ en haut avec son spectrogramme en bas.

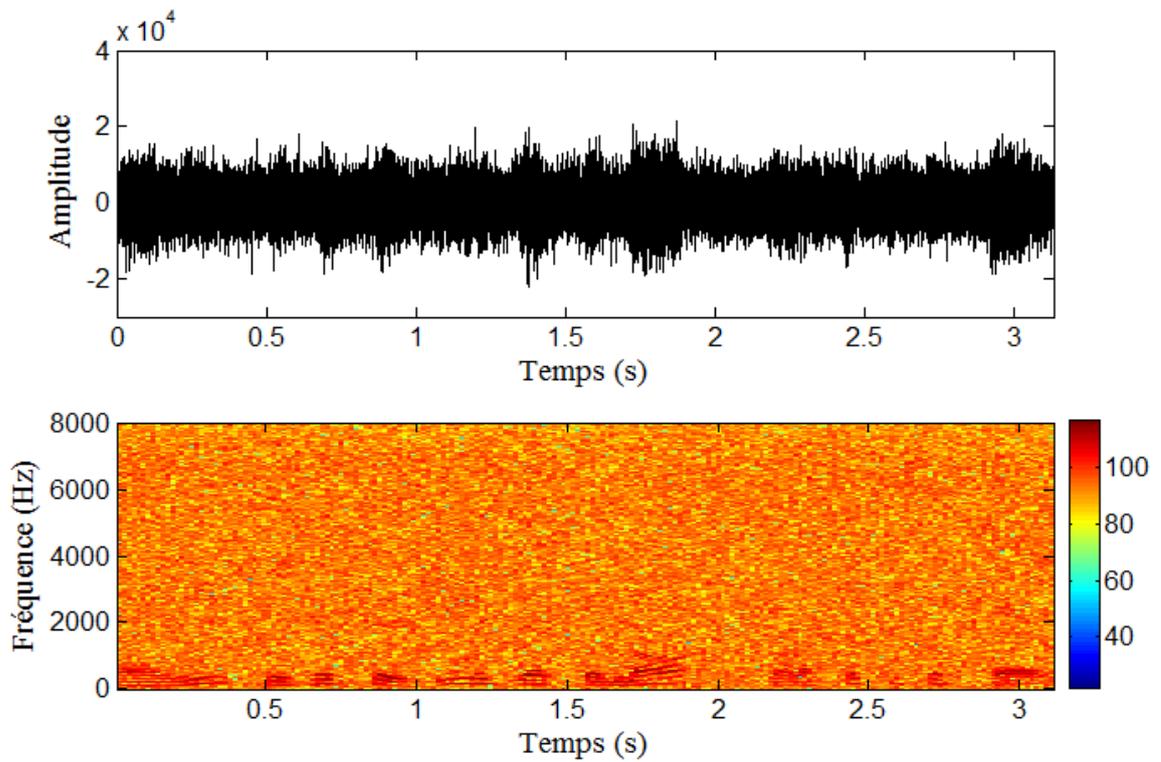


Figure IV.12: Signal de sortie $v_2(n)$ en haut avec son spectrogramme en bas.

A partir de la figure (IV.11), nous observons que le signal de parole en sortie $v_1(n)$ 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 $v_1(n)$.

D'après la figure (IV.12), nous remarquons que la parole est partiellement enlevée de la deuxième sortie $v_2(n)$, 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 Propos1 utilise un système de détection d'activité vocale automatique décrit précédemment dans le chapitre (III) 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 (IV.13) 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 $w_{21}(n)$.

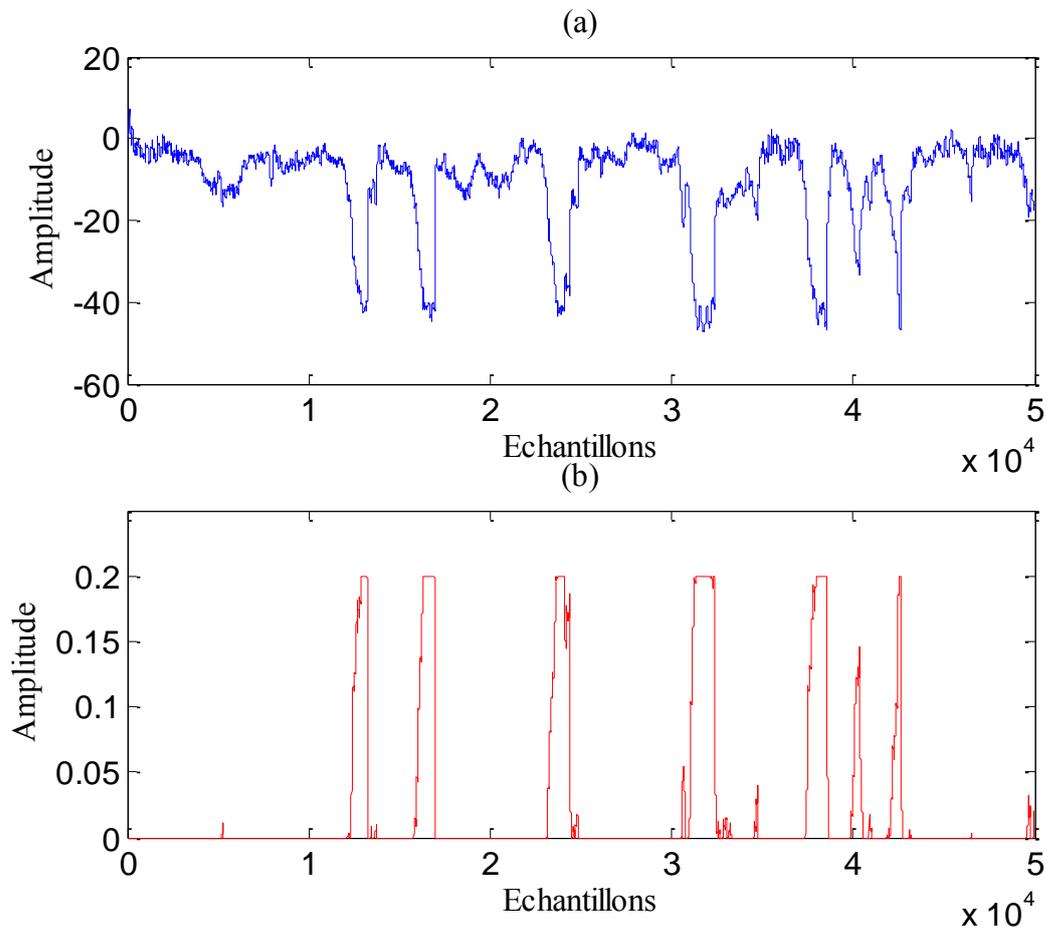


Figure IV.13 : (a) Le RSB_1 estimé au niveau de l'entrée primaire (Propos1) et (b) le pas d'adaptation $\mu_{w_{21}}(n)$ du filtre principal $w_{21}(n)$.

A partir des figures IV.13 (a et b), 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 respectivement sur les figures IV.14 (c et d).

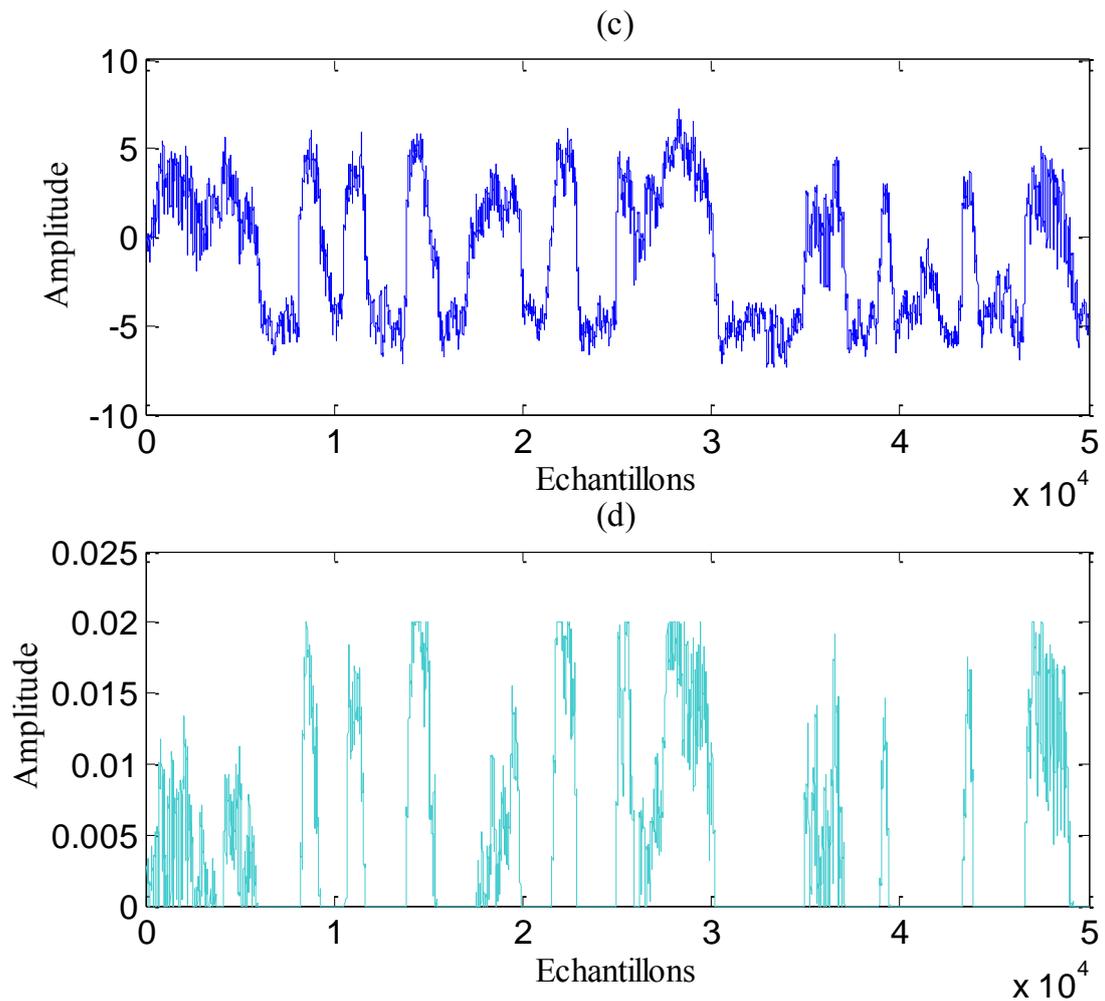


Figure IV.14 : (c) Le RSB_2 estimé au niveau de l'entrée de référence (Propos1) et (d) le pas d'adaptation $\mu_{w_{12}}(n)$ du filtre principal $w_{12}(n)$.

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 IV.15) l'évolution des pas d'adaptations ($\mu_{w_{12}}(n)$, $\mu_{w_{21}}(n)$) des filtres principaux.

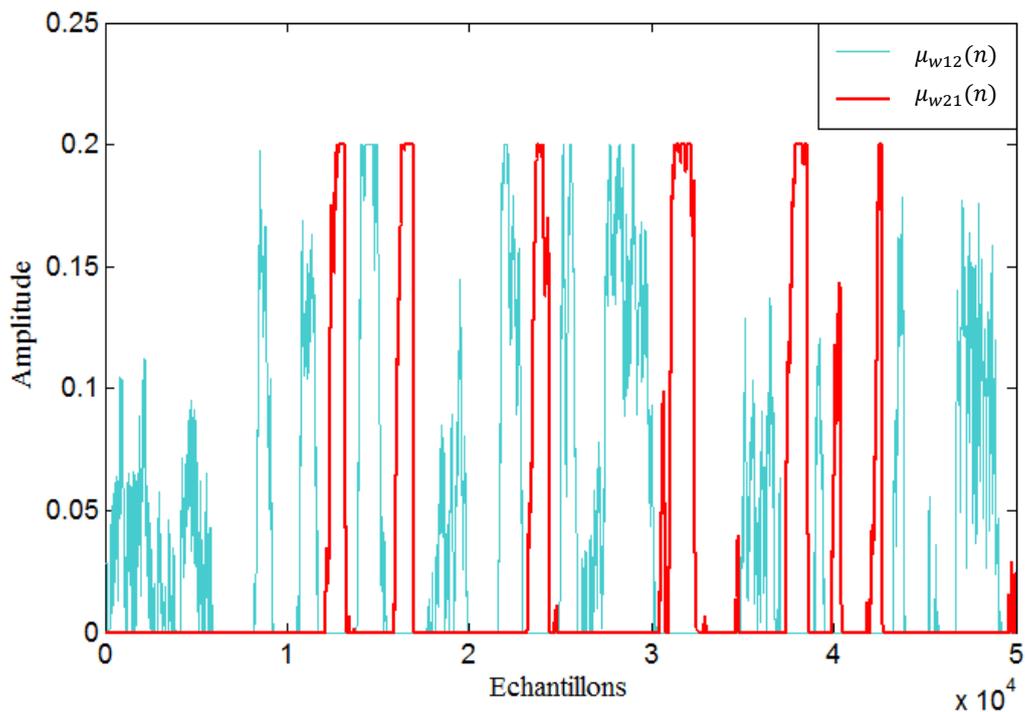


Figure IV.15 : Comportement des pas d'adaptations: $\mu_{w12}(n)$ (en bleu) et $\mu_{w21}(n)$ (en rouge).

A partir de cette figure, nous pouvons voir que l'ajustement du pas d'adaptation $\mu_{w21}(n)$ est désactivé pendant les périodes où l'ajustement du pas d'adaptation $\mu_{w12}(n)$ est permis, ces pas sont complémentaires. Par conséquent ces deux pas d'adaptation ($\mu_{w21}(n)$ et $\mu_{w12}(n)$) 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 $\mu_{w21}(n)$.

Nous représentons sur la figure (IV.16) les deux segmentations (DAV manuelle et automatique).

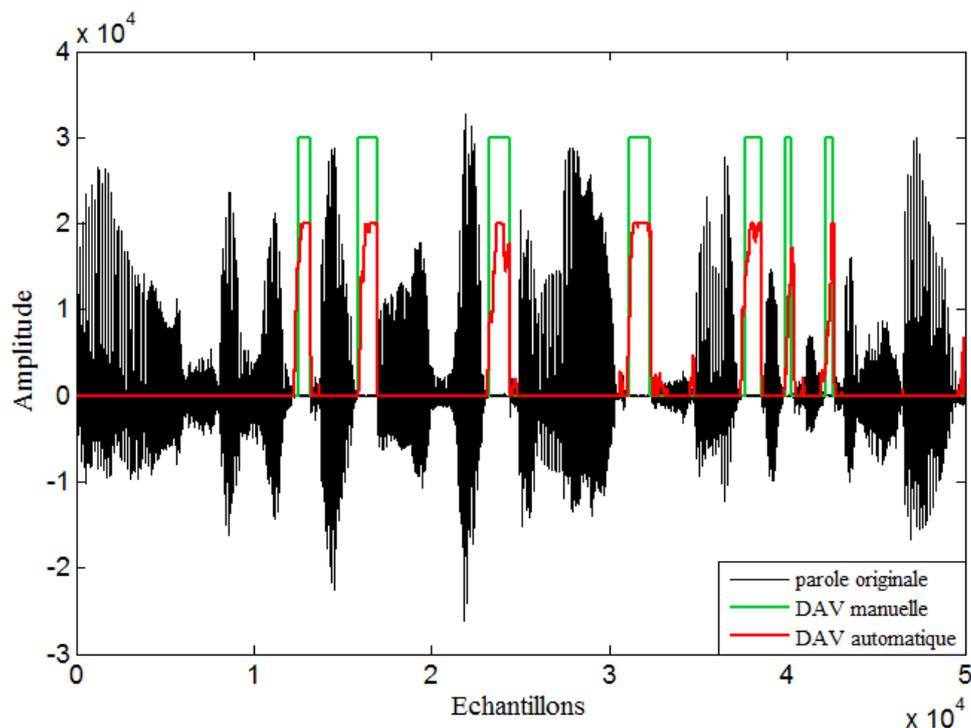


Figure IV.16 : Signal de parole utilisé avec les deux segmentations (manuelle et automatique (Propos1)), $F_e = 16$ KHz.

D'après la figure (IV.16), 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.

IV.7 Etude des performances de la première structure proposée (Propos1)

Pour améliorer le comportement de la première structure proposée (Propos1) dans le domaine de débruitage du signal de parole. Nous avons réalisé plusieurs simulations pour voir l'influence des paramètres (L , μ et RSB) sur les performances de cette structure. C'est-à-dire l'étude de performances de l'algorithme DNLMS en termes de vitesse de convergence.

Ces paramètres sont :

L : taille du filtre adaptatif,

μ : pas d'adaptation de l'algorithme DNLMS,

RSB_1, RSB_2 : rapport signal à bruit sur les deux entrées de la structure proposée.

IV.7.1 Effet de la taille du filtre adaptatif

Pour mieux voir l'effet de la taille du filtre L sur la structure proposée (Propos1), nous avons varié ce dernier ($L=32$; $L=64$; $L=128$). Les autres paramètres sont égaux 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 le résultat montré sur la figure (IV.17).

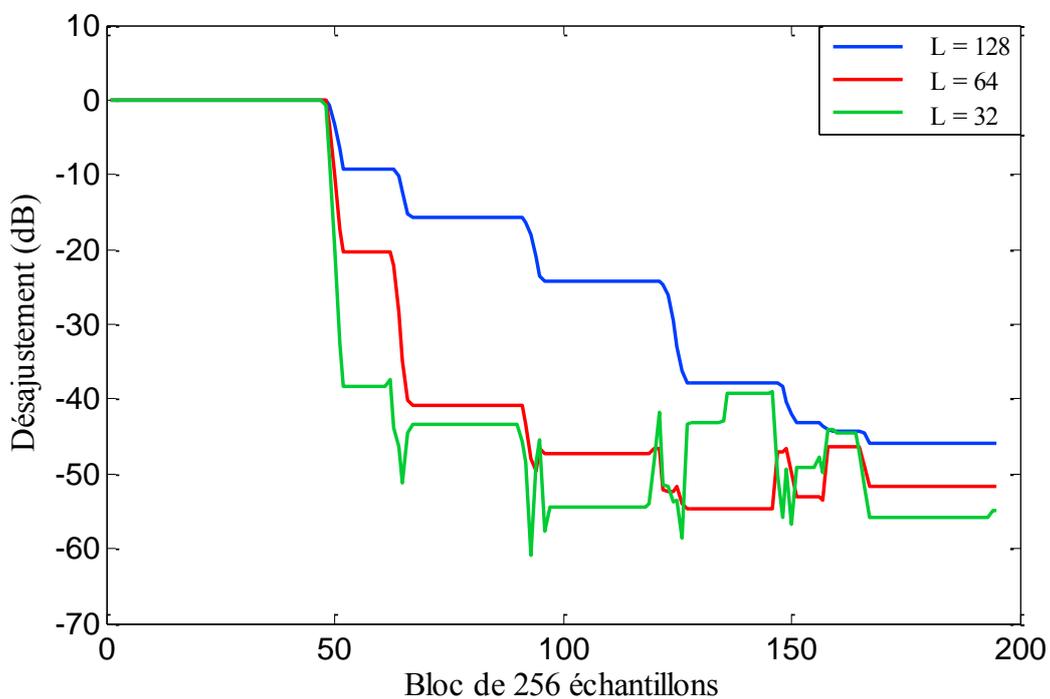


Figure IV.17 : Effet de la taille du filtre sur la première structure proposée (Propos1) pour $\mu=0.4$, $RSB_1=3$ dB et $RSB_2=0$ dB.

D'après cette figure, nous remarquons que l'algorithme DNLMS converge mieux lorsque la taille du filtre L est faible, ce qui montre que la vitesse de convergence de l'algorithme DNLMS est inversement proportionnelle à la taille du filtre adaptatif.

IV.7.2 Effet du pas d'adaptation

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

$RSB_1 = 3 \text{ dB}$ et $RSB_2 = 0 \text{ dB}$.

Les résultats du désajustement de cette expérience sont représentés sur la figure (IV.18):

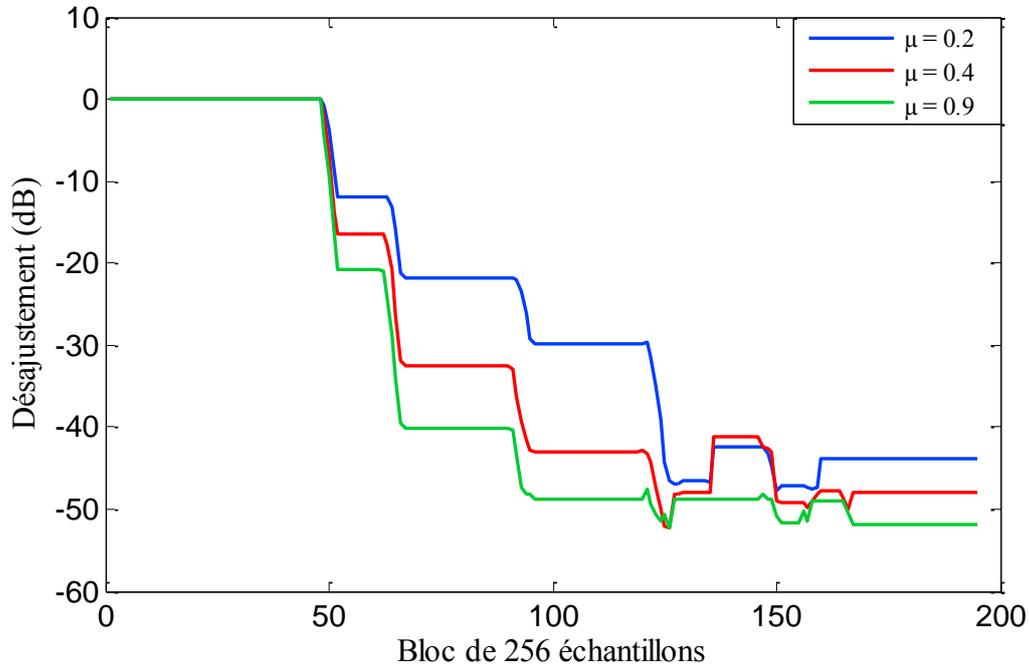


Figure IV.18 : Effet du pas d'adaptation sur la première structure proposée (Propos1) pour $L=64$, $RSB_1=3 \text{ dB}$ et $RSB_2=0 \text{ dB}$.

D'après cette figure, nous remarquons que l'algorithme DNLMS converge rapidement lorsque le pas d'adaptation μ est grand, ce qui montre que la vitesse de convergence de l'algorithme DNLMS est proportionnelle au pas d'adaptation.

IV.7.3 Effet du RSB du mélange d'entrée

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 proposée. Les paramètres de cette simulation sont :

L : taille du filtre = 64, μ : pas d'adaptation = 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 (IV.19).

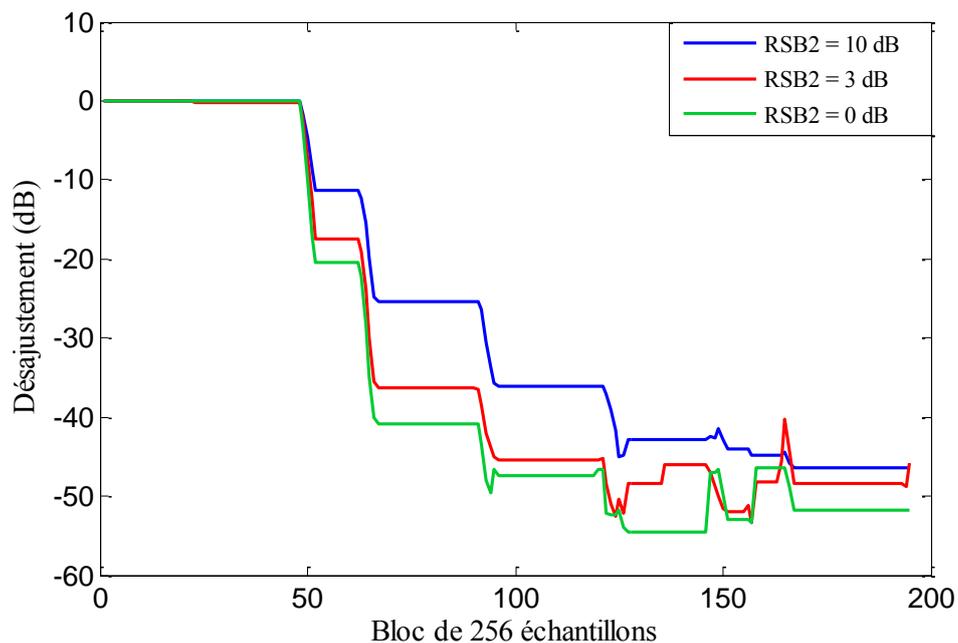


Figure IV.19 : Effet du RSB sur la structure proposée (Propos1) pour $L=64$, $\mu=0.4$ et $RSB_1=3$ dB.

D'après la figure (IV.19), nous avons remarqué que la vitesse de convergence de l'algorithme DNLMS vers l'optimum est obtenue lorsque le rapport signal à bruit est faible.

IV.8 Comparaison de la structure Propos1 avec deux structures de séparation de sources Réursive

Pour évaluer les performances de la structure proposée Propos1 qui utilise un système de détection d'activité vocale (DAV) automatique, nous avons réalisé une étude comparative de ses performances avec celles des deux structures de séparation de sources de type Réursive qui existent déjà dans la littérature. La première structure Réursive considérée correspond à une structure classique (BBSS), la seconde correspond à la structure d'Ikeda (Method1) [34].

Notons que la structure classique (BBSS) utilise une segmentation (DAV) manuelle dans son fonctionnement et la structure (Method1) utilise une DAV automatique.

Les paramètres de simulations pour la structure Réursive classique (BBSS) et la structure d'Ikeda (Method1) dans le cas des microphones espacés sont représentés dans le tableau IV.2.

Tableau IV.2 : Paramètres de simulations des structures BBSS et Method1.

Structures	Paramètres
BBSS	<p>Longueur des filtres: $L_{w12} = L_{w21} = 64$</p> <p>Pas d'adaptations des filtres : $\mu_{w12} = 0.4 ; \mu_{w21} = 0.4$</p>
Method1	<p>Longueur des filtres : $L_{w12} = L_{w21} = L_{wcont1} = L_{wcont2} = 64$</p> <p>Pas d'adaptation des filtres secondaires: $\mu_{wcont1} = 0.4 ; \mu_{wcont2} = 0.001$</p> <p>Pas d'adaptation des filtres principaux: $\mu_{w12min} = 0 ; \mu_{w12max} = 0.02 ;$ $\mu_{w21min} = 0 ; \mu_{w21max} = 0.2$</p> <p>Valeurs seuils du RSB estimés : $RSB_{1min} = -45 \text{ dB} ; RSB_{1max} = -15 \text{ dB} ;$ $RSB_{2min} = -13 \text{ dB} ; RSB_{2max} = -1 \text{ dB}$</p>

Sur la figure (IV.20), nous comparons les signaux de sorties obtenus avec les trois structures Propos1, BBSS et Method1. Dans cet exemple, nous montrons l'évolution temporelle des différents signaux utilisés en simulation (signal de parole originale, 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 trois structures fonctionnent bien, c'est-à-dire que le signal de parole en sortie est presque totalement débruité est très proche du signal de parole originale.

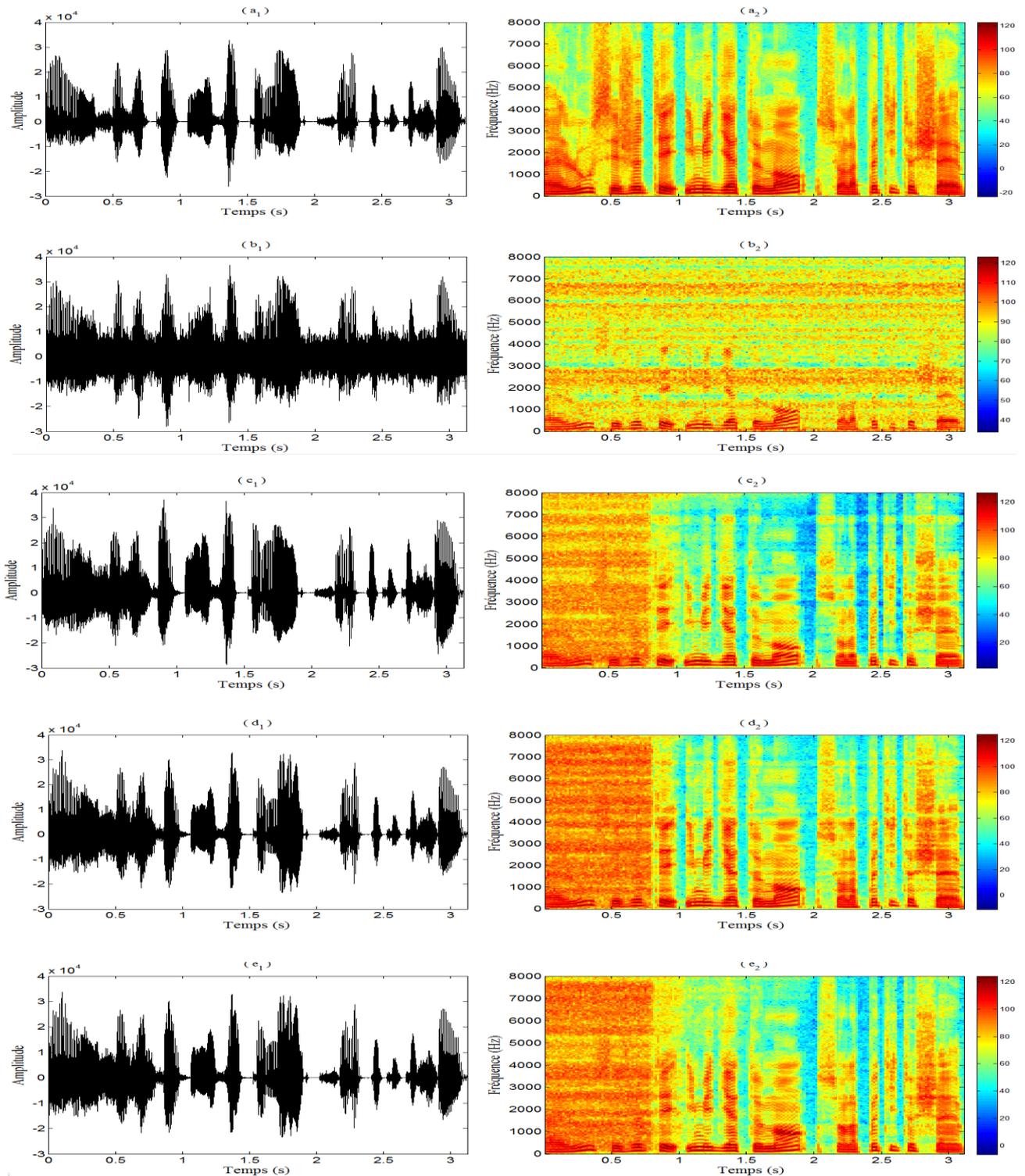


Figure IV.20 : (a₁, a₂) Signal de parole originale $s(n)$, (b₁, b₂) signal de mélange $p_1(n)$, (c₁, c₂) signal de sortie $v_1(n)$ de la structure (Propos1), (d₁, d₂) signal de sortie de la structure (BBSS), (e₁, e₂) signal de sortie de la structure (Method1).

Afin d'analyser les distorsions introduites par chacune des trois structures, nous avons représenté à la figure (IV.21) 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.

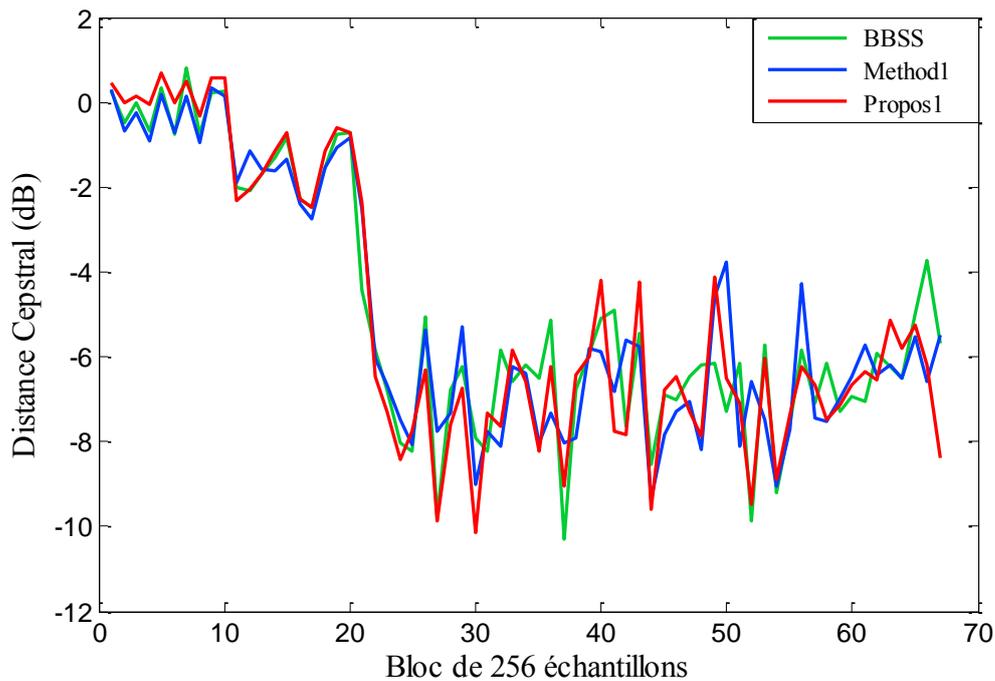


Figure IV.21 : Evolution du critère de la distance cepstrale des trois structures Propos1 (en rouge), Method1 (en bleu) et BBSS (en vert).

D'après cette figure, nous remarquons que les valeurs moyennes de la distance cepstrale sont de -7.86 dB, -7.65 dB et -7.80 dB respectivement pour les structures Propos1, Method1 et BBSS. Ce qui signifie que la distance cepstrale obtenu à partir de la structure proposée Propos1 est très proche de celles obtenus avec la structure classique BBSS et la structure d'Ikeda (Method1). Ces résultats montrent également le bon comportement de la structure proposée (Propos1).

La figure (IV.22) représente l'évolution temporelle du rapport signal à bruit pour les trois structures dans le cas des microphones espacés. 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, 50.56 dB et de 50.21 dB, respectivement pour la structure Propos1, la structure Method1 et la structure BBSS. Il existe donc un écart très faible entre les trois structures. Cela montre que la structure proposée (Propos1) fournit presque les mêmes performances que les autres structures (BBSS et Method1).

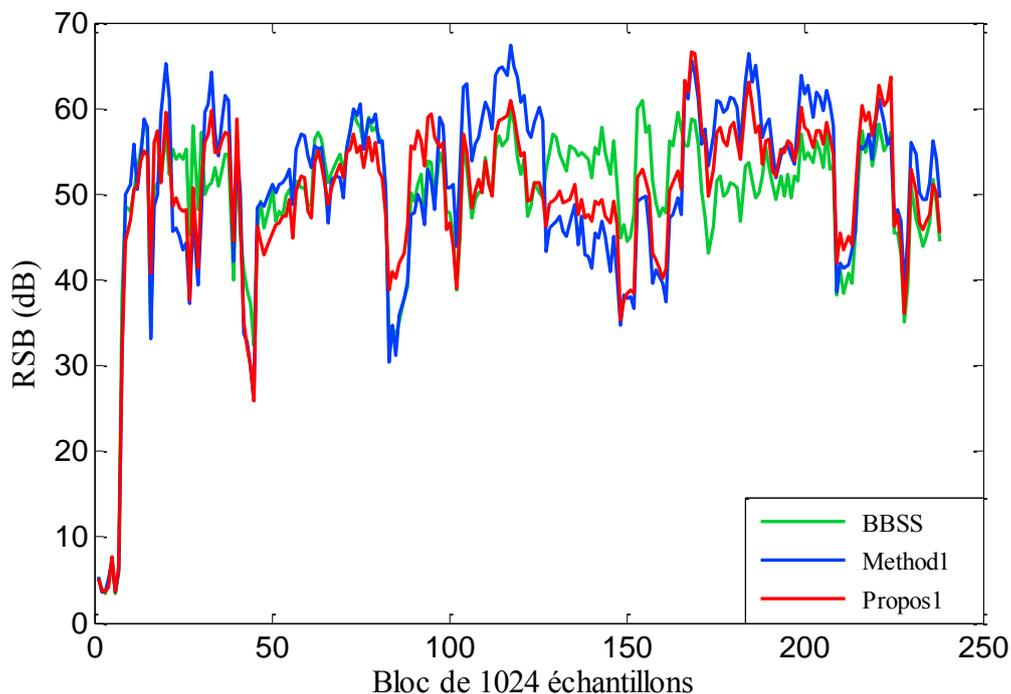


Figure IV.22 : Evolution du critère du rapport signal à bruit (RSB) pour les trois structures Propos1 (en rouge), Method1 (en bleu) et BBSS (en vert).

De manière à compléter l'analyse de comportement de la structure proposée Propos1, nous présentons sur la figure (IV.23) l'évolution temporelle du désajustement mesuré sur les coefficients du filtre adaptatif $w_{21}(n)$ pour les trois structures.

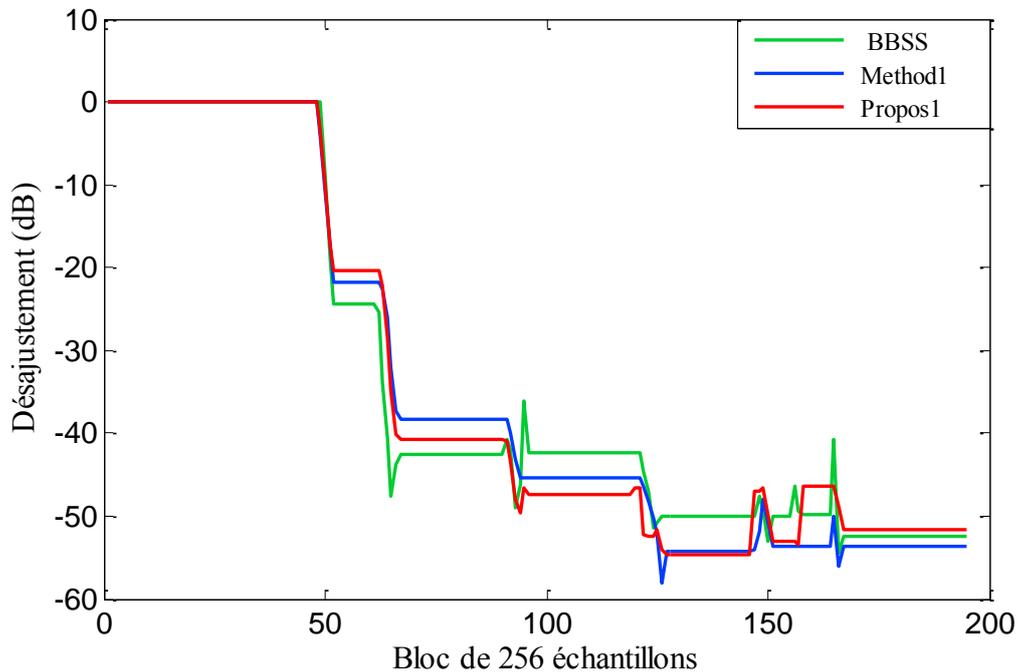


Figure IV.23 : Comparaison du désajustement obtenu avec les trois structures Propos1 (en rouge), Method1 (en bleu) et BBSS (en vert).

À partir du résultat de la figure (IV.23), nous notons que la vitesse de convergence est presque la même pour les trois structures. Ce qui démontre encore une fois le bon comportement de la nouvelle structure Propos1.

D'après les résultats obtenus avec la structure Récursive proposée (Propos1) qui se sont avérés très proches de ceux obtenus avec les structures Récursives qui existent dans la littérature (BBSS et Method1). Nous concluons que ces résultats sont très satisfaisants et confirment le bon comportement de la première structure de débruitage proposée dans ce mémoire.

Concernant la seconde structure proposée (Propos2), nous reprenons la même démarche d'étude et de simulation que celle initialement proposée (Propos1).

IV.9 Simulation de la deuxième structure proposée (Propos2)

Dans ce paragraphe, nous étudions la deuxième structure proposée (Propos2) qui correspond à une structure de séparation de sources Directe intégrant un bloc qui réalise la DAV, dont la description a été donnée au sein du chapitre III (voir figure III.10). Les signaux et les réponses impulsionnelles utilisés pour générer les signaux de mélange correspondent à ceux présentés au début du chapitre.

Dans cette étude, nous nous intéressons uniquement à la voie restituant le signal de parole utile (une seule des deux sorties de traitement est exploitée).

Les paramètres de simulations pour la structure (Propos2) dans le cas de microphones espacés sont représentés dans le tableau IV.3.

Tableau IV.3 : Paramètres de simulation de la structure Propos2.

Structure	Paramètre
Propos2	<p>Longueur des filtres : $L_{w12} = L_{w21} = L_{wcont1} = L_{wcont2} = 64$</p> <p>Pas d'adaptation des filtres secondaires: $\mu_{wcont1} = 0.4 ; \mu_{wcont2} = 0.001$</p> <p>Pas d'adaptation des filtres principaux: $\mu_{w12min} = 0 ; \mu_{w12max} = 0.02 ;$ $\mu_{w21min} = 0 ; \mu_{w21max} = 0.2$</p> <p>Valeurs seuils du RSB estimés : $RSB_{1min} = -40 \text{ dB} ; RSB_{1max} = -15 \text{ dB} ;$ $RSB_{2min} = -5 \text{ dB} ; RSB_{2max} = 2 \text{ dB}$</p>

Le signal de parole originale, le signal de mélange $p_1(n)$ et le signal disponible en sortie de la structure proposée (Propos2) sont représentés avec leurs spectrogrammes sur la figure (IV.24).

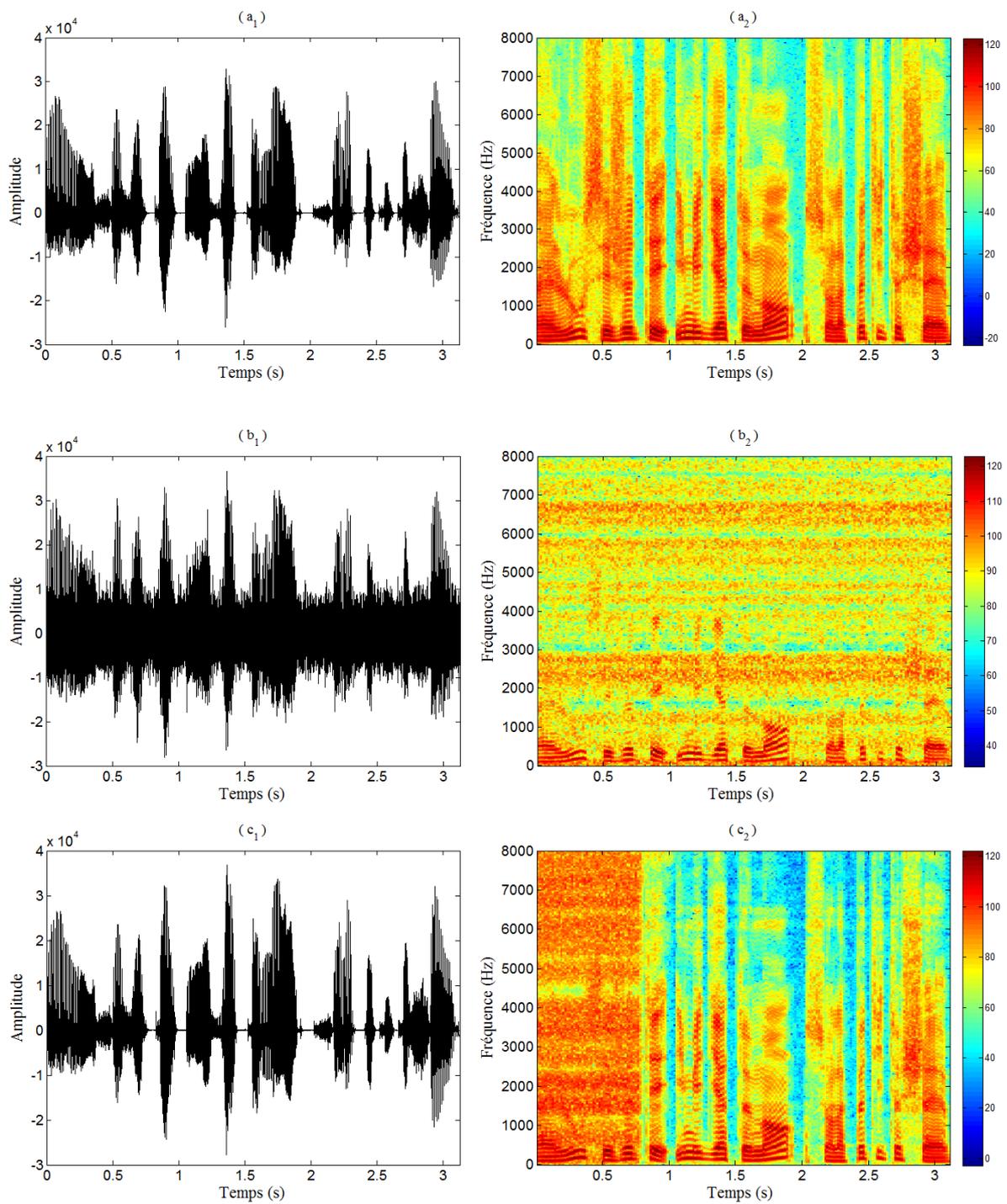


Figure IV.24 : (a₁, a₂) Signal de parole originale $s(n)$, (b₁, b₂) signal de mélange $p_1(n)$, (c₁, c₂) signal de sortie $u_1(n)$ de la structure (Propos2).

Nous pouvons également observer à partir de la figure IV.24, que le signal disponible en 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'une détection d'activité vocale (DAV) automatique comme indiqué auparavant.

L'évolution du rapport signal à bruit RSB_1 estimé 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 respectivement sur les figures IV.25 (a, b).

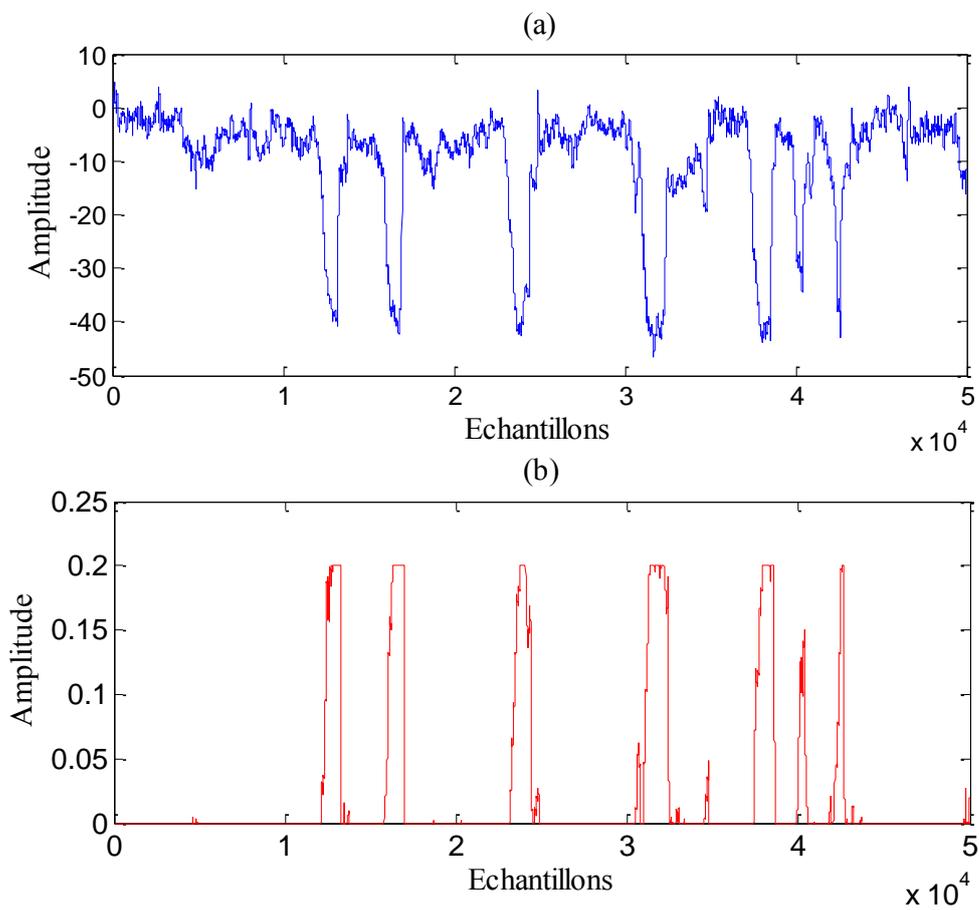


Figure IV.25 : (a) Le RSB_1 estimé au niveau de l'entrée primaire de la structure (Propos2) et (b) le pas d'adaptation $\mu_{w_{21}}(n)$ du filtre principal $w_{21}(n)$.

Selon la figure IV.25, nous remarquons que le pas d'adaptation $\mu_{w_{21}}(n)$ du filtre principal $w_{21}(n)$ est grand lorsque le RSB_1 est faible (cela signifie que le pas d'adaptation $\mu_{w_{21}}(n)$ prend des valeurs importantes lorsque le signal de parole est absent).

Nous représentons sur la figure (IV.26) les deux segmentations (DAV) manuelle et automatique.

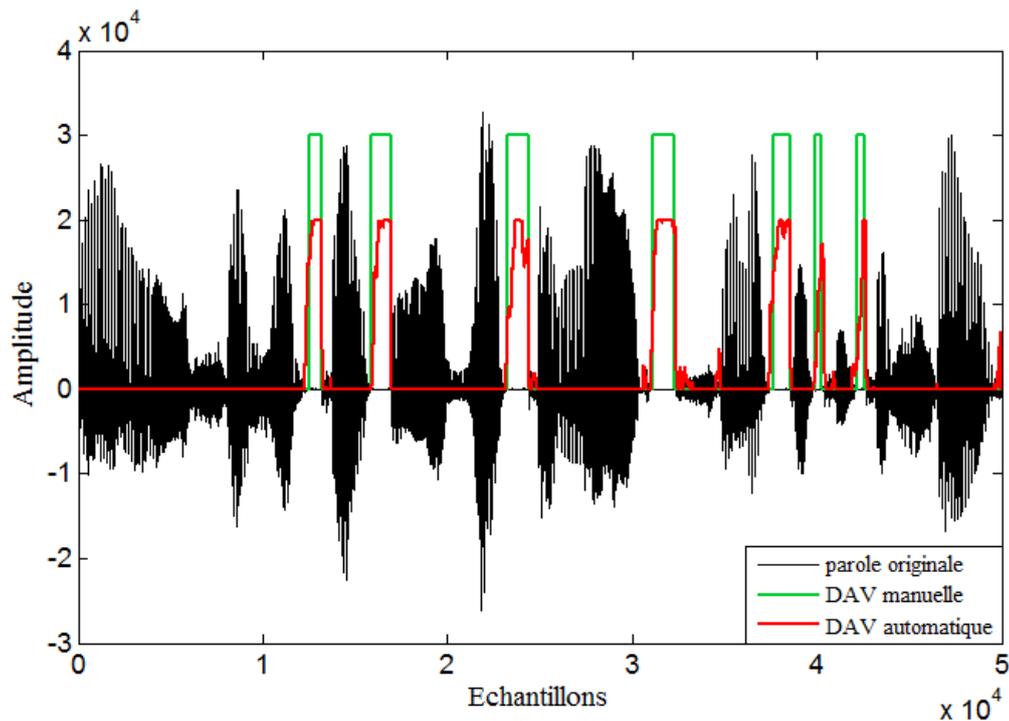


Figure IV.26 : Signal de parole utilisé avec les deux segmentations (manuelle et automatique (Propos2)), $F_e = 16$ KHz.

IV.10 Etude des performances de la deuxième structure proposée (Propos2)

Pour étudier le comportement de la deuxième structure proposée (Propos2), nous devons procéder aux simulations suivantes :

IV.10.1 Effet de la taille du filtre adaptatif

Dans cette simulation, nous avons évalué l'effet de la taille des deux filtres adaptatifs $w_{12}(n)$ et $w_{21}(n)$ sur les performances de la structure Propos2.

Les paramètres de cette simulation sont :

Le pas d'adaptation : $\mu = 0.4$,

Le RSB sur la voie 1 : $RSB_1 = 3$ dB,

Le RSB sur la voie 2 : $RSB_2 = 0$ dB,

Nous avons varié la taille du filtre L , en prenant trois valeurs : $L = 32$; $L = 64$; $L = 128$.

Pour réaliser cette comparaison, nous avons tracé l'évolution temporelle du désajustement pour les trois tailles et nous obtenons le résultat montré sur la figure (IV.27).

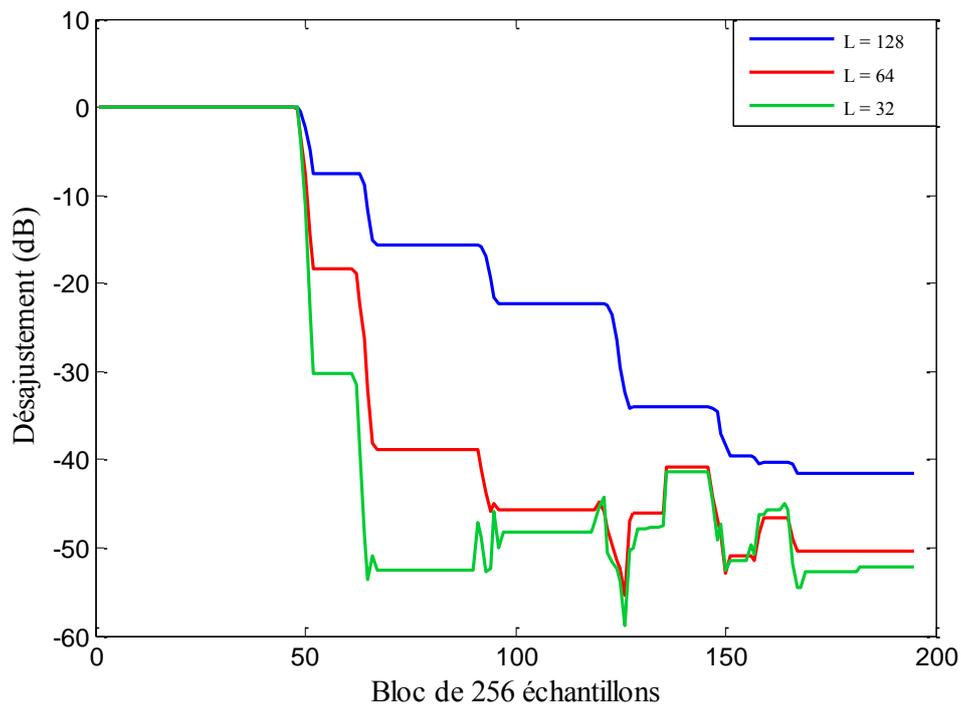


Figure IV.27 : Effet de la taille du filtre sur la structure proposée (Propos2) pour $\mu=0.4$, $RSB_1=3$ dB et $RSB_2=0$ dB.

Cette figure montre que la vitesse de convergence de l'algorithme DNLMS utilisé avec la structure Propos2 est inversement proportionnelle à la taille du filtre adaptatif.

IV.10.2 Effet du pas d'adaptation

Dans cette simulation, nous allons évaluer l'effet du pas d'adaptation de deux filtres adaptatifs $w_{12}(n)$ et $w_{21}(n)$ sur les performances de la structure Propos2. Pour ce faire, nous

avons pris : $L = 64$, $RSB_1 = 3$ dB et $RSB_2 = 0$ dB et nous avons varié le pas d'adaptation μ .
 Nous avons obtenu les résultats présentés sur la figure suivante :

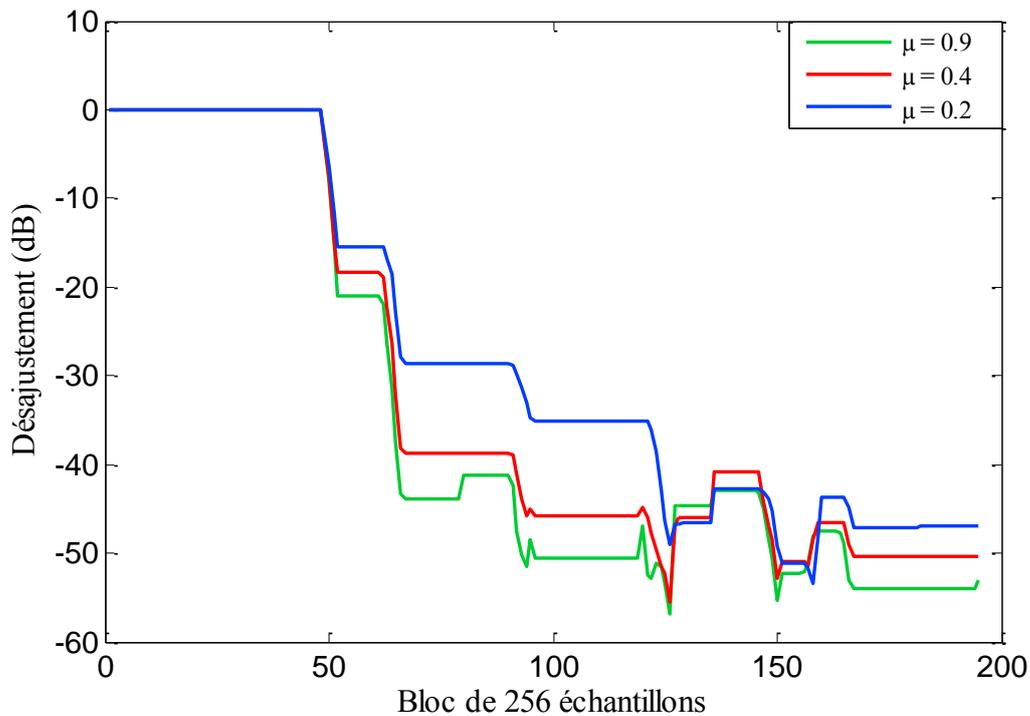


Figure IV.28 : Effet du pas d'adaptation sur la structure proposée (Propos2) pour $L=64$, $RSB_1=3$ dB et $RSB_2=0$ dB.

A partir des résultats de simulations illustrés dans la figure ci-dessus, nous remarquons bien que l'algorithme DNLMS converge mieux lorsque le pas d'adaptation μ est choisi grand.

IV.10.3 Effet du RSB du mélange d'entrée

Pour voir mieux l'effet du RSB_2 sur le comportement de la structure Propos2, nous avons fait les simulations illustrées dans la figure (IV.29), en prenant $\mu = 0.4$, $L = 64$ et $RSB_1 = 3$ dB.

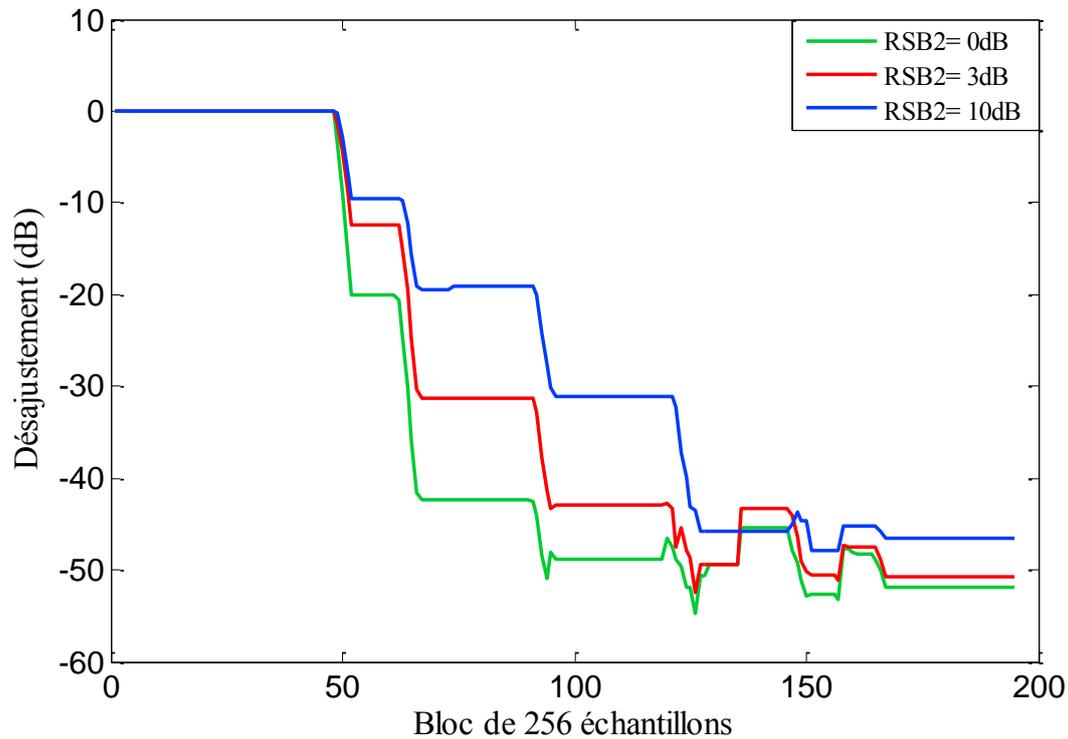


Figure IV.29 : Effet du RSB sur la structure proposée (Propos2) pour $L=64$, $\mu=0.4$ et $RSB_1=3$ dB.

D'après cette figure, nous remarquons que la vitesse de convergence de l'algorithme DNLMS vers l'optimum est obtenue lorsque le rapport signal à bruit RSB est faible.

IV.11 Comparaison de la structure Propos2 avec deux structures de séparation de sources Directe

Dans ce paragraphe, nous comparons la structure Propos2 (qui utilise un système de détection d'activité vocale DAV automatique) avec deux structures de séparation de sources de type Directe. La première structure Directe considérée correspond à une structure classique (FBSS), la seconde correspond à la structure (Method2) [35].

Les paramètres de simulations pour la structure Directe classique (FBSS) et la structure (Method2) dans le cas des microphones espacés sont représentés dans le tableau IV.4.

Tableau IV.4 : Paramètres de simulations des structures FBSS et Method2.

Structures	Paramètres
FBSS	<p>Longueur des filtres : $L_{w12} = L_{w21} = 64$</p> <p>Pas d'adaptation des filtres : $\mu_{w12} = 0.4 ; \mu_{w21} = 0.4$</p>
Method2	<p>Longueur des filtres : $L_{w12} = L_{w21} = L_{wcont1} = L_{wcont2} = 64$</p> <p>Pas d'adaptation des filtres secondaires: $\mu_{wcont1} = 0.4 ; \mu_{wcont2} = 0.001$</p> <p>Pas d'adaptation des filtres principaux:</p> <p>$\mu_{w_{12}min} = 0 ; \mu_{w_{12}max} = 0.02 ;$ $\mu_{w_{21}min} = 0 ; \mu_{w_{21}max} = 0.2$</p> <p>Valeurs seuils du RSB estimés :</p> <p>$RSB_{1min} = -35 \text{ dB} ; RSB_{1max} = -15 \text{ dB} ;$ $RSB_{2min} = -10 \text{ dB} ; RSB_{2max} = 0 \text{ dB}$</p>

Afin de comparer les performances de ces trois structures (Propos2, FBSS et Method2), nous présentons sur les figures suivantes les signaux de sortie temporels 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 et ceci dans le cas de configuration des microphones espacés.

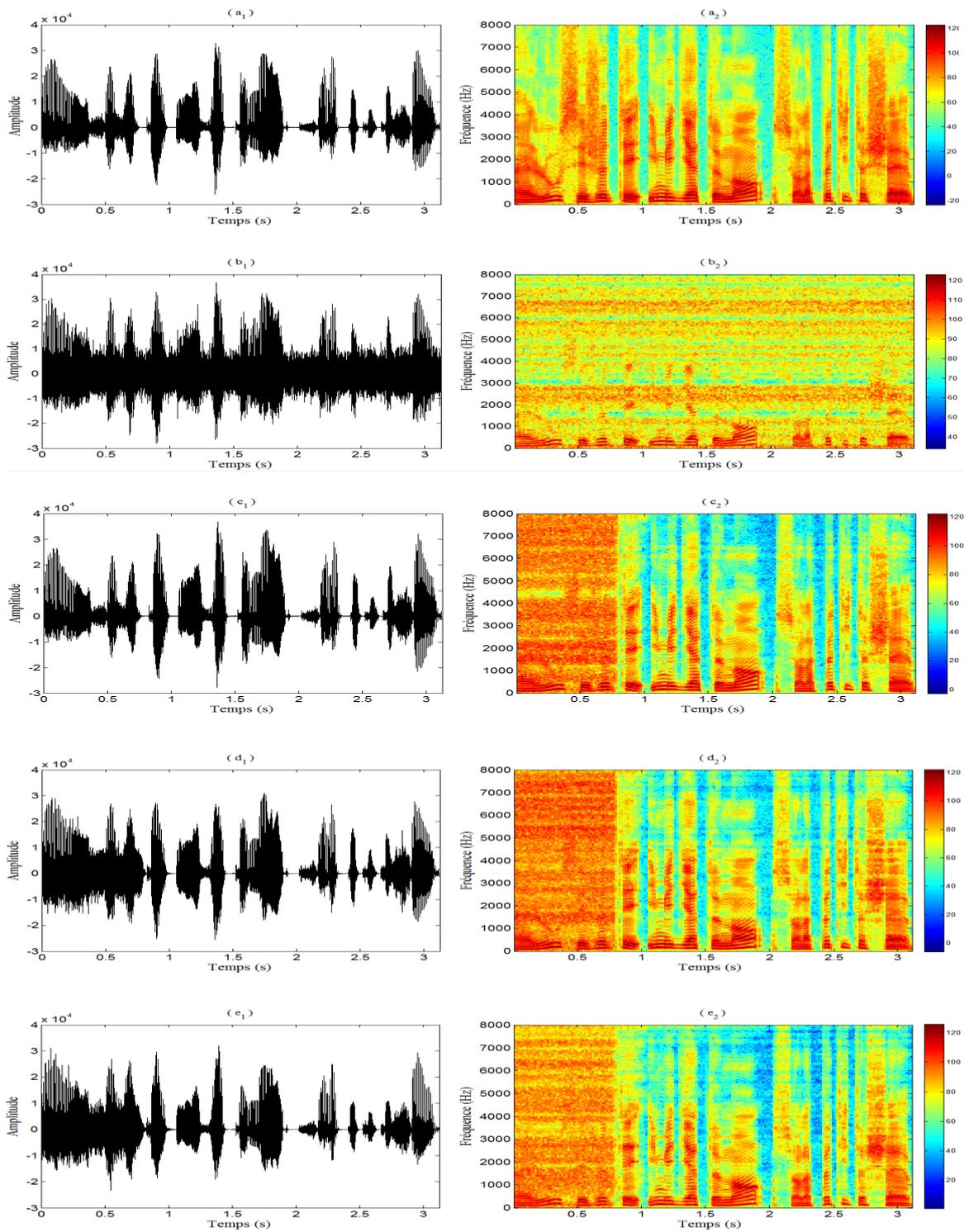


Figure IV.30 : (a_1, a_2) Signal de parole originale $s(n)$, (b_1, b_2) signal de mélange $p_1(n)$, (c_1, c_2) signal de sortie $u_1(n)$ de la structure Propos2, (d_1, d_2) signal de sortie de la structure FBSS, (e_1, e_2) signal de sortie de la structure Method2.

Nous pouvons observer à partir de la figure (IV.30) que les signaux disponibles en sortie de traitement à partir de ces trois structures (Propos2, FBSS et Method2) sont visuellement nettoyé du bruit et sont très 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 (Propos2, FBSS et Method2). Nous observons que les trois structures ont le même comportement en termes de distance cepstrale (DC).

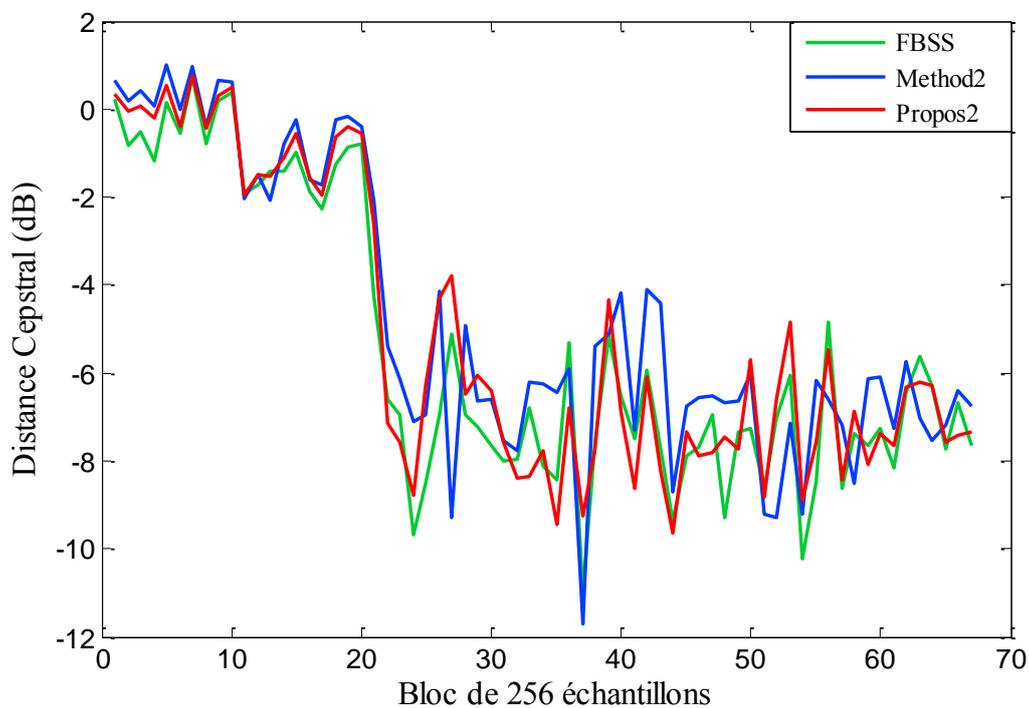


Figure IV.31 : Evolution du critère de la distance cepstrale des trois structures Propos2 (en rouge), Method2 (en bleu) et FBSS (en vert).

Sur la figure (IV.32) suivante, nous montrons l'évolution du critère du rapport signal à bruit (RSB) et ceci pour les trois structures. Nous remarquons bien que le comportement de la structure proposée (Propos2) est presque similaire à celui des deux autres structures (FBSS et Method2). Ceci confirme le bon comportement de la structure proposée (Propos2).

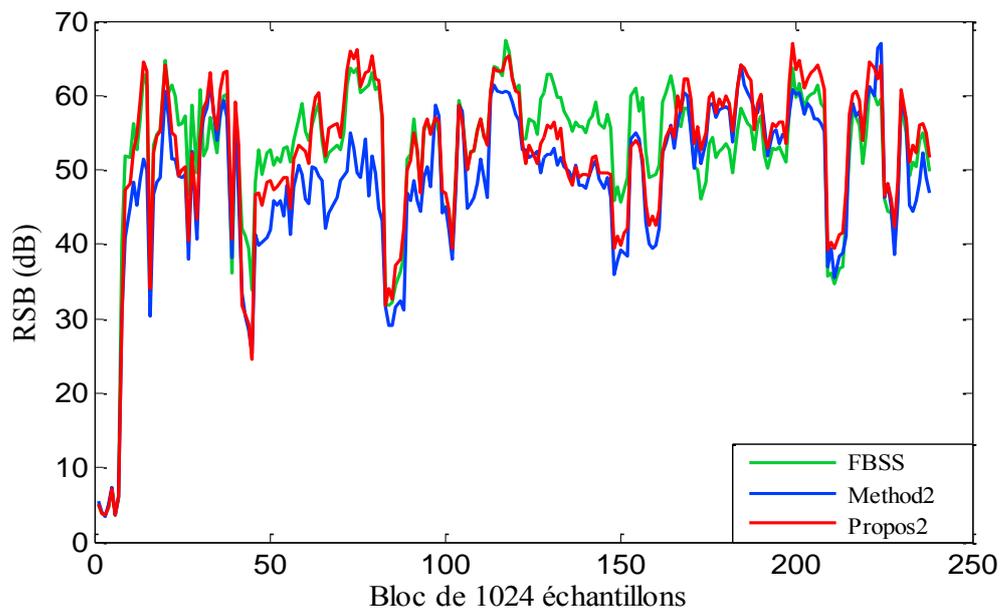


Figure IV.32 : Evolution du critère du rapport signal à bruit (RSB) pour les trois structures Propos2 (en rouge), Method2 (en bleu) et FBSS (en vert).

Afin de compléter les résultats précédents, nous présentons sur la figure (IV.33) l'évolution temporelle du désajustement mesuré sur les coefficients du filtre adaptatif $w_{21}(n)$ pour les trois structures (Propos2, Method2, FBSS).

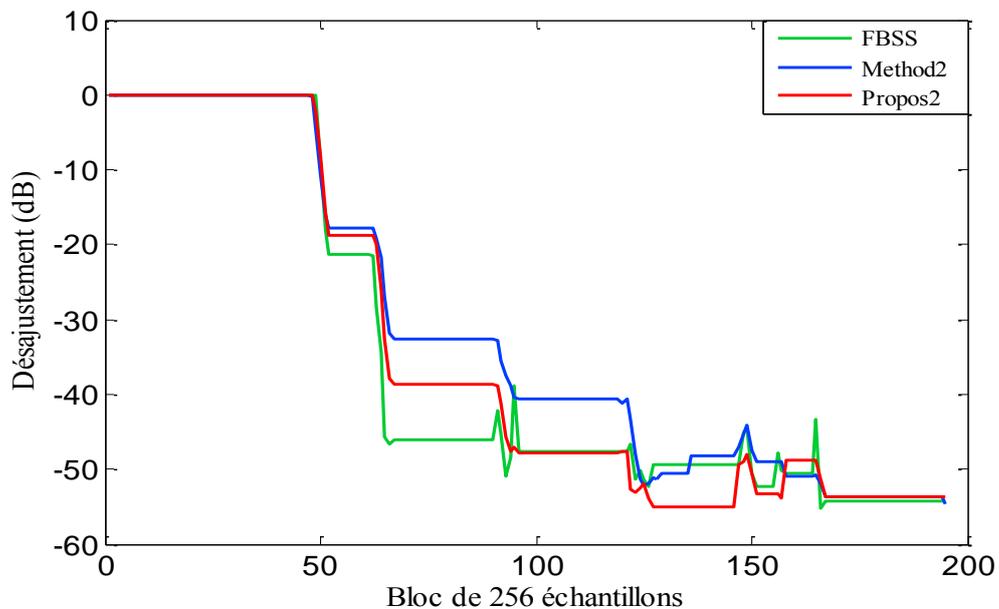


Figure IV.33 : Comparaison du désajustement obtenu avec les trois structures Propos2 (en rouge), Method2 (en bleu) et FBSS (en vert).

Cette figure montre bien la superposition presque totale des tracés de l'évolution temporelle du désajustement de chaque structures, ce qui démontre encore une fois les performances excellentes de la nouvelle structure (Propos2).

A partir des différents résultats présentés (signaux de sorties, distance cepstrale, rapport signal à bruit et le désajustement) nous pouvons conclure que la deuxième structure de débruitage proposée dans ce mémoire (Propos2) donne de bon résultats .

IV.12 Conclusion

Dans ce chapitre, nous avons décrit les signaux de tests ainsi que le modèle de mélange utilisé qui prend en compte l'espacement entre les microphones utilisés dans la prise du son. Les critères retenus pour l'analyse des performances ont également été présentés à savoir la distance cepstrale, le rapport signal à bruit en sortie de traitement et le désajustement des coefficients des filtres adaptatifs.

Nous avons ensuite présenté les résultats de simulations obtenus à partir des deux structures proposées (Propos1 et Propos2), ces résultats montrent bien que le débruitage du signal de parole avec ces deux structures a donné de bons résultats lorsque le signal de parole est segmenté automatiquement. Nous avons ainsi montré le comportement presque similaire de ces deux nouvelles structures en comparaison avec d'autres structures de la littérature.

Nous avons montré aussi que la vitesse de convergence de l'algorithme DNLMS appliqué aux nouvelles structures 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.

Conclusion générale

L'étude que nous avons présentée dans ce mémoire concerne le débruitage du signal de parole dans une voiture, en utilisant la structure de séparation de sources (Directe et Réursive) pour extraire le signal de parole à partir des observations bruitées dans le cas de microphones espacés.

Dans ce contexte, nous avons proposé deux nouvelles structures appelées *Propos1* et *Propos2*, qui correspondent à des versions améliorées des deux structures classiques de la séparation de sources (BBSS et FBSS). Ces nouvelles structures utilisent un système de détection d'activité vocale (DAV) automatique dans leur fonctionnement. Ce système est basé sur l'estimation des rapports signal à bruit (RSB) sur les deux voies de la structure de séparation.

Les simulations montrent bien le bon comportement de ces deux structures (*Propos1* et *Propos2*) employées avec l'algorithme du gradient stochastique normalisé double (DNLMS) qui contrôle l'adaptation des filtres de séparation.

Une étude comparative avec des structures de l'état de l'art a été effectuée, nous avons remarqué que les deux nouvelles structures proposées dans ce travail ont permis d'obtenir des performances plus proches que celles des structures de la littérature en termes de distance cepstrale (DC), le rapport signal à bruit (RSB) en sortie de traitement et le désajustement des coefficients des filtres adaptatifs.

Afin de conclure sur cette étude comparative, nous avons réalisé des tests d'écoute informels des signaux obtenus en sortie de traitement pour toutes les structures utilisées dans cette étude de comparaison. Ces tests ont confirmé le bon comportement des deux structures proposées dans ce mémoire.

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 des structures proposées. Nous avons remarqué que les deux nouvelles structures sont sensibles aux valeurs faibles du pas d'adaptation, aux tailles importantes des filtres de suppression croisés et aux valeurs importantes du RSB_2 .

Enfin, nous avons conclu que les deux nouvelles structures (Propos1 et Propos2) possèdent donc un bon comportement même lorsque les deux observations sont fortement bruitées (même avec un RSB négatif). Ces structures permettent de supprimer le bruit en sortie et par conséquent rehausser le signal de parole en sortie de traitement.

Bibliographie

- [1] F. Cottet, « Aide mémoire traitement de signal », Dunod, Paris, 2005.
- [2] S. Dos Santos, « Cours de traitement du signal, Signaux Déterministes (TS1) et Signaux Aléatoires (TS2) », École Nationale d'Ingénieurs du Val de Loire, 2008.
- [3] M. Bellenger, « Traitement Numérique du Signal », Collection CNET-ENST, MASSON, 1987.
- [4] G. Dimitris Manolakis, K. Vinay Ingle, M. Stephen Kogon, « Statistical and Adaptive Signal Processing », Artech House, 2005.
- [5] É. Tisserand, J. Pautex et P. Schweitzer, « Analyse et traitement des signaux, méthodes et applications au son et à l'image », 2^{ème} édition, Dunod, Paris, 2008.
- [6] G. Baudoin et J.F.- Bercher, « Éléments de traitement du signal », cours de traitement de signal version 0.89, École supérieur d'ingénieurs en électrotechnique et électronique, Université Paris-Est, 1998.
- [7] O. Sentieys, « Introduction au traitement numérique du signal », ENSSAT – Université de Rennes 1, IRISA- équipe de recherche R²D², 2003.
- [8] K. Meghriche, « Filtrage Adaptatif Utilisant les Statistiques d'Ordre Supérieur », Thèse de Doctorat d'Etat, École Nationale Polytechnique, 2006.
- [9] 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.

- [10] J.-F. Bercher et P. Jardin, « Introduction au filtrage adaptatif », ESIEE-Paris, I4-TTS, 2003.
- [11] 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.
- [12] G. Bellanger, « Adaptive Digital Filters », 2^{ème} édition, Marcel Dekker, Inc., New York, 2001.
- [13] 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.
- [14] 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.
- [15] A. Amehraye, « Débruitage perceptuel de la parole », Thèse de Doctorat, École Nationale Supérieure des Télécommunications de Bretagne, 2009.
- [16] 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.
- [17] C. Plapous, « Traitements pour la réduction de bruit. Application à la communication parlée », Thèse de Doctorat, Université de Rennes 1, 2005.
- [18] D. Alazard, « Introduction au filtre de Kalman », École nationale supérieure de l'aéronautique et de l'espace, Université de Toulouse CEDEX 4, 2005.

- [19] J. Bernussou, « Processus stochastique- filtrage », Cours ENSICA, Toulouse, 1994.
- [20] F. Ykhlef, « Réduction de bruit et contrôle s'écho pour les applications radio mobile et audioconférence » Thèse de Doctorat, École Nationale Polytechnique, 2008.
- [21] 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.
- [22] R. J. McAulay and M. L. Malpass, « Speech enhancement using a soft-decision noise suppression filter », In Proc. IEEE Int. Conf. Acoust, Speech, Signal Processing, vol. ASSP-28, no. 2, pp.137-145, 1980.
- [23] O. Cappé, « Techniques de réduction de bruit pour la restauration d'enregistrement musicaux », Thèse de Doctorat, École Nationale Supérieur des Télécommunications, Paris, 1993.
- [24] P. Howells, « Intermediate frequency side lobe canceller », US patent 3202990, Aug.24, 1965.
- [25] B. Widrow and M. E. Hoff Jr., « Adaptive switching circuits », in IRE WESCON Conv. Rec. , pt. 4, pp. 96-104, 1960.
- [26] 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.
- [27] B. Rivet, « La bimodalité de la parole au secours de la séparation de sources », Thèse de Doctorat, Institut National Polytechnique de Grenoble, 2006.
- [28] M. Akil, « Séparation de sources convolutes », Thèse de Doctorat, Université Joseph Fourier – Grenoble 1, 2008.

- [29] 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.
- [30] 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.
- [31] O.L. Frost, « An algorithm for linearly constrained adaptive array processing », Proceedings of the IEEE, vol.60, no.8, pp. 35-42, 1999.
- [32] L. Lepauloux, « Prise de son distante par système multi-microphone. Application à la communication parlée en environnement bruyant », Thèse de Doctorat, Université de Rennes 1, 2010.
- [33] 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.
- [34] S. Ikeda and A. Sugiyama, « An adaptive noise canceller with low signal- distortion in the present of crosstalk », IEICE Trans. Fundamentals, vol. E82-A, no.8, 1999.
- [35] 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.
- [36] 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.

