



HAL
open science

Traitement numérique du signal acoustique pour une aide aux malentendants

Thomas Fillon

► **To cite this version:**

Thomas Fillon. Traitement numérique du signal acoustique pour une aide aux malentendants. domain_other. Télécom ParisTech, 2004. English. NNT: . pastel-00001201

HAL Id: pastel-00001201

<https://pastel.hal.science/pastel-00001201>

Submitted on 18 Apr 2005

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

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



Thèse

présentée pour obtenir le grade de docteur
de l'École Nationale Supérieure des Télécommunications

Spécialité : **Signal et Images**

Thomas FILLON

Traitement Numérique du Signal Acoustique
pour une Aide aux Malentendants

Soutenue le 14 décembre 2004 devant le jury composé de

Alexandre Garcia
Pascal Scalart
Dirk Van Compernelle
André Gilloire
Yves Grenier
Régine Le Bouquin-Jeannès
Jacques Prado

Président
Rapporteurs
Examineurs
Directeur de thèse

Remerciements

Pour réaliser ce document et le travail qu'il présente, j'ai largement bénéficié de l'aide de nombreuses personnes. Je tiens à les remercier très sincèrement.

En premier lieu, je remercie les membres du jury de ma thèse pour l'intérêt qu'ils ont porté à mon travail et pour leur disponibilité. Je souhaite remercier particulièrement Régine Le Bouquin-Jeannès pour son aide précieuse et le soin qu'elle a apporté à la correction de mon manuscrit.

Je voudrais également exprimer toute ma reconnaissance à Jacques Prado qui a encadré mes recherches et qui s'est toujours soucié de m'offrir, de tout point de vue, les meilleures conditions de travail possibles. Je le remercie en particulier pour sa grande disponibilité, ses grandes qualités scientifiques et la confiance qu'il m'a accordée.

Travailler à l'ENST au département TSI a été très agréable et enrichissant et je tiens à remercier tous ceux qui ont contribué à créer cette atmosphère.

Merci à tous les permanents, stagiaires et thésards qui m'ont accompagné pendant ces années : du 1er étage au sous-sol, de Barrault à Dareau (et vice et versa). Merci à tous pour votre aide, votre soutien, votre sollicitude et surtout pour votre amitié.

Je remercie également très fortement pour leur soutien, leur patience et leur amour tous mes proches : Sophie, ma mère, ma grand-mère, mon frère Renaud et la famille Fillon-Gonzalez, Evelyne, Bernard, Marité et Marie.

Résumé

Pour un malentendant, le déficit auditif peut représenter un handicap majeur tant du point de vue physique que du point de vue social. Pour cette raison, la réhabilitation auditive constitue un champ d'application particulièrement important du traitement du signal audio et ses enjeux dépassent le seul contexte technologique.

C'est dans ce cadre, et en particulier dans celui de la réhabilitation auditive par port de prothèses externes, que se situe ce travail de thèse. L'audition étant un sens particulièrement performant et complexe, compenser les pertes auditives des malentendants est une tâche difficile et délicate. En outre, les appareils sont conçus à partir d'éléments miniaturisés de pointe, notamment les micro-processeurs, ce qui impose des contraintes de fonctionnement très strictes.

Au cours de nos recherches, nous avons travaillé sur la réalisation d'un module de rehaussement du signal de parole utilisant une technique d'*atténuation spectrale à court-terme* mise au point par Ephraïm et Malah. A l'issue de ce travail, nous avons proposé d'adapter cette méthode en réalisant l'analyse spectrale selon une échelle en fréquence conforme à notre perception auditive, l'échelle ERB (Equivalent Rectangular Bandwidth) qui modélise la résolution fréquentielle des filtres auditifs.

Dans un deuxième temps, nous avons étudié le problème de retour acoustique entre les transducteurs de la prothèse. Ce retour acoustique est source d'instabilité pour le système et aboutit fréquemment à la production d'un effet Larsen. Ce type de désagrément est extrêmement gênant et contraint l'audioprothésiste à limiter le gain d'amplification de la prothèse. De ce fait, la réhabilitation auditive ne peut pas être optimale. Pour pallier ce problème, nous proposons d'utiliser une méthode de filtrage adaptatif multi-délais basée sur le LMS (Least Mean Square) et réalisée dans le domaine de la transformée de Hartley. Du fait du retour acoustique, une prothèse auditive est un système électro-acoustique en boucle fermée. Les conditions d'application sont donc contraires aux hypothèses de convergence du LMS. Pour assurer la convergence de l'algorithme de descente de gradient, nous proposons une nouvelle méthode de définition du pas d'adaptation. Afin de vérifier le bon fonctionnement de la méthode proposée dans des conditions réalistes, nous avons mis en place une plate-forme de test en temps-réel basée sur un prototype de prothèses.

Abstract

Hearing impairment may be considered as a serious physical and social handicap. Alleviating hearing impairment is therefore a very challenging application for digital signal processing.

The framework of this thesis is the use and development of digital signal processing technique for use in hearing-aid devices.

Since the auditory system is a very effective and complex, compensating for the hearing-loss is a very challenging problem. Moreover, hearing-aid is based on miniaturized chips and electro-acoustic devices that force to be subject to certain technology constraints.

During our research, we have studied and developed a speech enhancement stage in the hearing-aid using a short-term spectral modification method, the Ephraïm and Malah suppression rule (EMSR). In order to perform the frequency analysis on a perceptually relevant frequency scale, we introduce an adaptation of the EMSR. The frequency scale used in this work, the ERB (Equivalent Rectangular Bandwidth) scale, models the resolution of the auditory filters.

Hearing-aid can be considered as a closed-loop electro-acoustical system that can become unstable because of the feedback. The main part of this thesis is devoted to feedback cancellation using multi delay adaptive filtering in the transform domain. The last part of this work consist in developing a real-time implementation of the feedback cancellation algorithm previously proposed using a prototype hearing aid.

Table des matières

Remerciements	i
Résumé	iii
Abstract	v
Table des matières	i
Table des figures	vii
Liste des tableaux	xi
Liste des abréviations	xiii
Introduction générale	1
I. Audition et réhabilitation auditive	3
Introduction de la première partie	7
1. Anatomie et physiologie de l'appareil auditif	9
1.1. Principe général	9
1.2. Mesure physique de l'intensité	10
1.3. L'oreille externe	11
1.3.1. Anatomie	11
1.3.2. Physiologie	11
1.4. L'oreille moyenne	12
1.4.1. Anatomie	12
1.4.2. Physiologie	12
1.5. L'oreille interne	14
1.5.1. Anatomie	14
1.5.2. Physiologie de la cochlée	18
1.6. Le nerf auditif	24
1.6.1. Courbe d'accord	24
1.6.2. Codage de la fréquence	26
1.6.3. Codage de l'intensité	27

1.7. Conclusions sur la physiologie de l'appareil auditif	27
2. Psychoacoustique et perception auditive	29
2.1. Les seuils auditifs	29
2.2. Intensité subjective	30
2.2.1. Échelle des phones	30
2.2.2. Échelle des sones ou échelle d'intensité subjective	32
2.3. Les filtres auditifs	33
2.3.1. Principe	33
2.3.2. Largeur des filtres auditifs	36
2.3.3. Forme des filtres auditifs	41
2.3.4. Pattern d'excitation	41
3. Pathologies de l'audition	45
3.1. Les différents types de surdit��	45
3.1.1. Les surdit��s de transmission	46
3.1.2. Les surdit��s de perception	46
3.2. Effets g��n��raux des surdit��s	47
3.3. Rehaussement des seuils auditifs	48
3.4. Le ph��nom��ne de recrutement	48
3.5. L'��largissement des filtres auditifs	49
4. R��habilitation auditive par port de proth��ses	51
4.1. Principe G��n��ral	51
4.2. Techniques de r��habilitation	52
4.2.1. Amplification et compression	52
4.2.2. R��duction du bruit	56
4.2.3. Traitements	57
4.3. Les diff��rents types d'appareils auditifs	58
4.3.1. Les contours d'oreille	59
4.3.2. Les intra-auriculaires	59
4.3.3. Le choix du type de proth��se	60
4.4. Technologie des proth��ses auditives	61
4.4.1. Transducteurs ��lectro-acoustiques	61
4.4.2. Autres voies de transduction	62
4.4.3. Processeurs et puces de traitement	63
Conclusion de la premi��re partie	65
II. R��duction du bruit	67
Introduction de la deuxi��me partie	71

5. Rehaussement de la parole	75
5.1. Techniques de rehaussement de la parole	75
5.1.1. Contexte	75
5.1.2. Choix du type de méthode	76
5.2. Modification spectrale à court-terme	77
5.2.1. Principes	77
5.2.2. Filtrage de Wiener Paramétrique	82
5.2.3. Le phénomène du bruit musical	83
5.3. Règle de suppression d'Ephraïm et Malah	86
5.3.1. Introduction	86
5.3.2. Description de la méthode	87
5.3.3. Réduction du bruit musical	88
6. Estimation du niveau de bruit	91
6.1. Méthodes d'estimation du niveau de bruit	91
6.1.1. Principes	91
6.1.2. Estimation discontinue	91
6.1.3. Estimation continue	92
6.2. Solution adoptée	93
6.2.1. Modèle élémentaire de DAV	93
6.2.2. Mise à jour de l'estimateur	95
6.2.3. Validation du choix de la méthode	95
7. Application de l'EMSR à une analyse en banc de filtres ERB	97
7.1. Motivations à l'utilisation de l'échelle des ERB	97
7.1.1. Filtres auditifs et perception des signaux bruités	97
7.1.2. Choix de l'échelle fréquentielle	98
7.1.3. Application de l'EMSR selon une analyse en échelle ERB	98
7.1.4. Rehaussement de la parole sur une échelle fréquentielle non-linéaire	100
7.2. Synthèse de bancs de filtres sur-échantillonnés en échelle ERB	101
7.3. Implémentation de l'EMSR pour une analyse banc de filtres	103
8. Résultats et discussion	107
8.1. Protocole d'évaluation	107
8.1.1. Motivations	107
8.1.2. Choix d'implémentations	107
8.1.3. Corpus de test	109
8.2. Amélioration du Rapport Signal à Bruit	109
8.2.1. Avec estimation du bruit	109
8.2.2. Sans estimation du bruit	111
8.3. Évaluation subjective	116
8.3.1. Analyse subjective	116
8.3.2. Tests subjectifs	117
8.3.3. Résultats	117

Conclusion de la deuxième partie	121
8.4. Conclusion	121
8.5. Perspectives	122
III. Annulation de retour acoustique	123
Introduction de la troisième partie	127
9. Retour acoustique dans les prothèses auditives	131
9.1. Description du système acoustique	131
9.2. Description du système électrique	133
9.3. Étude de stabilité	135
9.3.1. Étude de stabilité à temps continu	135
9.3.2. Étude de stabilité à temps discret	136
9.4. Analyse des conditions de stabilité dans une prothèse auditive	140
9.4.1. Facteurs pénalisant la stabilité dans une prothèse auditive	140
9.4.2. Variabilité des conditions	141
9.4.3. Distorsions	142
10. Techniques de compensation du retour acoustique	143
10.1. Méthodes classiques	143
10.1.1. Contrainte sur le chemin d'écho	143
10.1.2. Contraintes sur le gain	144
10.2. Solutions récentes	145
10.2.1. Description des méthodes	145
10.2.2. Prévention de l'accrochage	145
10.2.3. Compensation de l'écho	145
10.2.4. Annulation de l'écho	146
10.3. Annulation d'écho par filtrage adaptatif	148
10.3.1. Utilisation du gradient stochastique dans une prothèse auditive	148
10.3.2. Continuité de l'adaptation	150
10.3.3. Émission d'une séquence de bruit	151
10.3.4. Le délai	151
10.4. Synthèse	152
11. Annulation d'écho par filtrage adaptatif	153
11.1. LMS temporel par blocs	153
11.1.1. Principes	153
11.1.2. Calcul de l'erreur par convolution	154
11.1.3. Mise à jour du filtre adaptatif	154
11.2. LMS Rapide dans le domaine fréquentiel	155
11.2.1. Principes	155
11.2.2. Convolution rapide dans le domaine de la transformée de Fourier	156
11.2.3. Mise à jour du filtre adaptatif	158

11.3. FLMS dans le domaine de la transformée de Hartley discrète	159
11.3.1. La transformée de Hartley discrète	159
11.3.2. Convolution rapide dans le domaine de la transformée de Hartley .	160
11.3.3. Mise à jour du filtre adaptatif	161
12. Filtre à Délais Multiples Généralisé	163
12.1. Principes	163
12.2. Synthèse WOLA	165
12.3. Partitionnement des données	166
12.4. Implémentation du LMS	168
13. Modification du pas d'adaptation	169
13.1. Pas d'adaptation du LMS	169
13.1.1. Rôle du pas d'adaptation	169
13.1.2. Normalisation du gradient	169
13.2. Modification de la définition du pas d'adaptation	171
13.2.1. Seuil minimal de la puissance spectrale par sous-bande	171
13.2.2. Prise en compte de la puissance spectrale de l'erreur	172
14. Résultats - Simulations	175
14.1. Motivations	175
14.2. Simulations Matlab	175
14.2.1. Description de l'environnement	175
14.2.2. Mise en oeuvre du HD-GMDF	176
14.2.3. Résultats	177
14.3. Implémentation en temps-réel	181
14.3.1. Environnement de test	181
14.3.2. Résultats	182
14.3.3. Analyse des résultats	186
Conclusion de la troisième partie	187
Conclusion générale	191
Annexes	193
A. Calculs du gradient dans le domaine de Hartley	195
A.1. Expression de la fonction de coût dans le domaine de la THD	195
A.1.1. Expression de l'erreur dans le domaine de la THD	195
A.1.2. Expression de la fonction de coût sans contrainte	196
A.1.3. Expression de la fonction de coût avec contrainte	197
A.2. Gradient de la fonction de coût sans contrainte	198
A.2.1. Expression de la fonction de coût	198
A.2.2. Calcul de la dérivée	198
A.3. Gradient de la fonction de coût avec contrainte	200

A.3.1. Expression de la fonction de coût	200
A.3.2. Calcul de la dérivée	200
B. Développement d'une plate-forme de test	203
B.1. Motivations	203
B.2. Prototype de prothèse	203
B.3. Calibration du matériel	204
B.4. Mesure de réponses impulsionnelles	205
Bibliographie	207

Table des figures

1.1.	Schéma de l'appareil auditif humain	10
1.2.	Schéma de l'oreille moyenne	13
1.3.	Schéma de l'oreille interne	14
1.4.	Section axiale de la cochlée	15
1.5.	Canal cochléaire	16
1.6.	Organe de Corti	17
1.7.	Disposition des cellules ciliées	18
1.8.	Enveloppes des Ondes propagée de Békésy	19
1.9.	Représentation schématique de la tonotopie passive de la cochlée	20
1.10.	Réponse de la membrane basilaire à un son pur	21
1.11.	Tonotopie de la cochlée	22
1.12.	Réponse schématique de la membrane basilaire à un stimulus	23
1.13.	Amplification des vibrations de la membrane basilaire	23
1.14.	Courbe d'accord d'une fibre nerveuse	25
1.15.	Synchronisation temporelle des décharges d'une fibre	26
2.1.	Courbes d'isotonie pour l'écoute binaurale sur écouteurs	31
2.2.	Courbes d'isotonie pour l'écoute binaurale en champ libre	31
2.3.	Courbes de Pondérations A et C	32
2.4.	Courbe de sonie	33
2.5.	Seuil d'audibilité d'un signal en présence d'un bruit masquant	34
2.6.	Courbes d'effet de masque de bruits à bande étroite	35
2.7.	Courbes de sélectivité de fréquence psychoacoustique	36
2.8.	Masquage non-simultané	37
2.9.	Comparaison entre les bandes critiques en Bark et en ERB	40
2.10.	Comparaison Bark / ERB en fonction de la fréquence	40
2.11.	Comparaison Bark / ERB en fonction de la tonotopie	40
2.12.	Filtres auditifs à 250Hz	42
2.13.	Filtres auditifs à 1kHz	42
2.14.	Pattern d'excitation de deux sons purs de niveau modéré	43
2.15.	Pattern d'excitation de deux sons purs de fort niveau	44
3.1.	Courbes d'audiogramme en fonction de l'âge	47
3.2.	Courbes de sonie avec recrutement	49
4.1.	Schéma général d'une prothèse auditive	53

4.2.	Relation Entrée/Sortie dans une prothèse avec compression	54
4.3.	Calcul du gain pour compenser le recrutement	55
4.4.	Contour d'oreille	59
4.5.	Les différents types d'intra auriculaire	60
5.1.	Schéma général d'un système de débruitage	78
5.2.	Courbes de Gain pour 3 méthodes de Soustraction Spectrale	84
5.3.	Exemple de spectre de bruit musical	85
5.4.	Règle de suppression d'Ephraïm et Malah	88
5.5.	Règle de suppression d'Ephraïm et Malah	89
5.6.	Nature du bruit résiduel après rehaussement par l'EMSR	90
6.1.	Estimation du niveau de bruit, aspect temporel	94
6.2.	Estimation du niveau de bruit. Aspect fréquentiel	94
6.3.	Calcul du facteur de lissage	96
7.1.	Correspondance entre l'échelle ERB et les valeurs en Hertz	99
7.2.	Réponses en fréquence d'un banc de filtres ERB 32 canaux pour L=1000 .	102
7.3.	Application du gain spectral sur des blocs recouvrants	105
8.1.	Réponse en fréquence d'un banc de filtres uniforme sur 32 canaux	108
8.2.	Spectrogrammes des bruits employés	110
8.3.	Évaluation du RSB, avec estimation du bruit	111
8.4.	Comparaison des spectrogrammes des signaux rehaussés - Bruit blanc . . .	112
8.5.	Comparaison des spectrogrammes des signaux rehaussés - Bruit de voiture	113
8.6.	Comparaison des spectrogrammes des signaux rehaussés - Brouhaha	114
8.7.	Évaluation du RSB, sans estimation du bruit	115
8.8.	Scores subjectifs avec estimation du bruit	118
8.9.	Scores subjectifs sans estimation du bruit	118
9.1.	Schéma de la prothèse auditive sans bouclage.	131
9.2.	Schéma de la prothèse auditive avec bouclage.	133
9.3.	Schéma équivalent de la prothèse auditive avec bouclage.	134
9.4.	Illustration du critère de revers	137
9.5.	Modèle discret équivalent de la prothèse auditive avec bouclage.	137
9.6.	Hodographe indiquant les marges de gain et de phase	139
9.7.	Hodographe indiquant la marge de module	140
10.1.	Gabarit d'un filtre réjecteur réglable	144
10.2.	Schéma de la prothèse auditive avec bouclage et annulation d'écho.	147
10.3.	Annulation d'écho acoustique classique	149
10.4.	Annulation d'écho acoustique pour une prothèse	149
12.1.	Décomposition d'un filtre en sous-blocs	164
12.2.	Composition des vecteurs $\tilde{\mathbf{y}}_s^k$	167
13.1.	Définition de la puissance moyenne par sous-bandes	172

14.1. Modèles de chemin d'écho	176
14.2. Spectrogrammes des signaux de sortie pour un gain de 5 dB	178
14.3. Spectrogrammes des signaux en sortie pour un gain de 15 dB	180
14.4. Schéma de principe de l'implémentation temps-réel	181
14.5. Convergence en fréquence du filtre adaptatif, sans compression	183
14.6. Convergence en fréquence du filtre adaptatif, avec compression	184
14.7. Spectrogramme du signal en sortie de prothèse sans annulation d'écho . . .	185
14.8. Spectrogramme du signal en sortie de prothèse avec annulation d'écho . . .	185
B.1. Schéma de la plate-forme de test	204

Liste des tableaux

1.1. Correspondance entre niveau de pression acoustique et pression	11
2.1. Seuils d'audition monoauraux sous écouteurs	30
2.2. Normalisation de l'échelle des Barks et découpage en 25 bandes critiques	38
3.1. Tableau de classification audiométrique	45
4.1. Valeurs de gain et d'intensité maximales des prothèses	60
5.1. Filtrage de Wiener paramétrique	83
8.1. Paramètres d'implémentation des méthodes de rehaussement	108
8.2. Évaluation du RSB, avec estimation du bruit	110
8.3. Évaluation du RSB, sans estimation du bruit	115
8.4. Moyenne des classements avec estimation du bruit	117
8.5. Scores subjectifs sans estimation du bruit	117
14.1. Configuration de l'algorithme pour l'expérience.	182

Liste des abréviations

Pour des raisons de lisibilité, la signification d'une abréviation ou d'un acronyme n'est souvent rappelée qu'à sa première apparition dans le texte d'un chapitre. Par ailleurs, puisque nous utilisons toujours l'abréviation la plus usuelle, il est fréquent que ce soit le terme anglais qui soit employé, auquel cas nous présentons une traduction.

- BTE** *Behind-The-Ear* ; contours d'oreille, prothèse placée derrière le pavillon
- dB** décibel
- dB HL** *dB Hearing Level* ;
- dB SPL** *dB Sound Pressure Level*
- CAN** Convertisseur Analogique-Numérique
- CIC** *Completely-In-the-Canal* intra semi-profond ou profond *i.e.* prothèse placée dans le conduit
- CNA** Convertisseur Numérique-Analogique
- DSP** Digital Signal Processor ; Processeur de Traitement Numérique du Signal
- DAV** Détecteur d'Activité Vocale
- EBSB** Entrée Bornée - Sortie Bornée
- EQM** Erreur Quadratique Moyenne
- ERB** Equivalent Rectangular Bandwidth
- EMSR** Ephraïm and Malah Suppression Rule ; règle de suppression d'Ephraïm et Malah
- EMSR-BF** Implémentation de l'EMSR sur un banc de filtres uniforme
- EMSR-ERB** Implémentation de l'EMSR sur un banc de filtre en échelle ERB
- EMSR-FFT** Implémentation classique de l'EMSR par TFCT
- FC** Fréquence Caractéristique
- FDAF** Frequency Domain Adaptive Filter
- FFT** Fast Fourier Transform ; Transformée de Fourier rapide
- FHT** Fast Hartley Transform ; Transformée de Hartley rapide
- FLMS** Fast LMS ; implémentation fréquentiel Rapide du LMS temporel par blocs
- GMDF α** Generalized-MDF α
- GMDF** Generalized-MDF
- HD-GMDF α** Hartley-Domain-GMDF α
-

- ITC** *In-the-Canal* ; intra-conduit ou intra-canal, prothèse placée à l'entrée du conduit (*cf.* page 60)
- ITE** *In-the-Ear* ; intra-conque, prothèse placée dans la conque (*cf.* page 60)
- LIT** Linéaire Invariant dans le Temps
- LMS** Least Mean Square ; moindres carrés moyen, algorithme du gradient stochastique
- MDF** Multi-Delay Filter ; Filtre à délais multiples
- MMSE-STSA** Minimum Mean Square Error - Short Term Spectral Amplitude
- OLA** OverLap-Add
- OLS** OverLap-Save
- PBFDAF** Partitioned Block FDAF
- PE** Pattern d'Excitation
- RSB** Rapport Signal à Bruit
- STSA** Short-Term Spectral Amplitude
- TFCT** Transformée de Fourier à Court-Terme
- TFD** Transformée de Fourier Discrète
- THD** Transformée de Hartley Discrète
- WOLA** Weighted OverLap-Add
-

Introduction générale

Ce rapport de thèse présente notre travail sur le traitement numérique du signal dans le cadre de la réhabilitation auditive des malentendants. Nous nous plaçons dans le contexte de la réhabilitation par port de prothèses auditives. Le but de ce travail de recherche est de proposer des solutions technologiques innovantes et pertinentes.

L'audition est un de nos sens les plus performants. Ses mécanismes qui sont l'aboutissement de millions d'années d'évolution sont, à ce titre, extrêmement perfectionnés et complexes. En contrepartie, l'audition est un sens fragile qui subit l'attaque du temps de manière irréversible. La complexité des mécanismes qui participent à ses capacités est telle que la réhabilitation des déficits auditifs est bien plus difficile, par exemple, que la prise en charge des troubles de la vision.

Les solutions pour soulager les handicaps des malentendants nécessitent donc d'avoir recours à des techniques avancées de traitement du signal. La mise en oeuvre de calculs numériques dans les prothèses est une orientation technologique assez récente qui a été rendue possible par la miniaturisation des processeurs de traitement du signal. Les dernières générations de prothèses numériques ne se contentent plus, comme le faisaient leurs aînées, de reproduire les traitements des anciens modèles analogiques. Les technologies de traitement numérique du signal permettent d'offrir au malentendant une qualité de réhabilitation qui n'était pas envisageable auparavant.

Les prothèses auditives constituent, selon nous, un champ d'application particulièrement intéressant pour le traiteur de signal acoustique. Outre l'aspect humain, l'aspect technologique se révèle également passionnant et motivant. Il s'agit d'un domaine de recherche en pleine expansion et ouvert qui sollicite de nombreuses et diverses compétences de traitement du signal. Les perspectives d'innovation s'élargissent constamment au rythme des progrès conjoints de la physiologie de l'audition, de la psychoacoustique et de la miniaturisation des processeurs et des composants électro-acoustiques. Pour une revue de ces aspects, on peut se reporter à [53, 76].

Au cours de notre travail de recherche, nous nous sommes penchés sur deux problèmes particuliers. Le premier consiste à proposer une solution de rehaussement de la parole au sein des prothèses auditives. Les malentendants sont en effet assez sensibles au bruit. Ils sont, par exemple, très gênés dans leur compréhension lorsqu'un bruit ambiant vient

parasiter la parole de leur interlocuteur. Ce genre de situation leur demande alors un effort d'attention soutenue qui constitue une source de fatigue. La mise en place d'un algorithme de rehaussement de la parole se révèle donc tout à fait opportun.

Le second problème sur lequel nous nous sommes penchés est la compensation du phénomène de bouclage acoustique entre les transducteurs. Ce bouclage est cause d'instabilité pour le système de la prothèse et il peut engendrer ce que l'on nomme communément l'effet Larsen. Celui-ci se caractérise par un sifflement très désagréable. Pour limiter ce désagrément, il n'y a guère d'autre solution, pour le moment, que de limiter l'amplification apportée au signal, ce qui implique une restriction des capacités de réhabilitation auditive de la prothèse. Pour cette raison, la résolution de ce problème est, aujourd'hui, un des objectifs prioritaires des fabricants d'appareils auditifs.

Ce rapport de thèse est divisé en trois parties. La première présente, d'une part, un travail de synthèse de connaissances portant sur l'audition normale et pathologique et d'autre part, une revue des techniques de réhabilitation auditive.

Le rehaussement de la parole est abordé dans la deuxième partie du rapport. Pour aboutir à une solution efficace, nos travaux se sont orientés vers l'utilisation d'une technique très performante opérant dans le domaine fréquentiel et initialement introduite par Ephraïm et Malah. Nous proposons également une adaptation de cette méthode à une analyse par banc de filtres selon une résolution fréquentielle conforme à celle de l'oreille.

La dernière partie de ce document est consacrée à l'étude du phénomène de bouclage acoustique et à la présentation d'une solution originale. Elle constitue la majeure partie de ce travail de thèse. A l'issue de l'étude théorique, il apparaît que la mise en place d'un module d'annulation d'écho par filtrage adaptatif est tout à fait appropriée. La fin de cette partie présente les résultats d'expérimentations réalisées en temps-réel sur un prototype de prothèse.

Ces trois parties du rapport sont relativement indépendantes. En particulier, la lecture de la première partie n'est pas strictement nécessaire à la compréhension des deux suivantes. Elle s'inscrit dans un cadre plus large de traitement de signal pour les prothèses auditives qui dépasse les deux sujets abordés dans les deux dernières parties. Son intérêt tient dans le fait que l'audition est un sens très complexe et qu'il est nécessaire d'avoir une bonne connaissance de son fonctionnement et des pathologies qui peuvent l'affecter afin d'appréhender les difficultés inhérentes à la réhabilitation auditive. Celle-ci est un champ d'application difficile et délicat pour le traiteur de signal. La réhabilitation auditive ne se limite pas, par exemple, à une simple amplification du son et elle mérite une attention et une précaution très particulière.

Première partie .
Audition et réhabilitation auditive

Table des Matières

Introduction de la première partie	7
1. Anatomie et physiologie de l'appareil auditif	9
1.1. Principe général	9
1.2. Mesure physique de l'intensité	10
1.3. L'oreille externe	11
1.4. L'oreille moyenne	12
1.5. L'oreille interne	14
1.6. Le nerf auditif	24
1.7. Conclusions sur la physiologie de l'appareil auditif	27
2. Psychoacoustique et perception auditive	29
2.1. Les seuils auditifs	29
2.2. Intensité subjective	30
2.3. Les filtres auditifs	33
3. Pathologies de l'audition	45
3.1. Les différents types de surdité	45
3.2. Effets généraux des surdités	47
3.3. Rehaussement des seuils auditifs	48
3.4. Le phénomène de recrutement	48
3.5. L'élargissement des filtres auditifs	49
4. Réhabilitation auditive par port de prothèses	51
4.1. Principe Général	51
4.2. Techniques de réhabilitation	52
4.3. Les différents types d'appareils auditifs	58
4.4. Technologie des prothèses auditives	61
Conclusion de la première partie	65

Introduction de la première partie

Dans cette première partie, nous présentons les notions principales qu'il est nécessaire de connaître pour appréhender le domaine de la réhabilitation auditive. Dans les chapitres qui suivent, nous avons essayé de faire une synthèse claire et cohérente des connaissances, des résultats et des travaux des différents champs scientifiques concernés.

L'anatomie de l'oreille, et en particulier celle de la cochlée, est déterminante vis-à-vis de notre perception auditive. Il est donc nécessaire pour comprendre l'audition de décrire, dans un premier temps, les organes et les mécanismes physiologiques mis en jeu dans ce processus très complexe ainsi que leur rôle respectif. Le chapitre 1 est donc consacré à l'anatomie et à la physiologie de l'oreille. Au cours de ce chapitre, il se dégage l'idée que l'audition se base sur deux types de mécanismes : des mécanismes passifs et des mécanismes actifs. Au coeur de ce fonctionnement très particulier, ce sont des cellules particulières de l'oreille, appelées cellules ciliées externes, qui permettent à notre perception d'atteindre des performances étonnantes.

Les processus cognitifs mis en jeu étant très complexes, la seule étude de la physiologie du système auditif ne permet pas de comprendre entièrement son fonctionnement. A cet effet, le chapitre 2 couvre la psychoacoustique. Cette discipline permet de caractériser les performances du système auditif et de modéliser son fonctionnement. Une des modélisations principales introduite dans ce chapitre est la notion de *filtres auditifs* qui indique que l'oreille traite les sons d'une manière comparable à un banc de filtres.

Les deux premiers chapitres donnent les clés de lecture pour aborder le chapitre 3 qui présente les pathologies de l'audition. Les origines et conséquences de celles-ci sont alors décrites. Le déficit principal qui caractérise les surdités est le rehaussement des seuils auditifs. Les surdités de perception cependant sont généralement le résultat d'un dysfonctionnement des cellules ciliées externes qui induit donc la disparition de certains mécanismes actifs. En dehors du rehaussement des seuils auditifs, d'autres symptômes plus gênants et plus difficiles à prendre en charge apparaissent.

A ce sujet, le chapitre 4 présente les voies possibles de réhabilitation et les différentes techniques et technologies actuellement utilisées. En particulier, la technologie des prothèses auditives est présentée ainsi que les différents types de traitements qu'elles réalisent.

1. Anatomie et physiologie de l'appareil auditif

1.1. Principe général

Le système auditif humain possède des capacités remarquables. Sa sensibilité s'étend sur une plage de fréquences audibles allant de 20 Hz à 20 kHz et couvre un ensemble de pressions acoustiques de $20\ \mu\text{Pa}$ à $20\ \text{Pa}$ ce qui représente une incroyable dynamique¹ de $120\ \text{dB}$. En outre, les capacités de discrimination aussi bien en fréquence qu'en intensité sont excellentes. Cette sensibilité est étonnante au vu du faible nombre de cellules sensorielles dont dispose la cochlée (3500 cellules ciliées internes) en comparaison avec les millions de cellules photosensibles de l'œil.

Cette qualité « se paie » en retour au niveau de la complexité de fonctionnement de la chaîne de traitement que constitue l'oreille. L'appareil auditif, dont la structure est représentée dans la figure 1.1, est divisée en trois éléments de nature et de fonction différentes mais complémentaires :

- l'oreille *externe*,
- l'oreille *moyenne* et
- l'oreille *interne*.

Avant de rentrer dans le détail, nous allons tout d'abord indiquer de manière simplifiée le fonctionnement global de l'appareil auditif.

Captées par l'oreille externe, les vibrations acoustiques sont transmises par l'oreille moyenne au milieu liquidien de la cochlée, organe de l'audition de l'oreille interne. Au sein de la cochlée, les vibrations provoquent la mise en mouvement des liquides et des différentes membranes qui la constituent. Ces mouvements provoquent à leur tour l'inclinaison des stéréocils des cellules ciliées déclenchant ainsi l'activation des fibres nerveuses. Ces dernières transmettent alors un message électrique vers le cortex cérébral.

Dans l'oreille interne, ces mécanismes vibratoires ne sont pas seulement le fait de réactions mécaniques passives. En effet, des mécanismes actifs complexes viennent modifier et enrichir les propriétés mécaniques de la cochlée.

1. $20\ \mu\text{Pa}$ correspond à un déplacement de la membrane du tympan inférieur au diamètre d'un atome d'hydrogène!

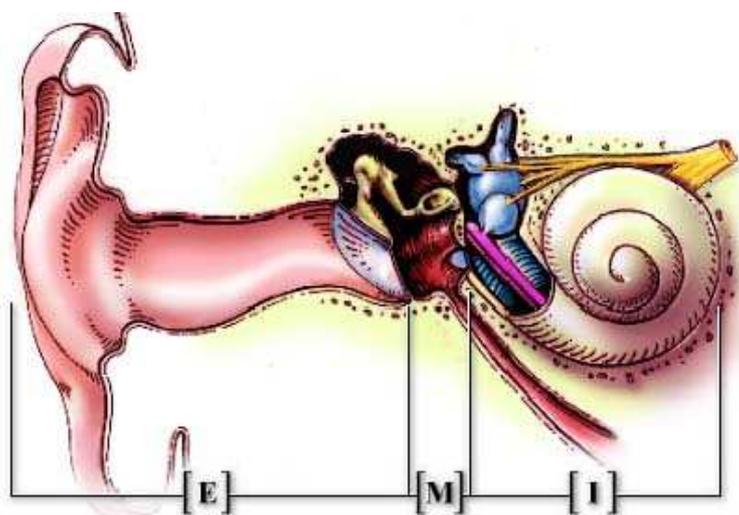


FIG. 1.1: Schéma de l'appareil auditif humain comprenant l'oreille externe [E], l'oreille moyenne [M] et l'oreille interne [I]. (image INSERM reproduit de [62])

1.2. Mesure physique de l'intensité

Le système auditif humain est sensible à la pression acoustique *i.e.* aux *variations* de la pression atmosphérique provoquées par une onde acoustique.

En acoustique, pour définir le niveau de pression, on utilise le *décibel en niveau de pression acoustique* ou *dB Sound Pressure Level* (dB SPL). Il est calculé à partir des valeurs de références en pression, P_0 , ou en intensité, I_0 , telles que :

$$\begin{aligned} P_0 &= 20 \mu Pa \\ I_0 &= 10^{-12} W/m^2 \end{aligned}$$

La *pression acoustique de référence* P_0 est choisie comme étant approximativement la plus petite valeur de pression acoustique audible. I_0 est l'intensité correspondante.

Le niveau de pression acoustique en décibels, L_{dB} , est alors défini en fonction de l'intensité acoustique, I , ou de la *pression efficace*, p , par la relation :

$$\begin{aligned} L_{dB} &= 10. \log \frac{I}{I_0} \\ &= 20. \log \frac{p}{P_0} \end{aligned} \tag{1.1}$$

Le tableau 1.1 indique quelques ordres de grandeur de niveau de *dB SPL* pour des exemples de bruits correspondants.

$dB SPL$	Pression (μPa)	Exemple de situation
130 – 140 dB	$63 \cdot 10^6 - 20 \cdot 10^7$	Seuil intolérable
120 dB	$20 \cdot 10^6$	Réacteur d'avion à 10 m
110 dB	$63 \cdot 10^5$	Atelier de chaudronnerie
100 dB	$20 \cdot 10^5$	Marteau-piqueur à 2 m
90 dB	$63 \cdot 10^4$	Atelier de tissage
80 dB	$20 \cdot 10^4$	Rue bruyante
60 dB	$20 \cdot 10^3$	Conversation vive
50 dB	$63 \cdot 10^2$	Musique douce
40 dB	$20 \cdot 10^2$	Conversation normale
30 dB	630	Résidence tranquille
20 dB	200	Studio d'enregistrement
10 dB	63	Laboratoire d'acoustique, minimum extérieur
5 dB	36	Chambre anéchoïque
0 dB	20	Niveau de référence

TAB. 1.1: Correspondance entre niveau de pression acoustique et pression

1.3. L'oreille externe

1.3.1. Anatomie

L'oreille externe est le premier maillon de la chaîne que constitue l'appareil auditif. Elle reçoit les vibrations acoustiques aériennes et les transmet à l'oreille moyenne. L'oreille externe est composée du *pavillon* et du *conduit auditif externe* (cf. figure 1.1).

1.3.2. Physiologie

L'oreille externe a une fonction d'*antenne acoustique* pour l'audition. Le pavillon amplifie de quelques décibels les fréquences avoisinant $5 kHz$ et le *conduit auditif externe* amplifie d'une dizaine de décibels celles autour de $2,5 kHz$. L'effet total du corps et de l'oreille externe engendre globalement une amplification qui varie entre 5 et 20 dB sur la plage des fréquences de 2 à $7 kHz$ [67, 59].

Chez l'homme, le pavillon n'étant pas mobile, son intérêt pour la localisation des sources est moindre que chez la plupart des autres mammifères. Néanmoins, le cerveau parvient à compenser, en partie, cet inconvénient. En effet, les ondes sonores transmises à l'oreille moyenne sont altérées par la combinaison entre les effets de diffraction acoustique sur le corps ou sur la tête et les effets de la géométrie de l'oreille externe. Pour chaque oreille, l'altération est fonction de la position de la source sonore par rapport au pavillon et par

rapport à l'axe du conduit auditif externe. Elle concerne principalement les hautes fréquences. L'oreille externe joue donc un rôle dans la localisation des sources sonores en transmettant au reste du système auditif différentes informations spectrales et temporelles témoignant de l'origine spatiale de la source et qui sont interprétées par le cortex cérébral. Plus précisément, pour une position particulière de la source sonore, la vibration acoustique parvient, à chacune des oreilles, affectée par une certaine fonction de transfert [9]. Ces fonctions de transfert varient selon les positions relatives de la source et de l'auditeur et correspondent aux *fonctions de transfert de tête* ou *Head-Related Transfer Functions* (HRTF) utilisées en traitement du signal dans le contexte de la spatialisation des sons [8, 42].

Cette faculté à localiser les sources sonores est relativement importante et intervient de manière capitale lorsque l'environnement est particulièrement bruyant en permettant à l'auditeur de focaliser son attention sur le locuteur. Cette aptitude que l'on désigne par le terme d'effet *cocktail-party* assure une bonne compréhension de la parole dans des situations à faible rapport signal à bruit.

En résumé, le rôle auditif de l'oreille externe est donc moindre que celui des autres organes de l'oreille. Celle-ci assure principalement une fonction de protection du tympan, d'amplification des sons (dans la zone 2 à 7 kHz) et permet la *localisation* des sources sonores.

1.4. L'oreille moyenne

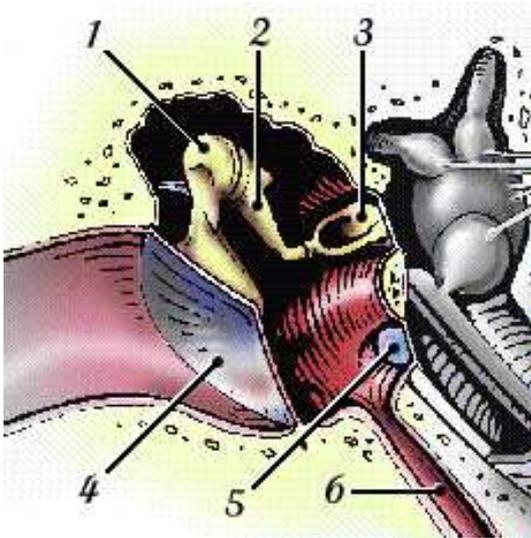
1.4.1. Anatomie

L'oreille moyenne est l'élément essentiel de la *transmission sonore*. Comme le montre le schéma 1.2, elle est composée principalement par la *membrane tympanique* et la *chaîne des osselets* ou chaîne *ossiculaire*.

1.4.2. Physiologie

Adaptation d'impédance

La fonction principale de l'oreille moyenne est de transmettre les vibrations aériennes vers l'oreille interne. Afin de parvenir à ce résultat, l'oreille moyenne réalise l'adaptation d'impédance nécessaire entre le milieu *aérien* et le milieu *liquidien* de l'oreille interne. Sans cette adaptation d'impédance, l'énergie acoustique serait réfléchiée dans sa quasi-totalité puisqu'il s'agit d'une transmission d'un milieu de basse impédance vers un milieu de haute



La chaîne ossiculaire comprend le marteau(1), l'enclume(2) et l'étrier(3). Le tympan(4) sépare le conduit auditif externe de la cavité de l'oreille moyenne qui est en relation avec la cavité buccale par la trompe d'Eustache(6). La fenêtre ovale, sur laquelle s'appuie la platine de l'étrier(3), et la fenêtre ronde(5) séparent l'oreille moyenne et l'oreille interne.

FIG. 1.2: Schéma de l'oreille moyenne (image INSERM reproduit de [62])

impédance [59, 3].

L'adaptation est réalisée en partie par un effet de levier mais surtout grâce au rapport de surface entre la surface du tympan et celle de la platine de l'étrier qui s'appuie sur la fenêtre ovale de la cochlée [65].

Ce rôle d'adaptation d'impédance est particulièrement efficace sur la plage des fréquences de la parole et permet à environ 46% de l'énergie d'être transmis [65].

Dans [54], Moore cite une autre propriété probable du système ossiculaire : la protection aux bruits internes (comme par exemple les bruits de mastication). Sans l'organisation particulière du système ossiculaire, ceux-ci seraient transmis à la cochlée par conduction osseuse.

Le réflexe stapédien

La deuxième fonction de l'oreille moyenne est de protéger, par réflexe, l'oreille interne aux sons de fort niveau, en particulier pour les basses fréquences. Ce réflexe se déclenche normalement en 150 ms pour des sons de plus de 80 dB HL (c'est à dire 80 dB au dessus du seuil auditif²) et en 25 à 35ms pour des niveaux plus élevés. Il met en jeu le muscle stapédien qui, en se contractant, augmente la rigidité de la chaîne tympano-ossiculaire et réduit ainsi la transmission sonore vers l'oreille interne [3, 62].

Cette protection est néanmoins très limitée :

2. voir la section 2.1 page 29 pour la définition des dB HL, *dB Hearing Level*

- en *niveau* : de 10 à 20 *dB* jusqu'à 2000 *Hz*
- en *durée* : au bout de *quelques minutes*, les muscles se relâchent après épuisement.
- en *fréquence* : aux sons en dessous de 1 à 2 *kHz*.
- en *nature* : il ne protège pas des bruits impulsifs.

Ce réflexe ossiculaire interviendrait également pour un locuteur ou un chanteur dans la réduction de la perception de sa propre voix.

1.5. L'oreille interne

1.5.1. Anatomie

L'oreille interne est l'organe principal de l'audition, car c'est le responsable de la *transduction* du signal acoustique en *message nerveux*. Son anatomie et sa physiologie très complexes sont à l'origine des capacités auditives très performantes dont disposent les mammifères.

La cochlée

L'oreille interne regroupe deux organes sensoriels distincts : le *vestibule*, organe de l'équilibre que nous ne décrirons pas ici et la *cochlée*, organe de l'audition.

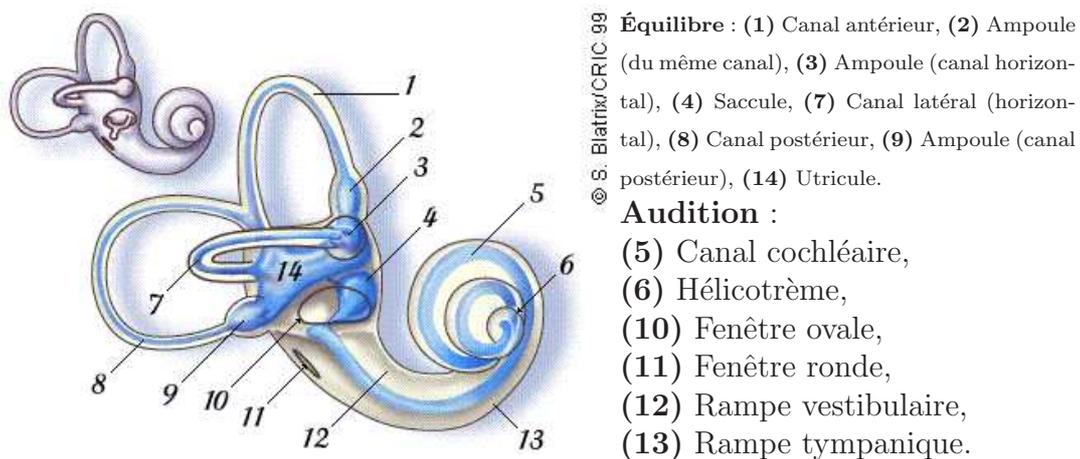


FIG. 1.3: Schéma de l'oreille interne (image INSERM reproduit de [62])

La cochlée est constituée d'un ensemble de trois tubes enroulés en spirale sur deux tours et demi (chez l'homme) et *remplis de liquides* (cf. figures 1.3 et 1.4 et 1.5) :

- les rampes tympanique et vestibulaire remplies par la *pérylymphe*,

— le canal cochléaire rempli par l'*endolymphe*.

Les vibrations mettent en mouvement le tympan et la chaîne des osselets. L'étrier, plaqué sur la *fenêtre ovale* (cf. figures 1.2 et 1.3), transfère la vibration au compartiment périlymphatique de la rampe vestibulaire selon le sens de circulation indiqué par les flèches de la figure 1.4.

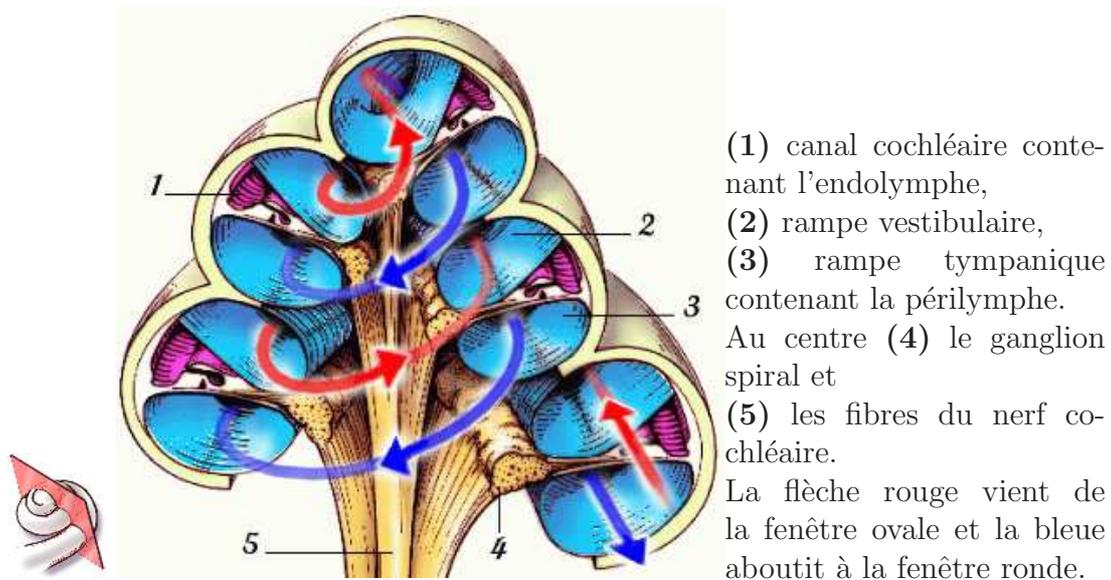


FIG. 1.4: Section axiale de la cochlée (image INSERM reproduit de [62])

C'est dans le canal cochléaire (indiqué par 1 sur le schéma 1.4 et représenté plus précisément sur la figure 1.5) que se trouve l'organe sensoriel essentiel de l'audition : l'*organe de Corti*.

L'organe de Corti

L'organe de Corti, qui repose sur la membrane basilaire, est l'organe sensori-nerveux de la cochlée. Il est composé des cellules sensorielles ou *cellules ciliées*, des fibres nerveuses qui leur sont connectées et des structures annexes ou de support. La figure 1.5 détaille son anatomie.

Il y a deux types de cellules ciliées :

- les *cellules ciliées internes* (CCI), dont le nombre par cochlée avoisine les 3500 et
- les *cellules ciliées externes* (CCE) qui sont environ 12500 par cochlée.

Les cellules ciliées sont ainsi nommées car leur extrémité supérieure, en contact avec l'endolymphe, porte une centaine de stéréocils en 3 rangées de tailles différentes (cf. la figure 1.5). Schématiquement, ces deux types de cellules (CCI et CCE) diffèrent par la forme de leur corps cellulaire et par l'arrangement des stéréocils (en ligne pour la CCI et en W pour la CCE) comme le montre la figure 1.7. Leur nombre est fixé très tôt dans le

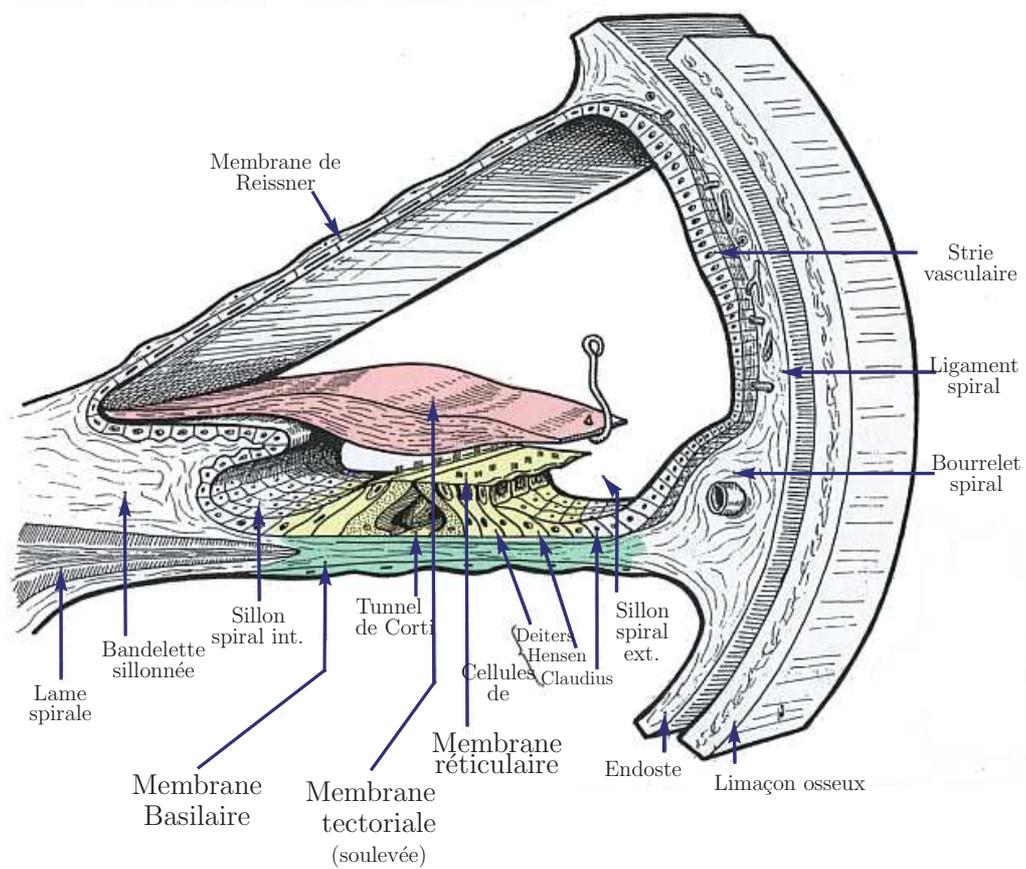


FIG. 1.5: Canal cochléaire. (reproduit de [48])

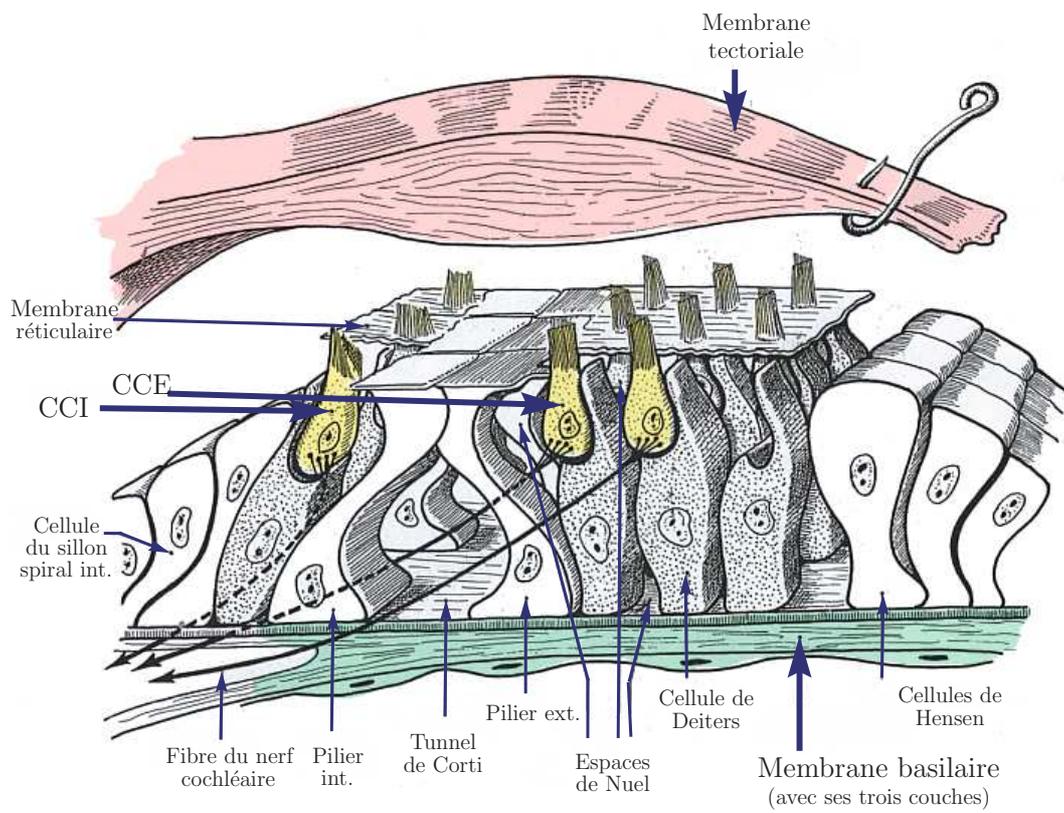


FIG. 1.6: Organe de Corti. (reproduit de [48])

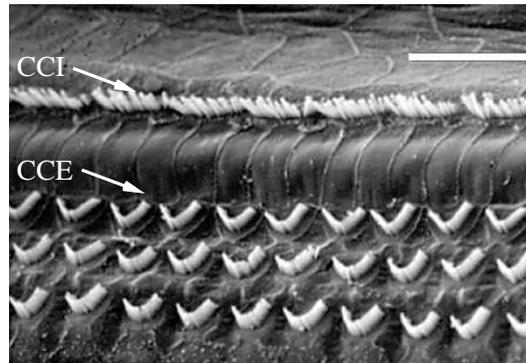


FIG. 1.7: Disposition des cellules ciliées. (photo Marc Lenoir, image INSERM reproduit de [62])

développement (10 semaines de gestation chez l'homme). Les cellules ciliées endommagées au cours de la vie ne sont pas remplacées.

On peut noter que les CCI sont pratiquement toutes identiques, alors que les CCE ont une taille variable et sont plus grandes à l'apex (*i.e.* l'extrémité) qu'à la base de la cochlée.

Innervation

Chaque CCI est *connectée* à environ 10 fibres nerveuses. Pour les CCE, le rapport est inverse, 10 CCE sont *contactées* par une même fibre. Les CCI sont les « véritables » cellules sensorielles puisqu'elles transmettent 95% des messages nerveux contre 5% pour les CCE. Par opposition, les CCE, reçoivent les messages *efférents* du système nerveux central.

1.5.2. Physiologie de la cochlée

La physiologie de la cochlée est très complexe et il n'y a pas à l'heure actuelle de consensus précis sur son fonctionnement. Dans cette section, nous présenterons néanmoins la théorie générale qui semble la plus admise dans la communauté scientifique.

Onde propagée ou Résonance ?

Deux grandes tendances se sont succédé pour expliquer le fonctionnement de l'oreille interne. A la fin du 19ème siècle, Helmholtz propose une théorie basée sur un phénomène de *résonance* de la cochlée selon la fréquence, à la manière des cordes d'une harpe [37]. En

1961, Georges von Békésy reçoit le prix Nobel pour sa théorie de l'*onde propagée* [11]. Au cours d'expériences menées sur des cochlées prélevées *post-mortem*, von Békésy montre que le mouvement vibratoire est transmis comme une onde qui se déplace de la base à l'apex. *Selon la fréquence du stimulus, cette onde produit une amplitude maximale en un point donné de la membrane basilaire* comme cela est représenté sur la figure 1.8.

La découverte de mécanismes actifs dans le fonctionnement de la cochlée est venue bouleverser cette dernière conception qui ne semble plus être qu'un épiphénomène. En effet, la finesse de la sélectivité fréquentielle de l'oreille humaine et la dynamique des sons perçus ne pouvaient être expliquées par le simple modèle de l'onde propagée de von Békésy.

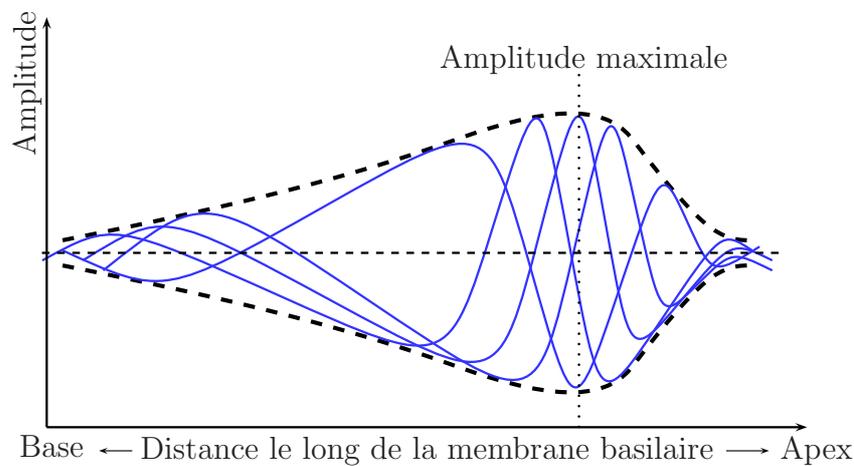


FIG. 1.8: Représentation schématique de l'enveloppe des ondes propagées selon le modèle de von Békésy

La tonotopie passive

La membrane basilaire devient de plus en plus large et de plus en plus flexible de la base de la cochlée à l'apex. Comme l'a montré von Békésy, cette structure particulière engendre une sensibilité préférentielle de chaque point de la cochlée à une tonalité spécifique. On parle alors d'une *organisation tonotopique*. La figure 1.9 décrit schématiquement cette *tonotopie* dite *passive* où les hautes fréquences mettent en mouvement la base de la cochlée tandis que les basses fréquences mobilisent l'extrémité (l'*apex*).

Les mécanismes actifs

L'organe de Corti n'est pas seulement un organe passif qui transforme les mouvements de la membrane basilaire en message nerveux. En effet, les CCE ont la propriété de

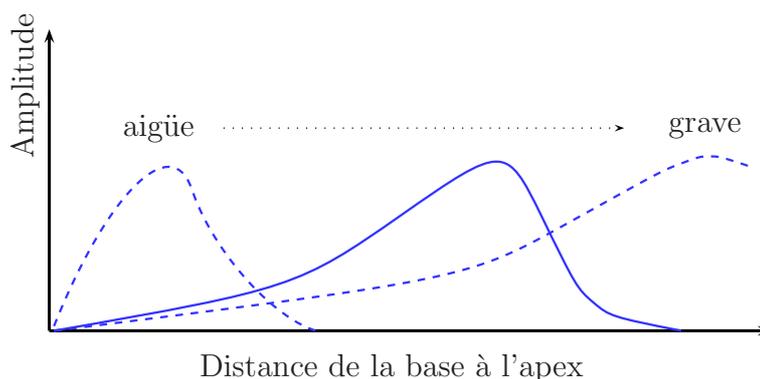


FIG. 1.9: Représentation schématique de la tonotopie passive de la cochlée. Chaque zone de la cochlée vibre préférentiellement pour une fréquence donnée.

pouvoir se *contracter*³, *amplifiant* ainsi le mouvement de la membrane basilaire. Pour des niveaux d'intensités faibles à modérés, le fonctionnement de l'organe de Corti peut schématiquement se résumer en 5 phases [62] :

1. Les vibrations sonores transmises à la périlymphe font onduler la membrane basilaire de haut en bas. La tonotopie passive mobilise la membrane basilaire de la base (sons aigus) à l'apex (sons graves).
2. Les stéréocils des CCE, implantés dans la membrane tectoriale, sont déplacés horizontalement. La CCE est alors dépolarisée.
3. Les CCE excitées (dépolarisées) se contractent (électromotilité). Du fait du couplage mécanique étroit entre CCE, membrane basilaire et lame réticulaire, ce mécanisme actif fournit de l'énergie amplifiant⁴ la vibration initiale ; en même temps il joue un rôle de filtre sélectif (*tonotopie active*).
4. La CCI est alors excitée, probablement par un contact direct avec la bande de Hensen de la membrane tectoriale.
5. La synapse entre CCI et la fibre du nerf auditif est activée et un message nerveux est envoyé au cerveau.

L'action des CCE est probablement commandée par le système nerveux central et enrichit le système passif mis en évidence par Békésy. On peut noter qu'une destruction ou une désactivation (par voie médicamenteuse) des CCE restaure les caractéristiques vibratoires de la membrane basilaire de l'étude de von Békésy. La figure 1.10 illustre l'action des CCE en représentant la réponse de la membrane basilaire à un son pur (ici 10 kHz) et en la comparant à la réponse observée par von Békésy. Le mécanisme actif amplifie d'environ 50 dB la vibration de la membrane basilaire et surtout il accorde cette vibration sur une portion beaucoup plus fine de l'organe de Corti, permettant ainsi une remarquable

3. Le mouvement des CCE est déclenché de manière électrique, on parle alors d'*électromotilité*. L'environnement liquidien fortement polarisé électriquement joue un rôle très important pour ce mécanisme (cf. [62])

4. Cette théorie présentant un « amplificateur cochléaire » est la théorie qui fait le plus l'unanimité dans la communauté scientifique, ce n'est cependant pas la seule.

tonotopie. Cet accord en fréquences (*tuning*) dépend étroitement de l'intégrité des CCE et se retrouve à l'identique au niveau des fibres du nerf auditif auxquelles il est fidèlement transmis par les CCI. La tonotopie passive décrite plus haut est alors remplacée par une tonotopie *active* aux valeurs légèrement décalées vers l'apex. Ce mécanisme actif opère pour des niveaux d'intensités sonores faibles à modérés. En effet, à partir de 90 dB SPL environ, on retrouve le fonctionnement passif décrit par Békésy.

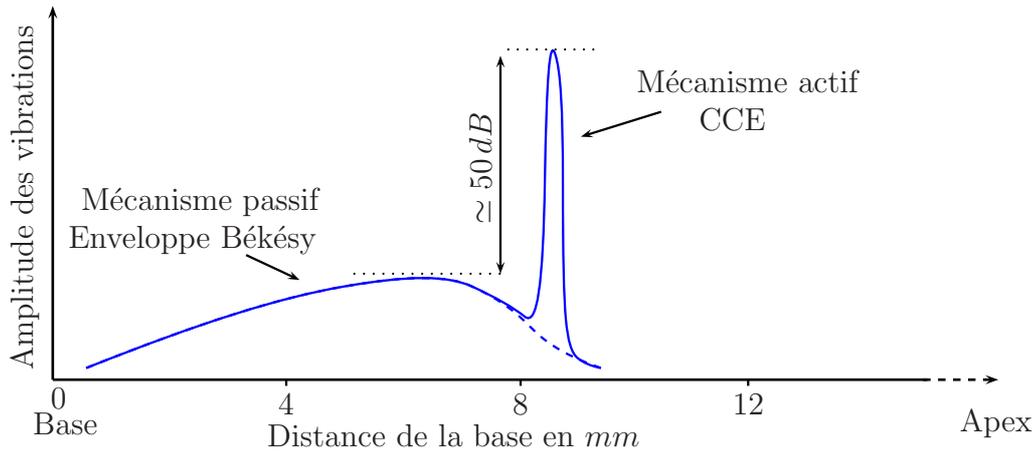


FIG. 1.10: Réponse de la membrane basilaire d'un cobaye à un son pur de 10 kHz , le mécanisme actif amplifie de 50 dB la vibration de la membrane basilaire par rapport à l'enveloppe passive de Békésy et affine la sélectivité fréquentielle.

Ces mécanismes actifs sont capitaux pour l'audition notamment en ce qui concerne :

- l'excellente sensibilité de l'oreille aux sons faibles,
- la finesse de la sélectivité fréquentielle de l'oreille et
- la non-linéarité de la réponse de la membrane basilaire à l'intensité sonore.

Les deux premiers points sont observables sur la figure 1.10. On note particulièrement la *spécificité fréquentielle* en chaque point de la cochlée. La figure 1.11 indique la relation entre la position d'un point de la membrane basilaire et la fréquence qui lui est associée dans ce mécanisme; on parle alors de Fréquence Caractéristique (FC) de la membrane basilaire en ce point.

Il y a donc un *codage* fréquence/position inhérent à l'anatomie de la cochlée. Cependant, le codage de l'information fréquentielle par la cochlée n'est pas que tonotopique, le phénomène du *verrouillage de phase* que nous décrirons plus loin semble aussi entrer en jeu.

La non-linéarité

En raison des mécanismes actifs, la relation entre l'amplitude des déplacements de la membrane basilaire et l'intensité sonore n'est pas linéaire. La *non-linéarité* de la mem-

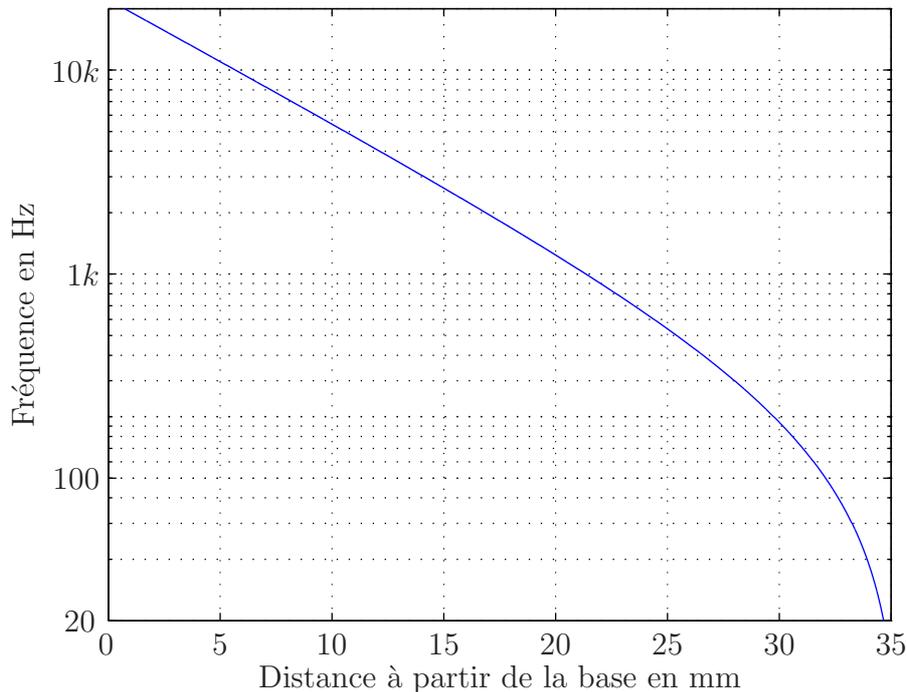


FIG. 1.11: Correspondance tonotopique entre la position sur la cochlée et la fréquence caractéristique

brane basilaire est telle qu'elle réalise un phénomène de *compression* et de *saturation*. Il est illustré sur le schéma 1.12. Ce schéma représente la réponse de la membrane basilaire à un son pur de 8 kHz émis à différentes intensités. La mesure de la réponse de la membrane basilaire est effectuée au point de FC 8 kHz . Par comparaison avec la réponse linéaire (indiquée en pointillé), on constate que le comportement de la membrane est compressif. L'interprétation la plus courante du phénomène observé est qu'à faible intensité (jusqu'à 20 dB SPL) les mécanismes actifs amplifient le mouvement de 40 à 50 dB , ce qui donne l'allure à peu près linéaire. Entre 30 et 90 dB SPL , on observe un *palier*⁵ pour lequel le gain diminue, réalisant ainsi la compression. Enfin aux intensités supérieures à 90 dB SPL , il n'y a plus de mécanismes actifs et on retrouve le comportement linéaire des mécanismes passifs. Les variations du gain correspondant à cette interprétation sont schématisées dans la figure 1.13. Ce modèle confirme le retour au fonctionnement passif de Békésy, décrit plus haut, pour des intensités sonores supérieures à 90 dB SPL .

Sur la figure 1.12, les vibrations de la membrane basilaire sont mesurées en un point de la cochlée ayant pour FC la fréquence du stimulus. Il est important de noter que plus on s'éloigne de ce point plus la réponse retrouve un caractère linéaire (disparition des mécanismes actifs).

Cette non-linéarité est principalement observable à la base de la cochlée, *i.e.* pour les

5. cette plage d'intensités correspond aux intensités usuelles de l'audition comme par exemple celles de la parole

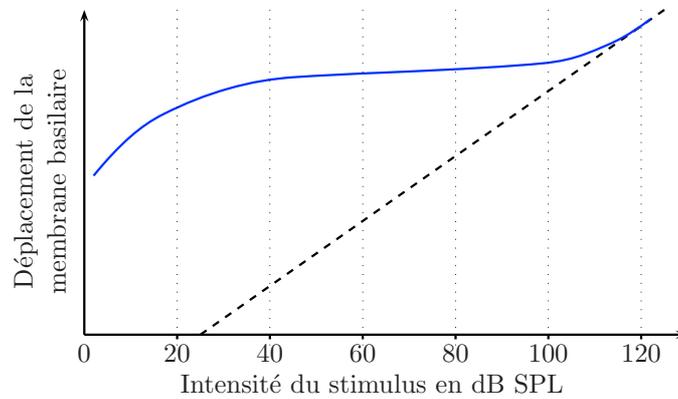


FIG. 1.12: Réponse schématique de la membrane basilaire à un stimulus. On stimule la cochlée avec un son pur à 8 kHz , la courbe tracée représente les déplacements observés sur la membrane basilaire au point de FC 8 kHz . (A partir de [41])

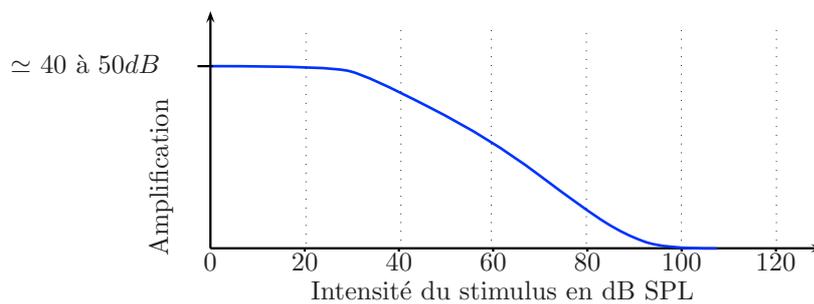


FIG. 1.13: Amplification des vibrations de la membrane basilaire par les mécanismes actifs selon l'intensité et d'après l'interprétation de la figure 1.12.

hautes fréquences. Dans les basses fréquences, la compression semble être moindre et elle apparaît pour d'autres fréquences de stimulus que la FC [54].

D'autres phénomènes sont également caractéristiques de cette non-linéarité :

- la suppression bitonale : la réponse à un premier ton décroît en présence d'un second ton de fréquence voisine.
- les sons de combinaison ou produits de distorsion : deux sons de fréquence f_1 et f_2 ($f_1 < f_2$) peuvent générer au niveau de la membrane basilaire une excitation correspondant à des sons de fréquences $f_2 - f_1$, $2f_2 - f_1$ ou $2f_1 - f_2$ et qui peuvent être perçus par l'auditeur (c'est le cas en particulier pour $2f_1 - f_2$) [59].

Remarque

Ces propriétés des mécanismes actifs sont capitales pour notre travail. En effet, la majorité des pathologies de l'audition pour lesquelles on envisage la mise en place de prothèses a pour source une déficience des cellules ciliées, en particulier des CCE, et correspond donc à un dysfonctionnement des mécanismes actifs. Nous reviendrons sur ce point dans le chapitre 3 consacré aux pathologies de l'audition.

1.6. Le nerf auditif

Schématiquement, les CCI stimulent les fibres afférentes tandis que les CCE sont stimulées par les fibres efférentes. Le rôle du système nerveux efférent est probablement un rôle de *feedback* qui influence les mécanismes actifs des CCE.

1.6.1. Courbe d'accord

Chaque fibre nerveuse devient sensible à un stimulus à partir d'une certaine intensité *seuil*. La figure 1.14 représente ce seuil de déclenchement en fonction de la fréquence du stimulus pour une fibre donnée. Une telle courbe s'appelle une *courbe d'accord*. On retrouve au niveau des courbes d'accord, la même sélectivité fréquentielle que celle déduite des mouvements de la membrane basilaire. Sur la figure 1.14, deux courbes d'accord ont été tracées : l'une obtenue avec les CCE intactes et l'autre après destruction des CCE. On note la similitude avec la réponse de la membrane basilaire de la figure 1.10. La spécificité fréquentielle se retrouve donc aussi au niveau de la fibre nerveuse. Par analogie avec la notion de Fréquence Caractéristique d'un point de la cochlée, on parle alors également de la Fréquence Caractéristique (FC) de la fibre. La FC d'une fibre nerveuse correspond à peu près à celle de la membrane basilaire à l'endroit où elle est connectée. En l'absence de CCE, le seuil de déclenchement de la fibre nerveuse est plus élevé de 40 dB à 50 dB et la spécificité fréquentielle de la fibre est atténuée.

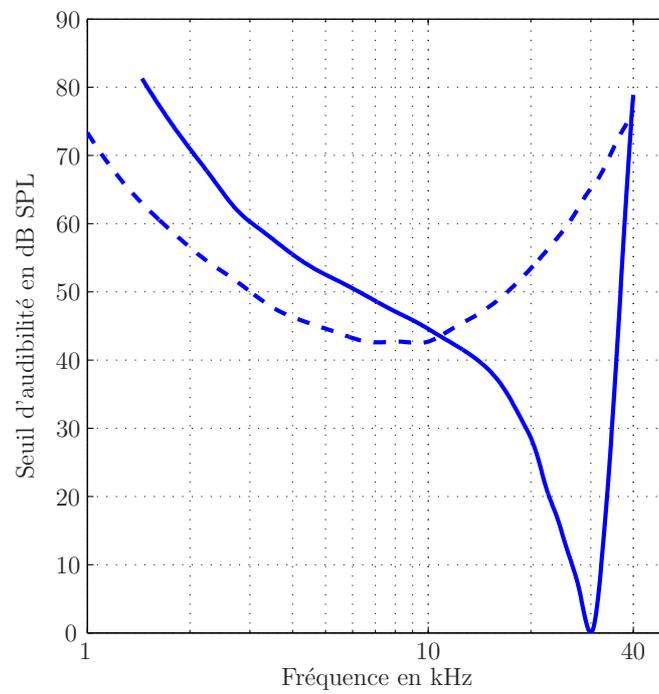


FIG. 1.14: Courbe d'accord d'une fibre nerveuse. Sous conditions normales (en trait plein) et après destruction des CCE (en pointillé). Retracer à partir de [66].

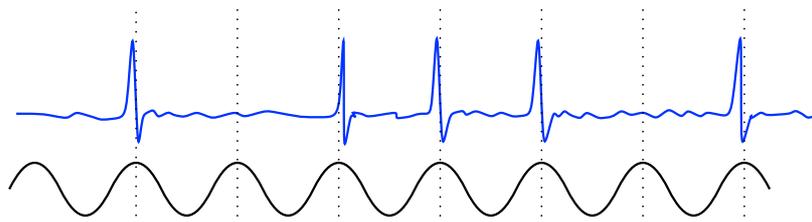


FIG. 1.15: Synchronisation temporelle des décharges d'une fibre (en haut) avec la phase avec la fréquence du stimulus, ici un son pur (en bas). Les décharges se produisent toujours sur la même phase

1.6.2. Codage de la fréquence

Deux phénomènes semblent entrer en jeu dans le codage de la fréquence des stimuli et constituent deux alternatives complémentaires :

- la tonotopie *i.e.* un codage place-fréquence et
- le *verrouillage de phase*.

L'hypothèse tonotopique a été présentée précédemment et se retrouve à l'identique au niveau de la membrane basilaire et des fibres du nerf auditif, pourtant un autre phénomène est observé qui *pourrait* aussi participer au codage de la fréquence.

En effet, pour une stimulation relativement basse fréquence, inférieure à 4 ou 5 kHz , on constate que, pour toute fibre nerveuse de FC arbitraire, les impulsions électriques sont *synchronisées* sur la phase du stimulus. La figure 1.15 illustre ce phénomène, l'impulsion électrique de la fibre est toujours déclenchée pour une même valeur de phase du stimulus. Quelquefois, les fibres ne peuvent pas répondre à chaque période pour les fréquences plus hautes, comme on le voit sur le schéma 1.15 pour les décharges absentes en position deux et six. Au delà de 5 kHz , il n'y a plus de verrouillage de phase car le rythme est trop élevé-vis à-vis de la période réfractaire des fibres qui est d'environ 1ms.

Le verrouillage de phase ou *phase-locking* est donc un phénomène *temporel* qui est potentiellement capable de remplir une certaine fonction dans le codage de la fréquence bien que cela ne soit pas encore acquis. Il est également possible que de telles informations sur la phase soient utiles pour la localisation des sources sonores.

Remarque

Pour certaines fibres, cette synchronisation est observable même à des niveaux d'intensités en dessous du seuil d'audibilité.

1.6.3. Codage de l'intensité

Sur chaque CCI est liée une dizaine de fibres nerveuses. Chacune de ces fibres possède un seuil de déclenchement propre et son taux de décharges par seconde augmente avec l'intensité du stimulus jusqu'à saturation. Comme la dynamique de chaque fibre n'est pas suffisante pour couvrir celle de l'audition, l'information sur l'intensité est probablement répartie sur les différentes fibres à différents niveaux. Le cortex cérébral interprète alors ces informations en niveaux d'intensité perçue.

Remarque

Ce sont directement les fibres nerveuses que les implants cochléaires viennent stimuler.

1.7. Conclusions sur la physiologie de l'appareil auditif

Nous venons de voir que l'appareil auditif humain a une structure très complexe. Cette complexité se retrouve au niveau de son fonctionnement. En particulier, on retiendra que des mécanismes actifs viennent enrichir la sensibilité et la sélectivité de l'oreille aux faibles niveaux sonores. Ces mécanismes dépendent étroitement de l'intégrité des CCE et leur action s'atténue lorsque l'intensité des stimuli croît laissant peu à peu place aux mécanismes passifs moins sensibles et moins sélectifs.

Le fonctionnement que nous venons de décrire n'est cependant pas suffisant pour décrire le fonctionnement global de l'appareil auditif car il ne prend pas en compte la manière avec laquelle le cerveau traite l'information nerveuse qu'il reçoit. Ce niveau du système auditif est bien évidemment encore plus complexe à étudier et on a alors recours à un autre domaine scientifique, la *psychophysique* et en particulier dans notre cas, la *psychoacoustique*.

2. Psychoacoustique et perception auditive

Après avoir décrit l'anatomie et la physiologie de l'appareil auditif, nous allons maintenant décrire ses capacités et ses propriétés perceptuelles dans des conditions normales de fonctionnement.

2.1. Les seuils auditifs

Comme on l'a vu dans le chapitre précédent, le comportement de la cochlée n'est pas identique pour les basses et les hautes fréquences, notamment dans la mise en jeu des mécanismes actifs. De même, les fonctions de transfert de l'oreille externe et de l'oreille interne varient selon la fréquence. Il en découle une variation des sensations auditives selon la hauteur tonale du stimulus. Les *seuils d'audition* ou *seuils auditifs* indiquent le niveau sonore minimal requis pour détecter un stimulus sonore en fonction de sa fréquence. Ainsi, on constate, par exemple, que le seuil de détection en *dB SPL* d'un son grave est plus haut que celui d'un son aigu.

Les seuils auditifs sont variables suivant les individus. Les valeurs moyennes de seuils auditifs pour des sujets normo-entendants sont indiquées dans le tableau 2.1 pour une mesure monoaurale sous écouteurs et pour dix fréquences. On notera particulièrement la différence relativement importante des niveaux de seuil entre les hautes et les basses fréquences. Les valeurs des seuils auditifs sont principalement fixées par la fonction de transfert globale de l'oreille externe et de l'oreille moyenne, excepté pour les basses fréquences où le seuil est vraisemblablement limité par le bruit interne de l'organisme. Ces valeurs sont utilisées dans le dépistage et la quantification des troubles auditifs. En effet, ils permettent de définir l'échelle des *dB Hearing Level* (*dB HL*) ou *dB HTL* (*Hearing Level* ou *Hearing Threshold Level*), qui indique la valeur en *dB* par rapport au seuil auditif. Cette échelle est normalisée selon les valeurs des seuils auditifs du tableau 2.1, ainsi, à une fréquence donnée, le niveau sonore de *0 dB HL* correspond au niveau sonore en *dB SPL* du seuil auditif. À titre d'exemple, un son pur à *500 Hz* d'intensité *31,5 dB SPL* a un niveau de *20 dB HL*. De même, un patient ayant un seuil d'audibilité à *20 dB HL*, à une fréquence donnée, a donc une perte auditive de *20 dB* à cette fréquence. On retrouve les courbes de seuil d'audition sur les courbes d'isotonie présentées dans la section suivante.

Fréquence (en Hz)	Seuil en dB	Fréquence (en Hz)	Seuil en dB
125	45	2000	9
250	25,5	3000	10
500	11,5	4000	9,5
1000	7	6000	15,5
1500	6,5	8000	13

TAB. 2.1: Seuils d'audition monoauraux sous écouteurs

2.2. Intensité subjective

2.2.1. Échelle des phones

Si le seuil auditif varie en fonction de la fréquence, il en est de même pour la sensation d'intensité. La sensation d'intensité est d'ailleurs identique au niveau des seuils auditifs pour toutes les fréquences. Afin d'obtenir une mesure d'intensité sonore à valeur subjective, l'échelle des phones a été construite. Par définition, le niveau sonore en phones d'un son est la valeur du niveau sonore en $dB SPL$ d'un son pur de $1 kHz$ évoquant la même sensation d'intensité. Ainsi, l'échelle des phones coïncide avec l'échelle des $dB SPL$ pour un son de $1 kHz$. L'échelle des phones indique des sensations d'intensité sonore équivalentes le long de la plage des fréquences audibles. Elle est naturellement illustrée par les courbes isophoniques ou courbes d'isotonie. La figure 2.1 présente les courbes d'isotonie sur écouteurs moyennées sur plusieurs sujets. Chaque courbe correspond à un niveau subjectif constant d'intensité en fonction de la fréquence et du niveau sonore en $dB SPL$. On désigne par *phone* l'unité d'intensité subjective.

La figure 2.2 représente les courbes isotoniques mesurées cette fois en champ libre pour une écoute binaurale. Leur allure est légèrement différente dans la plage de fréquences $1 - 8 kHz$ du fait des diffractions sur la tête et au niveau de l'oreille externe. Sur ces courbes, la courbe la plus basse (en pointillé) indique le seuil auditif.

Les courbes d'isotonie sont représentatives des limites du champ sonore de l'audition. En effet, si la courbe la plus basse correspond au seuil d'audition, les courbes hautes (110 à $120 dB$) indiquent le *seuil de douleur* à partir duquel le son est trop fort et peut endommager gravement l'oreille interne.

Remarque

Un exemple d'application de ces considérations psychoacoustiques en acoustique est l'utilisation de courbes de pondération pour les mesures d'intensité sonore, notamment pour les mesures de bruit. En effet, afin de compenser approximativement les différences subjectives entre fréquences, trois courbes de pondérations ont été définies. Chacune de ces courbes est utilisée pour une

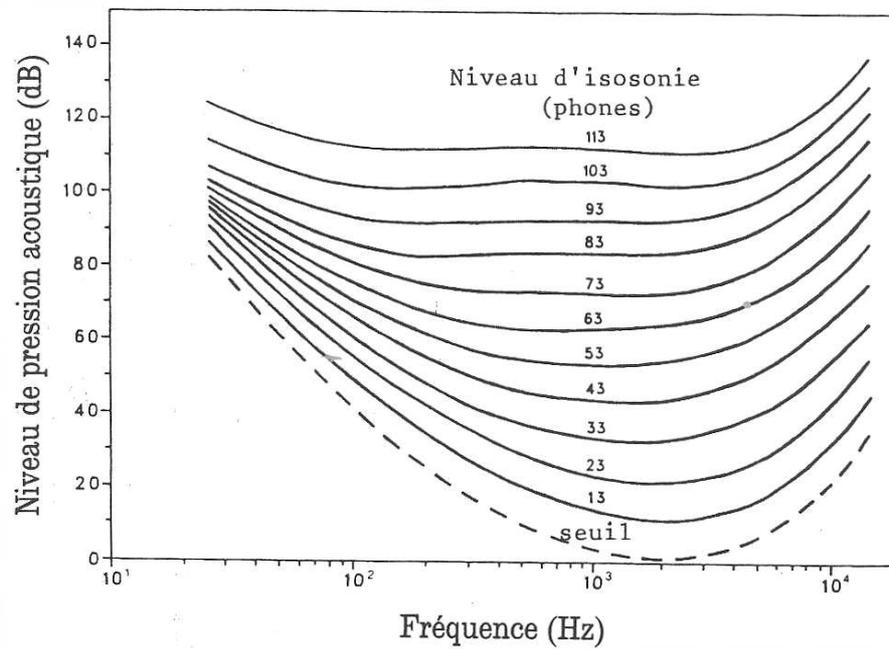


FIG. 2.1: Courbes d'isophonie pour l'écoute binaurale sur écouteurs. Reproduit de [9]

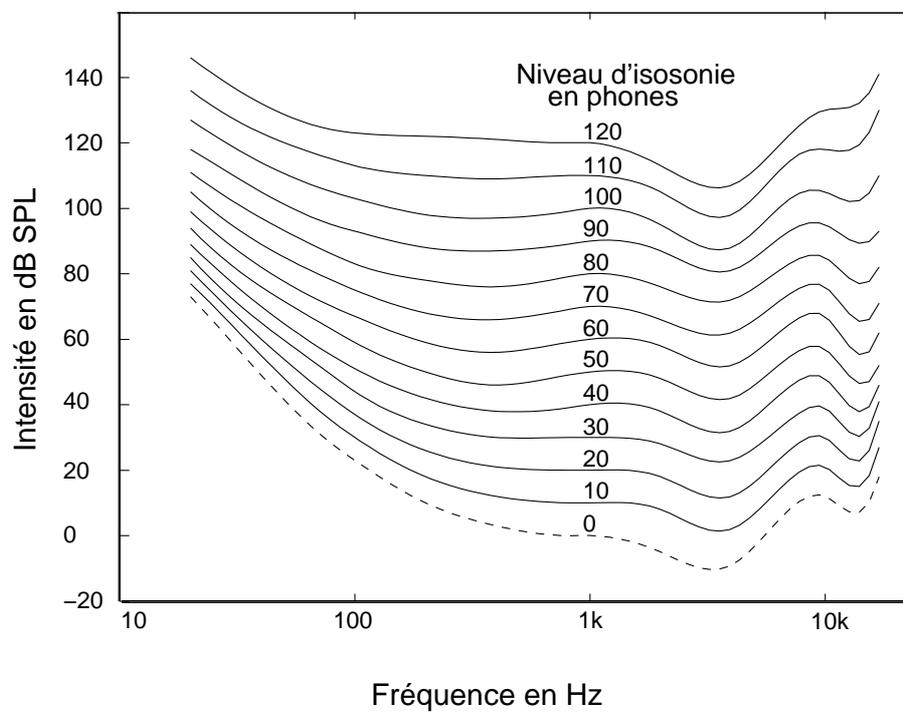


FIG. 2.2: Courbes d'isophonie pour l'écoute binaurale en champ libre. Reproduit de [9]

plage de niveau de bruit donnée et est construite à partir de la ligne isosonique de niveau correspondant :

- $dB(A)$ \longrightarrow 40 phones pour les bruits de faible niveau,
- $dB(B)$ \longrightarrow 70 phones pour les bruits de niveau moyen et
- $dB(C)$ \longrightarrow 100 phones pour les bruits de fort niveau.

La figure 2.3 indique la valeur des courbes de pondération A et C qui sont les plus couramment employées dans les sonomètres.

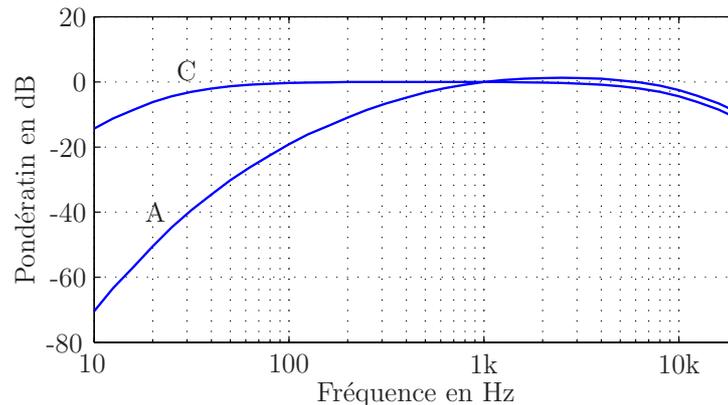


FIG. 2.3: Courbes de Pondérations A et C

2.2.2. Échelle des sones ou échelle d'intensité subjective

L'échelle des phones définie précédemment est encore une échelle relativement physique. En effet, si elle indique les niveaux de même sensation, elle ne renseigne pas sur la force de la sensation. L'échelle qui permet de définir réellement la sensation d'intensité perçue ou *sonie* est l'échelle des *sones*. Par convention, l'échelle des sones est définie par :

- la valeur 1 *sonie* est la sonie d'un son binaural de 40 phones et
- lorsque l'intensité *perçue* du son double, la sonie double.

Approximativement, on peut dire qu'au dessus de 40 phones, la sonie double pour toute augmentation de 10 dB du niveau en phones; la sonie S en sones s'exprime donc en fonction du niveau Ph en phones par la relation suivante :

$$S = 2^{\left(\frac{Ph-40}{10}\right)} \quad (2.1)$$

Pour les niveaux de sonie inférieurs à 1 (40 phones), la sonie croît plus rapidement. La figure 2.4 représente la valeur de la sonie en fonction du niveau d'intensité en phones. Sur cette figure, la courbe en pointillé illustre l'approximation réalisée par la formule 2.1.

L'allure non-linéaire des courbes de sonie dépend de deux phénomènes vus précédemment :

- la non-linéarité de la réponse de la membrane basilaire (*cf.* section 1.5.2 page 21),

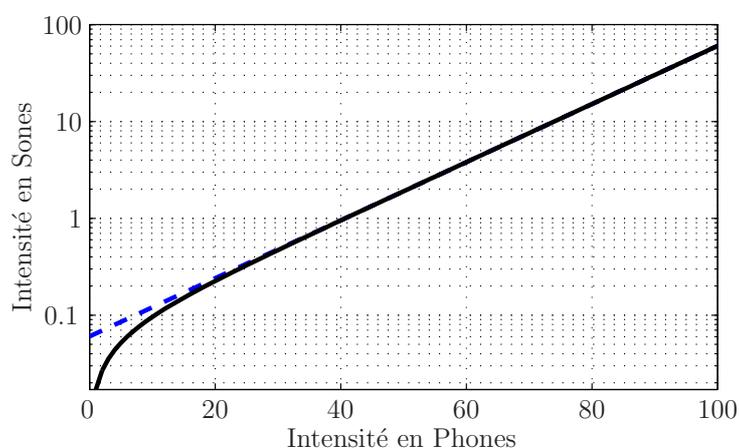


FIG. 2.4: Courbe de sonie pour un signal sinusoïdal à 1 kHz (d'après [54]). La courbe en trait plein indique la relation entre le niveau perçu en sones et l'intensité en $dB\ SPL$. La courbe en pointillé représente l'approximation réalisée par la formule (2.1)

— le codage de l'intensité dans le nerf auditif ou la réponse de fibres nerveuses (*cf.* section 1.6.3 page 27).

Ces résultats et ces définitions sont valables pour des sons purs présentés seuls. Dans le cas de sons complexes, il devient beaucoup plus difficile de définir la sonie car la perception de tels sons fait alors intervenir les notions de *filtres auditifs* et de phénomènes de *masquage*.

2.3. Les filtres auditifs

2.3.1. Principe

Lorsqu'on cherche à caractériser les lois psychoacoustiques pour des sons complexes, les observations montrent qu'il est alors pertinent de considérer que le système auditif réalise une analyse fréquentielle du signal à l'aide d'une série de filtres passe-bandes. Ces filtres, appelés *filtres auditifs*, se recouvrent continûment tout le long de la plage des fréquences audibles.

Afin de faire le lien avec la physiologie de l'oreille interne, on peut considérer que chaque région de la membrane basilaire est sensible à une certaine plage de fréquences et peut être ainsi assimilée à un filtre passe-bande.

Ce phénomène est particulièrement mis en évidence dans les expériences de masquage et de perception d'intensité.

Lorsque le système auditif tente de détecter un signal, on suppose qu'il utilise la sortie d'un filtre auditif dont la fréquence centrale est proche de celle du signal. Seul le bruit à l'intérieur de la bande passante détermine si le signal est détectable ou non. En réalité, plusieurs expériences ont montré que le système auditif utilise le filtre auditif dans lequel le rapport signal à bruit est le plus élevé (*off-frequency listening* ou *off-place listening*). Un premier exemple qui illustre ce propos est issu d'une expérience de Fletcher en 1940 : on considère un signal sinusoïdal masqué par un bruit de bande de largeur variable, le seuil d'audibilité du signal croît avec la largeur de bande du bruit jusqu'à ce que le spectre du bruit atteigne la bande atténuée du filtre. Dès lors le seuil reste approximativement constant. La figure 2.5 illustre cette expérience.

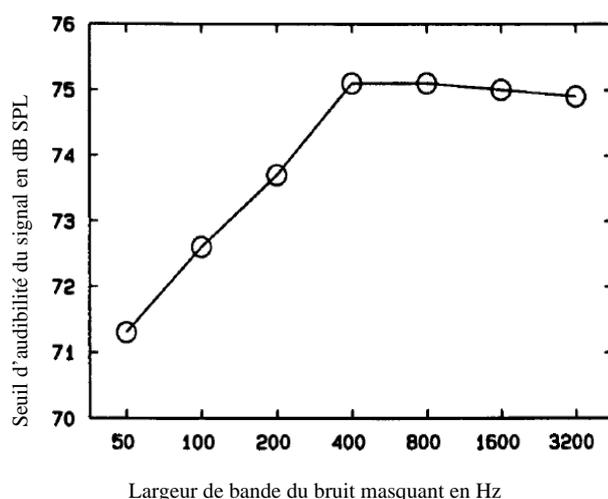


FIG. 2.5: Seuil d'audibilité d'un signal en fonction de la largeur de bande d'un bruit masquant. Le signal est un sinusoïde à 2 kHz masquée par un bruit de bande également centré sur 2 kHz et dont la largeur varie.

Masquage simultané De la même manière, lorsque deux sons purs sont présentés à l'oreille, le ton le plus faible ne sera audible que si le rapport des puissances est suffisamment élevé dans un des filtres auditifs. Ce phénomène est connu sous le nom de phénomène de *masquage* et révèle le rôle important des filtres auditifs pour la compréhension de la *sélectivité fréquentielle* du système auditif.

Schématiquement, supposons que la distance fréquentielle entre deux tons soit inférieure à une largeur de bande de filtre auditif et considérons un filtre auditif de fréquence centrale proche de celle du signal le plus faible (appelé signal *masqué* contrairement au premier appelé signal *masquant*). Les deux signaux se situent tous deux dans la bande passante du filtre qui a donc une faible influence sur le rapport des puissances des signaux. Le seuil d'audibilité du son masqué est donc relativement élevé.

Inversement, supposons que la distance fréquentielle entre les deux tons soit supérieure

à une largeur de bande de filtre auditif. Le masqueur se situe alors dans la bande atténuée du filtre et le rapport des puissances est donc accru. Par conséquent le seuil d'audibilité du signal masqué est relativement faible.

Ainsi, l'effet d'un signal masquant est prépondérant dans la bande critique auquel il appartient. Cependant s'il est de niveau acoustique suffisant, il peut également modifier les seuils d'audition dans les bandes adjacentes (*spread of masking*).

Ce mécanisme, décrit très simplement ici, illustre le pouvoir de séparabilité ou de résolution fréquentielle du système auditif.

La figure 2.6 donne un exemple de courbes d'effet de masque indiquant les seuils d'audibilité d'un son pur en fonction de sa fréquence pour un bruit de bande masquant présenté à différents niveaux d'intensité.

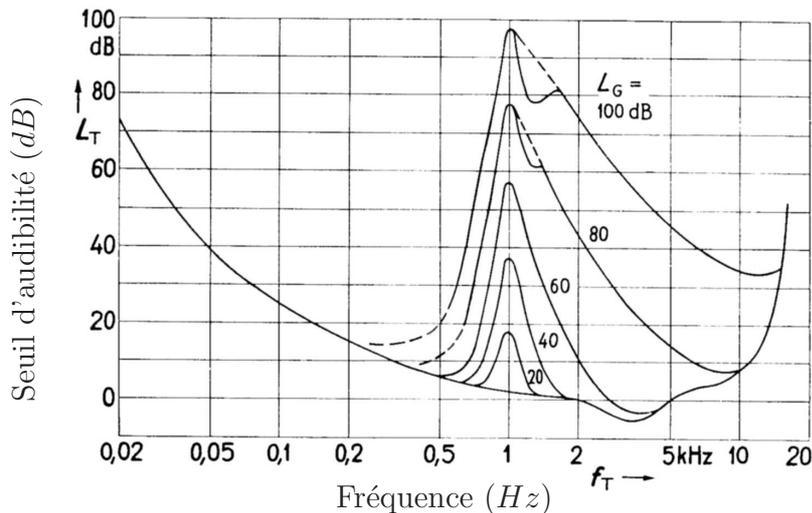


FIG. 2.6: Courbes d'effet de masque de bruits à bande étroite. Le signal masquant est un bruit de bande centré à 1 kHz de largeur 160 Hz présenté aux niveaux $L_G = 100, 80, 60, 40$ et 20 dB. D'après [81]

Dans l'expérience illustrée par la figure 2.6 le masqueur est fixe, tandis que le signal masqué varie en fréquence. Une alternative consiste à garder le signal constant tandis qu'on fait varier le masqueur. Les courbes ainsi obtenues sont appelées courbes de *sélectivité de fréquence psychoacoustique* (psychophysical tuning curves, PTC). Plusieurs exemples de PTC sont représentés sur la figure 2.7. Pour chaque courbe, le signal de référence est présenté à 10 dB SL (indiqué par un point) et le niveau maximal du signal masquant tel que le signal de référence soit juste audible est tracé en fonction de sa fréquence. On peut trouver plus de détail sur ce sujet dans [54].

Les courbes PTC de la figure 2.7 sont intéressantes dans la mesure où elles présentent une forte similitude avec les réponses de la membrane basilaire de la figure 1.10 (page 21) et surtout avec les courbes d'accord des fibres nerveuses de la figure 1.14 (page 25)

présentées dans le chapitre précédent. Elles permettent donc d'établir un lien direct entre les observations physiologiques et psychoacoustiques.

Enfin, le signal étant fixe, on peut supposer approximativement qu'un seul et même filtre auditif est sollicité. Chaque PTC représente donc approximativement la forme inverse du filtre auditif dont la fréquence centrale est celle du signal. Si on fait l'hypothèse que les filtres auditifs sont linéaires, on peut alors obtenir leurs formes en inversant l'allure des PTC obtenues.

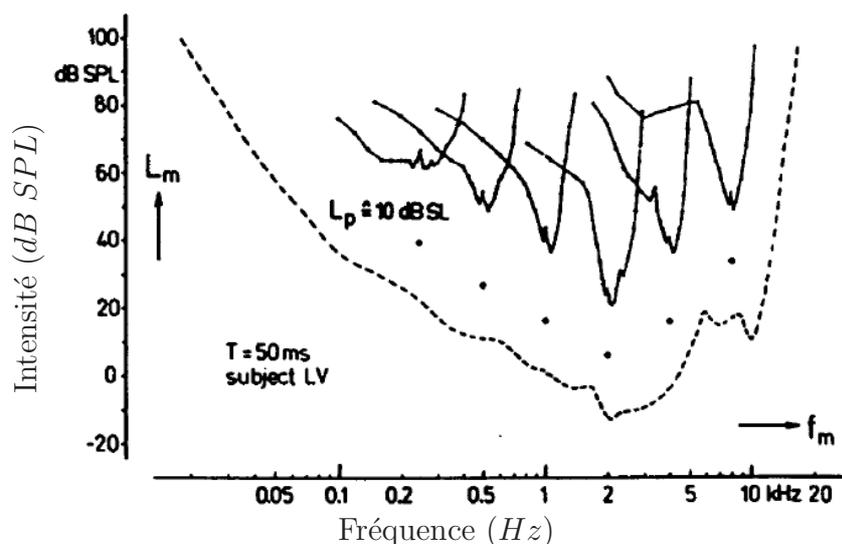


FIG. 2.7: Courbes de sélectivité de fréquence psychoacoustique (psychophysical tuning curves) cf.[54]

Remarque

Le phénomène de masquage a été récemment utilisé avec succès dans les applications de codage audio. L'originalité de ce type de codage est de faire en sorte que le bruit de quantification soit masqué par le signal utile.

Masquage non-simultané Le phénomène de masquage est également observable pour des sons non-simultanés. On parle alors de *masquage non-simultané* ou *masquage temporel*. Dans ce type de masquage, un son fort présenté *avant ou après* un stimulus peut masquer ce dernier. La figure 2.8 illustre ce phénomène.

2.3.2. Largeur des filtres auditifs

Historiquement, les filtres auditifs ont été caractérisés, en premier lieu, par leur largeur de bande sous l'hypothèse qu'ils pouvaient être modélisés par des filtres passe-bande

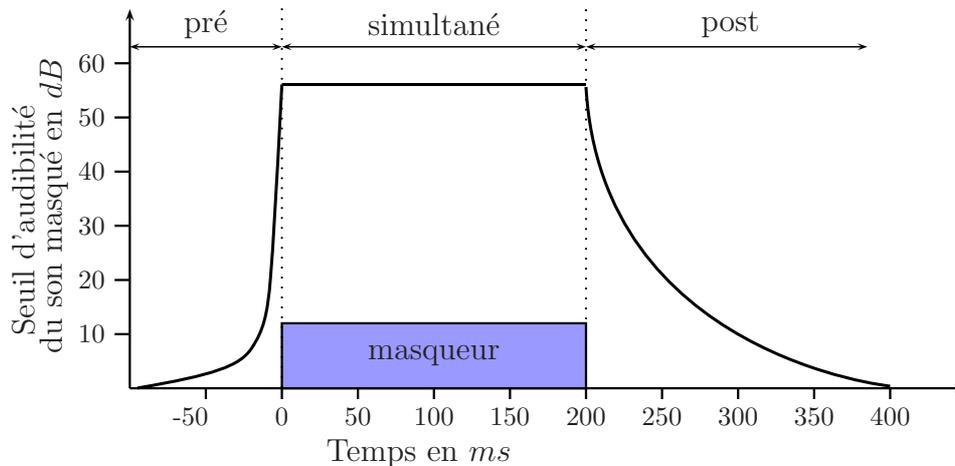


FIG. 2.8: Masquage non-simultané

idéaux, *i.e.* rectangulaires. Étant donné la complexité des phénomènes mis en jeu au niveau de la cochlée, cette hypothèse est évidemment très simplificatrice. En effet des mesures plus précises et plus rigoureuses des filtres auditifs ont montré leur caractère non-linéaire et en particulier leur dépendance vis-à-vis du niveau des stimuli. Les filtres auditifs ont en réalité une bande passante plus étroite pour les niveaux modérés que pour les niveaux forts.

Cependant, dans ce cadre très complexe, la modélisation par des passe-bande idéaux constitue une première approche pratique et elle permet de quantifier approximativement la résolution fréquentielle du système auditif.

De nombreuses autres expériences que l'expérience de Flechter utilisant des sons purs ou des bruits de bande permettent de mesurer la largeur de bandes des filtres auditifs selon la fréquence centrale.

La largeur d'un filtre est appelée *Bande Critique*. On se réfère souvent à la mesure des bandes critiques effectuée par Zwicker. Nous la citons en exemple, ici, car elle est très populaire et a été utilisée dans de nombreuses applications de traitement du signal faisant appel à des notions de psychoacoustique. Selon les mesures de Zwicker, les bandes critiques ont une largeur constante de 100 Hz pour des fréquences centrales jusqu'à environ 500 Hz . Au dessus de 500 Hz , la largeur des bandes critiques augmente avec la fréquence jusqu'à une largeur de 3500 Hz pour une fréquence centrale de 13500 Hz . La valeur des bandes critiques BW_c peut être calculée en fonction de la fréquence centrale f_c en Hz grâce à la formule suivante :

$$BW_c = 25 + 75 \cdot \left[1 + 1,4 \cdot (f_c/1000)^2 \right]^{0,69} \text{ Hz} \quad (2.2)$$

La figure 2.9 montre les valeurs de bandes critiques calculées à partir de cette formule.

Les valeurs de bandes critiques sont définies en tout point de la membrane basilaire,

cependant dans la pratique, notamment afin de définir un banc de filtres, il est souvent plus commode de considérer un nombre discret de filtres auditifs. A cet effet, le tableau 2.2 indique un découpage normalisé de l'échelle des fréquences en 24 bandes adjacentes. Ceci constitue alors une nouvelle échelle de fréquences dite *échelle Bark* allant de 1 à 25 Bark. Par définition, l'unité *Bark* est telle que chaque filtre auditif ait une largeur de 1 Bark. On obtient ainsi une échelle de fréquence cohérente avec la résolution fréquentielle du système auditif.

Fréquence Centrale	Bande Critique	No de bande
50	~ 100	1
150	100	2
250	100	3
350	100	4
450	110	5
570	120	6
700	140	7
840	150	8
1000	160	9
1175	190	10
1370	210	11
1600	240	12
1850	280	13
2150	320	14
2500	380	15
2900	450	16
3400	550	17
4000	700	18
4800	900	19
5800	1100	20
7000	1300	21
8500	1800	22
10500	2500	23
13500	3500	24
19500	-	25

TAB. 2.2: Normalisation de l'échelle des Barks et découpage en 25 bandes critiques

La formule qui permet alors de convertir une fréquence f en Hertz en sa valeur en Bark est la suivante [82] :

$$B(f) = 13. \arctan\left(\frac{0,76f}{1000}\right) + 3,5. \arctan\left(\left(\frac{f}{7500}\right)^2\right) \text{ Bark} \quad (2.3)$$

Plus récemment, une autre modélisation de la largeur de bande des filtres auditifs a été proposée par Moore et Glasberg dans [28] selon une méthode de mesure introduite par Patterson dans [58].

Dans leurs travaux, Moore et Glasberg ont mesuré des largeurs de filtre plus fines que celles obtenues par Zwicker, notamment pour les basses fréquences. A partir de ces résultats, les auteurs ont proposé une nouvelle mesure de la largeur des filtres auditifs appelée *largeur de bande rectangulaire équivalente* ou *Equivalent Rectangular Bandwidth* (ERB). Pour chaque filtre auditif, la valeur d'ERB est définie comme la largeur d'un filtre passe-bande idéal de même fréquence centrale qui, alimenté par un bruit blanc, produirait la même énergie en sortie. Typiquement, les valeurs d'ERB indiquent des bandes passantes dont la largeur varie entre 11% à 17% de la valeur de la fréquence centrale (donc à peu près à *facteur de qualité* Q constant).

La figure 2.9 montre une comparaison entre les largeurs des filtres décrits en bande critique ou en ERB. La mesure des ERB illustre donc une meilleure résolution fréquentielle du système auditif que celle indiquée par les valeurs de bandes critiques. La formule suivante donne la relation moyenne pour plusieurs sujets normo-entendants entre la fréquence centrale f en Hz et l'ERB en Hz :

$$ERB_N(f) = 24,7 \cdot \left(\frac{4,37f}{1000} + 1 \right) \quad (2.4)$$

Les valeurs de bandes critiques de la figure 2.9 sont calculées à partir de cette formule en comparaison avec les valeurs obtenues à partir de l'équation (2.2).

En outre, on peut définir une échelle des ERB de manière similaire à l'échelle Bark. Chaque numéro d'ERB correspond alors à une distance fréquentielle d'une Bande Rectangulaire Équivalente.

$$N^{\circ} ERB_N(f) = 21,4 \log_{10} \left(\frac{4,37f}{1000} + 1 \right) \quad (2.5)$$

Les formules (2.3) et (2.5) sont illustrées et comparées sur la figure 2.10. La figure 2.11 montre la même comparaison rapportée à la position sur la membrane basilaire à partir de l'apex (par rapport à la tonotopie de la membrane basilaire). On notera particulièrement sur cette dernière figure la correspondance quasi-linéaire entre l'échelle ERB_N et la distance en mm le long de la cochlée. Cette dernière remarque vient renforcer le lien entre la modélisation psychoacoustique des filtres auditifs et les observations physiologiques.

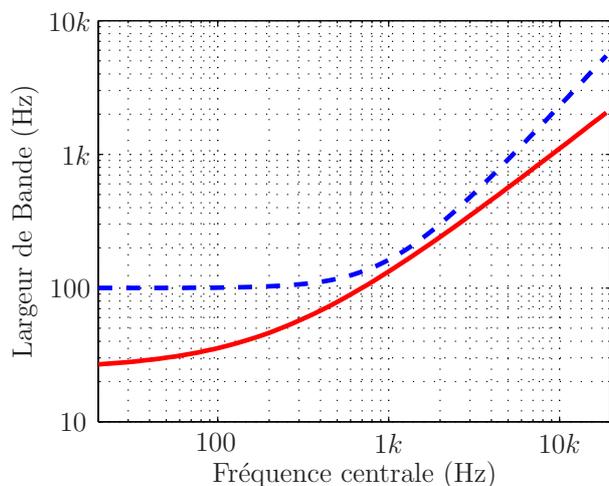


FIG. 2.9: Comparaison entre la largeur des bandes critiques en Bark et en ERB. Le terme *Critical Bandwidth* désigne généralement la largeur correspondant à l'échelle Bark (en pointillé), le terme ERB désigne la *bande rectangulaire équivalente* (en trait plein)

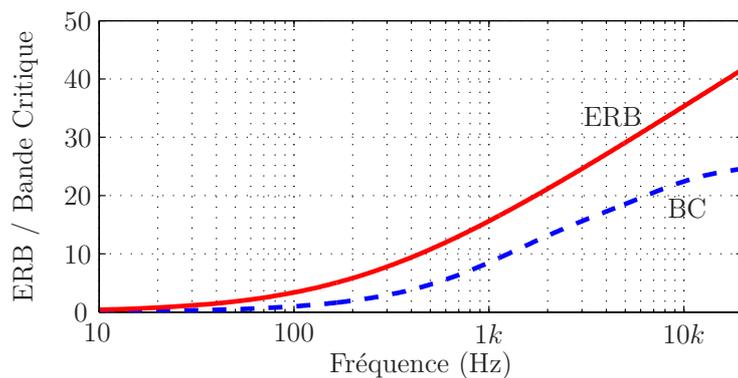


FIG. 2.10: Comparaison entre échelle Bark et numéro d'ERB en fonction de la fréquence

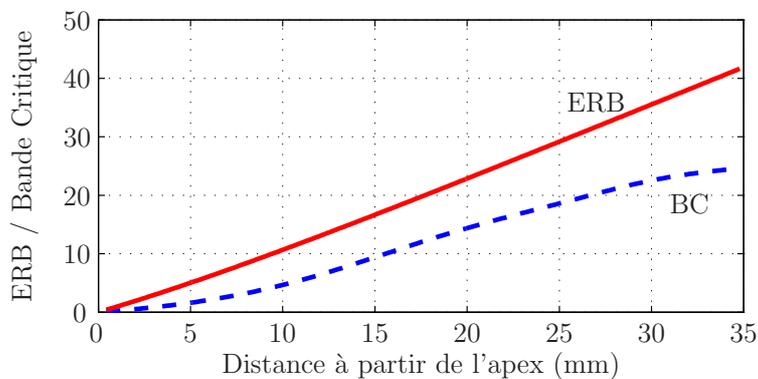


FIG. 2.11: Comparaison entre échelle Bark et numéro d'ERB en fonction de la position sur la membrane basilaire

2.3.3. Forme des filtres auditifs

La mesure rigoureuse du gabarit fréquentiel des filtres auditifs est une tâche difficile. Outre les non-linéarités des mécanismes mis en jeu, la principale difficulté est de s'assurer que les observations faites lors d'une expérience font appel au même filtre tout au long de cette expérience.

Dans les sections précédentes, nous avons vu que les filtres auditifs sont définis pour toute fréquence audible et que la largeur des filtres augmente avec la fréquence de telle sorte que cette largeur vaut approximativement 11% à 17% de la valeur de la fréquence centrale. Ces caractéristiques influent sur la résolution fréquentielle du système auditif qui est donc plus précis dans les basses que dans les hautes fréquences.

Néanmoins, cette description ne prend pas en compte tous les aspects complexes des filtres auditifs. En effet, du fait des non-linéarités des différents mécanismes du système auditif, la largeur des filtres est également dépendante de l'intensité du stimulus. Cette constatation est liée aux mécanismes actifs de la cochlée.

Schématiquement, on peut résumer ainsi les propriétés générales des filtres auditifs :

- Le sommet de chaque filtre est étroit et à peu près symétrique en fréquence.
- Les bords sont larges et asymétriques, en particulier leur pente du côté des basses fréquences est moins forte.
- plus le niveau du stimulus est élevé plus les filtres sont larges, en particulier pour les fréquences inférieures à la fréquence centrale.

Dans certaines modélisations, on inclut le gain des mécanismes actifs de l'oreille interne. Les figures 2.12 et 2.13 représentent des mesures de filtres auditifs à différents niveaux de stimulation respectivement pour des fréquences centrales de 250 Hz et 1 kHz. Sur ces schémas, les niveaux d'intensité du stimulus s'étendent de 30 dB SPL à 80 dB SPL par pas de 10 dB. On y remarque l'élargissement des filtres et la diminution du gain maximal lorsque l'intensité du stimulus croît.

2.3.4. Pattern d'excitation

La modélisation par les filtres auditifs permet également une représentation schématique de l'excitation de la membrane basilaire en réponse à un stimulus. Cette représentation est appelée Pattern d'Excitation (PE), anglicisme tiré de l'expression *excitation pattern*. Moore et Glasberg ont proposé dans [55] une méthode simple pour estimer le PE à partir de la sortie des filtres auditifs. Celui-ci est alors obtenu en traçant le niveau de sortie de chaque filtre en fonction de sa fréquence centrale.

La figure 2.14 montre un exemple de différents PE :

- celui d'un son pur à 500 Hz (10,7 en échelle ERB) et d'intensité 50 dB SPL indiqué par la courbe englobant la zone d'aire claire centrée à 10,7 ERB,

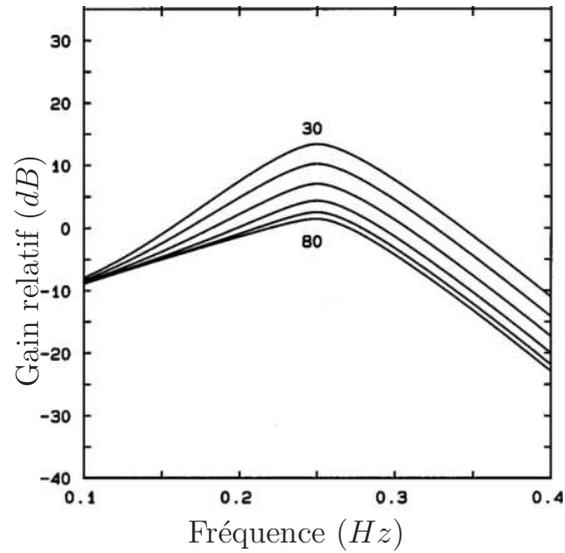


FIG. 2.12: Filtres auditifs à 250 Hz en fonction de l'intensité du stimulus. (reproduit de [29])

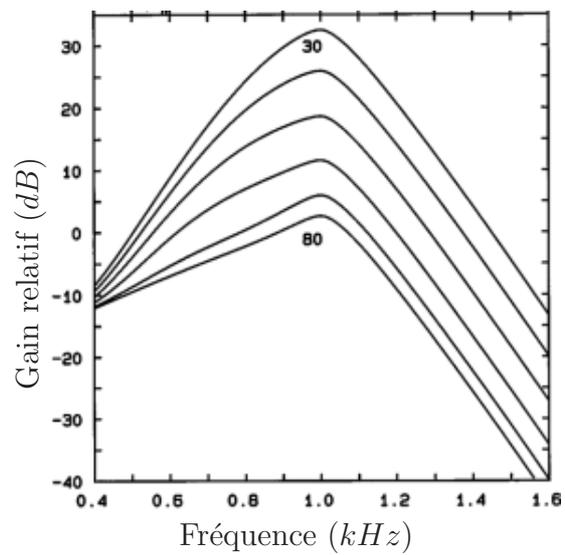


FIG. 2.13: Filtres auditifs à 1 kHz en fonction de l'intensité du stimulus. (reproduit de [29])

- celui d'un son pur à 1200 Hz (17 en échelle ERB) et d'intensité 20 dB SPL indiqué par la courbe englobant la zone d'aire foncée centrée à 17 ERB,
- et enfin le PE du son complexe composé des deux tons précédents, représenté par la courbe supérieure en trait gras.

On retrouve bien, sur les trois PE, la tonotopie de la membrane basilaire telle qu'elle a été présentée dans le chapitre 1.

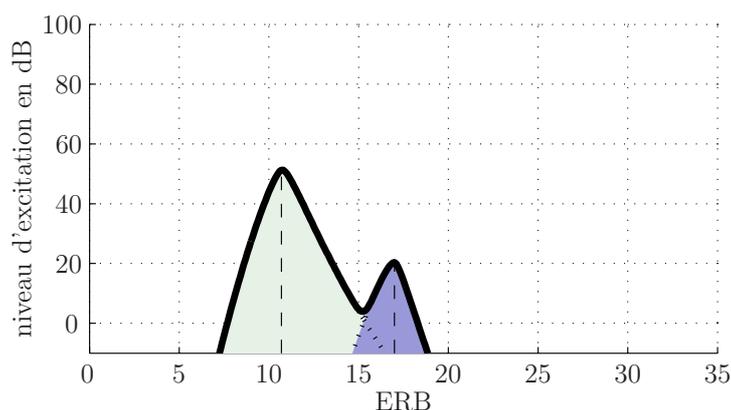


FIG. 2.14: Pattern d'excitation de deux sons purs de niveau modéré. $F_1 = 500\text{ Hz}$ (50 dB SPL) et $F_2 = 1200\text{ Hz}$ (20 dB SPL).

Le pattern d'excitation est un outil très intéressant car il illustre avec clarté la résolution fréquentielle du système auditif. Le PE du stimulus bitonal de la figure 2.14 présente en effet deux pics fréquentiels distincts correspondant à chacun des tons le composant. Dans ce cas, il n'y a pas de masquage et les deux tons sont audibles.

Considérons le même exemple que précédemment mais pour des tons purs rehaussés de 40 dB chacun. A ce niveau de stimulation, les filtres auditifs sont plus larges. Les différents PE obtenus pour ces nouveaux stimuli sont reproduits sur la figure 2.15. On remarque alors que le PE du premier ton indique une excitation beaucoup plus étendue sur la membrane basilaire que dans le premier exemple et surtout qu'en tout point de la membrane basilaire, le premier ton apporte plus d'énergie que le second. Du point de vue des PE, le second PE est complètement recouvert par le premier et le PE de la somme est pratiquement confondu avec le premier et n'indique plus qu'un seul pic fréquentiel. On voit donc ici clairement apparaître le phénomène de masquage.

La deuxième application pratique du PE est son utilisation dans les calculs de sonie. En effet, en appliquant une certaine fonction de compression à la courbe du PE (modélisant les non-linéarités de l'oreille interne), on obtient la courbe dite de *sonie spécifique partielle*. L'aire de la surface contenue sous cette courbe est une bonne approximation de la sonie [55].

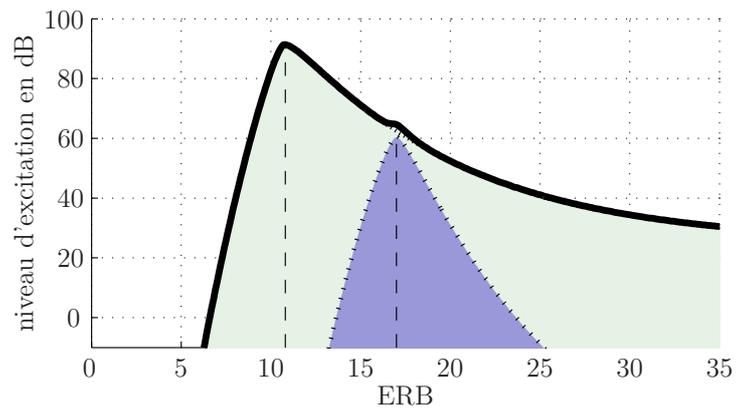


FIG. 2.15: Pattern d'excitation de deux sons purs de fort niveau. $F1 = 500\text{ Hz}$ (90 dB SPL) et $F2 = 1200\text{ Hz}$ (60 dB SPL)

3. Pathologies de l'audition

3.1. Les différents types de surdit 

Le premier diagnostic effectu  dans le cas d'une surdit  est la mesure du rehaussement des seuils auditifs   certaines fr quences d termin es¹ qu'on quantifie en terme de pertes en *dB*. Le diagramme des pertes en fonction de la fr quence est l'audiogramme (cf. exemple d'audiogramme sur la figure 3.1).

Pour exprimer le degr  de surdit  d'un malentendant, le Bureau International d'Audio-Phonologie (BIAP) a  tabli une classification audiom trique pr sent e dans le tableau 3.1 se basant sur la moyenne des pertes aux fr quences 500 *Hz*, 1000 *Hz*, 2000 *Hz* et 4000 *Hz*. Si l'oreille ne perçoit absolument pas une fr quence, on note la perte auditive   120 *dB*. Pour une perte moyenne de 120 *dB*, on consid re donc qu'aucun son n'est per u.

Seuil moyen	Classification	Effets
0-20 <i>dB HL</i>	Audition normale ou subnormale	�ventuellement atteinte tonale l�g�re sans incidence sociale
21-40 <i>dB HL</i>	Surdit� l�g�re	La parole est per�ue � voix normale, elle est difficilement per�ue � voix basse ou lointaine. Cependant La plupart des bruits familiaux sont per�us.
41-55 <i>dB HL</i> 56-70 <i>dB HL</i>	Surdit� moyenne (1er degr�) (2�me degr�)	La parole est per�ue si on �l�ve la voix. Le sujet comprend mieux en regardant parler. Quelques bruits familiers sont encore per�us.
71-80 <i>dB HL</i> 81-90 <i>dB HL</i>	Surdit� s�v�re (1er degr�) (2�me degr�)	La parole est per�ue � voix forte pr�s de l'oreille. Seuls les bruits forts sont encore per�us.
> 90 <i>dB HL</i>	Surdit� profonde	La parole n'est plus du tout per�ue. Seuls les bruits tr�s puissants sont per�us.

TAB. 3.1: Tableau de classification audiom trique

La mesure de l'audiogramme est r alis e par *conduction a rienne* ou par *conduction osseuse*, il permet alors de d terminer la nature de la surdit . Trois types de surdit  sont

1. 125, 250, 500, 1000, 2000, 4000 et 8000 *Hz*

ainsi distingués : les *surdités de transmission*, les *surdités de perception* et les *surdités mixtes* rassemblant les deux premières.

3.1.1. Les surdités de transmission

Les surdités de transmission ont pour origine une déficience de l'oreille moyenne ou externe soit au niveau du conduit auditif, du tympan ou de la chaîne ossiculaire (par exemple à la suite d'otites). Elles traduisent une mauvaise transmission entre le milieu aérien et la cochlée. Les seuils auditifs mesurés par transmission aérienne (sous écouteurs en général) sont donc rehaussés, surtout dans les basses fréquences.

Pour une surdité de transmission pure, l'oreille interne étant intacte, il n'y a pas de rehaussement des seuils auditifs lorsqu'on les mesure par conduction osseuse. De même le malentendant perçoit normalement sa propre voix.

Dans le cas d'une surdité mixte, on quantifie donc les pertes liées à une surdité de transmission par la différence entre les rehaussements des seuils auditifs par conduction aérienne et par conduction osseuse.

En général, la perte auditive liée à une surdité de transmission ne dépasse pas 60 dB (au delà de 60 dB SPL , l'énergie du stimulus est suffisante pour être transmise au milieu liquidien de la cochlée même si l'oreille moyenne ne remplit plus son rôle d'adaptation d'impédance).

3.1.2. Les surdités de perception

Les surdités de perception sont, quant à elles, d'origine cochléaire. Elles traduisent une atteinte des cellules ciliées externes ou internes ou du nerf auditif à quelque endroit de son trajet. Une combinaison de ces trois origines est également possible.

Une surdité de perception pure se traduit par un rehaussement des seuils auditifs identique par conduction osseuse et par conduction aérienne.

Les causes de l'atteinte de la cochlée sont multiples. Elles peuvent être soit génétiques soit acquises notamment suite à des surstimulations acoustiques (au dessus de 90 dB SPL) ou liées au vieillissement naturel de l'organe. L'oreille interne est l'organe sensoriel le plus fragile car les cellules ciliées, notamment les CCE, sont relativement sensibles et, comme cela a été précisé précédemment, car elles ne sont pas renouvelées. La perte de l'audition avec l'âge, appelée *presbyacousie*, est liée à la disparition progressive et irrémédiable des cellules ciliées et des neurones de la cochlée et est donc inéluctable et de nature beaucoup

moins bénigne que la presbytie². La figure 3.1 illustre par des courbes d'audiogramme, les conséquences progressives de la presbyacousie sur les valeurs des seuils auditifs.

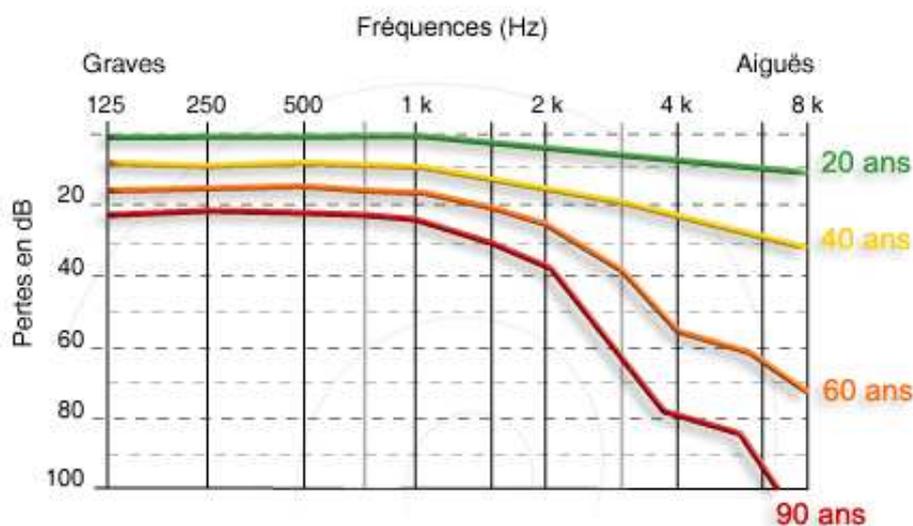


FIG. 3.1: Courbes d'audiogramme en fonction de l'âge. Ces courbes représentent des moyennes d'audiogrammes cliniques (en perte de dB) réalisées chez des sujets âgés de 20 ans à 90 ans. Source CRIC/INSERM [62].

3.2. Effets généraux des surdités

Le cas des surdités de transmission est relativement simple par rapport aux surdités de perception. Comme la transmission fonctionne de manière anormale, il y a une perte d'énergie transmise au milieu liquidien de la cochlée. Cette *atténuation* du signal acoustique entraîne donc naturellement un rehaussement des seuils auditifs.

Pour les surdités de perception, les causes physiologiques sont multiples et les mécanismes touchés sont très complexes. Dans ce cas, on est amené à distinguer les atteintes des CCE d'une part et celles des CCI ou des neurones du nerf auditif d'autre part.

Dans le cas d'une atteinte des CCE, ce sont les mécanismes actifs décrits précédemment qui sont touchés et donc, à travers ceux-ci, la sensibilité de l'oreille aux sons faibles et la finesse de sa sélectivité fréquentielle.

Dans le cas d'une atteinte des CCI ou des neurones du nerf auditif, la transduction du message acoustique ne se fait plus dans la zone touchée. On parle alors de *zone morte*. Un

² La presbytie correspond à un dysfonctionnement du cristallin et non pas à une atteinte des cellules sensorielles de la vision.

stimulus dont la fréquence correspond à la région de la cochlée atteinte ne peut donc être perçu que dans une région voisine. Or plus la fréquence caractéristique de cette région utile est éloignée de la fréquence du stimulus moins les mécanismes actifs agissent.

Bien que ces deux atteintes soient physiologiquement très distinctes, elles induisent donc toutes les deux une disparition locale des mécanismes actifs et des non-linéarités de l'oreille interne. Ceci entraîne des conséquences bien plus lourdes que dans le cas de surdité de transmission pour laquelle ces mécanismes sont toujours présents.

Dans la suite de cette section, nous allons présenter de manière concise les effets des pathologies de la cochlée. Une description complète peut être consultée dans l'ouvrage de Brian Moore [53].

3.3. Rehaussement des seuils auditifs

Une destruction des CCE sur une zone donnée de la cochlée entraîne la disparition des phénomènes non-linéaires sur cette zone. En particulier,

- les phénomènes de compression et d'amplification dus aux mécanismes actifs disparaissent alors et
- les filtres auditifs s'élargissent.

Une conséquence directe de l'atteinte des CCE est donc le rehaussement des seuils auditifs.

Si les CCI ou les neurones d'une région de la cochlée sont atteints, il n'y a plus de transduction possible du message acoustique en message électrique au niveau de la zone touchée. Si on considère un stimulus de fréquence correspondant à la région atteinte, la détection de ce stimulus ne peut avoir lieu que dans les zones voisines c'est-à-dire dans un filtre auditif pour lequel la fréquence de stimulation appartient à la bande atténuée. Là encore il y aura un rehaussement du seuil auditif et la disparition des non-linéarités puisque celles-ci ne sont observables qu'au point de la cochlée dont la fréquence caractéristique est celle du stimulus.

3.4. Le phénomène de recrutement

Schématiquement, on a vu que l'action des mécanismes actifs peut être modélisée, en partie, par l'apport d'un gain qui varie en fonction de l'intensité du stimulus (*cf.* section 1.5.2, en particulier le schéma 1.13 page 23). Ce gain variable permet d'étendre approximativement de 50 dB la dynamique de l'oreille interne par rapport à son fonctionnement passif. Une conséquence directe du dysfonctionnement de l'oreille interne est donc la réduction de la dynamique de l'oreille. *Il s'agit ici de la réduction de la dynamique des intensités physiques perçues et non de celle de la sensation d'intensité.* En effet, les

mécanismes actifs enrichissent les propriétés mécaniques de la membrane basilaire mais le mécanisme de transduction en message électrique au niveau des CCI reste identique.

Considérons un malentendant ayant une perte de 30 dB à une fréquence donnée auquel on présente un signal sinusoïdal à cette fréquence de niveau sonore juste audible, c'est-à-dire à 30 dB HL . Supposons que cette perte auditive soit uniquement due à une déficience des CCE. La sensation d'intensité perçue (sonie) par le malentendant est la même que celle évoquée par un stimulus de 0 dB HL pour un normo-entendant. Par contre, un signal particulièrement fort à la limite du seuil de douleur (100 dB SPL) présenté aux deux mêmes sujets provoque la même sensation d'intensité puisque à 100 dB SPL les mécanismes actifs n'entrent plus en jeu même chez le normo-entendant. L'échelle complète des sensations chez le malentendant est donc évoquée pour une dynamique plus réduite que pour un normo-entendant. Ce phénomène est connu sous le nom de recrutement de la sonie.

La figure 3.2 illustre ce phénomène en reprenant la courbe de sonie normale présentée initialement sur la figure 2.4 page 33. Sur cette figure, la fonction de sonie normale est rappelée (en trait plein) et deux courbes de sonie sous recrutement sont schématiquement indiquées (en tiret-pointillé) pour des pertes auditives respectives de 25 dB et 35 dB .

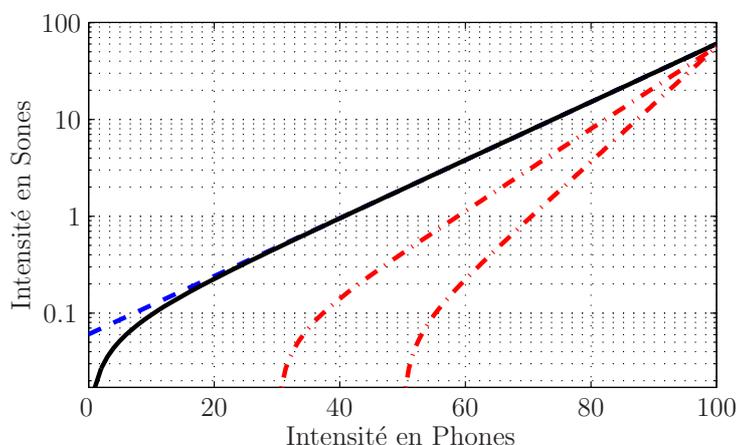


FIG. 3.2: Courbes de sonie avec recrutement pour un signal sinusoïdal à 1 kHz . Les deux courbes supérieures indiquent la fonction de sonie normale (trait plein) et son approximation (en pointillé) (cf. figure 2.4). Les deux courbes inférieures indiquent les courbes de sonie avec recrutement pour des seuils d'audibilité respectivement à 25 dB SPL et 35 dB SPL .

3.5. L'élargissement des filtres auditifs

Une autre conséquence du fonctionnement anormal de l'oreille interne est l'élargissement des filtres auditifs. Dans le cas d'une atteinte des CCE, on peut considérer schéma-

tiquement que la forme des filtres auditifs des malentendants correspond à celle des filtres auditifs des normo-entendants pour des stimuli de fort niveau.

L'élargissement des filtres auditifs est difficilement quantifiable. D'une part car les origines et la nature exacte du déficit auditif l'est aussi et d'autre part car la méthode de mesure des filtres auditifs est probablement peu adaptée aux conditions spécifiques de la surdité.

L'élargissement des filtres auditifs entraîne deux conséquences principales. D'une part, le phénomène de masquage est renforcé et d'autre part le bruit ambiant est plus gênant car les filtres sont moins sélectifs. Les malentendants éprouvent en effet de grandes difficultés à dissocier un signal tel que la parole en présence de bruit environnant ou dans un environnement réverbérant.

La résolution fréquentielle des malentendants est donc moins fine que pour les sujets normo-entendants. En particulier, le phénomène de masquage étant accru, il peut se révéler très gênant pour un signal de parole lorsqu'un formant d'une voyelle devient masquant pour un autre formant. Ainsi même avec une forte amplification sonore, l'intelligibilité de la parole peut être réduite.

De même, ce manque de finesse dans l'analyse fréquentielle des sons est un handicap pour la localisation. En effet, les informations spectrales et temporelles fines engendrées par les effets de filtrage et de diffraction de l'oreille externe et de l'oreille moyenne pouvant être perdues, la sensation de localisation en est profondément altérée [9, 53]. Une conséquence directe de cette déficience est la diminution des capacités dues à l'effet *cocktail-party*, les malentendants sont donc doublement gênés pour suivre une conversation dans un environnement bruyant.

4. Réhabilitation auditive par port de prothèses

4.1. Principe Général

La présentation des pathologies de l'audition qui précède montre que la surdité de perception est un handicap majeur qui témoigne d'atteintes rédhibitoires de l'oreille interne et de ses cellules sensorielles. La réhabilitation par port de prothèse est donc cruciale et particulièrement difficile à mettre en place. Elle ne peut conduire qu'à une amélioration partielle en référence à l'audition d'un normo-entendant.

Cependant, la surdité représente un tel handicap physique et social que toute amélioration même minime est appréciable pour son bénéficiaire.

Nous nous intéressons ici uniquement aux prothèses auditives externes et non aux prothèses dites « implantables » telles que les prothèses par voie osseuse, les implants d'oreille moyenne ou les implants cochléaires. Une prothèse auditive externe cherche à pallier les déficits du malentendant en adaptant de façon adéquate le signal acoustique parvenant à l'oreille. Les autres implants nécessitent une intervention chirurgicale et utilisent d'autres voies de transmission que la transmission par voie aérienne. Ces différents types de prothèse ne sont donc pas sujet au phénomène de Larsen engendré par le retour acoustique entre l'écouteur et le microphone comme c'est le cas pour les prothèses classiques.

Les *prothèses par voie osseuse* viennent stimuler directement la mastoïde et réalisent donc la transmission des sons vers la cochlée par voie osseuse. Ce type de prothèse est peu usité.

Les *implants d'oreille moyenne* ont comme particularité de stimuler directement la chaîne des osselets soit par un mécanisme piézoélectrique soit par induction électromagnétique. Il y a donc transmission directe des ondes acoustiques au milieu liquidien de la cochlée par action mécanique sur la chaîne ossiculaire. Ce type de prothèses est relativement nouveau et possède l'avantage par rapport aux prothèses classiques de laisser libre ou suffisamment aéré le conduit auditif.

Un *implant cochléaire* requiert la mise en place d'électrodes qui viennent directement stimuler les terminaisons du nerf auditif à l'intérieur de la cochlée. Les autres types de pro-

thèse continuent de stimuler les cellules sensorielles du malentendant, l'implant cochléaire, quant à lui, se substitue à ces cellules et réalise lui-même la transduction électrique. Son utilisation est donc réservée aux surdités profondes ou sévères pour lesquelles les cellules sensorielles sont très fortement ou totalement déficientes.

Le principe général de fonctionnement d'une prothèse auditive est représenté sur la figure 4.1. Ce schéma montre les principales étapes de traitement du signal effectuées dans la majorité des prothèses actuelles :

- En premier lieu, le signal est capté par le microphone puis numérisé par le Convertisseur Analogique-Numérique (CAN).
- Le signal est alors analysé dans un banc de filtres et séparé en différents canaux fréquentiels.
- Chaque canal est alors amplifié et compressé indépendamment (avec parfois des paramétrages différents en hautes ou basses fréquences).
- Suivant le modèle de prothèse, d'autres traitements peuvent être appliqués dans chaque bande (débruitage, mise en forme du signal, *etc.*).
- Enfin le signal est reconstruit à travers un banc de filtres de synthèse puis reconverti en analogique par le Convertisseur Numérique-Analogique (CNA).

Une majorité des surdités de transmission sont réversibles par intervention médicale (médicamenteuse ou opératoire). Si toutefois un appareillage est nécessaire, une simple amplification du signal est requise car une surdité de transmission peut être modélisée par une atténuation du signal.

La réhabilitation des surdités à l'aide d'une prothèse auditive concerne donc principalement les surdités de perception. Si la surdité est particulièrement lourde mais *sans atteinte du nerf auditif*, alors la mise en place d'un implant cochléaire est préférée. Dans la suite de ce document, nous nous concentrons donc sur la réhabilitation des surdités de perception légères, moyennes ou sévères.

4.2. Techniques de réhabilitation

4.2.1. Amplification et compression

Le premier module d'une prothèse auditive est le module d'amplification qui est dédié à la restauration des seuils auditifs. Le gain d'amplification doit donc atteindre de très hautes valeurs, jusqu'à 75 dB pour les surdités sévères. Cependant, comme nous le verrons dans la partie III, l'amplification est généralement restreinte à de plus faibles valeurs à cause du phénomène de Larsen.

Pour les surdités de perception, une simple amplification linéaire ne peut être envisagée à cause du recrutement. Il devient alors nécessaire d'introduire un module de compression, c'est-à-dire de *contrôle automatique du gain* (Automatic Gain Control, AGC).

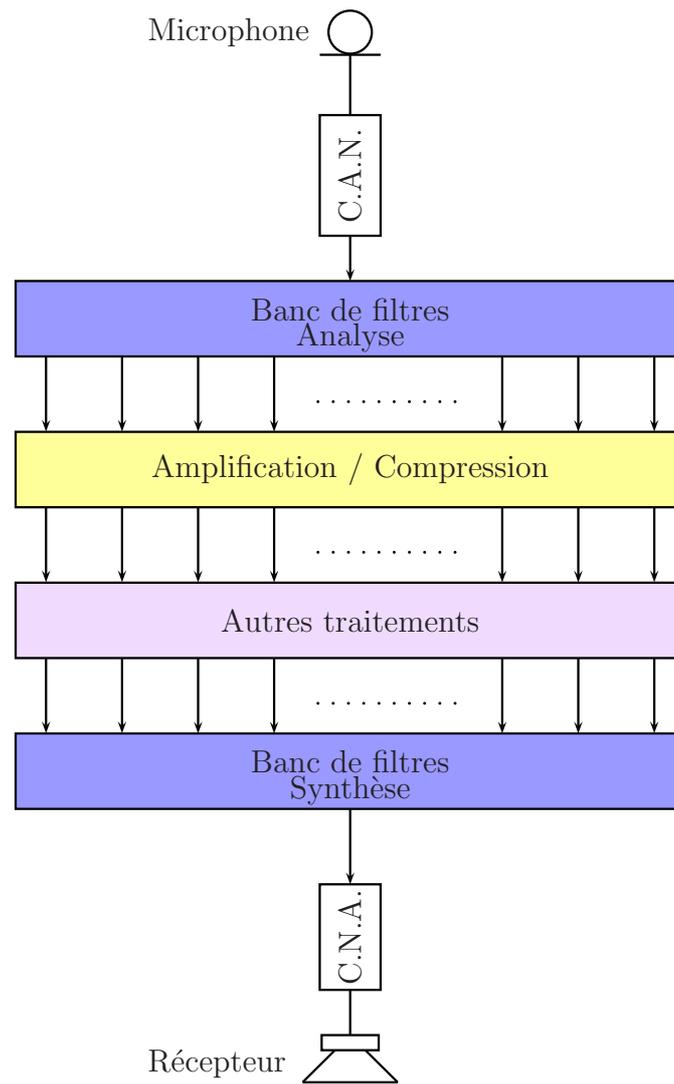


FIG. 4.1: Schéma général d'une prothèse auditive

La fonction première de ce module est d'empêcher les sons trop forts d'atteindre le tympan car dans ce type de surdité le seuil de douleur est identique ou proche de celui d'un normo-entendant ($110-120\text{ dB}$). Par la même occasion, cela maintient l'appareil hors des plages de saturation de l'amplificateur ou de l'écouteur et complète généralement de manière plus agréable le module d'écrêtage du signal de sortie. Un exemple schématique de relation Entrée/Sortie dans une prothèse avec compression est représenté dans la figure 4.2

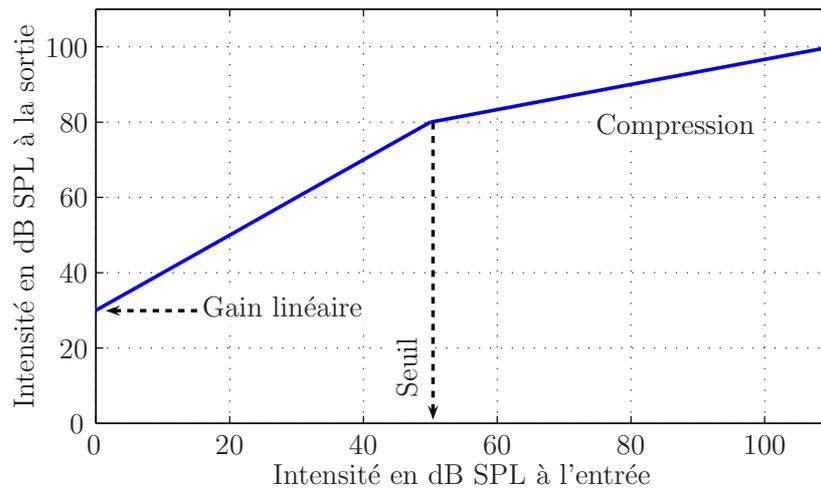


FIG. 4.2: Exemple schématique de relation Entrée/Sortie dans une prothèse avec compression. Dans cet exemple, un *gain linéaire* de 30 dB est appliqué jusqu'au *seuil de déclenchement* à 50 dB . Au delà de ce seuil un *taux de compression* de $3 : 1$ est appliqué.

Le second rôle de la compression peut être de restaurer les courbes de sonie pour compenser le *recrutement*. Dans ce cas, le seuil de déclenchement de la compression est relativement bas et les temps de réaction du compresseur rapides. L'amplification/compression consiste alors à apporter à l'oreille le *gain variable* que les mécanismes actifs déficients de l'oreille interne ne sont plus en mesure de fournir. Idéalement, si on se réfère aux courbes de sonie de la figure 3.2, le gain de la prothèse est fonction de l'intensité sonore du signal entrant et doit permettre de rattraper la courbe de sonie normale. La figure 4.3 indique théoriquement comment la valeur du gain peut être obtenue grâce à la courbe de sonie.

Les seuils auditifs varient selon les fréquences. Cette variation est caractérisée par l'audiogramme pour un certain nombre de fréquences normalisées en Hertz (125, 250, 1k, 2k 4k et 8k). De même le recrutement n'est pas identique pour toutes les fréquences. Si l'on souhaite une réhabilitation optimale, il est donc nécessaire de séparer le signal en différents canaux fréquentiels qui seront amplifiés et compressés individuellement selon les pertes. Ce point de vue est bien sûr idéaliste car dans la pratique les capacités calculatoires des prothèses sont limitées. Il est donc difficile de mettre en place dans les prothèses des bancs de filtres suffisamment sélectifs avec un nombre élevé de canaux et de s'assurer qu'il n'y a pas de recouvrement fréquentiel entre les différentes bandes après amplification.

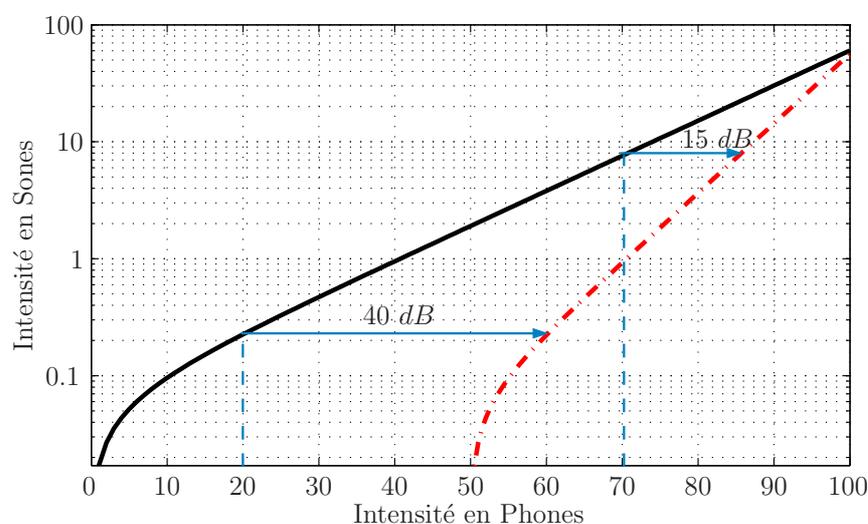


FIG. 4.3: Calcul du gain pour compenser le recrutement à partir des courbes de sonie de la figure 3.2. Pour un niveau donné du stimulus en phones, on repère le niveau en sones équivalent sur la courbe de sonie normale (en trait plein). Le gain d'amplification nécessaire pour rétablir une sensation de sonie correcte correspond donc à la distance horizontale en dB entre la courbe de sonie normale et la courbe avec recrutement pour un même niveau de sonie.

Les prothèses actuelles de dernière génération permettent généralement un traitement sur 16 à 32 canaux. Récemment la société Intrason a mis sur le marché un appareil fonctionnant sur 256 canaux.

Néanmoins, étant données les difficultés de mise au point d'un banc de filtres efficace, le bénéfice de l'augmentation du nombre de canaux n'a pas encore clairement montré sa supériorité. Cependant, la littérature sur le sujet prend rarement en considération les caractéristiques du banc de filtres d'analyse lors des comparaisons entre systèmes mono-canal, bi-canaux ou multi-canaux.

L'autre facteur déterminant de la compression est le choix des temps de réaction (temps de montée et de descente du compresseur). Si les temps de réaction sont trop courts, un effet désagréable de « pompage » du bruit de fond est audible pendant les périodes de silence. Par contre, si les temps de réaction sont trop longs, le système ne protège pas le porteur de certains bruits soudains et impulsifs comme, par exemple, le claquement d'une porte.

Ce facteur est décisif car le signal de parole est un signal de dynamique riche (30 dB), alternant silences, voyelles, et consonnes.

Dans certaines prothèses récentes, le réglage des paramètres de compression est différent dans les basses et les hautes fréquences. Dans les basses fréquences, on applique une compression dite *syllabique*, c'est-à-dire en accord avec le rythme de la parole, caractérisée

par des temps d'attaque et de descente très courts. Dans les hautes fréquences par contre, on garde un temps d'attaque court et le temps de descente est plus long, afin de protéger le malentendant des bruits soudains tout en conservant une bonne compréhension des consonnes.

4.2.2. Réduction du bruit

Si le choix des paramètres de compression est si délicat, c'est également qu'en dehors du rehaussement des seuils auditifs et du recrutement, l'oreille d'un malentendant est également caractérisée par un élargissement de ses filtres auditifs. Cet élargissement augmente le phénomène de masquage et réduit donc la sélectivité fréquentielle de l'oreille. L'autre conséquence est l'accroissement des difficultés à comprendre la parole au milieu du bruit environnant. L'amplification/compression des sons dans des bandes adjacentes doit prendre en compte ces paramètres afin de ne pas renforcer les problèmes de masquage.

Le traitement numérique du signal offre une possibilité supplémentaire pour traiter ce problème au travers de l'utilisation d'un module de *rehaussement du signal de parole* ou *réduction de bruit*.

Le module de réduction du bruit est excessivement important étant donné la grande difficulté qu'éprouvent les malentendants à comprendre distinctement la parole parmi un bruit de fond ou dans un environnement réverbérant. Son but est donc de maximiser le rapport signal à bruit.

Différents choix de mise en œuvre du rehaussement de la parole sont possibles. On distingue les stratégies que l'on dira *unilatérales* où chaque prothèse est indépendante de l'autre et les stratégies *bilatérales* où les prothèses sont reliées à un module de traitement unique opérant sur les bases d'algorithmes de traitement d'antenne.

Le dernier type de solutions utilise des supports de type lunettes ou colliers comportant un réseau de microphones et vise à focaliser l'antenne formée par les microphones de deux prothèses vers l'avant c'est-à-dire généralement vers le locuteur avec qui le malentendant converse. La réduction du bruit s'opère alors dans le domaine spatial par directivité du réseau de microphones ainsi formé.

Les désavantages des solutions bilatérales résident d'une part dans l'encombrement relatif et l'aspect non-esthétique de l'appareillage qu'elles nécessitent et d'autre part dans le fait que l'audibilité de l'ensemble des sons hors de la zone de focalisation est réduite quelles que soient la nature et l'importance de ces sons.

Remarque

Ces dernières remarques sur l'encombrement et l'aspect non-esthétique de l'appareillage sont valables pour d'autres algorithmes qui pourraient bénéficier

d'un traitement bilatéral ou d'un appareillage plus conséquent et qui leur permettrait de réaliser plus de calculs. Cependant, la discrétion des prothèses est requise pour une majorité de malentendants qui préfèrent que leurs appareils auditifs soient le moins visible possible.

Dans certaines prothèses récentes, deux à trois microphones sont placés sur une même prothèse et le traitement d'antenne est réalisé de manière unilatérale.

Les dimensions de ces réseaux de microphones laissent cependant supposer que la directivité des antennes qu'ils constituent ne peut pas être hautement sélective notamment dans les basses fréquences. D'après les résultats avancés par les fabricants de prothèses, un gain de 2 à 6 *dB* est réalisé dans la direction focalisée.

Malgré ce faible gain, une telle mise en œuvre de la réduction du bruit est jugée de manière assez positive par les malentendants et semble constituer une amélioration non négligeable.

Dernièrement un fabricant de prothèses a commercialisé un appareil effectuant le traitement d'antenne de manière adaptative, le diagramme de directivité évolue en fonction de la position estimée de la source de bruit.

Nous reviendrons en détail sur les traitements de réduction du bruit en mono-capteur dans la partie II de ce document, consacrée à ce sujet.

4.2.3. Traitements

L'utilisation du traitement numérique du signal permet d'envisager la mise en œuvre d'autres algorithmes dans une prothèse auditive.

Anti-Larsen

Le premier problème auquel le traitement du signal peut apporter une solution est le problème de retour acoustique entre les transducteurs de la prothèse. Ce retour est source d'instabilité pour le système pouvant mener à un effet Larsen qui se traduit donc par un sifflement de haute intensité en sortie de l'écouteur. Les solutions envisageables consistent soit à détecter les oscillations causées par l'instabilité et à réduire le gain de la prothèse lorsque l'effet Larsen se produit, soit à estimer le signal d'écho et le soustraire du signal entrant, ce qui revient à compenser le retour acoustique.

Ce sujet est traité plus en détails dans la partie III de ce document.

Mise en forme du signal

Le second point sur lequel il est possible de contribuer plus efficacement à la réhabilitation auditive par le biais d'algorithmes de traitement du signal est la mise en forme du signal dans le but de pallier certaines déficiences plus complexes liées à l'élargissement des filtres auditifs et à la perte des non-linéarités compressives de l'oreille interne.

Comme il n'est pas possible de rendre ses capacités à l'oreille déficiente, il s'agit de mettre en forme le signal du point de vue temporel ou fréquentiel afin de préserver des informations pertinentes qui pourraient être inaccessibles pour le système auditif du malentendant.

Parmi les diverses solutions envisagées, on peut citer le rehaussement fréquentiel proposé par Alain Goyé dans [32] qui consiste à renforcer le contraste fréquentiel du signal afin de pallier l'élargissement des filtres auditifs. Cet algorithme est justifié car il permet une maximisation du rapport signal à bruit au niveau de la sortie des filtres auditifs.

Christian Lorenzi et ses collaborateurs ont également proposé une mise en forme temporelle qui renforce l'amplitude des modulations présentes dans un signal de parole (*cf.* [2, 49]). Les modulations basses fréquences notamment autour de 2 à 4 Hz sont en effet des éléments acoustiques particulièrement caractéristiques du signal de parole dont il convient de préserver l'audibilité pour le malentendant.

D'autres auteurs ont également proposé des solutions qui modifient complètement la structure du signal en déplaçant certains pics spectraux du signal correspondant à des zones particulièrement touchées de la cochlée sur des zones plus épargnées. Il s'agit en général de compresser le spectre vers les basses fréquences pour compenser une perte très importante dans les hautes fréquences. On parle alors de *transposition* ou *compression spectrale*.

4.3. Les différents types d'appareils auditifs

Les premières solutions d'aide aux malentendants étaient relativement encombrantes. Parmi ces systèmes, certains se logeaient dans un boîtier ou sur une paire de lunettes. Même si a priori l'encombrement de ces appareils n'est pas extrême, il reste suffisamment gênant pour les malentendants qui souhaitent bénéficier d'appareils les plus discrets possibles.

La plupart des prothèses utilisées de nos jours intègrent dans un même appareil l'ensemble de ses éléments : l'alimentation, le circuit de traitement ainsi que le microphone et le récepteur. La miniaturisation de chacun des éléments est telle que les prothèses sont portables sur l'oreille, soit derrière le pavillon, soit dans le conduit.

4.3.1. Les contours d'oreille

Le type de prothèse le plus couramment utilisé est le *contour d'oreille*, appelé en anglais *Behind-The-Ear* (BTE), qui comme son nom l'indique se loge derrière le pavillon. Le microphone et le récepteur se trouvent donc également derrière le pavillon et le récepteur est couplé à l'oreille externe par un coude, un tube acoustique et un embout auriculaire comme le montre la figure 4.4.

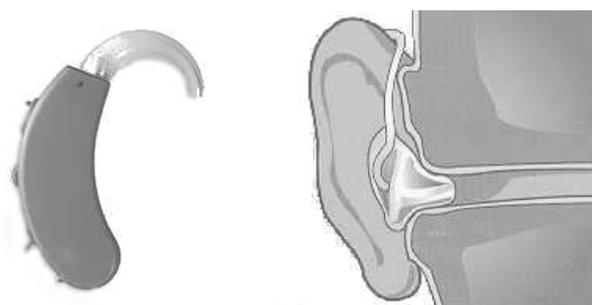


FIG. 4.4: Contour d'oreille

Les contours d'oreille sont les prothèses externes les plus volumineuses, ce sont donc celles qui peuvent contenir l'alimentation la plus puissante et sont d'ailleurs moins sujets au phénomène de Larsen car le microphone est relativement éloigné de l'embout auriculaire. Les BTE sont donc destinées plus particulièrement à fournir un gain important (gain maximal de 50 à 85 dB et intensités maximales en sortie de 120 à 140 $dB SPL$ suivant les modèles) et conviennent donc pour le traitement des surdités moyennes et sévères.

Les fabricants équipent de plus en plus les contours d'oreille de deux microphones qui leur offrent une certaine directivité.

4.3.2. Les intra-auriculaires

Les autres types de prothèse externe sont les intra-auriculaires. Pour ce type d'appareils, l'ensemble des éléments est intégré dans le coupleur auriculaire même. On distingue suivant la taille et l'emplacement dans l'oreille :

- l'intra-conque placé dans la conque, appelé en anglais *In-the-Ear* (ITE),
- l'intra-conduit ou intra-canal placé à l'entrée du conduit *In-the-Canal* (ITC) et
- l'intra semi-profond ou profond placé dans le conduit, appelé *Completely-In-the-Canal* (CIC).

Ces différents appareils sont représentés sur la figure 4.5.

Ces prothèses sont très discrètes, surtout dans le cas des CIC qui sont pratiquement invisibles. Du fait de leur taille réduite, leur puissance est moindre que celle des BTE et elles

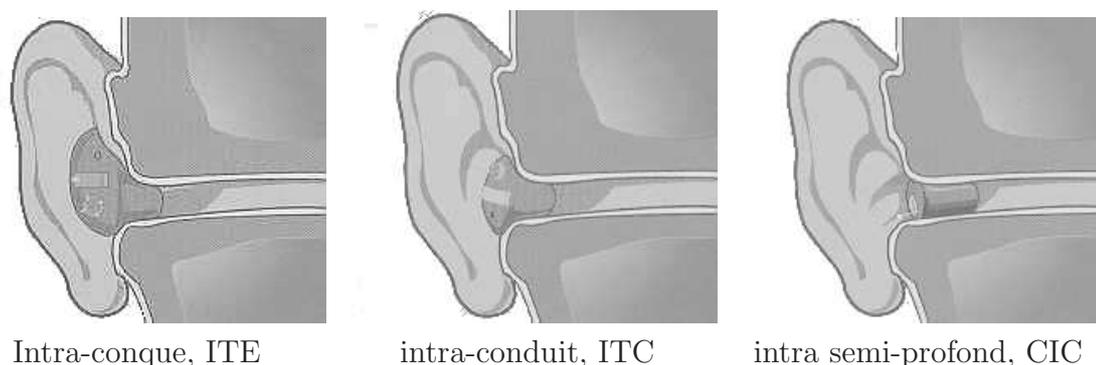


FIG. 4.5: Les différents types d'intra auriculaire

sont plus sujettes au phénomène de Larsen. Les valeurs de gain maximal d'amplification de ces appareils sont donc plus faibles. Le tableau 4.1 indique approximativement les valeurs moyennes de gain et d'intensités maximales annoncées par les principaux fabricants pour des appareils récents.

Types de prothèses	Gain maximal (dB)	Niveau maximal ($dB SPL$)
BTE	50 à 85	120 à 140
ITE	50 à 60	120 à 135
ITC	40 à 50	115 à 120
CIC	35 à 45	105 à 115

TAB. 4.1: Valeurs de gain et d'intensité maximales selon le type de prothèse. Ces valeurs correspondent à des valeurs moyennes annoncées par les principaux fabricants pour des appareils récents

Les valeurs de gain plus limitées des intra-auriculaires les destinent donc à la réhabilitation des surdités légères et moyennes.

4.3.3. Le choix du type de prothèse

Le choix du type de prothèse à utiliser pour une réhabilitation auditive est principalement dicté par le gain d'amplification maximal nécessaire pour pallier le rehaussement des seuils auditifs. Tant que la valeur du gain reste accessible à un intra-auriculaire, on choisit généralement l'appareil le plus discret. Si la surdité est très sévère ou profonde, le choix se porte sur un contour d'oreille. En général, pour un même fabricant, une série de prothèses est déclinée dans tous les types d'appareils à savoir BTE, ITE, ITC et CIC. En dehors des aspects de puissance, les traitements effectués par des prothèses d'une même série sont généralement identiques et ne rentrent donc pas en compte pour ce choix.

Dans quelques cas, on peut être cependant amené à opter pour un BTE ou un ITE

même si le degré de surdité ne le requiert pas. Ces appareils sont en effet plus faciles à manipuler que les intra de plus petite taille (ITC et CIC) et conviennent donc mieux pour des malentendants qui éprouveraient des difficultés de manipulation lors des réglages de l'appareil¹.

Remarque

Dans la section 1.3.2, nous avons signalé l'importance des effets de l'oreille externe pour la localisation. Étant donnée la position particulière du microphone dans le cas des BTE, les informations spectrales nécessaires à la localisation sont donc fortement dégradées. A un degré moindre, il en est de même pour les ITE car ils occupent tout ou partie de la conque. Pour les ITC et les CIC, la majorité des informations nécessaires à la localisation est conservée. De ce point de vue, les prothèses auditives et donc surtout les BTE contribuent à réduire les facultés de localisation. Toutefois il convient de pondérer cette affirmation car la surdité elle-même induit également une perte de ces facultés avec la diminution de la sélectivité fréquentielle.

4.4. Technologie des prothèses auditives

Comme de nombreuses applications biomédicales, les prothèses auditives sont récemment devenues le terrain d'évolutions technologiques majeures réunissant les avancées techniques de plusieurs domaines de recherche.

4.4.1. Transducteurs électro-acoustiques

Dans cette section, nous nous intéressons brièvement aux transducteurs des prothèses auditives. Afin de consulter une description plus complète de l'évolution et de l'état de l'art, on peut se reporter au chapitre consacré à ce sujet écrit par Mead C. Killion dans [46].

Les transducteurs des prothèses auditives ont principalement évolué vers la miniaturisation. En 1954, le plus petit microphone occupait 3cm^3 , de nos jours son équivalent occupe seulement $0,02\text{cm}^3$, il en est de même pour les écouteurs modernes. Si c'est la miniaturisation qui a principalement dicté l'évolution de ces instruments, leurs performances ont également grandement été améliorées parfois du fait même de la miniaturisation (*cf.* [46]). La conséquence de cette évolution est que, de nos jours, les transducteurs ne constituent généralement plus une limite pour les performances des prothèses auditives, notamment

1. Sur certains appareils, le malentendant a la possibilité de régler lui-même le volume ou de choisir convenablement le type de traitement à effectuer suivant la situation et l'environnement.

au niveau de la bande-passante (limitée par les capacités de calcul du processeur) et au niveau du bouclage entre transducteurs (limité par le couplage acoustique).

Dans le cadre de notre application, les tailles des transducteurs sont, bien entendu, un élément capital. Pour fixer un ordre de grandeur des dimensions courantes, on peut citer en exemple les caractéristiques de microphones omni-directionnels récents du fabricant *Knowles Electronics* :

- *série TO*, de forme parallélépipédique et de dimensions en mm $3,61 \times 3,61 \times 1,30$.
- *série FG*, de forme cylindrique avec $2,59\text{mm}$ de diamètre sur $2,59\text{mm}$ de hauteur.

Les transducteurs des prothèses offrent maintenant des réponses en fréquence très satisfaisantes. Les modèles d'écouteur pour les prothèses sont d'ailleurs également utilisés dans l'industrie du spectacle ou de la sonorisation (pour les « oreillettes » des présentateurs de télévision par exemple).

La diminution de la taille et les améliorations techniques récentes permettent aussi aux transducteurs de résister de manière suffisante aux chocs et autorisent donc leur intégration au sein des appareils portables que sont les prothèses auditives.

Les progrès en matière de transducteurs se jugent également sur leur susceptibilité au couplage magnétique ou vibratoire entre eux. La proximité du microphone et du récepteur est en effet propice au couplage qu'il soit acoustique, électrique, magnétique ou vibratoire. L'aspect acoustique est conditionné par la structure de la prothèse et l'aspect électrique est conditionné par la réalisation du circuit électronique. Les aspects magnétiques et vibratoires reposent quant à eux majoritairement sur la conception des transducteurs et dans le cadre des prothèses, un soin particulier est apporté par leurs fabricants à assurer une bonne isolation. Ainsi, dans les prothèses actuelles, la limitation du gain est principalement induite par le couplage acoustique.

Les transducteurs utilisés dans les prothèses sont magnétiques ce qui leur offre un excellent rendement du point de vue de la consommation tout en permettant une bonne qualité de transduction [46].

4.4.2. Autres voies de transduction

Transduction du signal entrant

La transduction du signal acoustique est un point particulièrement sensible des prothèses. Afin de transmettre un signal le moins corrompu possible au malentendant, certains appareils sont capables d'utiliser *en complément* d'autres moyens de transduction. Depuis quelques années, de nombreuses prothèses disposent, en parallèle du microphone, d'une bobine inductive qui permet au signal d'être transmis par induction électromagnétique sans l'intermédiaire du microphone. Ce système est notamment utilisé dans les cabines téléphoniques. Le son est alors transmis par induction par le combiné et la bobine

placée dans la prothèse se substitue au microphone dont le signal n'est alors plus transmis. Cela permet de s'affranchir des bruits parasites et des bruits d'environnement et de ne transmettre que le signal utile de la personne avec qui le malentendant est en communication. L'autre avantage est que la prothèse n'est alors plus sujette au phénomène de Larsen.

Des systèmes équivalents sont disponibles pour un usage domestique adaptable sur les téléviseurs ou d'autres appareils électroacoustiques.

Une alternative similaire plus récemment adoptée consiste à utiliser une transmission par ondes FM.

Transduction du signal sortant

Les *prothèses d'oreille moyenne*, issues d'une technologie encore plus récente, ne réalisent plus la transduction du signal électrique en signal acoustique mais viennent directement mettre en mouvement la chaîne ossiculaire (par induction électromagnétique ou par effet piézoélectrique) (une intervention chirurgicale est nécessaire pour fixer sur la chaîne ossiculaire la pièce qui est mise en mouvement). Le principal avantage de ce système est qu'il n'est également pas sujet au phénomène de Larsen et représente donc une amélioration sensible du confort pour le malentendant.

4.4.3. Processeurs et puces de traitement

Depuis la conception des premiers appareils auditifs numériques, le processeur de calcul est devenu l'élément clé des prothèses auditives.

Comme cela a été décrit précédemment, la réhabilitation auditive est très complexe et délicate. Afin d'être efficaces, les algorithmes utilisés pour mettre en œuvre cette réhabilitation requièrent donc des moyens de calcul importants. Pour l'instant, les processeurs intégrés dans les prothèses auditives restent encore modestes de ce point de vue. Toutefois, si on prend en compte leur taille et les contraintes de faible consommation, ils réalisent des prouesses techniques.

Plusieurs solutions sont utilisées pour la réalisation des calculs dans les prothèses : puces de traitement dédiés, réseaux programmables, ou processeurs de traitement du signal (Digital Signal Processor (DSP)). Si les solutions sur mesure étaient par le passé exclusivement employées, récemment des DSP basés sur des processeurs génériques sont devenus accessibles pour une utilisation dans les prothèses. Ils offrent l'avantage d'une plus grande souplesse d'évolution et de bénéficier des nombreux outils de développement déjà existants.

Les puces ou DSP utilisés dans les prothèses occupent maintenant seulement quelques mm^2 . Malgré cette taille très faible et les contraintes de faible consommation, leurs capacités de calculs atteignent la dizaine ou la centaine de millions d'instructions par seconde. Les contraintes imposent un fonctionnement en mode *virgule fixe* sur 16 bits (parfois 32 bits pour les calculs intermédiaires). La capacité des mémoires est également fortement limitée par la place disponible dans les appareils et permettent le stockage d'environ 8 à 16K mots en moyenne.

En ce qui concerne les produits dédiés, plusieurs algorithmes de traitement sont parfois implantés d'origine sur la puce et ne sont pas accessibles aux fabricants de prothèses. Ce peut être le cas notamment pour l'analyse/synthèse par banc de filtres.

Les capacités de calcul évoquées ci-dessus permettent déjà de réaliser des algorithmes de traitement nécessaires à la réhabilitation. Les prothèses numériques de dernières génération commercialisées offrent en effet de nombreux traitements parmi ceux que nous avons évoqués précédemment : amplification/compression, rehaussement du signal de parole, traitement du Larsen, gestion de microphones en réseau et mise en forme du signal. A l'heure actuelle, étant données les capacités de calcul encore réduites dont disposent les prothèses, on peut penser que ces algorithmes ne peuvent pas réaliser ces traitements de manière optimale. Cependant le traitement numérique du signal apporte des solutions très appréciables qui ne pouvaient pas être réalisées en analogique.

Notre étude nous laisse penser que le domaine du traitement du signal appliqué aux prothèses auditives se trouve probablement à un tournant décisif. Les prochaines générations de processeurs devraient permettre de réaliser des algorithmes plus performants et donc capables de répondre à des problèmes inhérents à la complexité du système auditif.

Pour illustrer ce propos, prenons l'exemple simple de l'amplification/compression. Les très hautes valeurs de gain nécessaires à la réhabilitation de surdit  induisent des problèmes de dynamique du signal lorsqu'on le traite sur 16 bits en virgule fixe. De m me, du fait de cette forte amplification appliqu e ind pendamment dans chaque bande de fr quence, les bancs de filtres d'analyse/synth se doivent  tre extr mement s lectifs si on veut s'assurer qu'il n'y ait pas de recouvrement inter-bandes. Il appara t donc clairement que cette  tape du traitement de la proth se sera fortement valoris e lorsque les capacit s de calcul des processeurs seront augment es.

Un autre point important en rapport avec ces remarques est la fr quence d' chantillonnage du signal. A l'heure actuelle certaines proth ses fonctionnent   une cadence maximale de l'ordre de 16 kHz , mais une majorit  reste   des fr quences d' chantillonnage de l'ordre de 8 kHz   11 kHz . Les tr s hautes fr quences sont donc inaccessibles   la r habilitation alors que ce sont celles qui sont principalement touch es. Cependant, ces fr quences sont peu importantes pour la compr hension de la parole qui demeure bien  videmment la priorit  en mati re de r habilitation.

Conclusion de la première partie

Au cours de cette première partie du document, nous avons abordé les concepts généraux qu'il est important de connaître dans le cadre de ce travail sur la réhabilitation des surdités au moyen d'algorithmes de traitement du signal.

En particulier, nous avons décrit le fonctionnement complexe du système auditif en insistant sur l'importance des *mécanismes actifs* de la cochlée et de leur caractère *non-linéaire* qui font de l'audition un sens particulièrement performant.

Dans un deuxième temps, nous avons exposé les origines et les conséquences physiologiques des surdités de perception en montrant qu'un des principaux symptômes était la disparition partielle ou totale des phénomènes non-linéaires.

Étudier sous le même angle des mécanismes actifs l'audition normale et l'audition pathologique, nous a permis, d'une part, d'énoncer de manière synthétique les difficultés diverses qu'éprouvent les malentendants et, d'autre part, d'en déduire logiquement les méthodes de réhabilitation envisageables.

Enfin, le contexte technologique a été précisé dans le but de définir les moyens et supports techniques disponibles pour effectuer les traitements envisagés ainsi que les contraintes inhérentes à cette application.

Au terme de cette première partie, le domaine des prothèses auditives apparaît, à notre sens, comme un champ d'application très intéressant pour le traiteur du signal. Le développement d'algorithmes d'aide aux malentendants nous semble très complet du point de vue technique car il touche beaucoup de sous-domaines du traitement du signal. Par ailleurs, il s'agit d'un champ d'application très ambitieux et prometteur qui met en jeu des appareils à la pointe de la technologie et qui vise la réhabilitation d'un déficit aux lourdes conséquences physiques et sociales.

Deuxième partie .
Réduction du bruit

Table des Matières

Introduction de la deuxième partie	71
5. Rehaussement de la parole	75
5.1. Techniques de rehaussement de la parole	75
5.2. Modification spectrale à court-terme	77
5.3. Règle de suppression d'Ephraïm et Malah	86
6. Estimation du niveau de bruit	91
6.1. Méthodes d'estimation du niveau de bruit	91
6.2. Solution adoptée	93
7. Application de l'EMSR à une analyse en banc de filtres ERB	97
7.1. Motivations à l'utilisation de l'échelle des ERB	97
7.2. Synthèse de bancs de filtres sur-échantillonnés en échelle ERB	101
7.3. Implémentation de l'EMSR pour une analyse banc de filtres	103
8. Résultats et discussion	107
8.1. Protocole d'évaluation	107
8.2. Amélioration du Rapport Signal à Bruit	109
8.3. Évaluation subjective	116
Conclusion de la deuxième partie	121
8.4. Conclusion	121
8.5. Perspectives	122

Introduction de la deuxième partie

Au cours de la première partie, nous avons évoqué les difficultés des malentendants à suivre une conversation dans un environnement bruyant ou réverbérant, selon l'origine et la sévérité de leur déficit auditif. Nous avons également vu que les causes de ces difficultés ne sont pas seulement liées à une hausse des *seuils auditifs* mais également à une *résolution fréquentielle* moins fine. Aux problèmes d'intelligibilité de la parole s'ajoutent aussi la fatigue auditive et les difficultés de concentration qui en découlent.

L'introduction d'un module de réduction du bruit dans les prothèses auditives est alors fortement souhaitable. Ce type de traitement présente, en outre, l'avantage de fournir un signal « propre » aux autres processus de traitement de l'appareil (en particulier avant l'amplification). Le module de réduction du bruit ou de *rehaussement de la parole* est une étape primordiale dans la chaîne de traitement d'une prothèse auditive.

Le domaine de la réduction du bruit est un domaine de recherche très actif et présent dans de nombreux champs d'applications du traitement du signal (téléphonie, restauration d'enregistrements audio anciens, analyse des signaux bio-médicaux, *etc.*).

La grande majorité des algorithmes utilisés à cet effet se fonde sur le principe de *soustraction spectrale* ou sur celui du *filtrage de Wiener*. Ces méthodes visent à atténuer le spectre à court-terme des observations bruitées [19] et elles parviennent de cette manière à réduire efficacement le niveau de bruit de fond. En contrepartie, ces techniques produisent, au sein du bruit résiduel, des artéfacts audibles ayant un spectre particulièrement tonal et désignés, de ce fait, par le terme de *bruit musical*.

La nature tonale du bruit musical est très gênante du point de vue auditif et elle l'est d'autant plus dans le cadre des prothèses auditives. En effet, un tel bruit résiduel pourrait être considéré comme un signal utile par un des autres processus implantés et être amplifié au détriment de composantes issues du signal source. Il est donc rigoureusement nécessaire de prévenir l'apparition du bruit musical.

A cet effet, des méthodes de réduction du bruit plus sophistiquées ont été mises au point pour parvenir à une réduction du bruit efficace tout en limitant la présence du bruit musical. Parmi ces techniques, celle proposée par Ephraïm et Malah dans [23] réalise le compromis visé de façon très convenable [14]. Au vu de ces résultats satisfaisants,

cette méthode est devenue une référence aussi bien pour le rehaussement de la parole que pour la restauration d'enregistrements anciens [13]. En outre, sa complexité, pourtant plus élevée que celle d'une méthode par atténuation spectrale à court-terme, reste acceptable. Ainsi, la règle de suppression d'Ephraïm et Malah (Ephraïm and Malah Suppression Rule (EMSR)) semble un candidat particulièrement opportun dans le contexte de la réhabilitation auditive.

Au cours de cette deuxième partie, nous étudierons également l'adaptation de l'EMSR à une analyse spectrale différente du cadre classique de la Transformée de Fourier Discrète (TFD) prévu par Ephraïm et Malah. En effet, l'analyse du signal bruité sur laquelle se base l'EMSR est habituellement réalisée à partir d'une *Transformée de Fourier à Court-Terme à temps discret*. Or, au cours du travail de thèse d'Alain Goyé [32], il a été observé que l'EMSR pouvait être adaptée à une analyse par banc de filtres uniforme sur 32 canaux avec décimation critique tout en conservant des résultats convenables. A partir de cette observation et dans le but d'adopter une approche en accord avec les règles de la psychoacoustique, nous proposons, ici, d'adapter la règle de suppression d'Ephraïm et Malah à un banc de filtres non-uniforme et non décimé réalisant une analyse spectrale selon l'échelle de fréquence des ERB sur 32 canaux. En effet, le fonctionnement de l'oreille humaine est souvent modélisé par un banc de filtres d'analyse ayant une résolution non-linéaire en fréquence que l'échelle des ERB approche. Même si cette vision est quelque peu rapide et simplificatrice, il n'en demeure pas moins que, du point de vue perceptif, une analyse du signal selon une échelle telle que celle des Barks ou des ERB est plus justifiée qu'une analyse par TFCT à résolution linéaire. Aussi, il nous a semblé intéressant de réaliser la réduction du bruit en accord avec la résolution des filtres auditifs de l'oreille interne.

L'organisation de cette deuxième partie peut être décrite de la manière suivante. Dans le chapitre 5, nous décrirons les principes généraux des méthodes de *Modification Spectrale à court-terme* et nous nous attacherons plus particulièrement à l'étude de l'EMSR et de ses qualités vis-à-vis du bruit musical.

Dans le chapitre 6, nous présenterons les méthodes usuelles d'estimation du niveau de bruit dans chaque sous-bande. A partir des travaux de la littérature, nous dégagerons une méthode simple et efficace qui nous semble convenir au contexte des prothèses auditives.

Dans le chapitre 7, l'adaptation de l'EMSR à une analyse fréquentielle uniforme en échelle ERB sera abordée. Nous y présenterons la structure du banc de filtres choisie et nous proposerons une adaptation simple de l'EMSR à cette structure.

Dans le chapitre 8, une étude comparative des résultats obtenus par les trois méthodes proposées (TFCT, analyse par banc de filtres uniformes et par banc de filtres non-uniformes) sera présentée.

A chaque étape de cette étude, nous nous attacherons à préciser clairement les motivations des choix technologiques sous-jacents, à la lumière des objectifs et des contraintes inhérents à l'utilisation de tels algorithmes dans un processus de réhabilitation auditive.

Publication Cette partie de notre travail et notamment la comparaison des performances du rehaussement de la parole selon le type d'analyse fréquentielle utilisé a donné lieu à la publication d'un article dans la revue *Speech Communication* :

- Thomas FILLON et Jacques PRADO, *Evaluation of an ERB frequency scale noise reduction for hearing aids : a comparative study*, *Speech Communication*, Volume 39, Numéros 1-2, January 2003, Pages 23-32.
-

5. Rehaussement de la parole

5.1. Techniques de rehaussement de la parole

5.1.1. Contexte

Dans cette étude, nous nous restreignons au rehaussement du signal en situation *mono-capteur*, comme c'est le cas dans la majorité des prothèses. Dans un contexte multi-capteurs, deux ou trois microphones sont placés sur l'appareil et le rehaussement du signal est réalisé par traitement d'antenne en focalisant l'antenne acoustique ainsi formée dans une direction privilégiée. En général, le diagramme de directivité correspondant est fixe et privilégie les sources frontales car la source utile émane généralement d'un locuteur placé en face du malentendant. Dans certaines prothèses très récentes, le traitement d'antenne est réalisé de manière *adaptative* afin que les lobes de *réjection* du diagramme de directivité restent dans l'axe *estimé* de la source du bruit¹.

Nous avons choisi de limiter notre étude au cas mono-capteur pour plusieurs raisons. Premièrement et simplement car c'est le contexte le plus courant en traitement du signal quel que soit le champ d'application concerné. De nombreuses solutions pertinentes ont donc été étudiées et proposées et il apparaît plus commode de se diriger vers des techniques ayant fait leur preuve dans d'autres domaines afin de les adapter au contexte des prothèses auditives.

Le deuxième point est lié à la nature même des solutions multi-capteurs. Étant donné la structure très rudimentaire de l'antenne acoustique que forme le réseau de microphone, le bénéfice potentiel en terme de rapport signal à bruit est inéluctablement limité à quelques décibels. Ce type de technique requiert également une adaptation du malentendant puisque le rehaussement est généralement limité aux signaux provenant de sources frontales. Le malentendant doit donc se tourner vers la source. Ceci n'est pas particulièrement gênant lors d'une conversation, d'autant plus qu'un malentendant a naturellement tendance à s'aider de la lecture labiale pour améliorer sa compréhension. Néanmoins, dans d'autres situations le traitement se révèle peu naturel et ne prend pas en compte le cas

1. Ce type de méthodes adaptatives estime la direction de la source de bruit pour réduire son influence et non celle de la source utile pour la privilégier car elles reposent sur l'hypothèse que le bruit est quasi-stationnaire et lentement variable dans le temps relativement à la source utile qui est généralement non-stationnaire (signal de parole).

où la source utile n'est pas dans la zone frontale.

Le contexte des prothèses auditives nous oblige également à envisager un algorithme rapide qui opère en temps-réel tout en demeurant modeste en coût de calcul.

5.1.2. Choix du type de méthode

Le domaine du rehaussement de la parole en situation mono-capteur est riche en solutions et fait appel à de nombreux concepts de traitement du signal.

Dans le choix de la méthode à utiliser, plusieurs critères entrent en jeu dont, au premier rang, la complexité et le caractère *temps-réel* qui imposent, dans notre cas, les contraintes les plus fortes.

Ces contraintes nous incitent à écarter les solutions telles que les méthodes dites *paramétriques*. Ces méthodes, comme leur nom l'indique, se basent sur une paramétrisation ou un codage du signal de parole. Souvent, les paramètres calculés pour modéliser le signal sont estimés de manière robuste au bruit et tendent à ne conserver que le signal utile. De même, si ces paramétrisations sont efficaces, on peut supposer qu'elles concentrent les informations sur le signal utile sur un nombre restreint de coefficients. En ne conservant que les coefficients contenant cette information utile pour la reconstruction du signal, on peut ainsi réduire le niveau de bruit. On notera qu'il est possible de rapprocher ces méthodes des algorithmes de décomposition en sous-espaces signal et bruit. En raison de leur complexité, le choix d'utiliser ce type d'algorithmes de réduction du bruit n'est pas compatible avec l'architecture actuelle de traitement du signal dans les prothèses auditives. En dehors de leur complexité, les méthodes paramétriques induisent également des modifications sur la structure du signal ce qui peut causer des distorsions. Si ces distorsions ne posent en général pas de problèmes pour un normo-entendant, il n'en est pas de même pour un malentendant.

Un principe plus ou moins similaire sous-tend les solutions de réduction du bruit utilisant une analyse du signal par ondelettes suivi d'un seuillage. Ce type de méthodes, introduit et développé par Donoho dans [20], se montre assez efficace dans de nombreux contextes comme, par exemple, le traitement d'images. En traitement de la parole, l'utilisation d'une analyse en ondelettes paraît très attrayante car elle peut permettre une analyse fréquentielle plus proche de l'analyse effectuée par les filtres auditifs de l'oreille interne. En utilisant une décomposition en bandes appropriée, on peut même approcher de manière convenable des échelles fréquentielles issues des travaux de la psychoacoustique comme l'échelle des Barks ou l'échelle ERB. Cependant, il n'y a pas à notre connaissance de travaux qui révèlent une utilisation convaincante d'une telle méthode de débruitage par analyse en ondelettes pour le rehaussement de la parole. Des essais que nous avons menés à ce sujet, nous ont confortés dans de telles conclusions : le bruit résiduel des signaux rehaussés obtenus possède une sonorité très artificielle et très gênante. Sans doute

le choix du type d'ondelettes et de leur longueur est-il primordial ? Cette question mérite, selon nous, de l'attention et une étude approfondie mais ceci dépasse malheureusement le cadre du présent travail de thèse.

Parmi les autres solutions envisageables, les méthodes de modifications spectrales à court-terme constituent une famille d'algorithmes de référence pour le rehaussement de la parole. Outre leur simplicité d'implémentation, ces méthodes sont également très efficaces pour réduire le bruit de fond. De plus, comme nous le soulignerons plus loin, leur structure est particulièrement adaptée à une implantation dans un processus de réhabilitation auditive. Nous avons donc choisi de nous tourner vers ce type de solutions pour proposer une méthode de rehaussement de la parole que l'on peut envisager d'implanter dans une prothèse auditive.

5.2. Modification spectrale à court-terme

5.2.1. Principes

Cette section est consacrée à la description des méthodes de *rehaussement de la parole* opérant dans le domaine fréquentiel au moyen de *modification spectrale à court-terme*. L'idée de base de ces méthodes est d'atténuer plus ou moins fortement les composantes spectrales du signal observé en fonction de l'estimation du niveau de bruit. Pour désigner cette famille d'algorithmes, on parle également d'*atténuation spectrale à court-terme*. Dans la description de ces techniques de rehaussement, que nous allons développer, nous avons choisi d'adopter, en partie, le formalisme utilisé dans [19]. On peut d'ailleurs se reporter à cette référence afin de consulter une description plus complète ainsi qu'à la thèse d'O. Cappé [13] qui présente une étude théorique claire et rigoureuse de ces mêmes méthodes et de leurs effets sur le signal rehaussé.

Afin d'aborder le principe de fonctionnement général de la modification spectrale à court-terme, considérons un signal d'observation, y , composé d'un signal de parole, x , corrompu par un *bruit additif*, b . Pour chaque indice temporel discret, n , le signal d'observation bruité, y , est donné par :

$$y[n] = x[n] + b[n] \quad (5.1)$$

Nous désignons par $\hat{x}[n]$ le signal rehaussé en sortie qui est une estimation de $x[n]$.

On suppose, ici, que les signaux de parole et de bruit ne sont *pas corrélés* entre eux. Ni $\{x[n]\}_{n \in \mathbb{Z}}$ ni $\{b[n]\}_{n \in \mathbb{Z}}$ ne sont des processus stationnaires, néanmoins, nous faisons l'hypothèse que le bruit est quasi-stationnaire tandis que le signal de parole peut-être considéré comme quasi-stationnaire sur des trames temporelles de courte durée de l'ordre de 10 à 30 *ms*.

Les différentes étapes du processus de rehaussement d'un signal par atténuation spectrale à court-terme sont indiquées par le schéma représenté sur la figure 5.1 et développées dans les paragraphes suivants.

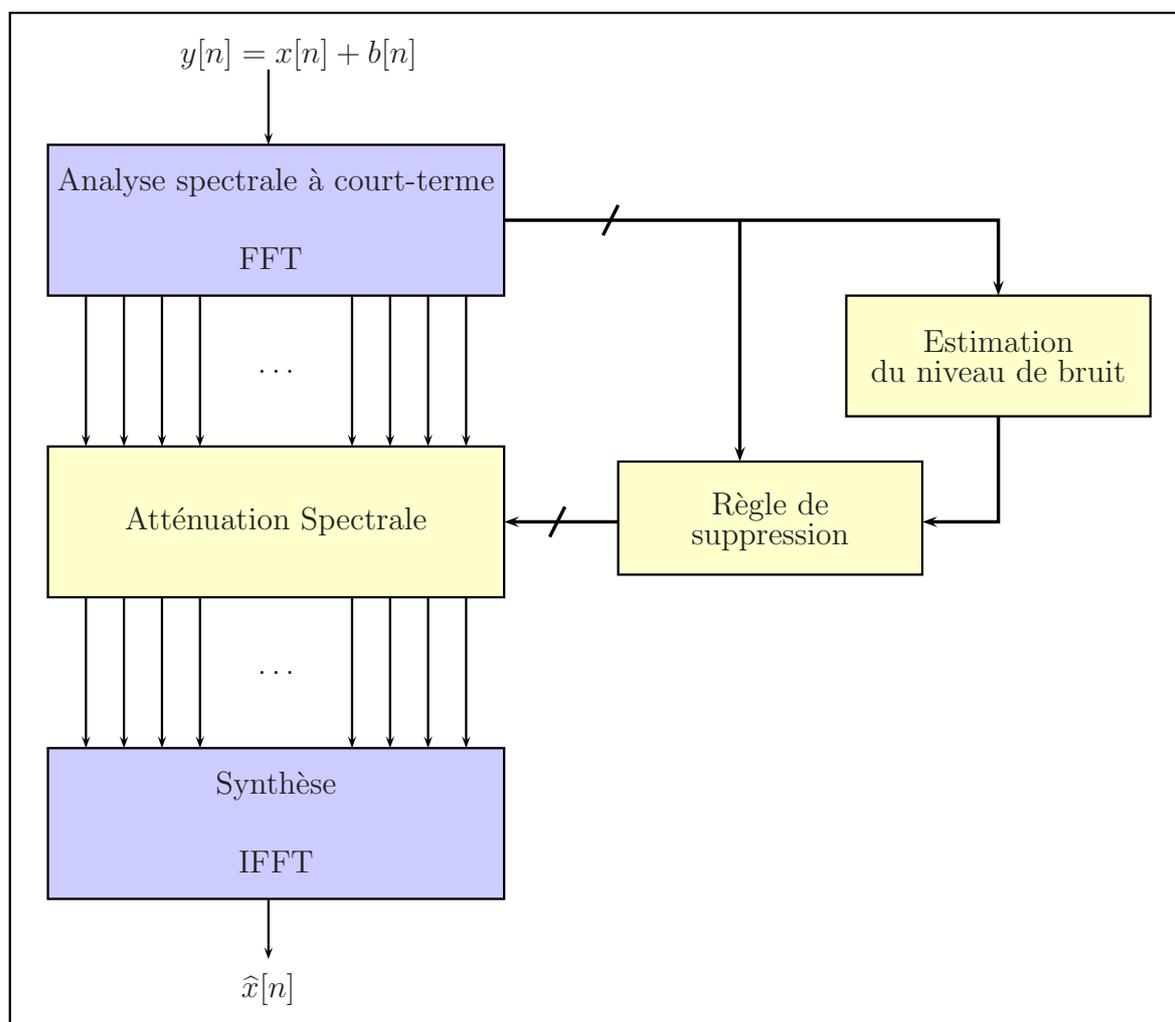


FIG. 5.1: Schéma général d'implémentation d'une méthode de débruitage par atténuation spectrale

Analyse spectrale à court-terme

Afin de travailler dans le domaine spectral, on a recours à la *transformée de Fourier à court-terme à temps discret*, que nous désignerons simplement par le terme Transformée de Fourier à Court-Terme (TFCT) dans la suite de ce document.

Le signal d'observation, $y[n]$, est divisé en blocs recouvrants de taille N au moyen d'une *fenêtre d'analyse* $\{w[p]\}_{p \in [0; N-1]}$ satisfaisant les critères d'inversibilité de la TFCT. Soit

R le pas d'avance entre chaque bloc, qui indique également le *taux de recouvrement* α_R entre blocs successifs par la relation :

$$\alpha_R = \frac{N}{R} \quad (5.2)$$

Soit \mathbf{y}_s le vecteur formé par fenêtrage des échantillons pour le bloc d'indice temporel s :

$$\mathbf{y}_s = \begin{bmatrix} y_{[sR]} \\ \vdots \\ y_{[sR+N-1]} \end{bmatrix} \odot \begin{bmatrix} w_{[0]} \\ \vdots \\ w_{[N-1]} \end{bmatrix} \quad (5.3)$$

où \odot représente le *produit de Schur* qui est la multiplication terme à terme des composantes de chaque vecteur.

La TFCT du signal est ensuite calculée à chaque indice temporel de bloc s et pour chaque indice fréquentiel k . On note \mathbf{W}_F la *matrice de transformation de Fourier discrète* de dimension M ($M \in \mathbb{N}$) et de coefficients :

$$\mathbf{W}_F[k, l] = \exp\left(-2j\pi \frac{kl}{M}\right) \quad (5.4)$$

et \mathbf{W}_F^{-1} la matrice de transformation inverse :

$$\mathbf{W}_F^{-1}[k, l] = \frac{1}{M} \exp\left(2j\pi \frac{kl}{M}\right) \quad (5.5)$$

Sous ces conditions, la TFCT à l'indice de bloc s est donnée par le vecteur, \mathbf{y}_s de taille M tel que :

$$\mathbf{y}_s = \mathbf{W}_F \cdot \mathbf{y}_s \quad (5.6)$$

Remarque

La taille de la transformée M est choisie comme une puissance de 2 afin de faciliter l'emploi de transformées rapides. En pratique, la taille des blocs, N , est généralement définie comme une puissance de 2 afin de simplifier l'implantation et on a donc $M = N$. Toutefois, il est également possible de considérer :

- N quelconque et M la puissance de 2 supérieure la plus proche de N ou
- N puissance de 2 et $M = 2N$ afin d'augmenter la précision fréquentielle.

Dans le cas où N est inférieur à M , il est alors nécessaire de compléter le vecteur \mathbf{y}_s de coefficients nuls pour effectuer le zero-padding.

Le spectre à court-terme se décompose alors pour chaque indice fréquentiel k ($k \in [0, M - 1]$) en amplitude et phase selon l'expression :

$$\mathbf{y}_s[k] = |\mathbf{y}_s[k]| \cdot e^{j\Phi_s[k]} \quad (5.7)$$

Estimation des densités spectrales de puissance

A partir de $|\mathcal{Y}_s[k]|$, on calcule une estimation de la *densité spectrale de puissance* du signal observé aux indices fréquentiels discrets k que l'on note $\mathcal{P}_s^y[k]$. Parallèlement, une méthode *aveugle* d'estimation du niveau de bruit permet d'évaluer la densité spectrale de puissance de la séquence de bruit $\{b[n]\}_{n \in \mathbb{N}}$. On note $\mathcal{P}_s^b[k]$ cette estimation.

Les estimateurs utilisés pour le calcul de ces densités spectrales de puissance sont, bien entendu, des facteurs déterminants pour le rehaussement. Pour le signal observé, on a généralement recours au *périodogramme* :

$$\mathcal{P}_s^y[k] = |\mathcal{Y}_s[k]|^2 \quad (5.8)$$

Pour l'estimation du bruit, la méthode est plus complexe car elle doit fonctionner sans connaissance *a priori* précise sur le bruit. Nous reviendrons sur la nature de la méthode d'estimation dans le chapitre 6. Nous supposons dans la suite de ce chapitre que l'estimation calculée est satisfaisante.

Atténuation spectrale

Dans les méthodes de rehaussement par atténuation spectrale, seule l'amplitude spectrale à court-terme est manipulée. On peut également formuler cette affirmation en considérant que les méthodes d'atténuation spectrale permettent de construire une estimation du spectre du signal source (sans le bruit) en dissociant l'estimation de l'amplitude et celle de la phase, l'estimation de la phase du signal rehaussé étant alors simplement la phase du signal observé.

A partir des variables $\mathcal{P}_s^y[k]$ et $\mathcal{P}_s^b[k]$, l'algorithme de rehaussement détermine une valeur d'atténuation à appliquer sur l'amplitude spectrale dans chaque bande de fréquences. Notons $G_s[k]$ le gain correspondant à l'atténuation pour l'indice de bloc s et la bande fréquentielle k . La règle qui détermine la valeur de $G_s[k]$ en fonction de $\mathcal{P}_s^y[k]$ et $\mathcal{P}_s^b[k]$ est appelée *règle de suppression du bruit*, ou simplement *règle de suppression*. Dans la section 5.2.2, nous détaillerons les expressions exactes des règles de suppression les plus couramment utilisées.

L'estimation de l'amplitude spectrale du signal rehaussé, $|\widehat{\mathcal{X}}_s[k]|$, est alors simplement déduite à partir de la valeur de $|\mathcal{Y}_s[k]|$ selon la formule :

$$|\widehat{\mathcal{X}}_s[k]| = G_s[k] \cdot |\mathcal{Y}_s[k]| \quad (5.9)$$

La phase du signal d'observation bruité n'est, quant à elle, pas modifiée car il est très difficile en pratique de manipuler ce paramètre ou de faire une estimation de la phase du bruit. En outre, dans [23], Ephraïm et Malah montrent que la conservation de la phase

du signal bruité permet de conserver l'optimalité de l'estimation de l'amplitude spectrale. Enfin, on peut remarquer que l'oreille est approximativement insensible à ce type de modification de la phase des signaux.

Au terme de ces traitements, le spectre à court-terme estimé du signal rehaussé, $\widehat{\mathcal{X}}_s$, s'exprime par la relation :

$$\widehat{\mathcal{X}}_s[k] = |\widehat{\mathcal{X}}_s[k]| \cdot e^{-j\Phi_s[k]}, \quad \forall k \in [0, M-1] \quad (5.10)$$

Synthèse

Le signal temporel estimé est reconstruit par *addition-recouvrement* en générant d'abord le bloc d'échantillons temporels rehaussés, $\widehat{\mathbf{x}}_s$, tel que :

$$\widehat{\mathbf{x}}_s = \mathbf{W}_F^{-1} \cdot \widehat{\mathcal{X}}_s[k] \quad (5.11)$$

Puis, pour tout $n \in \mathbb{N}$, le signal rehaussé $\widehat{x}[n]$ est construit par addition des α_R blocs d'indice temporel s contenant n , définis formellement comme les indices s appartenant à l'ensemble B_n tel que :

$$s \in B_n \Leftrightarrow \exists p \in [0, N-1], \quad n = sR + p$$

$\widehat{x}[n]$ est donc finalement obtenu selon la formule :

$$\widehat{x}[n] = \frac{1}{W_0} \sum_{s \in B_n} \widehat{\mathbf{x}}_s[n-sR] \quad (5.12)$$

où W_0 exprime la contrainte de synthèse de la TFCT par addition-recouvrement définie par :

$$W_0 = \sum_{p=0}^{N-1} w[p]$$

Interprétation par banc de filtres

Dans cette section, nous avons présenté le principe général qui structure les techniques de rehaussement de la parole par atténuation spectrale à court-terme comme celui de méthodes réalisant une estimation de l'amplitude spectrale à court-terme du signal source. Une interprétation alternative consiste à considérer la TFCT comme un banc de filtres d'analyse. L'action du rehaussement peut alors être envisagée comme un traitement par bande de fréquences qui consiste à atténuer les canaux relativement à la valeur estimée du Rapport Signal à Bruit (RSB) dans chaque canal.

Cette interprétation est non seulement intéressante pour appréhender ces méthodes de rehaussement de la parole, mais aussi parce qu'elle s'intègre parfaitement aux méthodes

de traitement du signal réalisées dans les prothèses auditives. En effet, le module d'amplification/compression déjà présent dans les prothèses s'appuie sur le même principe de traitement : décomposition du signal par banc de filtres, analyse de la puissance dans chaque bande, amplification/compression par canal et synthèse.

En outre, le module d'amplification/compression revient à effectuer une amplification avec un gain variable en fonction de la puissance du signal dans chaque canal fréquentiel de la même manière que le rehaussement consiste à y appliquer le gain spectral $G_s[k]$ défini par la règle de suppression en fonction des niveaux relatifs de puissance du signal et du bruit estimé.

L'analogie soulignée ici peut donc permettre d'envisager d'intégrer le rehaussement et l'amplification du signal au sein d'un même module de traitement.

Cette constatation est une des justifications majeures qui nous amène à proposer l'implantation de telles méthodes de rehaussement dans une prothèse auditive.

5.2.2. Filtrage de Wiener Paramétrique

Après avoir présenté de manière générale le principe de fonctionnement des méthodes d'atténuation spectrale à court-terme, nous allons définir plus précisément les règles de suppression correspondant à ces algorithmes. Les techniques de rehaussement de la parole les plus couramment employées sont des généralisations de techniques comme *le filtrage de Wiener à court-terme* et les méthodes dites de *soustraction spectrale* [19]. Ces méthodes ont été largement étudiées et sont employées pour de nombreuses applications.

Cette famille d'algorithmes peut être rassemblée sous une même catégorie appelée *filtrage de Wiener paramétrique* ou *soustraction spectrale paramétrique* car on peut les exprimer sous une formulation paramétrique générale. Le gain spectral correspondant à cette paramétrisation est alors défini en fonction des densités spectrales de puissance estimées, $\mathcal{P}_s^y[k]$ et $\mathcal{P}_s^b[k]$, par la formule :

$$G_s[k] = \left[1 - \left(\frac{\mathcal{P}_s^b[k]}{\mathcal{P}_s^y[k]} \right)^{\frac{\gamma}{2}} \right]^{\beta} \quad (5.13)$$

où γ et β sont deux paramètres fixes. A partir de cette généralisation et selon les différentes valeurs prises par le couple (γ, β) , indiquées dans le tableau 5.1, on retrouve les différentes méthodes de réduction du bruit populaires que sont la *soustraction spectrale en puissance* ou *en amplitude* et le *filtrage de Wiener à court-terme*.

Considérons la variable $R_s^{post}[k]$ définie par :

$$R_s^{post}[k] = \frac{\mathcal{P}_s^y[k]}{\mathcal{P}_s^b[k]} - 1 \quad (5.14)$$

Méthode	γ	β
Soustraction spectrale en puissance	2	$\frac{1}{2}$
Soustraction spectrale en amplitude	1	1
Filtrage de Wiener à court-terme	2	1

TAB. 5.1: Filtrage de Wiener paramétrique et méthodes correspondantes pour quelques valeurs de (γ, β)

Sous les hypothèses de décorrélation du signal source et du bruit, $R_s^{post}[k]$ peut être vue comme une estimation du *RSB instantané* pour le bloc d'indice temporel s . Nous le notons, ici, *RSB a posteriori* afin d'être compatible avec les notations que nous introduirons dans la section consacrée à l'algorithme de débruitage d'Ephraïm et Malah. La règle de suppression de l'équation (5.13) peut alors être reformulée selon l'expression suivante :

$$G_s[k] = \left[1 - \left(R_s^{post}[k] + 1 \right)^{-\frac{\gamma}{2}} \right]^\beta \quad (5.15)$$

Cette reformulation de la règle de suppression pour les méthodes de filtrage de Wiener paramétrique montre que le gain spectral ne dépend que du RSB instantané, exprimé par la variable $R_s^{post}[k]$. La *règle de suppression* de l'équation (5.15) est donc une règle *locale* dans le sens qu'elle ne se définit qu'à partir d'une information instantanée estimée à l'indice temporel de bloc s .

La figure 5.2 présente l'évolution du gain spectral $G_s[k]$ en fonction du rapport $R_s^{post}[k]$ comme défini par la formule (5.15) pour les trois méthodes de débruitage déjà évoquées ici.

5.2.3. Le phénomène du bruit musical

En utilisant des valeurs de (γ, β) issues du tableau 5.1, les algorithmes de filtrage de Wiener paramétrique parviennent à réduire de manière très efficace le niveau de bruit de fond. Néanmoins, le bruit résiduel qui subsiste dans le signal rehaussé après modification spectrale est très gênant du point de vue perceptif. Ce phénomène est caractéristique des méthodes d'atténuation spectrale à court-terme. Du fait du caractère tonal particulier de cet artéfact, il est désigné par le terme de *bruit musical*. Son spectre à court-terme correspond approximativement à une distribution aléatoire de pics spectraux.

L'origine du bruit musical est la variance des estimateurs locaux de la densité spectrale des signaux (cf. [14, 13]). En effet, comme le spectre à court-terme du bruit fluctue autour de sa valeur moyenne, son amplitude atteint à certains instants et pour certains indices fréquentiels des valeurs largement supérieures à la moyenne. Considérons le cas où le signal

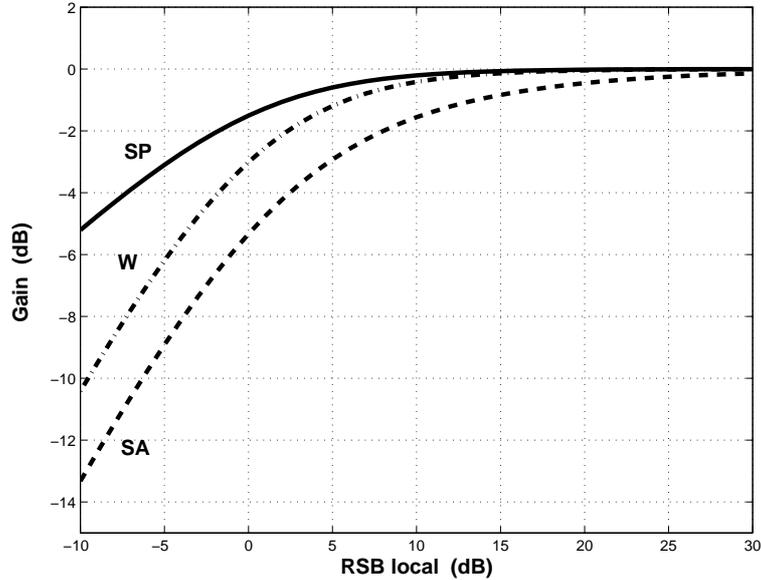
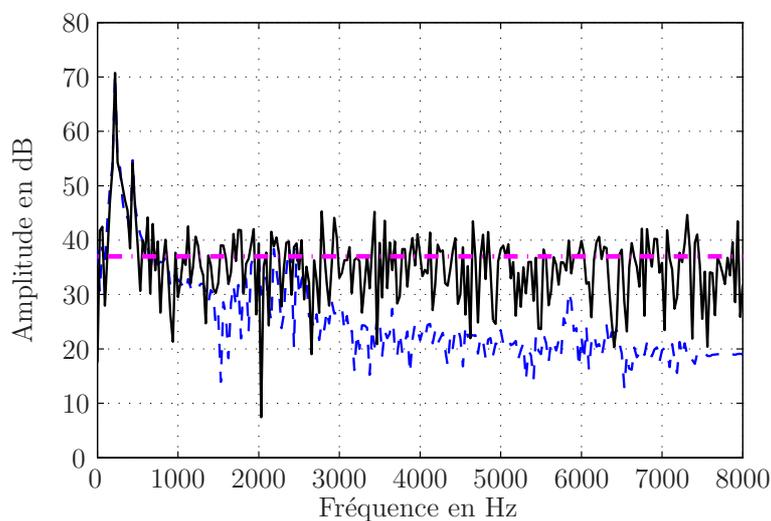


FIG. 5.2: Gain par bande de fréquence en fonction de l'estimation du RSB instantané, $R_s^{post}[k]$. Les 3 courbes correspondent respectivement au gain pour la Soustraction en Puissance (SP), la Soustraction en Amplitude (SA) et le filtrage de Wiener (W).

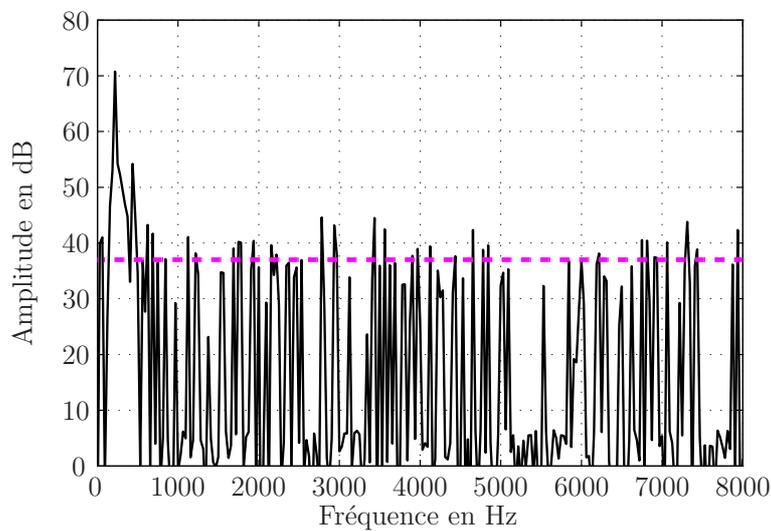
est fortement bruité dans une certaine zone du spectre où le rapport signal à bruit est, par conséquent, relativement bas. Sporadiquement, du fait de la variance du bruit, l'amplitude spectrale du signal atteint des valeurs largement supérieures au niveau moyen estimé du bruit conduisant donc à une surestimation locale et instantanée du rapport signal à bruit, $R_s^{post}[k]$. Selon la règle de suppression définie dans l'équation (5.15), la bande de fréquences correspondante est alors traitée comme du signal utile et est relativement moins atténuée que les composantes fréquentielles voisines. De manière sporadique, des pics fréquentiels isolés se dégagent donc du *spectre atténué*, engendrant ainsi le phénomène de bruit musical.

La figure 5.3 illustre cette explication. Sur cette figure, on constate que, dans le signal bruité (Fig.5.3(a)), le bruit est prédominant dans la partie haute du spectre et prend parfois des valeurs supérieures à la moyenne. Les pics fréquentiels correspondant à ces valeurs particulières se retrouvent alors dans le spectre rehaussé sous la forme de pics fréquentiels (Fig.5.3(b)).

Une telle tonalité du bruit résiduel est problématique du point de vue perceptif et peut se révéler très gênante dans les étapes suivantes de traitement du signal d'une prothèse auditive. En effet, considérons, par exemple, que la prothèse contienne un module de rehaussement fréquentiel, le traitement appliqué au signal aura alors pour effet de renforcer le caractère tonal du bruit résiduel. Encore une fois, il faut nuancer cette affirmation en la replaçant dans le contexte d'une déficience auditive : le but du rehaussement est de reconstruire un son de telle manière qu'il soit perçu par le malentendant comme il l'aurait été sans traitement par un normo-entendant. Cependant, il faut remarquer que tout traitement d'un son issu d'un bruit résiduel sera fait au détriment d'un son issu du signal de parole.



(a) Spectre à court-terme d'une trame de signal corrompu par un bruit blanc additif (trait plein). En trait pointillé sont tracés le spectre à court-terme du signal source (---) et le niveau moyen du bruit (-·-). Dans les hautes fréquences, on note que le signal est fortement bruité.



(b) Spectre à court-terme rehaussé (trait plein). Le niveau moyen du bruit est indiqué en pointillé (-·-). On note dans la partie haute fréquence, l'allure tonale du spectre correspondant au bruit musical.

FIG. 5.3: Exemple de spectre de bruit musical

Il est également important de noter que les malentendants éprouvent des difficultés pour percevoir intelligiblement la parole ce qui leur demande donc de grands efforts de concentration et qui induit une fatigabilité plus grande que pour un normo-entendant. De ce point de vue, la nature tonale du bruit musical est fortement gênante et il apparaît nécessaire dans le cadre des prothèses auditives de se prévenir d'un tel artéfact.

Par conséquent, il convient, même au prix d'une complexité légèrement plus élevée, de recourir à une règle de suppression plus sophistiquée. Parmi ces méthodes, la règle de suppression d'Ephraïm et Malah que nous allons développer dans la prochaine section semble être un choix particulièrement judicieux de par sa capacité à se prévenir de l'apparition de bruit musical tout en restant très efficace dans la réduction du bruit.

5.3. Règle de suppression d'Ephraïm et Malah

5.3.1. Introduction

Dans [23, 24], Ephraïm et Malah ont introduit différentes règles de suppression redéfinissant chacune le gain spectral $G_s[k]$ de l'équation (5.9). Ces règles ont été construites à partir de l'hypothèse que les TFCT des signaux source et bruit sont des variables aléatoires *gaussiennes et indépendantes*.

Dans [23], l'estimateur de l'amplitude spectrale est conçu pour minimiser l'Erreur Quadratique Moyenne (EQM) sur le spectre d'amplitude. Cet estimateur est couramment désigné dans l'article original comme l'estimateur de l'amplitude spectrale à court-terme au sens de l'EQM, *Minimum Mean Square Error - Short Term Spectral Amplitude (MMSE-STSA)* ou plus couramment dans la littérature comme la règle de suppression d'Ephraïm et Malah, *Ephraïm and Malah Suppression Rule (EMSR)*.

Par la suite Ephraïm et Malah ont élaboré d'autres règles de suppression à partir des mêmes hypothèses. Dans [24], l'estimateur présenté minimise l'EQM du logarithme du spectre (Minimum Mean-Square Error Log-Spectral Amplitude estimator (MMSE-LSA)). Plus récemment, dans [50], Malah et al. ont proposé un estimateur correspondant à une modification du MMSE-LSA appelé *Multiplicatively-modified LSA*.

L'ensemble de ces règles conduit à un rehaussement du signal de parole très satisfaisant aussi bien du point de vue de la réduction du bruit que vis-à-vis du bruit musical.

Afin de proposer une solution adéquate pour le rehaussement de la parole au sein d'une prothèse auditive, nous avons choisi de nous concentrer sur le plus simple estimateur proposé par Ephraïm et Malah, à savoir l'EMSR. Ce choix est motivé d'une part car l'EMSR est un algorithme couramment utilisé et qui a fait ses preuves dans de nombreuses applications et d'autre part car les estimateurs qui dérivent de l'EMSR ne nous paraissent

pas apporter une amélioration suffisamment importante du point de vue du rehaussement compte tenu du surcoût de complexité qu'ils engendrent.

Le développement qui mène à l'expression de l'EMSR a été présenté dans [23], dans le cadre de ce rapport de thèse, nous nous contentons de rappeler sa définition.

5.3.2. Description de la méthode

Notons $G_s^{EM}[k]$ le gain spectral défini par l'EMSR pour la $k^{\text{ème}}$ bande de fréquences et le bloc d'indice s . La particularité de cette règle de suppression par rapport aux règles classiques d'atténuation spectrale par filtrage de Wiener paramétrique est que $G_s^{EM}[k]$ n'est plus défini simplement en fonction du rapport signal à bruit local ou RSB *a posteriori*, $R_s^{post}[k]$, mais également en fonction d'une variable $R_s^{prio}[k]$, dite rapport signal à bruit *a priori*, dont l'expression est :

$$R_s^{prio}[k] = (1 - \alpha_{EM}) \cdot \max(R_s^{post}[k], 0) + \alpha_{EM} \cdot \frac{|\widehat{\mathcal{X}}_{s-1}[k]|^2}{\mathcal{P}_s^b[k]} \quad (5.16)$$

où l'on rappelle que $|\widehat{\mathcal{X}}_{s-1}[k]|$ désigne l'amplitude spectrale à court-terme du signal rehaussé dans la trame précédente et où α_{EM} est le paramètre contrôlant la mise à jour de R_s^{prio} (on considère en général α_{EM} proche de 1).

Ici, afin de ne pas prendre en compte d'éventuelles valeurs négatives de R_s^{post} , la mise à jour de R_s^{prio} ne prend en compte que la partie positive de $R_s^{post}[k]$, $\max(R_s^{post}[k], 0)$.

D'après les équations (5.9) et (5.8), on peut récrire la formule (5.16) exprimée en fonction de \mathcal{P}_{s-1}^y :

$$R_s^{prio}[k] = (1 - \alpha_{EM}) \cdot \max(R_s^{post}[k], 0) + \alpha_{EM} \cdot \frac{G_s^{EM}[k-1] \cdot \mathcal{P}_{s-1}^y[k]}{\mathcal{P}_s^b[k]} \quad (5.17)$$

A partir des deux variables, $R_s^{post}[k]$ et $R_s^{prio}[k]$, l'EMSR définit alors le gain spectral $G_s^{EM}[k]$ par :

$$G_s^{EM}[k] = \frac{\pi}{2} \cdot \sqrt{\frac{1}{1 + R_s^{post}[k]} \times \frac{R_s^{prio}[k]}{1 + R_s^{prio}[k]}} \times \mathcal{M} \left(\left(1 + R_s^{post}[k]\right) \times \frac{R_s^{prio}[k]}{1 + R_s^{prio}[k]} \right) \quad (5.18)$$

où la fonction $\mathcal{M}(\cdot)$ est définie par :

$$\mathcal{M}(u) = \exp\left(-\frac{u}{2}\right) \times \left((1+u) \cdot \mathcal{I}_0\left(\frac{u}{2}\right) + u \cdot \mathcal{I}_1\left(\frac{u}{2}\right) \right) \quad (5.19)$$

Dans l'équation (5.19), $\mathcal{I}_0(\cdot)$ et $\mathcal{I}_1(\cdot)$ sont respectivement les fonctions de *Bessel* modifiées d'ordre zéro et du premier ordre.

Les figures 5.4 et 5.5 montrent les valeurs prises par G_s^{EM} selon celles de R_s^{post} et R_s^{prio} .

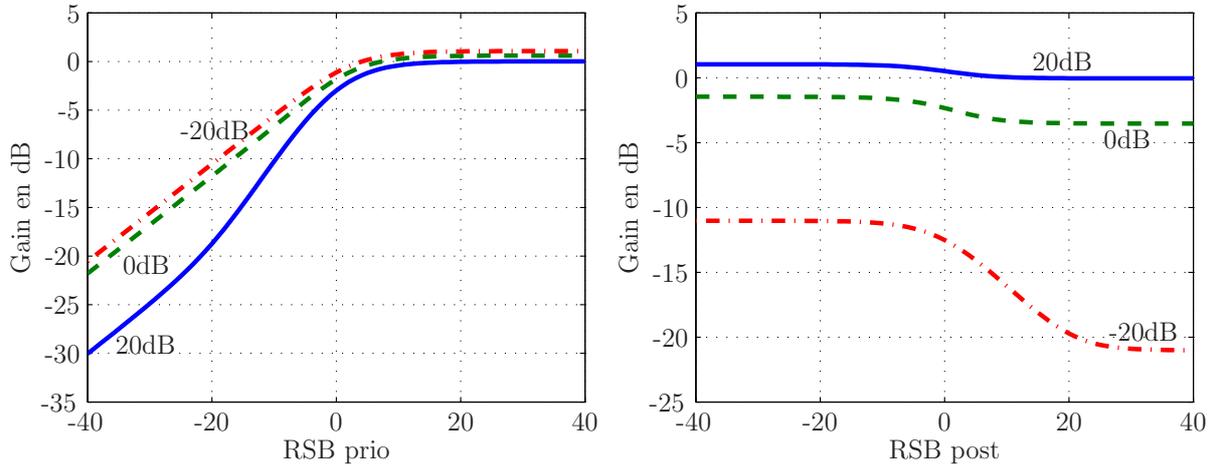


FIG. 5.4: Règle de suppression d'Ephraïm et Malah. A gauche, variation du gain en fonction du RSB *a priori* pour trois valeurs du RSB *a posteriori* (20dB, 0dB, -20dB). A droite, variation du gain en fonction du RSB *a posteriori* pour trois valeurs du RSB *a priori* (20dB, 0dB, -20dB). On notera que le RSB *a priori* est le paramètre principal.

5.3.3. Réduction du bruit musical

L'originalité de l'EMSR tient dans la prise en compte du RSB *a priori*. Dans [14], O. Cappé souligne que le RSB *a priori* est le paramètre dominant de l'équation (5.18), comme on peut le constater sur les figures 5.4 et 5.5. Il montre également qu'il est la clé du mécanisme qui empêche l'apparition de bruit musical.

Nous avons indiqué que la règle de suppression de l'équation (5.15) engendre le phénomène du bruit musical notamment car c'est une règle de décision *instantanée* qui ne prend en compte que le RSB *a posteriori* et qui est donc, de ce fait, sensible à la variance de l'estimateur.

Au contraire l'EMSR prend également en compte un nouveau paramètre, R_s^{prio} , défini en fonction des trames précédentes. Cette prise en compte du passé permet d'éviter que certains des pics en fréquence locaux du spectre du bruit ne soient considérés comme étant issus du signal source, comme c'est le cas du filtrage de Wiener paramétrique. Dans [14], O. Cappé distingue deux interprétations possibles de $R_s^{prio}[k]$:

- 1^{er} cas : si $R_s^{post}[k]$ est inférieur à 0dB, alors $R_s^{prio}[k]$ correspond à une version lissée de $R_s^{post}[k]$,

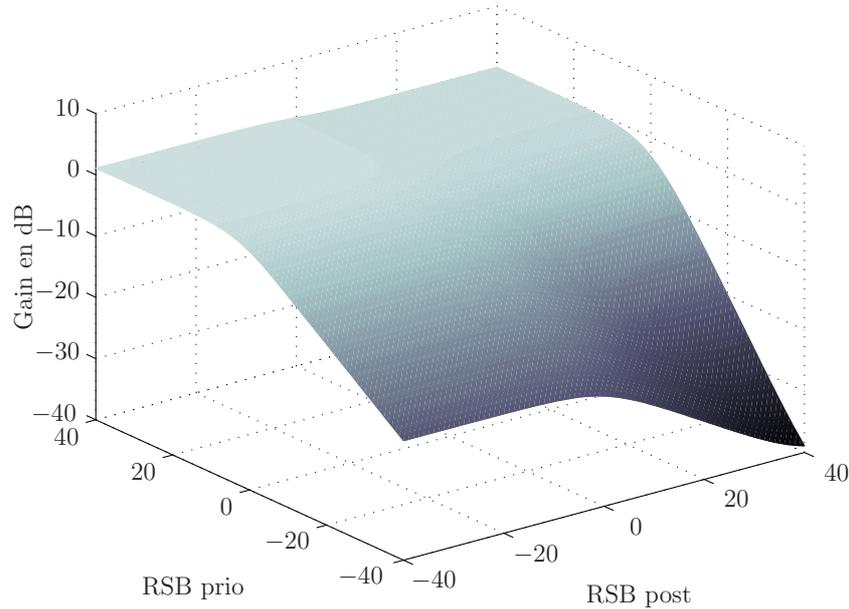


FIG. 5.5: Règle de suppression d'Ephraïm et Malah

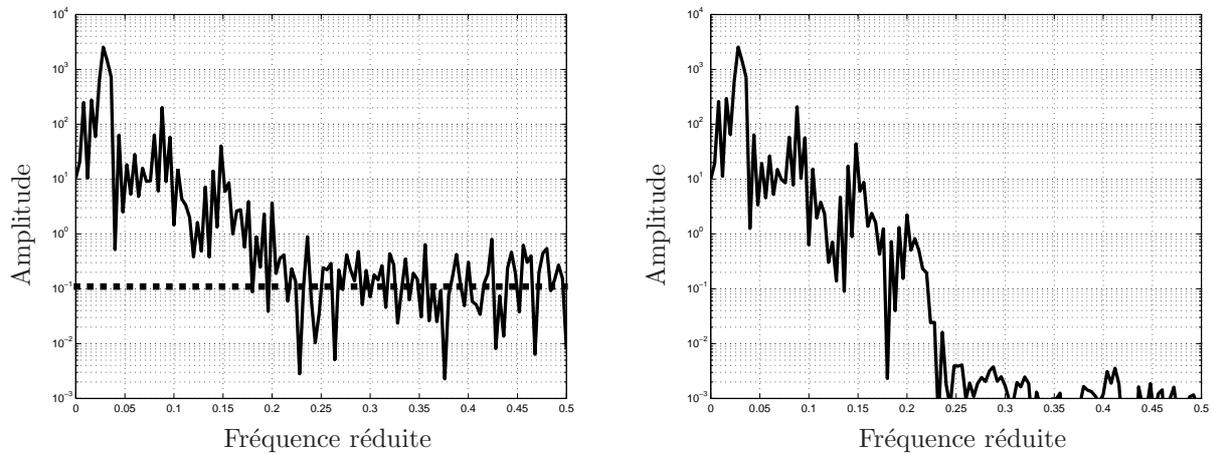
- 2nd cas : si $R_s^{post}[k]$ est supérieur à 0dB, alors la valeur de $R_s^{prio}[k]$ suit celle de $R_s^{post}[k-1]$.

Le lissage correspondant au 1^{er} cas (lorsque le niveau de bruit est élevé) permet ainsi de limiter la variance de l'estimateur qui est source du bruit musical comme on l'a souligné précédemment. Sa non-linéarité est également un facteur important de l'efficacité de l'EMSR car elle permet dans le 2nd cas, lorsque le signal a un niveau supérieur au niveau de bruit, d'adopter un comportement adéquat vis-à-vis de la non-stationnarité des signaux. La prise en compte du RSB *a priori* solutionne donc de manière particulièrement adaptée le problème du bruit musical.

Les expérimentations pratiques de rehaussement de la parole grâce à l'EMSR viennent corroborer ces dernières considérations théoriques et montrent que cet algorithme est une solution très appréciable.

Afin de ne pas se restreindre uniquement au signal de parole qui reste néanmoins la priorité dans le cadre de la réhabilitation auditive des malentendants, il convient de souligner que l'efficacité de l'EMSR a été prouvée aussi bien dans le cadre de rehaussement du signal de parole que dans le cadre de la restauration d'enregistrements phonographiques musicaux anciens [13].

En terme de coût de calcul, la règle de suppression (5.18) apparaît relativement complexe du fait des fonctions qui interviennent dans le calcul. On peut cependant simplifier



(a) Spectre d'un signal bruité corrompu par un bruit blanc additif (trait plein). En trait pointillé est indiqué le niveau de la densité spectrale du bruit.

(b) Spectre du signal rehaussé en utilisant l'EMSR.

FIG. 5.6: Nature du bruit résiduel après rehaussement en utilisant l'EMSR. On peut comparer cette figure à la figure 5.3 page 85. En particulier, on voit que le bruit musical n'apparaît pas dans la partie hautes fréquences.

la mise en œuvre en calculant par avance les valeurs de G_s^{EM} en fonction des deux paramètres de RSB et en réunissant ces valeurs dans une matrice mise en mémoire. On substitue ainsi au calcul une lecture dans un registre.

6. Estimation du niveau de bruit

6.1. Méthodes d'estimation du niveau de bruit

6.1.1. Principes

L'estimation de la puissance spectrale du bruit est un des éléments critiques de toute opération de rehaussement de la parole. Ainsi, l'efficacité des méthodes de rehaussement par atténuation spectrale à court-terme décrites dans le chapitre précédent repose sur une estimation préalable correcte du niveau de bruit.

Cette estimation doit fournir une information précise et fiable sur la *densité spectrale de puissance* du bruit, c'est-à-dire une information à la fois sur le niveau du bruit et sur son contenu spectral.

En l'absence de connaissances *a priori* sur la densité spectrale de puissance réelle du bruit, la seule hypothèse de départ des méthodes d'estimations du niveau de bruit est de considérer que celui-ci est *stationnaire* ou *quasi-stationnaire*. De manière plus rigoureuse, cela revient principalement à considérer que les statistiques du bruit de fond varient lentement par rapport à celles du signal source.

A partir de ces hypothèses, deux types de méthodes sont envisageables pour estimer la densité spectrale de puissance du bruit :

- soit la mise à jour de l'estimation se fait uniquement dans les périodes de silence, c'est-à-dire dans les trames temporelles où le signal source est supposé absent,
- soit l'estimation est réalisée de manière continue au moyen d'une règle de mise à jour prenant en compte la nature quasi-stationnaire du bruit de fond.

6.1.2. Estimation discontinue

Dans les différents contextes applicatifs de réduction du bruit, le signal source n'est, en général, pas présent dans toutes les trames temporelles. Dans le cadre du traitement de la parole, notamment, le signal est régulièrement ponctué de silences au sein desquels seul le bruit de fond persiste. Si l'on se place dans le contexte de la restauration d'enregistrements

sonores anciens, on peut de même exploiter les périodes de silences en début ou en fin de l'enregistrement. Sous l'hypothèse de quasi-stationnarité du bruit de fond, il est alors commode d'utiliser ces périodes de silence pour mettre à jour l'estimation de la densité spectrale de puissance du bruit. On suppose alors que le bruit de fond conserve les mêmes statistiques en dehors de ces périodes.

Dans ce cas de figure, la difficulté consiste alors à définir avec précision si le signal source est présent ou non à un instant donné et requiert, par conséquent, l'utilisation d'un Détecteur d'Activité Vocale (DAV).

Lorsque le signal de parole est détecté, il n'y a donc pas de mise à jour de l'estimation de la densité spectrale de puissance. Inversement, lorsque le signal de parole est absent, la mise à jour est effectuée, notamment selon la méthode du périodogramme lissé.

L'estimation de $\mathcal{P}_s^b[k]$ est obtenue à partir de $\mathcal{P}_s^y[k]$ qui est lui même généralement défini selon la méthode du périodogramme (cf. équation (5.8)). A chaque indice de trame s , la valeur de $\mathcal{P}_s^b[k]$ est mise à jour selon une équation récursive du premier ordre, la mise à jour est contrôlée par le paramètre de lissage α (compris entre 0 et 1).

$$\mathcal{P}_s^b[k] = (1 - \alpha) \cdot \mathcal{P}_s^y[k] + \alpha \cdot \mathcal{P}_{s-1}^b[k] \quad (6.1)$$

L'utilisation d'un DAV est très pertinente non seulement pour le rehaussement de la parole mais aussi pour d'autres modules de traitement implantés dans une prothèse auditive. En effet, son utilisation serait également justifiée dans le cadre de l'annulation adaptative de retour acoustique afin de détecter les périodes de silence qui sont propices à engendrer la divergence du filtre adaptatif. Une autre application appréciable d'un DAV est qu'il donne une indication sur la nature du signal à traiter et pourrait permettre d'envisager la mise en place de stratégies différenciées de traitement selon la nature du signal en entrée de la prothèse (parole, musique, bruit de fond, *etc.*), ceci afin de répondre de manière plus adaptée aux besoins du malentendant.

Malheureusement, au regard des performances actuelles des processeurs utilisés dans les appareils auditifs, l'implantation d'un DAV performant et précis est encore trop coûteuse sur le plan calculatoire pour être envisagée. Lorsque la technologie le permettra, son emploi pourrait se révéler d'un très grand intérêt.

6.1.3. Estimation continue

L'estimation continue du niveau de bruit suppose que le signal de parole apparaît de manière *transitoire* au milieu du bruit de fond. En exploitant cette hypothèse, on peut considérer approximativement que toute hausse instantanée du niveau au-dessus de la valeur moyenne estimée du bruit témoigne de la présence de signal.

Ainsi, il est alors possible d'estimer le niveau du bruit à partir d'un lissage récursif du

premier ordre comparable à celui de l'équation (6.1) en employant des valeurs différentes du paramètre de lissage α suivant que la trame correspond à une attaque ou non.

$$\alpha = \begin{cases} \alpha_a & \text{si } \mathcal{P}_s^y[k] > \mathcal{P}_s^b[k] \\ \alpha_d & \text{si } \mathcal{P}_s^y[k] < \mathcal{P}_s^b[k] \end{cases}$$

Pour réaliser l'estimation souhaitée, il convient d'employer une constante de temps longue pour les périodes d'attaque et plus courte pour les périodes de descente. Dans l'exemple de la figure 6.1, le temps d'attaque correspondant à α_a est de 4s tandis que le temps de descente, correspondant à α_d est de 1ms¹. Cette figure illustre la mise en oeuvre de l'estimateur de niveau de bruit explicité ci-dessus sur un segment de signal de parole. Pour un indice fréquentiel k arbitraire, l'évolution de $\mathcal{P}_s^y[k]$ est indiquée en trait plein, tandis que celle de $\mathcal{P}_s^b[k]$ est indiquée en trait pointillé. L'estimation du niveau de bruit est mise en oeuvre après une phase d'initialisation sur 40 trames. Les valeurs hautes de $\mathcal{P}_s^y[k]$ correspondent à la présence de parole au milieu d'un bruit de fond (bruit blanc de densité spectrale de puissance d'amplitude 10). On constate que la valeur de $\mathcal{P}_s^b[k]$ tend progressivement vers le niveau du bruit de fond. En particulier, c'est la faible valeur du temps de descente (1ms) qui permet cette adaptation. Le temps d'attaque est réglé de telle sorte que l'estimateur puisse s'adapter à une éventuelle hausse du niveau de bruit sans pour autant être trop sensible à une hausse sporadique.

La figure 6.2 reprend le même exemple dans le domaine fréquentiel pour un indice de trame arbitraire s suffisamment grand pour que l'estimateur ait pu converger vers des valeurs convenables. On constate qu'à cet instant, l'estimation de la densité spectrale de puissance du bruit est relativement proche de sa valeur réelle (*i.e.* 10).

6.2. Solution adoptée

6.2.1. Modèle élémentaire de DAV

Comme cela a été précisé précédemment, l'implantation d'un DAV performant au sein d'une prothèse auditive n'est pour le moment pas encore techniquement envisageable. Toutefois, si l'on étudie sous un autre angle les méthodes d'estimation continue du niveau de bruit que nous venons de présenter, on peut considérer que celles-ci ont finalement recours à un modèle de DAV rudimentaire. En effet, cette approche empirique du traitement du signal audio qui consiste à observer l'enveloppe du signal en sous-bande et à dissocier les moments d'attaque et de descente du signal revient à faire une hypothèse *a priori* sur la présence du signal et donc à utiliser un DAV élémentaire.

1. Le temps d'attaque (temps de réponse à -3dB), τ_{att} , permet de définir la valeur du paramètre α_a tel que $\alpha_a = 0,5^{T_e/\tau_{att}}$ où T_e est la période d'échantillonnage.

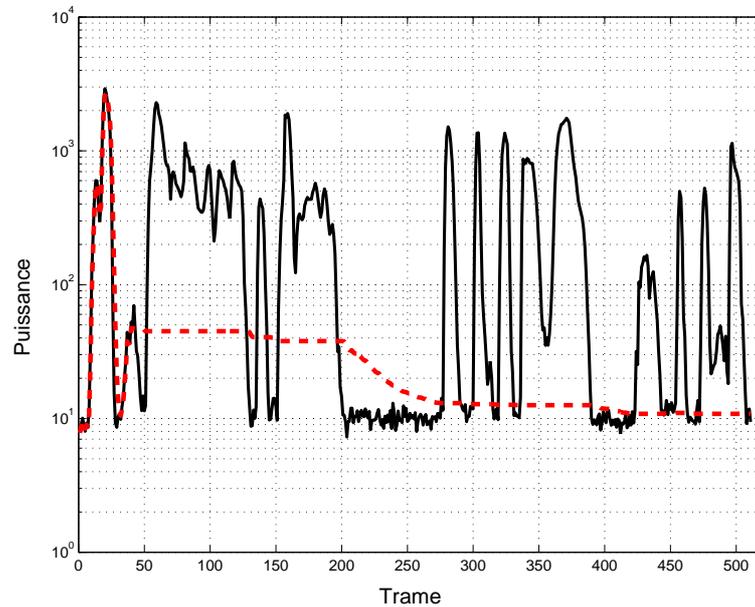


FIG. 6.1: Estimation du niveau de bruit, aspect temporel. $\mathcal{P}_s^y[k]$ est indiquée en trait plein, tandis que $\mathcal{P}_s^b[k]$ est indiquée en trait pointillé. Le bruit est un bruit blanc dont la densité spectrale de puissance a pour valeur 10. La phase d'initialisation sur les 40 premières trames est un simple lissage du signal.

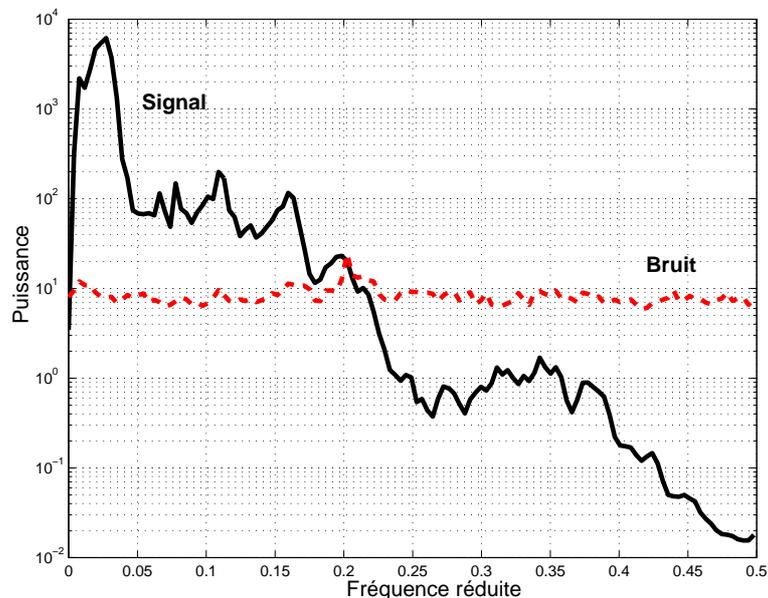


FIG. 6.2: Estimation du niveau de bruit, aspect fréquentiel. Les valeurs du périodogramme du signal source (sans bruit) sont indiquées en trait plein, tandis que les valeurs de $\mathcal{P}_s^b[k]$, sont indiquées en trait pointillé. Le bruit est un bruit blanc dont la densité spectrale de puissance a pour valeur 10.

Cette hypothèse est fortement liée à l'hypothèse de stationnarité du bruit de fond. Dans le cadre de notre application, il est, de toute manière, difficile d'envisager d'autres hypothèses plus rigoureuses sur les statistiques du signal source ou sur celles du bruit. Malgré tout, ce type d'estimation du niveau de bruit parvient néanmoins de manière assez correcte à estimer le niveau moyen du bruit de fond dans chaque sous-bande et permet donc de réduire au moins *la partie stationnaire* du bruit dans le signal rehaussé.

Dans la littérature et, en particulier, dans [50], un autre modèle rudimentaire de DAV est exploité dont la démarche est assez semblable à la détermination des attaques sur l'enveloppe. Son principe se base sur l'observation de la valeur moyenne de $R_s^{post}[k]$ sur les M indices fréquentiels k pour le bloc courant que l'on note \bar{R}_s^{post} :

$$\bar{R}_s^{post} = \frac{1}{M} \cdot \sum_{k=0}^{M-1} R_s^{post}[k] \quad (6.2)$$

Rappelons que R_s^{post} est une estimation du RSB pour la trame courante. Si \bar{R}_s^{post} a une valeur nulle ou quasiment nulle, la trame peut être considérée comme une trame de bruit. Il est donc possible en considérant un certain seuil, \bar{R}_{seuil} , proche de 0, de prendre une décision sur la présence éventuelle du signal de parole :

- si $\bar{R}_s^{post} > \bar{R}_{seuil}$, le signal de parole est présent,
- si $\bar{R}_s^{post} < \bar{R}_{seuil}$, seul le bruit est présent.

Dans [50], Malah et al. proposent de choisir \bar{R}_{seuil} entre 0,3 et 1. Cette valeur légèrement supérieure à 0 permet de tenir compte d'une éventuelle hausse du niveau de bruit.

6.2.2. Mise à jour de l'estimateur

Dans ce même article [50], les auteurs développent une manière très efficace de mettre à jour l'estimation du spectre du bruit. La mise à jour se fait grâce à une équation de récurrence régie par un facteur de lissage. Ce facteur de lissage s'exprime de manière simple à partir du RSB dans le bloc courant.

$$\alpha_b = 1 - 0,2 \cdot |\bar{R}_s^{post}| \quad (6.3a)$$

$$\text{en ajoutant la contrainte : } 0,8 < \alpha_b < 0,98 \quad (6.3b)$$

Les variations des valeurs de α_b en fonction de celles de $|\bar{R}_s^{post}|$ sont indiquées sur la figure 6.3.

6.2.3. Validation du choix de la méthode

Différents tests d'évaluation ont été menés sur les différentes techniques d'estimation du niveau de bruit. Ils ne seront pas présentés ici mais ils nous ont permis de déterminer que la méthode proposée dans [50] offre la possibilité d'aboutir à des résultats satisfaisants

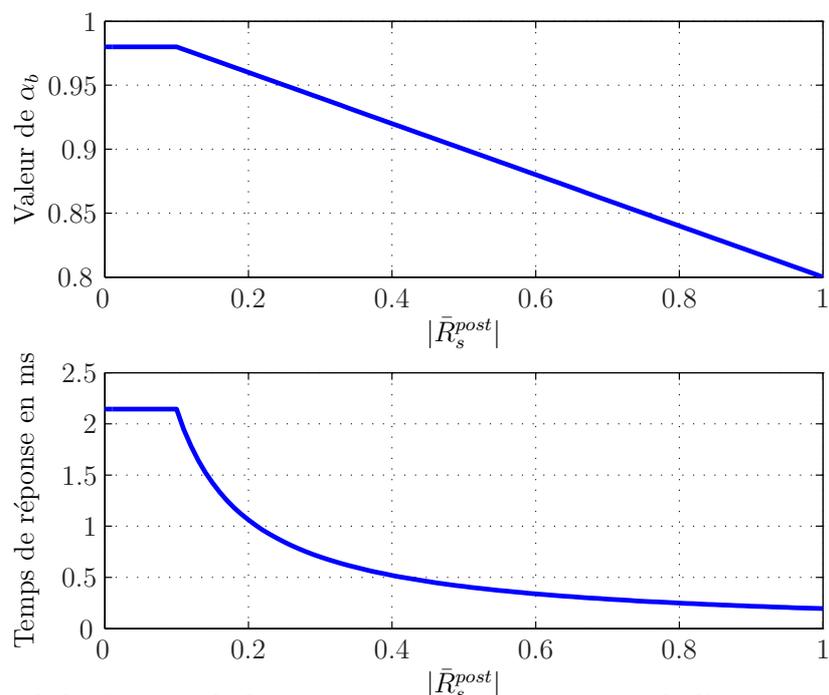


FIG. 6.3: Calcul du facteur de lissage α_b pour la mise à jour de l'estimation du niveau de bruit en fonction de \bar{R}_s^{post} . En haut, la figure indique les valeurs de α_b selon la formule (6.3). En bas, les temps de réponse à $-3dB$ correspondants sont indiqués (en ms).

de rehaussement tout en conservant un rapport efficacité / coût de calcul très intéressant. Lors de ces évaluations, les méthodes d'estimation du niveau de bruit n'ont pas été évaluées, *de manière directe*, *i.e.* sur leur précision d'évaluation de la densité spectrale de puissance du bruit mais de manière indirecte lors d'une utilisation conjointe avec la méthode de rehaussement d'Ephraïm et Malah.

Ce choix d'estimation est volontairement applicatif et vise à considérer l'efficacité de ces estimateurs uniquement dans le cadre du rehaussement de la parole.

7. Application de la règle de suppression d'Ephraïm et Malah à une analyse en banc de filtres ERB

7.1. Motivations à l'utilisation de l'échelle des ERB

7.1.1. Filtres auditifs et perception des signaux bruités

Dans la section 5.2.1, nous avons indiqué que les méthodes d'atténuation spectrale à court-terme peuvent être interprétées comme des méthodes se basant sur une analyse par bancs de filtres. Dans le cadre général du filtrage de Wiener paramétrique comme dans celui de la règle de suppression d'Ephraïm et Malah (EMSR), l'analyse spectrale est réalisée au moyen d'une TFCT.

Ceci implique donc naturellement que le signal est analysé à partir d'une échelle linéaire et uniforme en fréquence. Cependant, on a vu dans la partie I de ce document que le système auditif humain ne perçoit pas les signaux selon une échelle uniforme. Notamment, notre système auditif effectue une analyse plus précise dans les fréquences graves que dans les bandes hautes du spectre. Cette idée se retrouve à l'origine dans la littérature psychoacoustique sous le concept des *filtres auditifs*. Comme nous l'avons précisé dans la partie I, la largeur de ces filtres varie en fonction de leur fréquence centrale d'une manière comparable à celles des filtres à facteur de qualité Q constant.

Ceci est à l'origine de la notion de largeur de bande critique [54, 81]. La largeur de bande critique désigne, pour une fréquence centrale donnée, la largeur de la bande passante du filtre auditif correspondant selon la terminologie introduite par Zwicker.

Les concepts de bandes critiques et de filtres auditifs sont de plus en plus utilisés dans de nombreux champs du traitement du signal audio notamment pour le calcul des effets de masquage pour les codeurs audio perceptifs, en traitement de la parole ou encore pour le calcul de la sonie dans les mesures de niveau sonore. En outre, l'utilisation de la psychoacoustique a été fortement popularisée par les excellentes performances des codeurs audio perceptifs employés dans la couche audio des récents formats de compression MPEG.

La capacité de notre système auditif à distinguer un signal utile dans un environnement bruyant est intimement liée à cette notion de filtres auditifs. En effet, la détection d'un signal corrompu par un bruit additif est dépendante du seul RSB au sein du filtre auditif dont les caractéristiques (fréquence centrale et largeur de bande) maximisent ce RSB.

Dès lors, dans le contexte du rehaussement de la parole, il nous semble alors particulièrement intéressant d'étudier la possibilité de procéder à l'analyse du signal en employant une échelle conforme aux lois de la psychoacoustique plutôt qu'une échelle linéaire en fréquence comme le fait l'analyse par TFCT.

7.1.2. Choix de l'échelle fréquentielle

La largeur des bandes critiques des filtres auditifs a été mesurée sur la totalité de la bande audible à partir d'expériences classiques de masquage [81]. Ces résultats ont permis de définir une nouvelle échelle de fréquence pour laquelle une distance d'*une unité* correspond à *une largeur bande critique*. L'unité ainsi définie porte le nom de *Bark* et l'échelle de fréquence correspondante est l'*échelle Bark* que l'on retrouve notamment dans les applications de codage audio citées plus haut [57, 35]. On peut également noter l'utilisation de l'échelle Bark et d'une autre échelle fréquentielle issue de la psychoacoustique, l'échelle des *mels*, dans le domaine du traitement de la parole [63, 12].

Moore et Glasberg au cours d'expériences visant à déterminer la forme des filtres auditifs ont mis au point une autre échelle de fréquence appelée l'échelle Equivalent Rectangular Bandwidth (ERB) [55]. La figure 7.1 illustre la correspondance entre l'échelle ERB et les valeurs en Hertz. À notre connaissance, il semble que l'échelle ERB, plus récente, soit également plus juste et plus rigoureuse que l'échelle Bark du point de vue psychoacoustique. Cette affirmation est appuyée par son utilisation dans les modèles utilisés pour les calculs de sonie et de *pattern d'excitation*. L'utilisation de l'échelle ERB nous semble donc préférable par rapport à celle de l'échelle Bark.

7.1.3. Application de l'EMSR selon une analyse en échelle ERB

Actuellement, dans une prothèse auditive, l'implémentation de la méthode de rehaussement de la parole par application de l'EMSR peut se révéler inadaptée si la taille de la TFCT dépasse les 128 à 256 points à une fréquence d'échantillonnage de 16 kHz . Pour une décomposition fréquentielle sur plus d'une centaine de coefficients, les différentes raisons de cette inadaptation sont les suivantes :

- L'EMSR induit une complexité de calcul relativement importante étant données les capacités actuelles des processeurs dédiés aux prothèses auditives.
 - Le délai de traitement par blocs est alors proche de la limite accessible de 10 à 20ms.
 - Quelquefois, les processeurs de signaux dédiés hébergent des DFT ou des bancs de
-

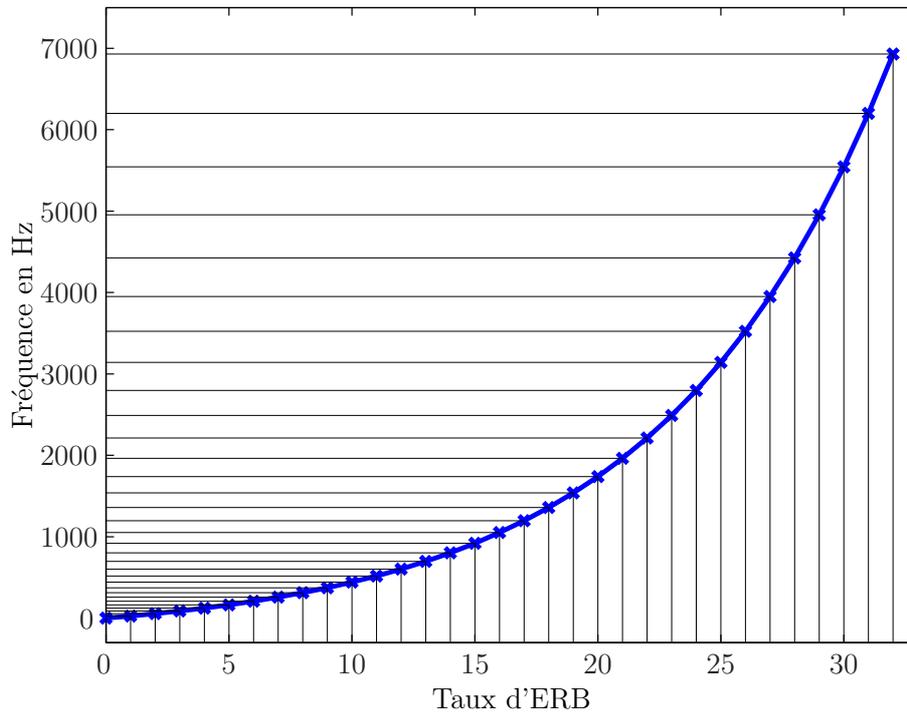


FIG. 7.1: Correspondance entre l'échelle ERB et les valeurs en Hertz

filtres d'analyse de petite taille (par exemple 32 canaux).

A cet effet, lors de son travail de thèse [32], Alain Goyé a implanté l'EMSR sur un banc de filtres uniforme à 32 canaux. Selon l'auteur, les résultats subjectifs obtenus étaient tout à fait convenables et de l'ordre de ceux réalisés en employant une analyse TFCT sur 256 canaux à la même fréquence d'échantillonnage de 16 kHz . Au regard de la théorie, ces résultats sont difficilement explicables car Ephraïm et Malah ont conçu leur règle de suppression sur des hypothèses d'indépendance statistique des coefficients relatifs de la TFCT du signal de parole et du bruit. Cette indépendance et l'hypothèse de gaussianité des coefficients sont d'autant plus utilisables que la taille de la TFCT est grande. Ainsi, selon la théorie, l'EMSR est censée être d'autant plus efficace que la taille de la TFCT est grande.

Sur la base de ces résultats, il apparaît que l'implémentation de l'EMSR à partir d'une analyse en banc de filtres sur un nombre réduit de canaux est réalisable. Nous avons souhaité étudier la possibilité d'utiliser une analyse par banc de filtres non-uniformes sur un nombre également restreint de canaux. La résolution fréquentielle choisie est une résolution uniforme sur l'échelle des ERB.

Dans certaines publications très récentes, on trouve l'idée de concevoir une règle de suppression selon des critères psychoacoustiques [75, 77]. Malgré l'intérêt d'une telle démarche, notre propos, ici, est simplement de réaliser l'analyse des signaux, l'estimation du niveau et le rehaussement selon une échelle fréquentielle répondant à des critères psychoa-

coustiques. Les phénomènes de masquage et de discrimination des sons sont intimement liés à la notion de filtres auditifs qui conditionnent les facultés de perception et d'extraction d'informations extrêmement performantes de notre système auditif. Il paraît donc intéressant d'envisager une analyse et une estimation du signal bruité en accord avec ce fonctionnement interne du système auditif.

Outre cet intérêt perceptif pour le rehaussement, la démarche choisie ici présente également un intérêt vis-à-vis des autres modules de traitement du signal de la prothèse. En effet, nous avons précédemment évoqué la compatibilité d'architecture entre les méthodes de rehaussement de la parole par atténuation spectrale à court-terme et les autres modules de traitement du signal d'une prothèse auditive. Or, pour ces autres traitements, le choix de l'utilisation d'une analyse fréquentielle uniforme sur l'échelle ERB est fortement appréciable.

7.1.4. Rehaussement de la parole sur une échelle fréquentielle non-linéaire

Dans [34], les auteurs comparent plusieurs méthodes de rehaussement de la parole réalisées sur un banc de filtres en ondelettes de 70 canaux et sur un banc de filtres de 256 canaux non-uniformes à une implémentation classique au moyen d'un TFCT sur 256 canaux. Dans cet article, les deux bancs de filtres à résolution non-uniforme avaient été construits de manière à approximer une résolution uniforme en échelle Bark. Le banc de filtres à ondelettes se découpe en 7 octaves de 10 voies. Le banc de filtres non-uniforme a été synthétisé par modification d'un banc de filtres polyphase classique. La synthèse de ces filtres consiste à remplacer les délais par des filtres passe-tout, réalisant un *warping* du signal au moyen d'une transformée bi-linéaire [70, 38].

Les méthodes de rehaussement choisies sont l'EMSR et la soustraction spectrale classique.

Selon les auteurs, l'utilisation d'une résolution fréquentielle uniforme en échelle Bark permet d'obtenir, du point de vue perceptif, des résultats meilleurs que les méthodes classiques de soustraction spectrale. Néanmoins, la présence de bruit musical est encore constatée. Par contre, les auteurs signalent que ces deux types d'analyse améliorent légèrement la qualité sonore lorsqu'on utilise la règle de suppression d'Ephraïm et Malah.

7.2. Synthèse de bancs de filtres sur-échantillonnés en échelle ERB

Afin de synthétiser le banc de filtres souhaité, nous nous sommes tournés vers une solution générale proposée par M. Goodwin dans [31]. Dans cette étude, l'auteur présente une méthode de synthèse de banc de filtres non-uniforme sans échantillonnage critique. Nous avons conservé la méthode proposée à l'exception du choix de la fonction de fenêtrage utilisée dans le calcul des réponses impulsionnelles des filtres. Goodwin propose l'utilisation de fenêtre en cosinus sur-élevé (*Raised Cosine Pulse*), dans notre cas, une fenêtre de Blackman nous a paru plus opportune car elle offre une meilleure résolution fréquentielle.

Dans la pratique, il est très difficile de définir un banc de filtres à résolution non-uniforme qui puisse être sous-échantillonné de manière critique sans générer des problèmes d'aliasing ou des distorsions. La synthèse de tels filtres est encore un domaine de recherche récent et largement ouvert. La réalisation d'un banc de filtres satisfaisant les conditions de sous-échantillonnage critique, bien que particulièrement attractif pour le domaine des prothèses auditives, dépasse le cadre de ce travail et n'a donc pas été abordée. Il en découle que le type de banc de filtres choisi représente un coût de calcul trop prohibitif pour être utilisé actuellement dans des appareils auditifs. Cette étude nous semble néanmoins intéressante afin de dégager la faisabilité du rehaussement de la parole avec une résolution en échelle ERB. Il convient donc d'envisager cette partie du travail de thèse comme une étude préliminaire.

Le banc de filtres est défini selon une répartition des fréquences limites uniformes selon l'échelle ERB.

Soit $\{E_0, E_1, \dots, E_k, \dots, E_K\}$ un ensemble équi-réparti de fréquences exprimées en échelle ERB tel que :

$$\begin{cases} E_0 = 0 \\ E_K \text{ est la valeur en ERB correspondant à la demi fréquence d'échantillonnage } \frac{F_e}{2} \\ E_0 < E_1 < \dots < E_k < \dots < E_K \\ \forall k \in [1; K], E_k - E_{k-1} = \frac{E_K - E_0}{K}. \end{cases}$$

A partir de cet ensemble de fréquences, on calcule alors les valeurs en Hertz correspondantes et les pulsations discrètes associées qui forment alors à leur tour l'ensemble $\{\Phi_0, \Phi_1, \dots, \Phi_k, \dots, \Phi_K\}$ avec $\Phi_0 = 0$ et $\Phi_K = \pi$.

Pour tout $k \in [1; K]$, on définit, Δ_k , la largeur de bande du $k^{\text{ème}}$ filtre telle que :

$$\Delta_k = \Phi_k - \Phi_{k-1}$$

ainsi que sa fréquence centrale, ω_k , telle que :

$$\omega_k = \frac{\Phi_k + \Phi_{k-1}}{2}$$

A partir de ces différents paramètres, on peut alors calculer explicitement les expressions des réponses impulsionnelles de chaque filtre du banc.

$$\forall p \in \mathbb{Z}, h_k[p] = \frac{\Delta_k}{\pi} \operatorname{sinc}\left(\frac{\Delta_k p}{2}\right) f[p] \quad (7.1)$$

$f[p]$ représente la fonction de fenêtrage qui définit la longueur de la réponse impulsionnelle. Comme nous l'avons déjà indiqué plus haut, nous avons choisi d'utiliser une fenêtre de Blackman de taille $2L + 1$ dont la définition est :

$$f[p] = \begin{cases} 0,4266 + 0,4965 \cos\left(\frac{2\pi p}{2L}\right) + 0,076 \cos\left(\frac{4\pi p}{2L}\right) & \text{si } p \in [-L; L] \\ 0 & \text{ailleurs} \end{cases}$$

Les filtres d'analyse utilisés sont des filtres à phase linéaire, symétriques et de taille impaire ($2L + 1$).

La figure 7.2 indique les réponses en fréquence d'un tel banc de filtres construit avec $K = 32$ canaux et de demi-longueur $L = 1000$. L est fixé à une valeur très importante afin d'assurer des zones de transition très nettes entre bandes, notamment pour les basses fréquences pour lesquelles la largeur de bande-passante est très fine. En sortie du banc de

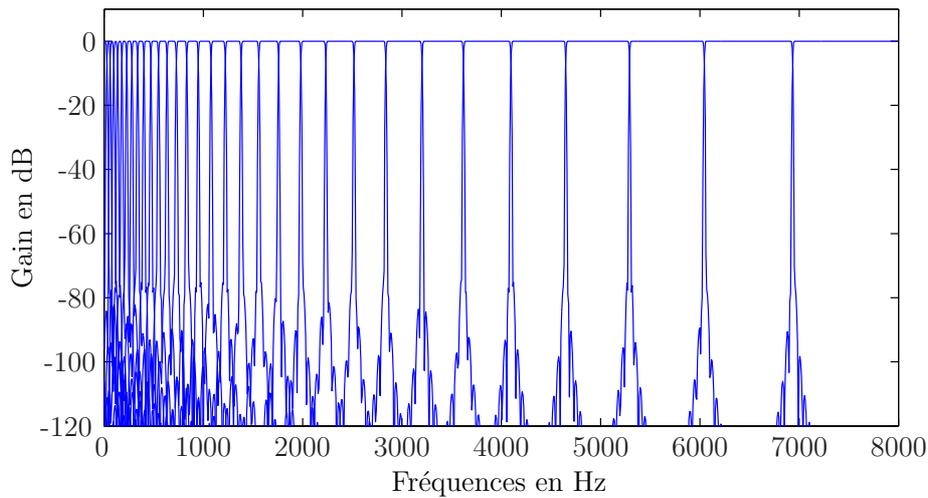


FIG. 7.2: Réponses en fréquence d'un banc de filtres ERB 32 canaux pour $L=1000$

filtres d'analyse, le signal d'entrée $y[n]$ est ainsi décomposé en K signaux de sous-bande

$y_k[n]$ ($k \in [1; K]$).

$$y_k[n] = \sum_{p=-L}^L h_k[p] \cdot y[n+p] \quad (7.2)$$

La contrainte imposée pour la synthèse des filtres est la reconstruction parfaite du signal d'entrée par simple addition des K signaux de sous-bandes :

$$y[n] = \sum_{k=1}^K y_k[n] \quad (7.3)$$

7.3. Implémentation de l'EMSR pour une analyse banc de filtres

En se basant sur l'interprétation « bancs de filtres » des méthodes d'atténuation spectrale à court terme, nous présentons, dans cette section, une méthode d'implémentation de l'EMSR sur une structure bancs de filtres sans décimation critique. Compte tenu de la structure fréquentielle des méthodes d'atténuation à court terme, cette adaptation est réalisée de manière très simple et intuitive.

Le gain spectral de l'EMSR, $G_s^{EM}[k]$, est calculé à partir de la puissance estimée de chaque sous-bande selon une analyse à court-terme et un découpage des signaux de sous-bandes en trames. De plus, afin de ne pas générer d'*effet de blocs*, on introduit un recouvrement temporel entre blocs successifs auxquels on applique une fonction de fenêtrage $\{w[p]\}_{p \in [0; N-1]}$. En procédant de cette manière, l'analogie avec une analyse par TFCT et une synthèse WOLA est encore plus directe.

En se basant sur le formalisme introduit pour regrouper le signal en trames recouvrantes dans l'équation (5.3), page 79, et dans la description de la synthèse WOLA, page 81, on construit les vecteurs des signaux en sous-bandes en entrée :

$$\mathbf{y}_s^k = \begin{bmatrix} y_k[sR] \\ \vdots \\ y_k[sR+N-1] \end{bmatrix} \odot \begin{bmatrix} w[0] \\ \vdots \\ w[N-1] \end{bmatrix} \quad (7.4)$$

L'estimation de la puissance du signal dans chaque sous-bande est faite directement à partir de ces vecteurs. Pour ne pas multiplier les notations de ce document, on redéfinit alors le vecteur \mathcal{P}_s^y des estimations du spectre en puissance pour le bloc temporel d'indice s par :

$$\mathcal{P}_s^y[k] = \sum_{p=0}^{N-1} \mathbf{y}_s^k[p]^2 \quad (7.5)$$

où l'on rappelle que k est l'indice de sous-bande, N est la taille des blocs et R est le pas d'avance entre deux blocs successifs.

Dès lors qu'on dispose, pour chaque trame s et pour chaque sous-bande fréquentielle d'indice k , de l'estimateur de la puissance spectrale $\mathcal{P}_s^y[k]$, la suite des calculs des différents paramètres nécessaires à la définition du gain spectral de l'EMSR (RSB *a priori* et *a posteriori*), $G_s^{EM}[k]$, et de l'estimation de la puissance du bruit $\mathcal{P}_s^b[k]$ s'opère de manière rigoureusement identique au cas classique présenté précédemment.

A partir de cette nouvelle expression de $\mathcal{P}_s^y[k]$, on calcule l'expression du gain spectral $G_s^{EM}[k]$ selon la formule (5.18).

On construit ensuite le signal rehaussé en multipliant dans chaque sous-bande le bloc s courant par le gain spectral $G_s^{EM}[k]$. Les signaux rehaussés de sous-bandes sont alors naturellement estimés par :

$$\hat{\mathbf{x}}_s^k = G_s^{EM}[k] \cdot \mathbf{y}_s^k \quad (7.6)$$

La reconstruction du signal est effectuée selon un modèle inspiré de la méthode d'*addition-recouvrement* Weighted OverLap-Add (WOLA) comme cela est schématisé sur la figure 7.3.

En optant pour l'utilisation d'une fenêtre de *Hanning* avec un taux de recouvrement des blocs de 50% ($R = \frac{N}{2}$), on peut alors se passer de l'utilisation d'une fenêtre de synthèse en sortie. De manière similaire à l'équation (5.12) :

$$\hat{x}[n] = \frac{1}{W_0} \sum_{s \in B_n} \hat{\mathbf{x}}_s^k[n-sR] \quad (7.7)$$

où B_n est dans notre cas l'ensemble des deux trames successives qui se recouvrent et qui contiennent l'indice temporel n .

Finalement, comme le suggère la condition de reconstruction du signal de l'équation (7.3), le signal rehaussé est construit par addition des signaux rehaussés de sous-bandes :

$$\hat{x}[n] = \sum_{k=1}^K \hat{x}_k[n] \quad (7.8)$$

On notera que les signaux manipulés, ici, sont réels. Il n'y a donc plus dans ce cas de notion de phase et d'amplitude comme dans le cas classique.

Cette implémentation peut être facilement extrapolée au cas d'un banc de filtres uniforme à décimation critique. La définition de la taille de blocs considérée doit être modifiée et on doit ajouter l'étape d'interpolation et le banc de filtres de synthèse avant la reconstruction finale par sommation des signaux de sous-bande de l'équation (7.8).

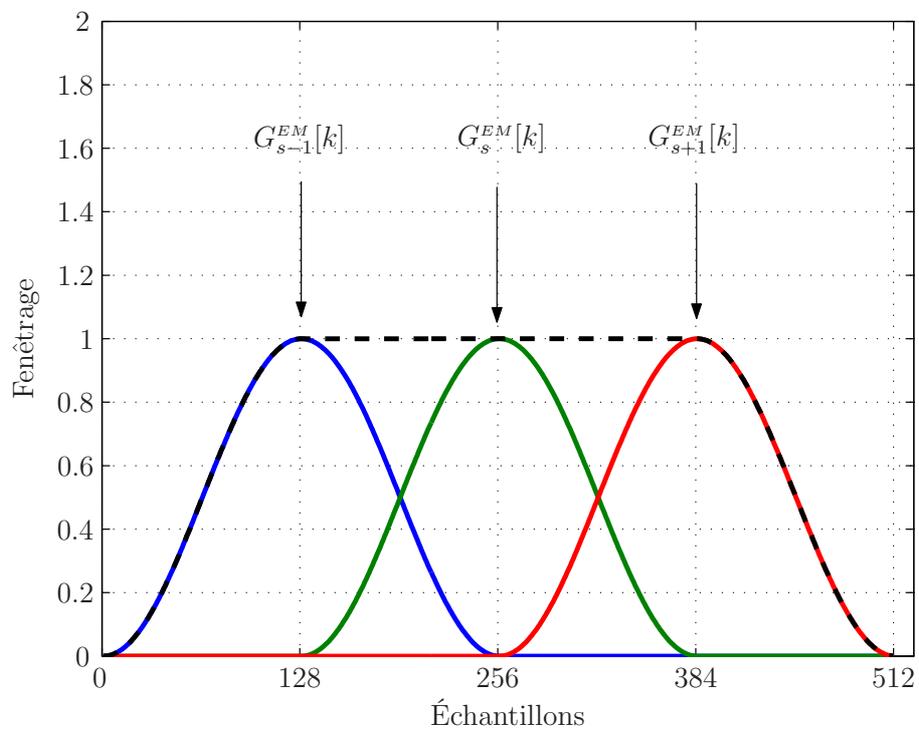


FIG. 7.3: Application du gain spectral sur des blocs recouvrants

8. Résultats et discussion

8.1. Protocole d'évaluation

8.1.1. Motivations

Lors de cette étude, nous avons cherché à caractériser l'efficacité de l'EMSR en échelle ERB. A cette fin, nous avons décidé de comparer cette méthode à l'implémentation classique par TFCT. De plus, afin de fournir un élément de comparaison supplémentaire, nous avons considéré une implémentation de l'EMSR à partir d'une analyse en banc de filtres classique (avec décimation maximale).

Le protocole d'évaluation consiste à comparer ces différents algorithmes sur une tâche de rehaussement d'un ensemble de signaux corrompus artificiellement par un bruit additif.

Comme le niveau de bruit introduit dans le signal est connu, il est alors possible d'évaluer de manière objective le rehaussement apporté par chacune des méthodes à travers l'augmentation du RSB. Cependant, une telle mesure objective ne renseigne pas nécessairement sur la qualité subjective réelle du rehaussement. A cet effet, nous avons décidé de mener une série de tests d'évaluation perceptive.

8.1.2. Choix d'implémentations

Afin de simplifier la lecture de ce document, nous ferons désormais référence aux trois implémentations de l'EMSR à travers les dénominations :

- *EMSR-FFT*, pour l'implémentation classique avec TFCT
- *EMST-BF*, pour l'implémentation de l'EMSR par banc de filtres uniforme à décimation maximale
- *EMSR-ERB*, pour l'implémentation de l'EMSR par banc de filtres en échelle ERB.

Pour les trois méthodes, l'analyse du signal s'accompagne d'un traitement des données par blocs, fenêtrés au moyen d'une fenêtre de Hanning et se recouvrant selon un taux de 50%.

Pour cette étude, l'analyse fréquentielle à court-terme précédant l'EMSR-FFT est réalisée au moyen d'une TFCT sur 256 points.

Les deux implémentations en banc de filtres, l'EMSR-BF et l'EMSR-ERB sont réalisées à partir d'une décomposition du signal en 32 sous-bandes et la taille des blocs est de $20ms$.

Pour l'EMSR-BF, le banc de filtres utilisé est un banc de filtres en cosinus-modulé classique à décimation critique (filtre pseudo-QMF)[74, 60]. Les réponses en fréquences de ce banc de filtres sont reproduites à titre indicatif sur la figure 8.1.

Pour chaque méthode, les différents paramètres d'implémentations employés sont explicités dans le tableau 8.1.

Implémentation	EMSR-FFT	EMSR-ERB	EMSR-BF
Nombres de sous-bandes	129	32	32
Taille des fenêtrage d'analyse (éch.)	256	320	320
Durée des fenêtrage d'analyse (ms)	16	20	20
Taux de recouvrement	50%	50%	50%
Fonction de fenêtrage	Hanning	Hanning	Hanning
Longueur des filtres (RIF)	-	2401	640

TAB. 8.1: Paramètres d'implémentation des méthodes de rehaussement

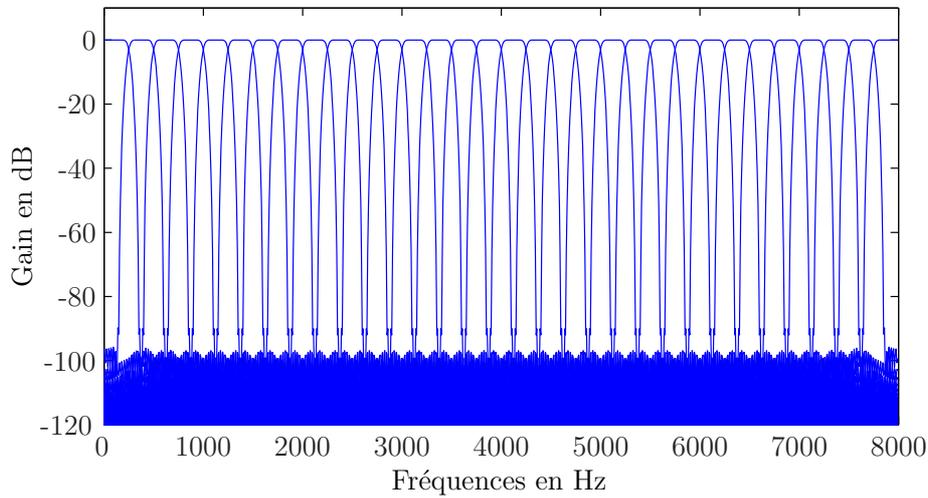


FIG. 8.1: Réponse en fréquence d'un banc de filtres uniforme sur 32 canaux

La fréquence d'échantillonnage étant de $16kHz$, le banc de filtres en échelle ERB décompose le signal en 32 canaux sur l'ensemble de la plage de fréquences $[0; 8kHz]$. Comme $8kHz$ correspond à une valeur de $33,3$ ERB, chaque sous-bande possède ainsi une largeur quasiment équivalente à celle d'un filtre auditif (1 ERB).

Cette étude consiste à appliquer ces différents algorithmes à un même ensemble de signaux corrompus par un bruit additif et à comparer leur performance.

8.1.3. Corpus de test

Les signaux à rehausser sont des signaux bruités artificiellement à partir de signaux de parole (phrases en français) auxquels nous avons ajouté différents types de bruit à différents niveaux de RSB.

Tous les signaux traités sont échantillonnés à la fréquence de 16 kHz et quantifiés sur 16 bits.

Les signaux de bruits utilisés sont issus de la base de données *Noisex-92* disponible sur le site internet du *DSP group* de l'université de *RICE* (USA) au sein de la banque d'informations *SPIB* [64].

Ces bruits sont des bruits large-bande et sont constitués respectivement :

- du bruit blanc,
- du bruit de voiture (prise de son dans l'habitacle du véhicule),
- du bruit de voix (bruit d'ambiance d'une salle de réfectoire) appelé, ici, *bruit de brouhaha*.

Pour chaque type de bruit, on synthétise les signaux bruités en ajoutant le bruit au signaux originaux de manière à atteindre les niveaux de RSB suivants :

- pour le bruit blanc : 10, 20 et 30 *dB* de RSB,
- pour le bruit de voiture : -10, 0 et 10 *dB* de RSB et
- pour le bruit de brouhaha : 0, 10, 20 *dB* de RSB.

Comme on peut le constater, les valeurs de RSB choisies pour ces signaux sont de niveaux distincts. En particulier, pour le bruit de voiture, les niveaux de RSB sont 10dB plus bas que pour le brouhaha et 20dB plus bas que pour le bruit blanc. Cette différence est justifiée par la différence de pouvoir masquant de ces bruits sur le signal. Pour chaque type de bruit, les trois niveaux de RSB choisis correspondent approximativement à une sensation subjective équivalente de gêne auditive due au bruit.

Afin de donner une indication plus précise de la coloration spectrale des bruits utilisés, nous avons reproduit sur la figure 8.2 les spectrogrammes respectifs du bruit de voiture et du bruit de brouhaha.

8.2. Amélioration du Rapport Signal à Bruit

8.2.1. Avec estimation du bruit

Comme les signaux bruités ont été synthétisés artificiellement, on peut calculer le RSB final du signal rehaussé dans chacun des cas et le comparer au niveau initial avant rehaussement. Les résultats obtenus sont rassemblés dans le tableau 8.2 et illustrés sous forme

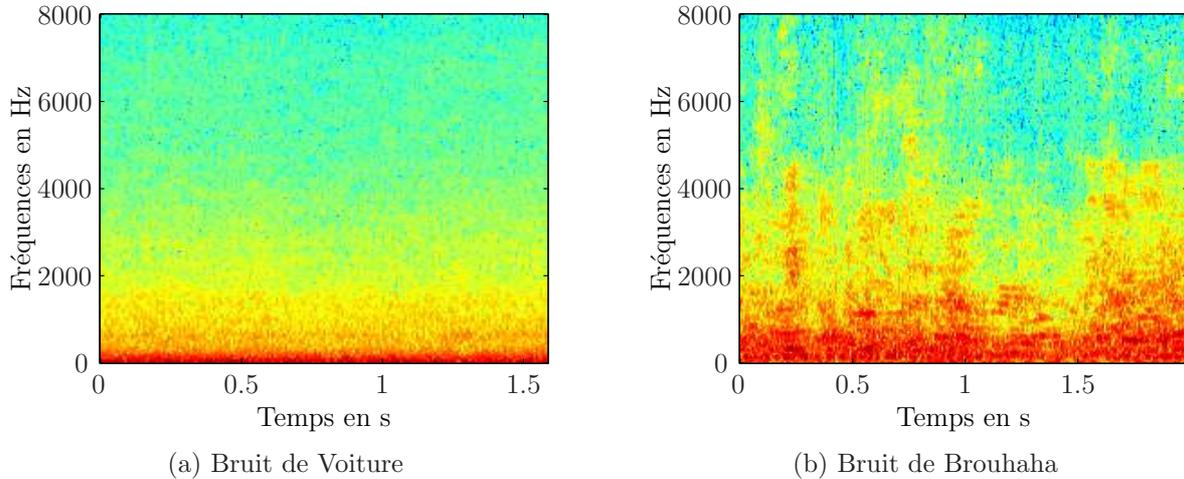


FIG. 8.2: Spectrogrammes des bruits employés

graphique dans la figure 8.3. Il convient de noter que les essais ont été réalisés sans aucune optimisation particulière pour chaque méthode. En effet, pour chacune d'elles, nous avons conservé la paramétrisation classique initiale de l'EMSR-FFT et de la méthode d'estimation du niveau de bruit présentée dans le chapitre 6.

Méthode \ RSB en dB	Bruit blanc			Brouhaha			Bruit de voiture		
	10	20	30	0	10	20	-10	0	10
EMSR-FFT	15,2	23,2	31,9	1,2	10,0	21,0	3,4	10,9	20,8
EMSR-BF	13,1	20,9	24,5	-1,6	5,6	18,5	2,8	7,9	9,0
EMSR-ERB	15,4	24,3	33,2	1,6	12,0	21,6	-4,1	6,1	16,1

TAB. 8.2: Évaluation du RSB pour les trois implémentations, avec estimation du bruit

Dans la grande majorité des cas, le RSB a été substantiellement amélioré, confirmant ainsi la qualité de l'EMSR à réduire le bruit de fond et la possibilité de l'adapter à une structure en banc de filtres. En première analyse, les résultats obtenus montrent en effet que les implémentations de l'EMSR en structure bancs de filtres 32 canaux (EMSR-BF ou EMSR-ERB) donnent, en général, des résultats quantitatifs du même ordre que l'EMSR-FFT sur 256 points de TFCT. Cependant, on peut noter que, dans la majorité des cas, l'implémentation en échelle ERB offre des résultats légèrement meilleurs que la méthode classique et de quelques dB supérieurs à l'alternative en banc de filtres uniformes. La seule exception est le cas de bruit de voiture où la méthode classique se montre nettement plus efficace.

Ce résultat est surprenant étant donné que l'analyse ERB apporte une meilleure résolution dans les fréquences graves. On aurait pu s'attendre à ce qu'elle se comporte mieux face à un bruit à spectre basse fréquence.

Ces résultats sont également illustrés à l'aide de spectrogrammes sur les figures 8.4,

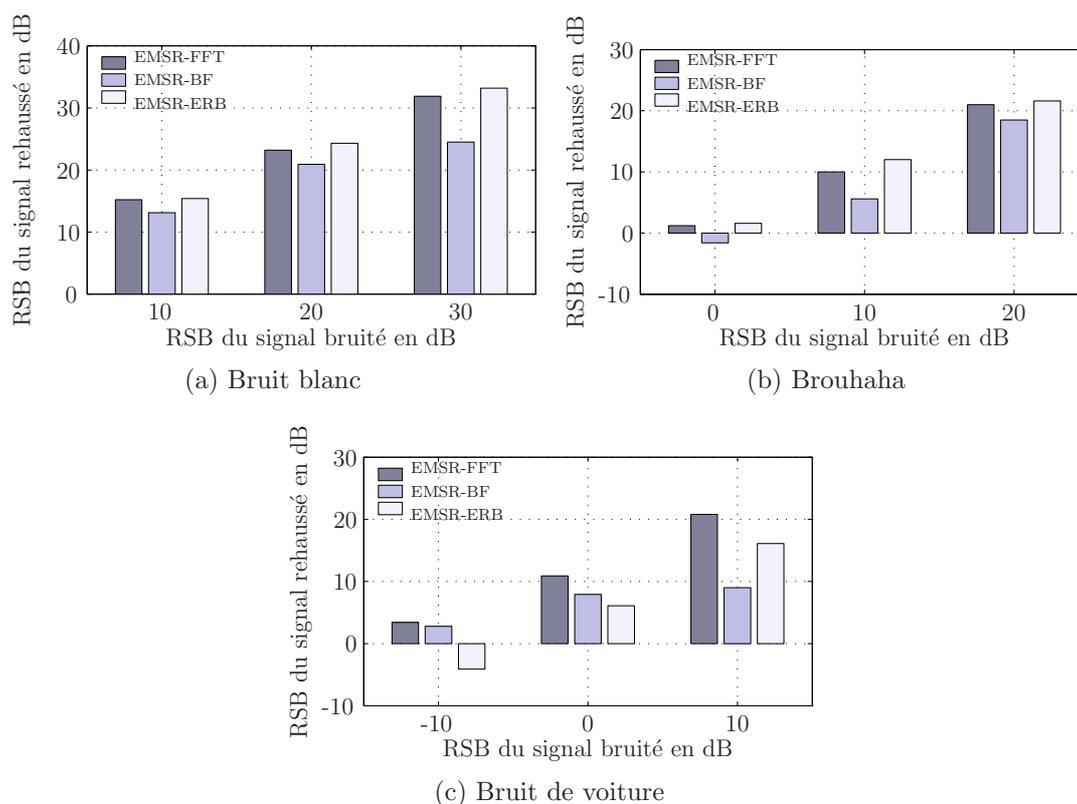


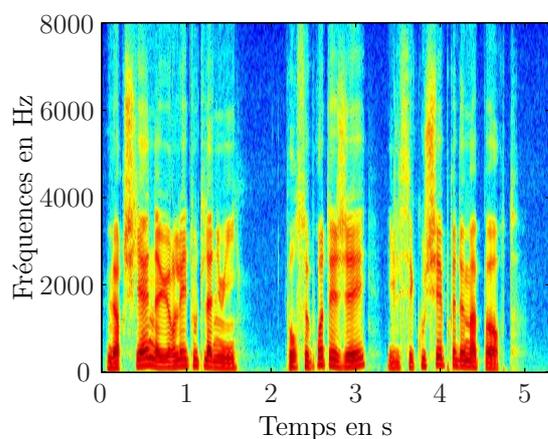
FIG. 8.3: Évaluation du RSB pour les trois implémentations, avec estimation du bruit

8.5 et 8.6 présentant les cas de signaux corrompus respectivement par du bruit blanc, du bruit de voiture et du brouhaha.

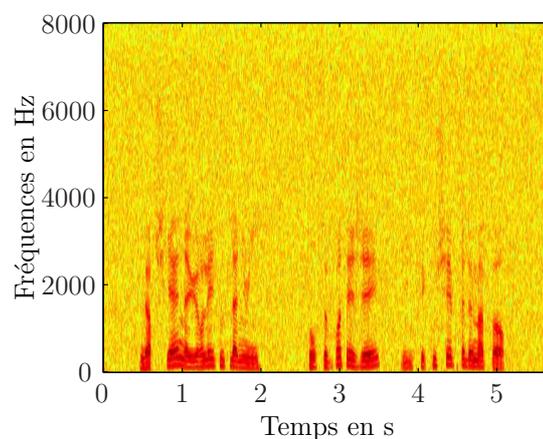
8.2.2. Sans estimation du bruit

Comme les caractéristiques du bruit ajouté au signal source sont connues, il est possible de faire les essais en utilisant non plus une estimation de la puissance du bruit mais la vraie valeur de cette puissance. Les résultats obtenus alors sont quantitativement différents des précédents comme on peut le voir dans le tableau 8.3. En effet, on peut voir que l'EMSR-ERB est alors plus efficace que les autres méthodes dans le cas de bruit de voiture et qu'elle est aussi efficace que la méthode classique pour du brouhaha. Dans le cas de bruit blanc, c'est la méthode classique qui est plus efficace. Cependant dans ces trois conditions de bruit, la différence entre ces deux méthodes est relativement faible. Par contre, l'EMSR-FFT et l'EMSR-ERB sont à chaque fois plus performants que l'implémentation en banc de filtres uniformes.

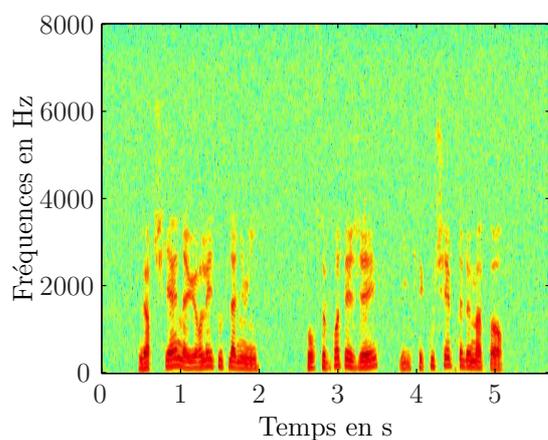
Ces derniers résultats sont plus intuitifs que les précédents puisqu'ils expriment la meilleure résolution de la technique ERB en basse fréquence.



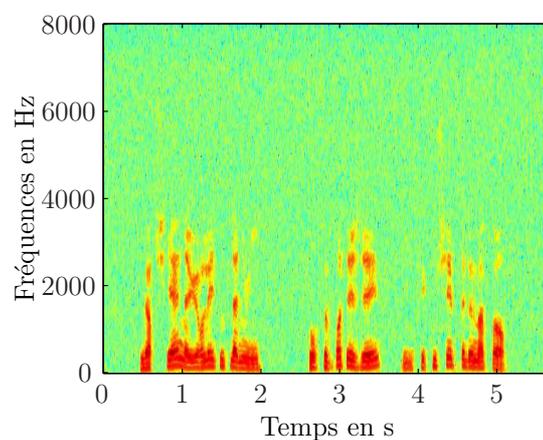
(a) Signal de parole source (locuteur français masculin)



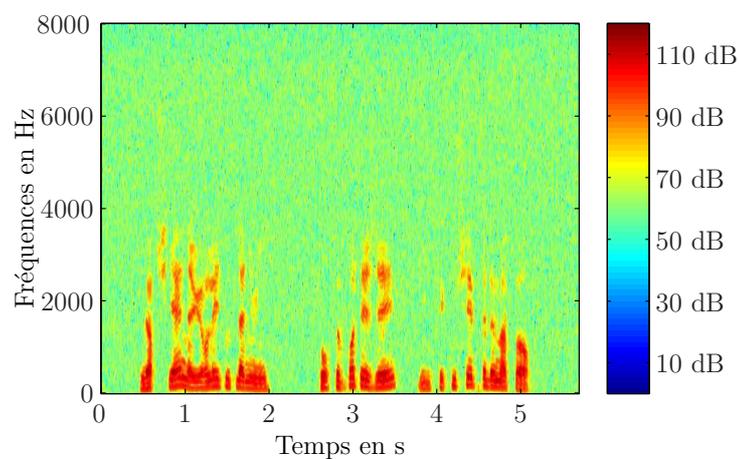
(b) Signal de parole corrompu par un bruit blanc additif (RSB=10dB)



(c) Signal rehaussé (EMSR-FFT)

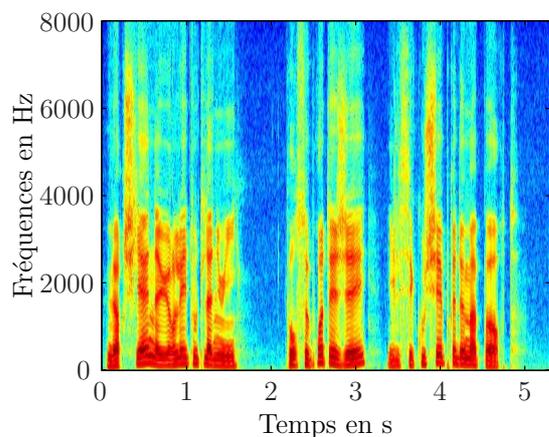


(d) Signal rehaussé (EMSR-BF)

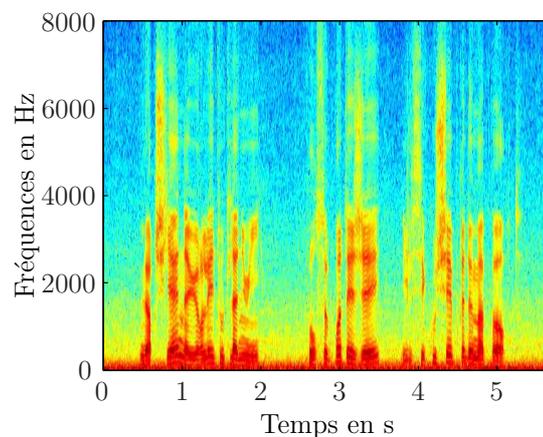


(e) Signal rehaussé (EMSR-ERB)

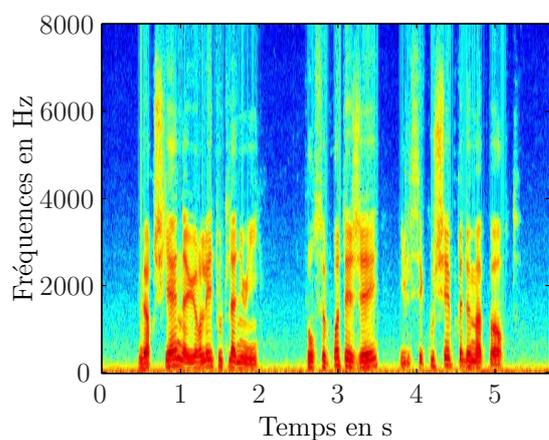
FIG. 8.4: Comparaison des spectrogrammes des signaux rehaussés correspondant au signal corrompu par un bruit blanc à 10 dB de RSB indiqué dans le tableau 8.2.



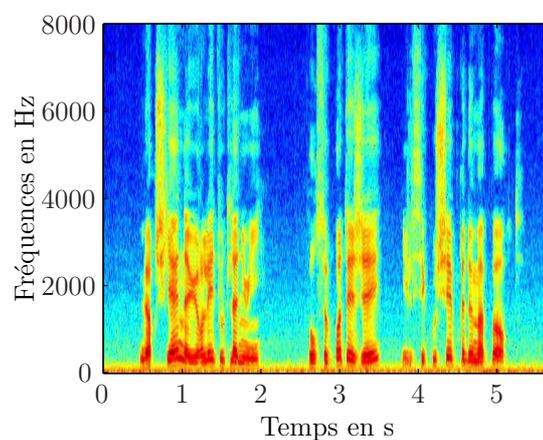
(a) Signal de parole source (locuteur français masculin)



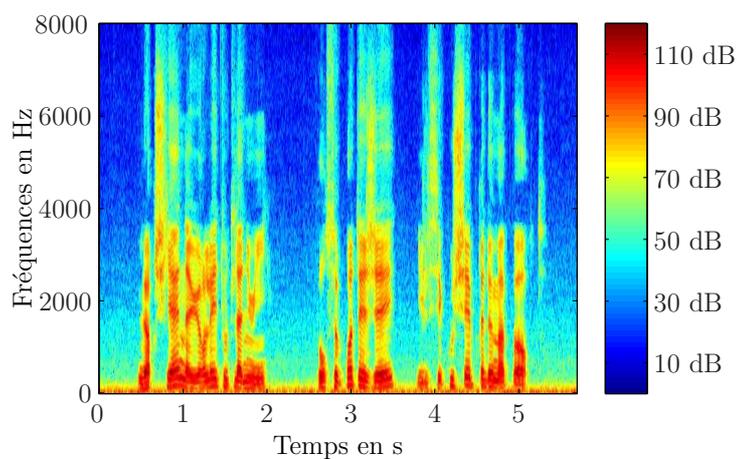
(b) Signal de parole corrompu par un bruit de voiture additif (RSB=0dB)



(c) Signal rehaussé (EMSR-FFT)

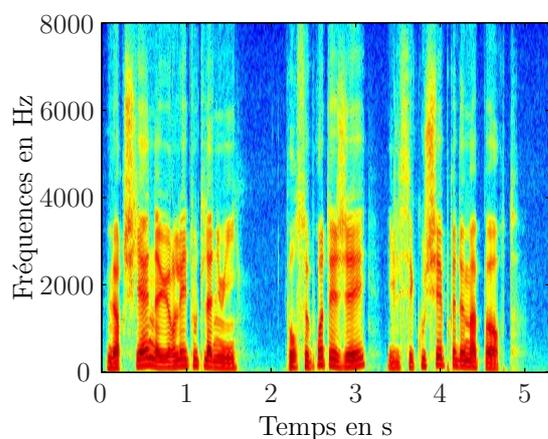


(d) Signal rehaussé (EMSR-BF)

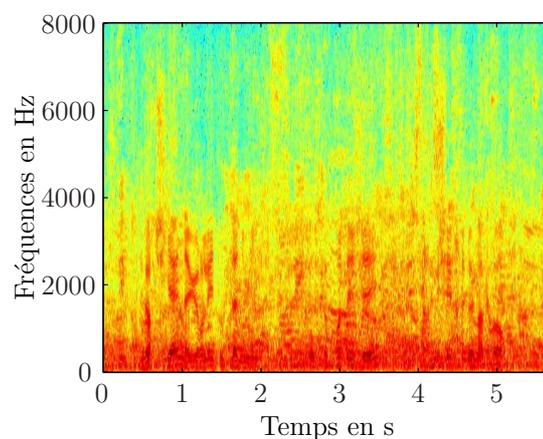


(e) Signal rehaussé (EMSR-ERB)

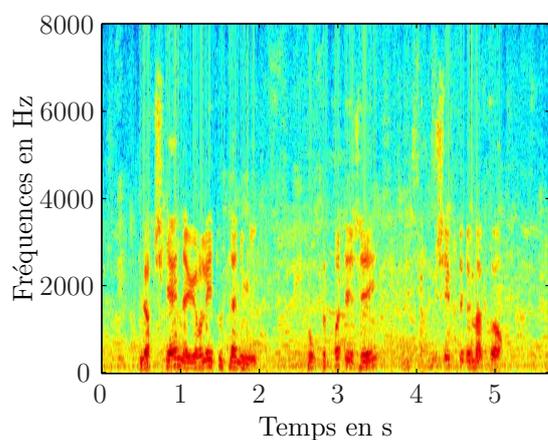
FIG. 8.5: Comparaison des spectrogrammes des signaux rehaussés correspondant au signal corrompu par un bruit de voiture à 0 dB de RSB indiqué dans le tableau 8.2.



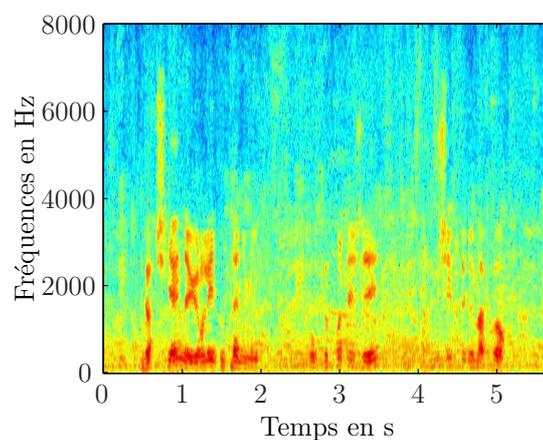
(a) Signal de parole source (locuteur français masculin)



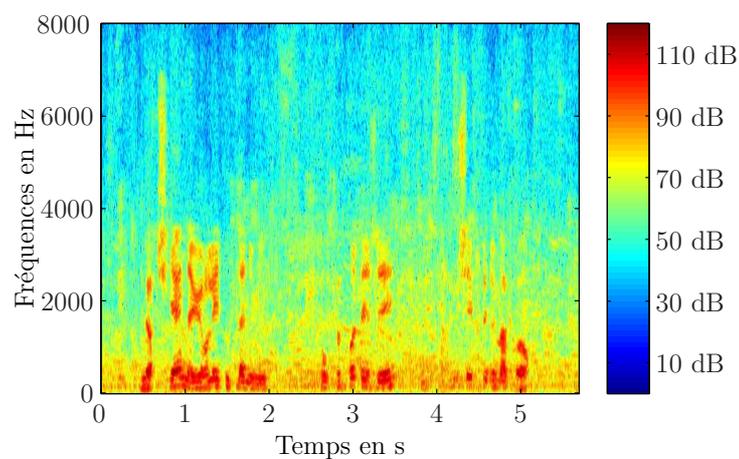
(b) Signal de parole corrompu par un bruit aditif de type brouhaha (RSB=0dB)



(c) Signal rehaussé (EMSR-FFT)



(d) Signal rehaussé (EMSR-BF)



(e) Signal rehaussé (EMSR-ERB)

FIG. 8.6: Comparaison des spectrogrammes des signaux rehaussés correspondant au signal corrompu par un bruit de brouhaha à 0 dB de RSB indiqué dans le tableau 8.2.

Méthode \ RSB	Bruit blanc	Brouhaha	Bruit de voiture
	10	10	0
EMSR-FFT	17,1	13,0	12,1
EMSR-BF	15,2	11,0	7,9
EMSR-ERB	16,0	13,0	13,6

TAB. 8.3: Évaluation du RSB pour les trois implémentations, sans estimation du bruit

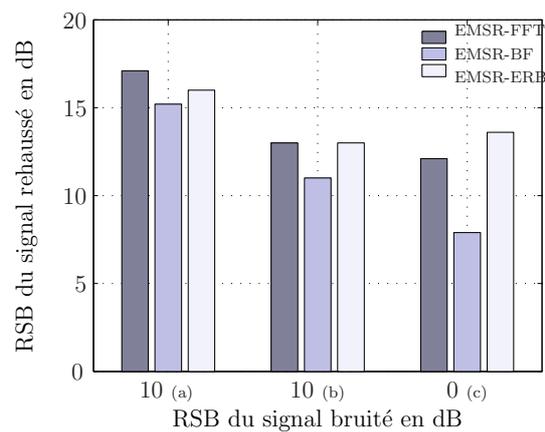


FIG. 8.7: Évaluation du RSB pour les trois implémentations, sans estimation du bruit. (a) Bruit blanc à 10 dB de RSB. (b) Brouhaha à 10 dB de RSB. (c) Bruit de voiture à 0 dB de RSB.

Ceci nous montre que la méthode ERB est performante et que les résultats relativement moins bons dans le cas où l'on estime le niveau de bruit de manière aveugle sont peut-être dus à une paramétrisation inadaptée de l'estimation du niveau de bruit. Par conséquent, il serait souhaitable, dans un travail futur, de définir de manière rigoureuse une paramétrisation adéquate pour ce type d'analyse.

Les résultats quantitatifs que nous venons de présenter ont l'avantage d'être facilement exploitables et se basent sur des outils d'analyse définis de manière claire et reproductible. Cependant, si la mesure du RSB est un bon indicateur de la quantité de bruit effectivement soustrait au signal, cette mesure n'en demeure pas moins un indicateur peu représentatif du point de vue psychoacoustique.

Afin de présenter une étude qualitative, parallèlement à l'étude quantitative ci-dessus, nous avons réalisé une série de tests d'écoute subjectifs que nous présentons dans la section suivante.

8.3. Évaluation subjective

8.3.1. Analyse subjective

Du point de vue perceptif, le rehaussement réalisé par chacune des trois méthodes est à peu près du même ordre de performance. On peut cependant noter quelques différences de qualité sonore.

En utilisant rigoureusement la même paramétrisation de l'EMSR et de l'estimation aveugle du bruit dans les trois méthodes, on obtient les résultats qualitatifs suivants :

- Dans l'implémentation EMSR classique par TFCT sur 256 points, le signal rehaussé contient de légères distorsions mais le niveau de bruit a été fortement diminué. De plus, le bruit de fond résiduel possède, à l'écoute, la même coloration spectrale que l'original. On peut cependant noter la présence d'artéfacts dans ce bruit de fond.
- Dans l'implémentation ERB, le niveau de bruit final est un peu plus important que pour l'implémentation classique mais sans distorsions audibles. Dans ce cas, le bruit résiduel semble plus concentré dans les basses fréquences que le bruit original ce qui lui confère parfois une coloration plus naturelle à l'oreille.
- Dans l'implémentation sur banc de filtres uniforme, le rehaussement obtenu est comparable à l'implémentation classique avec toutefois une moins bonne réussite dans la suppression du bruit. Cependant, dans quelques exemples, la qualité sonore du bruit résiduel paraît plus agréable à l'écoute.

On peut supposer que les implémentations bancs de filtres engendrent un bruit résiduel plus agréable car la largeur des bandes ne donne plus au bruit résiduel ce caractère tonal qui caractérise le bruit musical. Cet aspect positif est obtenu en contrepartie d'une moins bonne réduction du bruit.

On peut également noter que la méthode classique se comporte mieux que les autres face à du bruit blanc. Dans d'autres conditions de bruit, les résultats sont plus nuancés et moins évidents à analyser.

8.3.2. Tests subjectifs

Les résultats des tests perceptifs sont rassemblés dans les tableaux 8.4 et 8.5.

Ces tests ont consisté en un classement préférentiel entre les trois méthodes de rehaussement par les auditeurs. Les extraits sonores ont été présentés de manière aléatoire à un panel de onze auditeurs auquel on a demandé de classer les trois extraits par ordre de préférence et d'assigner les notes suivantes suivant les préférences d'écoute :

- 1 pour l'extrait préféré,
- 2 pour l'extrait moyen et
- 3 pour l'extrait le moins apprécié.

Une notation égale est admise si le choix est difficile à effectuer entre deux extraits.

8.3.3. Résultats

Les tableaux 8.4 et 8.5 présentent la moyenne des classements par ordre de préférence obtenus pour chaque niveau de RSB. Les extraits employés pour ces tests perceptifs sont les mêmes que ceux définis pour l'évaluation quantitative et pour lesquels le RSB avait été calculé et présenté dans les tableaux 8.2 and 8.3.

Méthode \ RSB	Bruit blanc			Brouhaha			Bruit de voiture		
	10	20	30	0	10	20	-10	0	10
EMSR-FFT	1,2	1,4	1,4	1,7	1,7	2,5	2,5	1,4	1,3
EMSR-BF	1,9	1,9	1,8	1,7	2,5	1,3	1,5	2,5	2,8
EMSR-ERB	2,7	2,7	2,8	2,6	1,7	2,1	2	2,2	1,9

TAB. 8.4: Tests d'écoute subjectifs. Moyenne des classements avec estimation du bruit

Méthode \ RSB	Bruit blanc	Brouhaha	Bruit de voiture
	10	10	0
EMSR-FFT	1,6	1,8	1,9
EMSR-BF	2	2,8	3
EMSR-ERB	2	1,4	1,1

TAB. 8.5: Tests d'écoute subjectifs. Scores subjectifs sans estimation du bruit

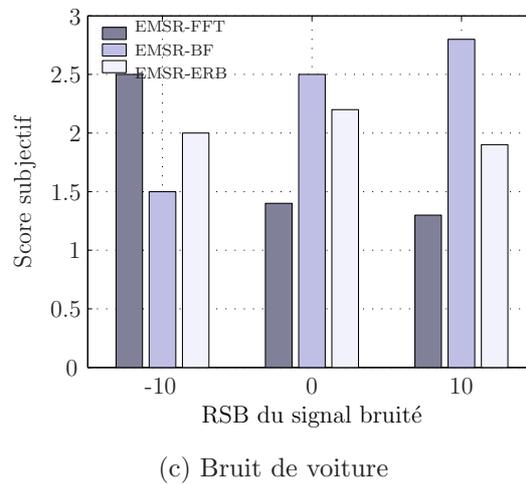
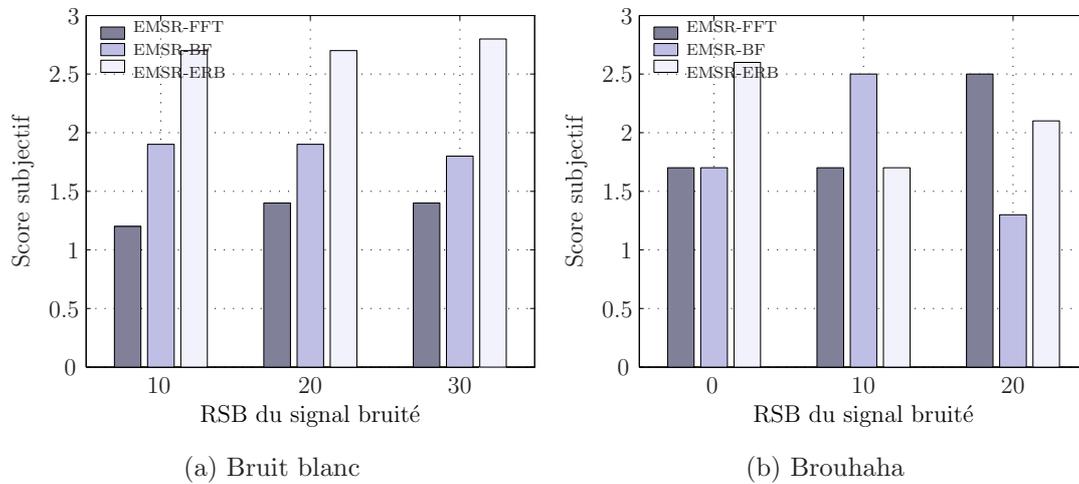


FIG. 8.8: Tests d'écoute subjectifs. Scores subjectifs avec estimation du bruit

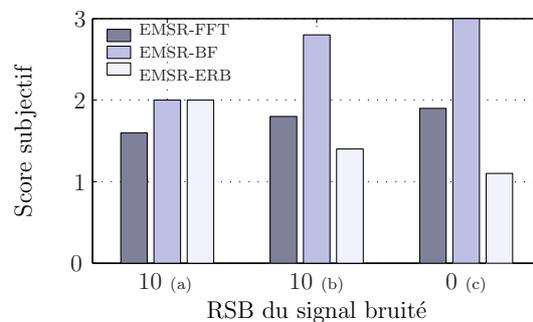


FIG. 8.9: Tests d'écoute subjectifs. Scores subjectifs sans estimation du bruit. (a) Bruit blanc à 10 dB de RSB. (b) Brouhaha à 10 dB de RSB. (c) Bruit de voiture à 0 dB de RSB.

A partir des résultats du tableau 8.5, on remarque que la méthode ERB est préférée aux deux autres pour le bruit de voiture et le bruit de parole lorsque les vraies valeurs de niveau de bruit sont utilisées. Avec l'estimation aveugle du bruit décrite au chapitre 6, les résultats sont plus nuancés comme le montre le tableau 8.4.

Cette remarque corrobore celle faite précédemment pour les tests quantitatifs, à savoir que l'estimation aveugle du bruit ne semble pas tout à fait adaptée dans sa paramétrisation au cas ERB. Or, d'après les bons résultats obtenus dans les conditions où le niveau de bruit est connu, on peut penser qu'une optimisation de la méthode pourrait conduire à de très bons résultats, dépassant ceux de la méthode classique.

Conclusion de la deuxième partie

8.4. Conclusion

Au cours de cette deuxième partie du rapport de thèse, nous avons présenté les méthodes de rehaussement du signal de parole par modification spectrale à court-terme dont la structure est particulièrement appropriée à une implantation dans une prothèse auditive. Parmi ces méthodes, nous avons désigné la règle de suppression d'Ephraïm et Malah (EMSR) comme étant une solution préférable dans le contexte de la réhabilitation auditive pour les malentendants.

Suite à cela, nous avons également proposé une adaptation de l'EMSR à une structure par banc de filtres à résolution uniforme selon l'échelle ERB. Ce choix a été motivé par le souhait d'adopter une analyse conforme aux lois de la psychoacoustique.

Au regard des résultats quantitatifs obtenus, il apparaît que la méthode proposée permet d'obtenir une amélioration du RSB du même ordre de grandeur que la méthode de référence, *i.e.* l'EMSR. Dans certains cas, la méthode ERB permet même d'atteindre des valeurs de RSB supérieures.

En outre, les tests perceptifs entrepris ont confirmé la capacité de cette implémentation particulière à réaliser un rehaussement correct. Ces résultats montrent également que le rehaussement pourrait être encore amélioré si le réglage des paramètres de contrôle de l'estimation du niveau de bruit était reformulée de manière appropriée à l'analyse selon l'échelle ERB.

L'étude comparative que nous venons de présenter montre que l'analyse du signal selon l'échelle ERB convient tout à fait pour la décomposition du signal en sous-bandes avant application de la règle de suppression du bruit.

D'autres tests que nous avons réalisés en utilisant l'EMSR à partir d'une modélisation des filtres auditifs basée sur les filtres *Gammachirp* introduits par Irino et Patterson dans [39, 40] se sont révélés très performants vis-à-vis de la qualité perceptive du rehaussement de la parole. Nous n'avons pas présenté ces résultats car la complexité de cette modélisation des filtres auditifs est extrêmement élevée.

La performance de l'EMSR appliquée sur une échelle ERB est très intéressante car, comme les autres méthodes d'atténuation spectrale à court-terme, l'EMSR partage avec les modules d'amplification et de compression d'une prothèse auditive le principe d'appliquer un gain variable à chaque sous-bande. Or, l'analyse en échelle ERB est particulièrement avantageuse pour l'amplification et la compression puisque l'échelle ERB est issue directement de la modélisation des filtres auditifs. Son utilisation permettrait alors d'ajuster avec précision les paramètres de compensation des pertes auditives.

8.5. Perspectives

Un important travail reste à accomplir d'une part pour améliorer le rehaussement ainsi que l'estimation du niveau de bruit et d'autre part pour réduire la complexité et notamment permettre un sous-échantillonnage critique des signaux de sous-bandes afin de pouvoir envisager son implémentation sur une prothèse auditive.

Dans le but d'améliorer le rehaussement, il nous semble intéressant d'envisager de prendre en compte pour la règle de suppression et pour l'estimation du niveau de bruit la probabilité de présence du signal comme cela est suggéré dans les travaux récents d'Israël Cohen [17, 18] ou par l'intermédiaire d'un Détecteur d'Activité Vocale. En outre, la probabilité de présence du signal est une variable intéressante à plus d'un titre dans le cadre de la réhabilitation des déficits auditifs. Elle peut permettre d'adapter les traitements effectués sur le signal en fonction de la présence ou non d'un signal de parole.

Du fait de la complexité élevée de l'analyse en banc de filtres ERB par rapport aux capacités des DSP utilisées dans les prothèses, il n'est pas envisageable, pour le moment, d'implanter cet algorithme dans un appareil auditif portable. Aussi, l'autre partie du travail qui reste à mener concerne la définition du banc de filtres en échelle ERB. Il faut parvenir à une solution à décimation maximale qui soit de faible complexité et qui conserve un délai de traitement court.

Dans ce domaine, les perspectives demeurent largement ouvertes et il convient de ne pas négliger les études sur l'analyse par ondelettes même si pour l'instant peu d'études sur des applications à des signaux audio se basant sur une analyse en ondelettes se soient montrées particulièrement convaincantes. Néanmoins, certains travaux récents comme [16] proposent des solutions de débruitage par paquet d'ondelettes qui semblent intéressantes mais qui demeurent encore coûteuses en temps de calcul.

Troisième partie .
Annulation de retour acoustique

Table des Matières

Introduction de la troisième partie	127
9. Retour acoustique dans les prothèses auditives	131
9.1. Description du système acoustique	131
9.2. Description du système électrique	133
9.3. Étude de stabilité	135
9.4. Analyse des conditions de stabilité dans une prothèse auditive	140
10. Techniques de compensation du retour acoustique	143
10.1. Méthodes classiques	143
10.2. Solutions récentes	145
10.3. Annulation d'écho par filtrage adaptatif	148
10.4. Synthèse	152
11. Annulation d'écho par filtrage adaptatif	153
11.1. LMS temporel par blocs	153
11.2. LMS Rapide dans le domaine fréquentiel	155
11.3. FLMS dans le domaine de la transformée de Hartley discrète	159
12. Filtre à Délais Multiples Généralisé	163
12.1. Principes	163
12.2. Synthèse WOLA	165
12.3. Partitionnement des données	166
12.4. Implémentation du LMS	168
13. Modification du pas d'adaptation	169
13.1. Pas d'adaptation du LMS	169
13.2. Modification de la définition du pas d'adaptation	171
14. Résultats - Simulations	175
14.1. Motivations	175
14.2. Simulations Matlab	175
14.3. Implémentation en temps-réel	181

Conclusion de la troisième partie	187
Conclusion générale	191

Introduction de la troisième partie

Dans une prothèse auditive, une partie du signal sonore émis par l'écouteur peut être captée en retour au niveau du microphone réalisant ainsi un *bouclage acoustique*. Ceci induit qu'une prothèse est un système électro-acoustique fermé et, à ce titre, potentiellement instable.

En pratique l'instabilité est très fréquente et elle est à l'origine d'un phénomène courant en électro-acoustique, l'*effet Larsen*¹, qui se caractérise par un sifflement puissant émis par l'écouteur.

Utilisé par certains guitaristes, gênant pour les ingénieurs du son qui sonorisent les concerts, l'effet Larsen se révèle particulièrement contraignant dans le cadre des prothèses auditives. Il se produit, en effet, de manière très fréquente et c'est un phénomène extrêmement désagréable du fait de la proximité entre l'écouteur et le tympan.

Il est très difficile de parvenir à limiter les conditions favorisant l'instabilité de la prothèse. Le travail de l'audioprothésiste peut y contribuer, mais sa marge de manœuvre reste limitée. La seule précaution efficace consiste alors à réduire l'amplification apportée au signal ce qui a pour conséquence de limiter la réhabilitation du déficit auditif. Pour les mêmes raisons, les prothèses les plus discrètes, pour lesquelles les transducteurs sont les plus rapprochés, ne peuvent pas toujours être utilisées selon le degré de surdité. Or la discrétion de l'appareillage est un critère sur lequel les malentendants sont particulièrement attentifs.

De ce fait, le phénomène de bouclage acoustique constitue, pour le moment, le facteur limitant principal pour la réhabilitation auditive par port de prothèse auditive. En réponse à ce problème, le traitement numérique du signal peut apporter des solutions qui n'étaient jusqu'alors pas envisageables.

Cette dernière partie du rapport de thèse s'attache à présenter notre travail sur le problème du bouclage acoustique qui aboutit à la conception d'une solution originale.

Le chapitre 9 présente une étude théorique complète sur le bouclage acoustique dans les prothèses auditives. A partir du formalisme Laplacien et de l'analogie avec un sys-

1. du nom du physicien danois Søren Larsen.

tème linéaire asservi par une boucle de régulation, la modélisation discrète du système est établie. Cette modélisation permet de définir clairement les différentes conditions d'instabilité tout en précisant de manière pratique leur incidence sur le fonctionnement d'une prothèse.

Le chapitre 10 est consacré à une synthèse de l'état de l'art au sujet de la compensation du retour acoustique. Celle-ci révèle que la solution qui prévaut clairement consiste à mettre en œuvre un module d'annulation d'écho par filtrage adaptatif. Il s'agit en particulier d'utiliser un algorithme de type *gradient stochastique*. En s'appuyant sur l'étude de stabilité, la section 10.3 explicite les conditions d'application des algorithmes adaptatifs pour l'annulation du retour acoustique dans les prothèses. Il apparaît, notamment, que la situation de bouclage acoustique renforce les difficultés de convergence du processus adaptatif. Par contre, la stabilisation du système requiert une estimation précise du chemin d'écho.

Ces conclusions nous amènent à suggérer l'utilisation d'un algorithme adaptatif rapide très performant. A ce titre, nous proposons d'utiliser un algorithme développé initialement dans le cadre de l'estimation de réponses impulsionnelles de salles de grandes dimensions. Cet algorithme est présenté en détail dans le chapitre 12. Son originalité est de décomposer le filtre adaptatif initial en sous-blocs de moindre taille ce qui est tout à fait approprié dans notre contexte. Pour des raisons pratiques, nous proposons d'implémenter cette méthode non pas dans le domaine de la transformée de Fourier mais dans celui de la transformée de Hartley qui présente pour notre application quelques avantages que nous précisons au cours de ce chapitre.

En complément, nous avons développé une nouvelle méthode de gestion du pas d'adaptation de l'algorithme ce qui conduit à une amélioration des capacités de convergence. Le chapitre 13 rappelle, dans un premier temps, la définition et le rôle fondamental du pas d'adaptation dans les méthodes rapides de filtrage adaptatif. Dans un second temps, la solution originale proposée est explicitée et justifiée.

A l'issue de cette étude théorique, le chapitre 14 présente les résultats des simulations et des expériences réalisées. En particulier, nous avons mis au point une plate-forme de tests en temps-réel utilisant un prototype de prothèse qui permet de réaliser ces expériences dans des conditions réalistes. Les résultats obtenus montrent que les solutions développées dans le cadre de ce travail de thèse conviennent et permettent de réduire sensiblement les problèmes d'instabilité au sein des prothèses auditives.

Au terme de cette troisième partie, la conclusion s'attache à préciser l'apport de notre travail pour le problème du bouclage acoustique. Les perspectives d'amélioration et d'ouverture sont également suggérées.

Publication Cette partie de notre travail a donné lieu à la publication de deux articles de congrès :

- Thomas FILLON et Jacques PRADO, *Acoustic Feedback Cancellation for hearing-aids, using Multi-Delay Filter*, NORSIG-2002, 5th Nordic Signal Processing Symposium, October 4-7, 2002, On board Hurtigruten.
- Thomas FILLON et Jacques PRADO, *Annulation d'écho acoustique dans les prothèses auditives*, congrès GRETSI 03, 8 - 11septembre 2003, Paris.

9. Retour acoustique dans les prothèses auditives

9.1. Description du système acoustique

En raison du bouclage acoustique entre les transducteurs¹, la prothèse auditive peut être considérée comme un système en boucle fermée. A ce titre, le système électro-acoustique ainsi formé est donc potentiellement instable. Du point de vue théorique, cette instabilité provoque des auto-oscillations qui correspondent à ce que l'on nomme communément l'*effet Larsen*.

Supposons, pour commencer, qu'il n'y ait pas de retour acoustique entre l'écouteur et le microphone. Le système équivalent est illustré sur le schéma de la figure 9.1. La relation entre le signal source capté par le microphone, $s_0(t)$, et le signal émis par l'écouteur, $y_0(t)$, est une simple relation d'entrée/sortie en boucle ouverte.

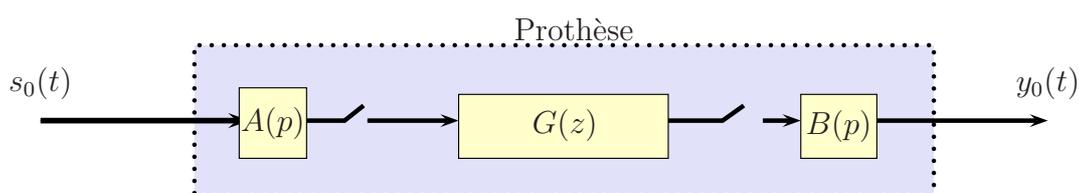


FIG. 9.1: Schéma de la prothèse auditive sans bouclage. $A(p)$ est la fonction de transfert du microphone, $B(p)$ celle de l'écouteur et $G(z)$ la fonction de transfert discrète de la puce de traitement qui opère la compensation des pertes auditives.

Nous faisons, ici, l'*hypothèse* que l'on peut modéliser la réponse de chaque module de la prothèse par une fonction de transfert Linéaire Invariante dans le Temps (LIT). Nous considérons :

- $A(p)$, la fonction de transfert du microphone,

1. En dehors du couplage acoustique, les couplages électriques, magnétiques et mécaniques interviennent également. Cependant, les progrès technologiques récents au niveau de la conception des prothèses et de leurs composants sont tels que ces autres bouclages peuvent être considérés comme négligeables par rapport au bouclage acoustique.

- $B(p)$, la fonction de transfert de l'écouteur et
- $G(p)$ qui représente l'équivalent dans le domaine continu de $G(z)$, la fonction de transfert discrète de la prothèse.

Pour simplifier l'étude, nous supposons également que l'ensemble du traitement audio est réalisé de manière numérique. Ce n'est cependant généralement pas le cas, les modules d'amplification, de compression et d'écêtage étant notamment réalisés dans le domaine analogique pour éviter les problèmes de dynamique des signaux numériques et d'aliasing dus aux effets des non-linéarités. Cette hypothèse n'est pas gênante dans le cadre de cette étude théorique puisque nous considérons que les conversions analogique-numérique et numérique-analogique sont réalisées de manière idéale.

Soit $\mathcal{T}_0(p)$, la fonction de transfert du système, que l'on peut caractériser par la relation suivante :

$$\mathcal{T}_0(p) = \frac{Y_0(p)}{S_0(p)} = A(p).G(p).B(p) \quad (9.1)$$

Remarque

Dans le cadre de cette partie théorique et afin d'explicitier clairement les causes du phénomène de Larsen, nous supposons que toutes les fonctions de transfert décrites sont de type Linéaire Invariant dans le Temps (LIT). Une fois de plus, ce n'est bien sûr pas le cas puisque d'une part le traitement au sein de la prothèse est non-linéaire² et fonction du signal d'entrée et puisque d'autre part le chemin d'écho est variable dans le temps. Cette simplification, comme les précédentes, est néanmoins nécessaire pour appréhender les causes du phénomène de Larsen.

La fonction de transfert de l'équation (9.1) est la fonction de transfert *idéale* dans le sens où la prothèse réalise exactement le traitement souhaité pour compenser les pertes auditives. En réalité, le système ne peut pas être considéré comme ouvert en raison de l'existence d'un chemin de retour acoustique entre les transducteurs par vibrations mécaniques internes ou par retour aérien. Le signal source $s_0(t)$ est ainsi corrompu par un signal d'écho additif $u_0(t)$. Ce signal d'écho est issu de l'écouteur et transmis à travers le canal acoustique entre les transducteurs comme cela est représenté sur la figure 9.2. Le signal capté par le microphone en entrée de la prothèse n'est donc plus $s_0(t)$ mais $x_0(t)$ tel que $x_0(t) = s_0(t) + u_0(t)$.

$C(p)$ étant la fonction de transfert du canal acoustique du chemin d'écho, on en déduit les relations suivantes entre les différents transformées de Laplace des signaux $s_0(t)$, $x_0(t)$, $u_0(t)$ et $y_0(t)$:

$$\begin{cases} X_0(p) = S_0(p) + U_0(p) \\ Y_0(p) = A(p)G(p)B(p)X_0(p) \\ U_0(p) = C(p)Y_0(p) \end{cases} \quad (9.2)$$

2. Sans oublier les non-linéarités des transducteurs, notamment de l'écouteur.

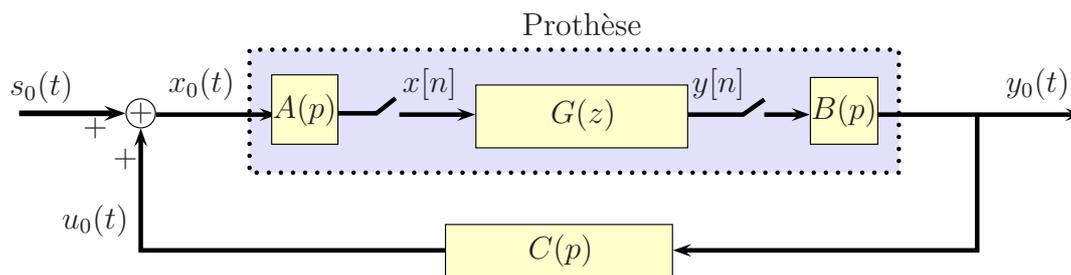


FIG. 9.2: Schéma de la prothèse auditive avec bouclage.

La fonction de transfert globale $\mathcal{T}_0(p)$ du système peut donc être modélisée par la relation suivante :

$$\mathcal{T}_0(p) = \frac{Y_0(p)}{S_0(p)} = \frac{A(p)G(p)B(p)}{1 - A(p)G(p)B(p)C(p)} \quad (9.3)$$

Il s'agit de l'équation d'un système en *boucle fermée*. Sous certaines conditions, ce système peut donc devenir instable et causer le phénomène de Larsen.

Remarque

L'analogie entre le système considéré ici et un système asservi par une boucle de régulation est évidente. Par conséquent, nous utiliserons, pour cette étude, des outils d'automatique des systèmes asservis et le vocabulaire correspondant. Il convient cependant de souligner que nous sommes en présence ici d'une boucle de rétroaction positive, et non négative comme c'est le cas en automatique. En effet le signal de sortie s'ajoute au signal source à l'entrée du système.

9.2. Description du système électrique

Afin de simplifier les notations et afin de se concentrer sur les données observables, nous allons considérer un système équivalent au système décrit précédemment du point de vue de l'étude de stabilité. Nous nous plaçons désormais au niveau des signaux électriques.

Soit $x(t)$ le signal électrique issu du microphone, que l'on sépare en une partie issue du signal acoustique source $s_0(t)$ et une partie issue du signal d'écho $u_0(t)$.

En effet, en développant l'expression de $X(p)$:

$$\begin{aligned}
X(p) &= X_0(p)A(p) \\
&= \left(S_0(p) + U_0(p) \right) A(p) \\
&= S_0(p)A(p) + U_0(p)A(p)
\end{aligned}$$

et en posant :

$$\begin{cases} S(p) = S_0(p)A(p) \\ U(p) = U_0(p)A(p) \end{cases}$$

on obtient :

$$X(p) = S(p) + U(p)$$

On définit alors les signaux $s(t)$ et $u(t)$ de transformées de Laplace respectives $S(p)$ et $U(p)$. Ainsi, $x(t)$, le signal transmis par le microphone, se décompose en somme de $s(t)$ et $u(t)$, signaux correspondant respectivement aux composantes source et écho en sortie du microphone.

Par analogie avec un système linéaire asservi, on peut décrire les signaux caractéristiques du système considéré comme étant :

- $s(t)$: la *consigne*,
- $x(t)$: l'*écart* et
- $y(t)$: la *sortie*.

La figure 9.3 indique le schéma correspondant à ce système.

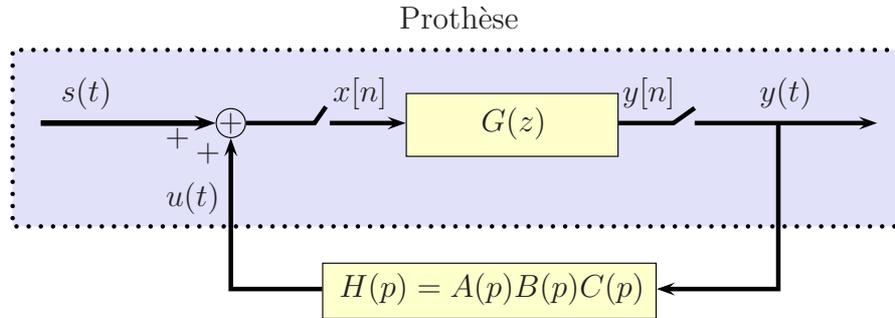


FIG. 9.3: Schéma équivalent de la prothèse auditive avec bouclage.

Les relations suivantes régissent alors les différents signaux :

$$\begin{cases} X(p) = S(p) + U(p) \\ Y(p) = G(p)X(p) \\ U(p) = A(p)C(p)B(p)Y(p) \end{cases} \quad (9.4)$$

On définit $H(p)$ telle que :

$$H(p) = A(p)C(p)B(p) \quad (9.5)$$

$H(p)$ est donc la fonction de transfert du chemin d'écho *global* intégrant la réponse de l'écouteur et du microphone.

On obtient finalement pour le système considéré sur la figure 9.3 la fonction de transfert, $\mathcal{T}(p)$, telle que :

$$\mathcal{T}(p) = \frac{Y(p)}{S(p)} = \frac{G(p)}{1 - G(p)H(p)} \quad (9.6)$$

L'avantage de considérer ce système plutôt que le précédent est que sa fonction de transfert met directement en évidence, au numérateur, la fonction de transfert $G(p)$ de la prothèse et que l'on s'affranchit par la même occasion de l'identification des fonctions de transfert du microphone et de l'écouteur. D'autre part seuls les signaux électriques sont observables.

On peut noter que si l'on suppose que le système est correctement calibré³, la valeur de $|A(p)B(p)|$ est *de l'ordre de 1* pour tout $p = j\omega$. Ceci implique, en particulier, que le chemin d'écho acoustique $C(p)$ est le facteur principal dans la définition du chemin d'écho global $H(p)$.

9.3. Étude de stabilité

9.3.1. Étude de stabilité à temps continu

Dans la suite de ce document, par analogie avec un système asservi, nous noterons $\mathcal{T}_{BO}(p)$ la fonction de transfert *en boucle ouverte* :

$$\mathcal{T}_{BO}(p) = -G(p)H(p) \quad (9.7)$$

L'équation (9.6) définissant $\mathcal{T}(p)$ se récrit alors :

$$\mathcal{T}(p) = \frac{Y(p)}{S(p)} = \frac{G(p)}{1 + \mathcal{T}_{BO}(p)} \quad (9.8)$$

Remarque

Le signe « - » qui apparaît dans la définition de la fonction de transfert en boucle ouverte, $\mathcal{T}_{BO}(p)$, est introduit afin d'avoir une écriture compatible avec celle d'un système asservi. Nous sommes cependant bien en présence d'une boucle de rétroaction positive (cf. remarque précédente page 133).

3. Nous considérons, ici, qu'une calibration correcte de la prothèse est telle que si le traitement de l'appareil est un simple gain unitaire $G(p) = 1$ alors la fonction de transfert globale $A(p)G(p)B(p)$ doit être très proche d'un gain unitaire. Dans la pratique, une telle calibration n'est cependant pas évidente comme nous le verrons dans l'annexe B.

D'un point de vue théorique, le système sera stable au sens Entrée Bornée - Sortie Bornée (EBSB) si la fonction de transfert $\mathcal{T}(p)$ n'a pas de pôle instable, c'est-à-dire si $1 + \mathcal{T}_{BO}(p)$ n'a pas de racines à droite de l'axe imaginaire.

Le critère géométrique de Nyquist [33] permet de définir une règle renseignant sur la stabilité du système :

Critère géométrique de Nyquist :

Le système sera stable si, dans le diagramme de Nyquist, $\mathcal{T}_{BO}(p)$ tourne autour du point critique $(-1; 0)$ dans le sens direct P fois où P est le nombre de pôles instables de $\mathcal{T}_{BO}(p)$.

Dans le cas qui nous concerne, il apparaît clairement que $\mathcal{T}_{BO}(p)$ est stable comme mise en série de systèmes stables. Le nombre P de pôles instables du critère géométrique de Nyquist est donc nul. Grâce à cette propriété, on peut donc considérer un critère de stabilité plus simple [33], le critère du revers :

Critère du revers :

Si la fonction de transfert en boucle ouverte $\mathcal{T}_{BO}(p)$ d'un système asservi ne possède aucun pôle à partie réelle positive, alors ce système est stable en boucle fermée si, en parcourant le lieu de Nyquist de la fonction de transfert en boucle ouverte dans le sens des ω croissants, on laisse toujours le point critique à gauche de la courbe.

Le critère du revers peut être illustré de manière synthétique à partir du tracé du lieu de Nyquist ou hodographe comme cela est schématisé dans la figure 9.4.

9.3.2. Étude de stabilité à temps discret

Conditions d'instabilité

Les solutions que nous allons employer pour stabiliser le système seront implémentées dans le domaine discret. Dans cette section, en nous basant sur le formalisme Laplacien décrit précédemment, nous allons donc exprimer d'une part le modèle discret équivalent au système étudié et d'autre part ses conditions de stabilité.

A cet effet, nous supposons que les convertisseurs Analogiques-Numériques (CAN) et Numériques-Analogiques (CNA) sont idéaux.

On définit les versions échantillonnées respectives $s[n]$, $x[n]$, $u[n]$ et $y[n]$ des signaux $s(t)$, $x(t)$, $u(t)$ et $y(t)$. De même, on définit $H(z)$ la fonction de transfert discrète du

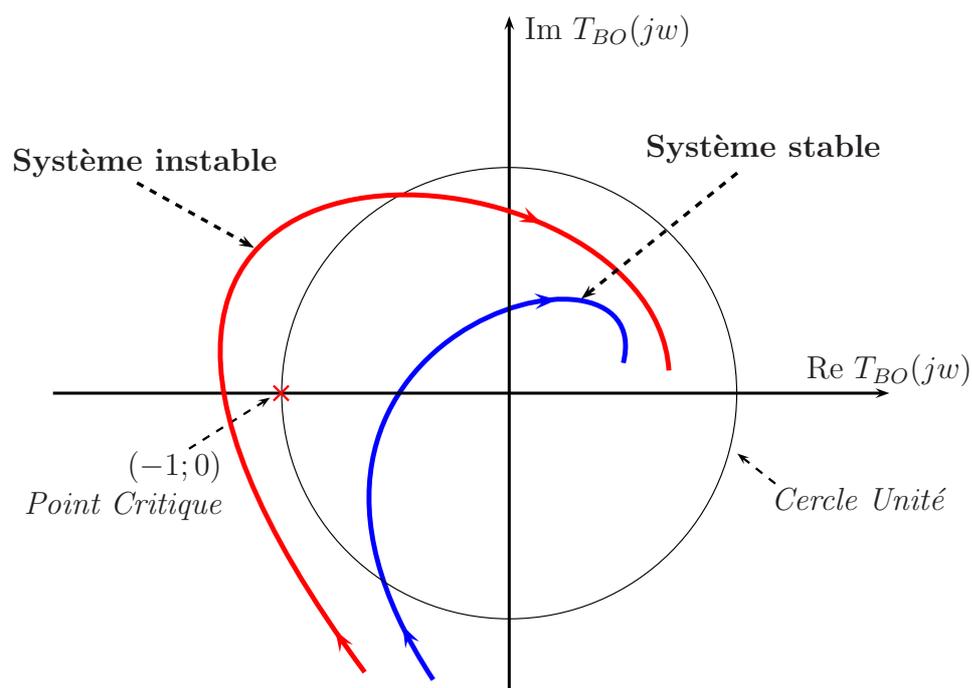


FIG. 9.4: Illustration du critère de revers. La figure présente les hodographes de deux systèmes l'un stable, l'autre instable.

chemin d'écho, équivalente au modèle continu $H(p)$ défini par la formule (9.5).

Le modèle discret équivalent de la prothèse auditive avec bouclage est représenté sur la figure 9.5.

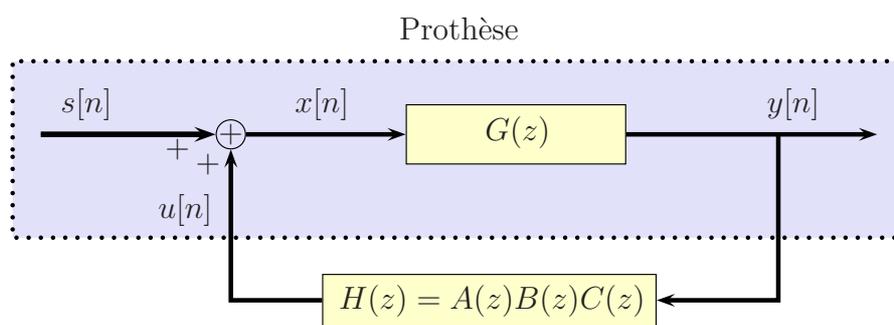


FIG. 9.5: Modèle discret équivalent de la prothèse auditive avec bouclage.

Pour ce système et d'après l'expression de $\mathcal{T}(p)$ définie dans (9.6), la fonction de transfert discrète correspondante, $\mathcal{T}(z)$, s'écrit alors :

$$\mathcal{T}(z) = \frac{Y(z)}{S(z)} = \frac{G(z)}{1 - G(z)H(z)} \quad (9.9)$$

Il apparaît alors que le système sera instable sous les conditions du *critère de Nyquist* suivant :

Conditions d'instabilité de Nyquist :

Le système est instable, s'il existe au moins une valeur z_0 sur le cercle unité pour laquelle :

$$\begin{cases} |G(z_0).H(z_0)| > 1, \text{ et} \\ \arg(G(z_0).H(z_0)) = 2m\pi \text{ rad}, m \in \mathbb{N}. \end{cases} \quad (9.10)$$

Si une telle valeur $z_0 = e^{j\omega_0}$ existe, il y a des oscillations entretenues à la pulsation ω_0 . Dans un système électroacoustique comme celui considéré ici, cela entraîne l'émission d'un fort sifflement tonal caractéristique de l'effet Larsen⁴.

Marges de robustesse

Nous rappelons, ici, les différents outils qui permettent de caractériser la *robustesse* du système vis-à-vis de la stabilité. Pour plus de précisions, on peut se reporter à [47] et [7].

- la *marge de gain*,
- la *marge de phase*,
- la *marge de retard* et
- la *marge de module*.

Marge de gain La marge de gain, ΔG , indique la valeur maximale de gain que l'on peut introduire dans la boucle avant d'atteindre les conditions limites de stabilité.

En d'autres termes, c'est l'inverse du module de $\mathcal{T}_{BO}(e^{-j\omega})$ pour la pulsation ω à laquelle la phase est de -180° . Si plusieurs pulsations satisfont cette relation de phase, alors ΔG est déterminée à la pulsation qui minimise sa valeur.

A titre d'exemple, la marge de gain est indiquée sur l'hodographe de la figure 9.6.

Marge de phase De même que pour la marge de gain, la marge de phase, $\Delta\Phi$, indique le déphasage additionnel maximal que l'on peut introduire dans le système avant d'atteindre les conditions limites d'instabilité.

La marge de phase est indiquée sur l'hodographe de la figure 9.6.

4. Les non-linéarités du système d'émission, en particulier la saturation et l'écrêtage, viennent limiter le niveau du sifflement de sortie

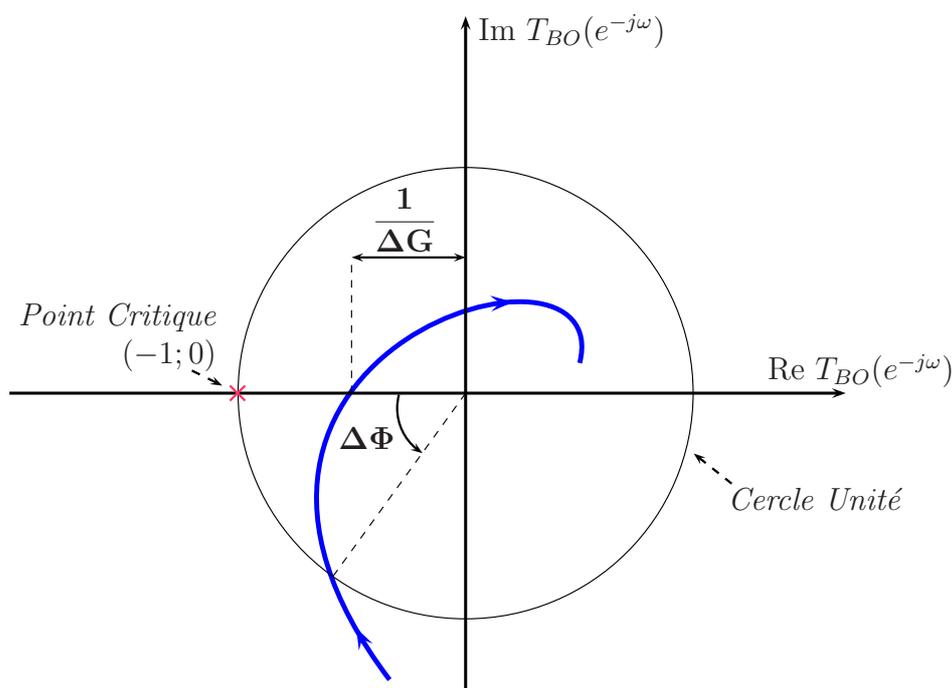


FIG. 9.6: Hodographe indiquant les marges de gain et de phase, ΔG et $\Delta\Phi$

Marge de retard À partir de la marge de phase, on définit la marge de retard, $\Delta\tau$, qui indique le retard minimal que l'on peut introduire avant d'atteindre les conditions limites de stabilité⁵.

$$\Delta\tau = \frac{\Delta\Phi}{\omega_{cr}}$$

où ω_{cr} est la pulsation à laquelle la marge de phase est définie.

Marge de module La marge de module ou marge *gain-phase*, ΔM , est la distance minimale en dB entre l'hodographe de T_{BO} et le point critique $(-1; 0)$. Elle est représentée sur l'hodographe de la figure 9.7

Condition suffisante de stabilité

Dans un cas pratique comme le nôtre, les conditions de Nyquist ne sont pas aisément utilisables. En effet, d'une part la fonction de transfert du chemin d'écho, $H(z)$, est variable dans le temps et d'autre part, le système est non-linéaire ; il est donc difficile de garantir la stabilité à partir de telles conditions.

Il paraît plus pratique, dans notre cas, de se baser sur une *condition suffisante de stabilité* déduite des conditions de Nyquist, énoncé comme suit :

5. La marge de phase est une notion utile vis à vis du retard introduit par le traitement par blocs

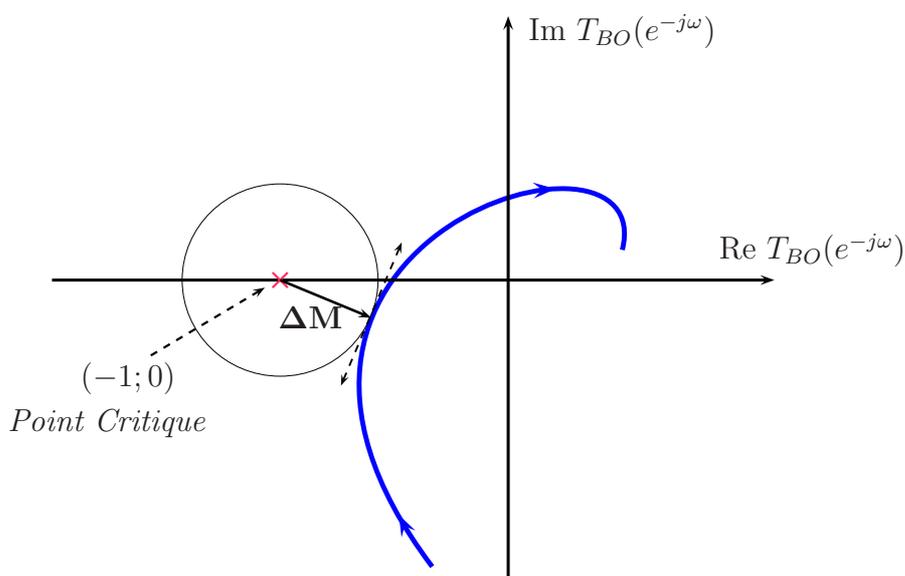


FIG. 9.7: Hodographe indiquant la marge de module ΔM

Condition suffisante de stabilité :

$$|G(z).H(z)| < 1, \forall z \in \mathcal{C}; \text{ où } \mathcal{C} \text{ est le cercle unité.} \quad (9.11)$$

Cette dernière condition est très précieuse pour la compréhension des différents facteurs qui ont un impact sur la stabilité du système et sur le comportement d'une prothèse auditive. Dans la suite de ce document, nous y aurons régulièrement recours.

9.4. Analyse des conditions de stabilité dans une prothèse auditive

9.4.1. Facteurs pénalisant la stabilité dans une prothèse auditive

Comme on vient de le voir, une prothèse auditive est un système électroacoustique potentiellement instable. La condition suffisante de stabilité (9.11) indique clairement le lien entre l'instabilité et les modules respectifs des fonctions de transfert de la prothèse et du canal acoustique entre les transducteurs, $|G(z)|$ et $|H(z)|$.

Dans une prothèse auditive, plusieurs facteurs sont défavorables à la condition suffisante de stabilité compte tenu de leur impact sur la valeur de $|G(z)|$ ou sur celle de $|H(z)|$.

Dans le but de compenser les pertes auditives d'un malentendant, la mission première

d'une prothèse auditive est d'amplifier le signal. Cela induit nécessairement de fortes valeurs pour le facteur $|G(z)|$ car, comme nous l'avons précisé dans la première partie de ce document, les gains d'amplifications peuvent atteindre des valeurs proches de 75dB pour les surdités sévères (*cf.* la section 4.2, page 52).

Parmi ces facteurs, la forte amplification appliquée au signal pour compenser les pertes auditives induit une grande valeur de $|G(z)|$. De même, pour le chemin d'écho, $|H(z)|$ est d'autant plus important du fait :

- de la proximité entre les transducteurs électroacoustiques,
- d'une éventuelle mauvaise étanchéité de l'embout et
- de la présence d'événements plus ou moins larges pour éviter l'effet d'occlusion.

Sans solutions adéquates, le seul moyen de se prévenir de l'effet Larsen serait donc de réduire le gain d'amplification de la prothèse.

Or, si le gain est réduit en-dessous du seuil nécessaire pour compenser la hausse des seuils d'audibilité du malentendant, la prothèse ne remplit plus intégralement son rôle de réhabilitation.

9.4.2. Variabilité des conditions

Dans l'étude théorique que nous avons présentée précédemment, nous avons supposé que le système pouvait être modélisé à partir de fonctions de transfert *Linéaires et Invariantes dans le Temps*. S'il en était ainsi, il serait relativement aisé de stabiliser le système. Cependant, nous sommes, en réalité, en présence d'un système non-linéaire et variant dans le temps.

En effet, d'une part les équipements électro-acoustiques employés ont un caractère non-linéaire (transducteurs ou amplificateur) et d'autre part le traitement réalisé dans la prothèse est hautement non-linéaire puisqu'il s'agit principalement d'un module d'amplification et de compression. La valeur de $|G(z)|$ est donc variable en fonction de la puissance du signal d'entrée.

En outre, l'environnement acoustique est fortement variable et ses caractéristiques évoluent suite à l'approche d'objets près de la prothèse, par exemple, ou suite à des mouvements de mâchoire. Comme le montrent plusieurs études sur le sujet [36, 73], les variations du chemin de retour acoustique sont loin d'être négligeables. Dans [73] les auteurs indiquent, en effet, qu'une hausse de 20dB de la fonction de transfert en boucle ouverte est mesurable lorsqu'on approche un combiné téléphonique de la prothèse.

Pour empêcher le phénomène de Larsen, il devient alors nécessaire de faire fonctionner le système avec des marges de robustesse assez larges.

Le gain d'amplification de la prothèse doit donc être limité fortement. Ainsi, en plus

de la gêne occasionnelle due aux auto-oscillations, *le problème d'écho constitue l'un des facteurs majeurs limitant la réhabilitation auditive* par port de prothèse puisqu'il restreint le gain maximum apporté par l'appareil.

9.4.3. Distorsions

L'équation (9.10) du système nous éclaire sur un autre problème causé par le bouclage acoustique dans une prothèse auditive. En effet, même si cette condition suffisante de stabilité est satisfaite, mais que $|G(z).H(z)|$ n'est pas négligeable, alors la fonction de transfert du système n'est pas la fonction de transfert souhaitée $G(z)$. Le bouclage acoustique peut donc également engendrer des distorsions du signal par rapport au signal traité idéalement.

Par exemple, si pour une pulsation ω_0 particulière, $|G(z_0).H(z_0)|_{z_0=e^{j\omega_0}}$ est proche de 1 alors la fonction de transfert possède une résonance forte à la pulsation ω_0 uniquement due au bouclage.

Ainsi, un traitement efficace du problème de bouclage doit non seulement parvenir à respecter la condition suffisante de stabilité (9.11) mais également permettre d'atteindre une *condition de moindre distorsion* :

$$|G(z).H(z)| \ll 1, \forall z \in \mathcal{C} \quad (9.12)$$

10. Techniques de compensation du retour acoustique

10.1. Méthodes classiques

Afin de favoriser la condition suffisante de stabilité (9.11), deux stratégies principales sont envisageables pour la réduction du phénomène de bouclage. La première consiste à agir sur le chemin d'écho afin de réduire $|H(z)|$, la seconde contraint le gain de la prothèse afin de réduire $|G(z)|$.

10.1.1. Contrainte sur le chemin d'écho

Le premier type de solution repose en fait sur le travail de l'audioprothésiste. Sur ce point, son travail consiste à assurer une bonne *étanchéité acoustique* par une réalisation précise de l'embout auriculaire (pour une prothèse de type contour d'oreille) ou de la coque (pour une prothèse intra-auriculaire) par rapport au conduit auditif du malentendant.

Cette étanchéité acoustique est toutefois limitée. En effet, il convient de préserver le confort du malentendant, notamment en empêchant l'effet d'*occlusion*. Si le conduit auditif est bouché, il y a alors un phénomène d'amplification et de résonance des sons internes tels que la voix ou les bruits de mastication. C'est ce phénomène que l'on nomme l'effet d'occlusion. Il est très gênant et les malentendants s'en plaignent fréquemment. Aussi, une bonne aération du conduit auditif est nécessaire. L'audioprothésiste¹ réalise donc un ou des événements (ou percages) de diamètre adapté pour permettre à l'air de circuler.

En trouvant un compromis satisfaisant, l'audioprothésiste parvient à réaliser une importante réduction du retour acoustique entre transducteurs.

1. Pour plus de précisions sur le travail de l'audioprothésiste, on peut se reporter à [15]

10.1.2. Contraintes sur le gain

Dans certains systèmes analogiques ou dans certains systèmes numériques de première génération, un filtre réjecteur permet de réduire le phénomène de Larsen pour une fréquence fixe comme le représente la figure 10.1. Le réglage de la fréquence rejetée est effectuée par l'audioprothésiste lors de la mise en place de l'appareil. En pratique, ce réglage se fait selon les étapes suivantes :

- l'embout auriculaire est introduit dans le conduit auditif du patient,
- un phénomène de Larsen est alors généralement constaté,
- l'audioprothésiste ajuste ensuite, pas à pas, la fréquence centrale de la bande rejetée jusqu'à la disparition du sifflement.

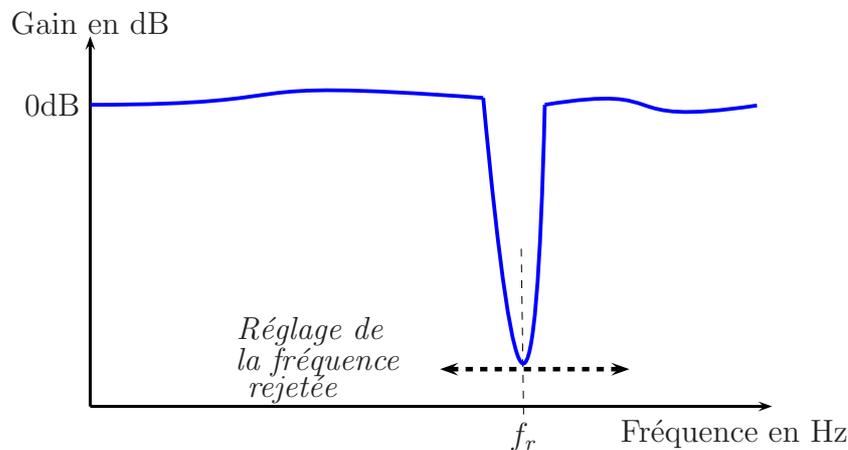


FIG. 10.1: Gabarit d'un filtre réjecteur réglable. La fréquence centrale de la bande rejetée, f_r , est ajustée sur la fréquence d'accrochage lors de la mise en place de l'appareil.

Cette solution bien qu'améliorant les performances de la prothèse possède deux inconvénients. En premier lieu, elle offre très peu de précision dans la correction apportée (aussi bien sur la fréquence centrale que sur la largeur de la bande rejetée). Ainsi, *la réhabilitation auditive dans la bande de fréquence concernée est très fortement compromise*. Ceci est particulièrement gênant car les problèmes de Larsen interviennent souvent dans les fréquences hautes qui sont, par ailleurs, souvent touchées par le déficit auditif. Le deuxième point est que ce type de solution ne permet de satisfaire les conditions de stabilité que dans une seule configuration puisque la variabilité du chemin d'écho n'est pas prise en compte. Elle se révèle inefficace et inadaptée dès lors que la fréquence d'accrochage varie.

Woodruff et al. ont proposé l'utilisation d'un filtre fixe dans [78]. Une telle solution permet plus de précision dans la correction du chemin d'écho mais reste insuffisamment efficace à cause de la variabilité du chemin d'écho.

A la lumière de ces exemples, il apparaît nécessaire de fournir une solution qui parvienne à s'adapter aux variations du chemin d'écho. C'est donc principalement cet aspect que les techniques présentées récemment proposent de résoudre.

10.2. Solutions récentes

10.2.1. Description des méthodes

Récemment, d'autres approches ont été envisagées afin de pallier le problème d'instabilité des prothèses. Dans le cadre de ce document, nous ne détaillerons pas toutes les solutions, en voici néanmoins une liste non exhaustive.

On distingue notamment :

- ▷ les solutions qui empêchent l'accrochage par rétroaction :
 - variation dans le temps du délai et de la phase
 - re-allocation en fréquence
- ▷ les versions adaptatives des solutions classiques :
 - filtre réjecteur adaptatif
 - filtrage inverse adaptatif
- ▷ enfin, la dernière famille de solutions qui tend à annuler le signal d'écho au travers d'une estimation adaptative du chemin d'écho.

Pour avoir plus de détails sur ces méthodes, on se reportera à [10] et [43].

10.2.2. Prévention de l'accrochage

Les solutions qui visent à empêcher l'accrochage ne sont pas suffisamment robustes étant données les conditions difficiles imposées par la configuration du système en boucle fermée. La re-allocation de fréquence consiste à déplacer vers une autre zone du spectre les signaux initialement présents autour de la fréquence d'accrochage initiale. Il s'agit en général de « compresser » le spectre du signal vers les basses fréquences. Cette technique peut permettre de réduire efficacement le problème de Larsen. Elle ne garantit cependant pas la stabilité si la fréquence d'accrochage se déplace vers les basses fréquences. D'autre part, elle modifie complètement la perception auditive du malentendant.

10.2.3. Compensation de l'écho

Les méthodes qui emploient un filtre inverse ou un filtre réjecteur (adaptatif ou fixe) ne sont pas optimales du point de vue de la réhabilitation auditive. Leur principe repose en effet sur l'introduction d'un filtre en sortie de prothèse. Considérons $H_c(z)$ la fonction de transfert d'un tel filtre. La réponse du système, $\mathcal{T}(z)$, devient alors :

$$\mathcal{T}(z) = \frac{G(z)H_c(z)}{1 - G(z)H(z)H_c(z)} \quad (10.1)$$

Le but visé est, ici, de stabiliser le système, c'est-à-dire de construire $H_c(z)$ tel que la stabilité soit assurée. Si l'on s'en tient aux conditions de stabilité de Nyquist, le système est peu robuste. Si l'on considère la condition suffisante de stabilité alors le filtre $H_c(z)$ doit garantir la condition suivante :

$$|H_c(z)| < \frac{1}{|G(z).H(z)|}, \forall z \in \mathcal{C}; \text{ où } \mathcal{C} \text{ est le cercle unité.} \quad (10.2)$$

Ceci implique principalement que $H_c(z)$ atténue les fréquences pour lesquelles la fonction de transfert en boucle ouverte $G(z).H(z)$ est de module élevé. Le terme dominant de $\mathcal{T}(z)$ est donc $G(z)H_c(z)$ qui, contrairement à $G(z)$, n'est pas optimal au sens de la réhabilitation auditive puisque certaines fréquences seront atténuées pour assurer la stabilité.

10.2.4. Annulation de l'écho

Les solutions au problème de retour acoustique présentées précédemment ne permettent pas de garantir une compensation optimale des pertes auditives. Pour remédier à cela, les travaux les plus récents dans le domaine proposent d'utiliser des méthodes d'annulation d'écho par filtrage adaptatif ([44, 45, 51]) comme dans les applications d'annulation d'écho acoustique classiques.

En effet, la mise en place d'un annuleur d'écho par filtrage adaptatif permet, en théorie, d'obtenir une solution optimale du point de vue de la réhabilitation.

L'introduction d'un filtre adaptatif modifie le système en boucle fermée que constitue la prothèse auditive. La figure 10.2 représente le système ainsi réalisé. Le filtre adaptatif estime alors le chemin d'écho et soustrait le signal d'écho estimé au signal entrant compensant ainsi le signal d'écho réel.

Considérons $\widehat{H}(z)$ la fonction de transfert du filtre adaptatif conçue pour estimer la fonction de transfert du chemin d'écho $H(z)$. La fonction de transfert globale du système, $\mathcal{T}(z)$, devient alors :

$$\mathcal{T}(z) = \frac{G(z)}{1 - G(z).(\widehat{H}(z) - H(z))} \quad (10.3)$$

Soit en notant $\Delta H(z) = H(z) - \widehat{H}(z)$:

$$\mathcal{T}(z) = \frac{G(z)}{1 - G(z).\Delta H(z)} \quad (10.4)$$

A partir de la nouvelle expression de $\mathcal{T}(z)$ et des conditions de Nyquist de l'équation (9.10), on en conclut qu'en présence du filtre adaptatif la *condition suffisante* de

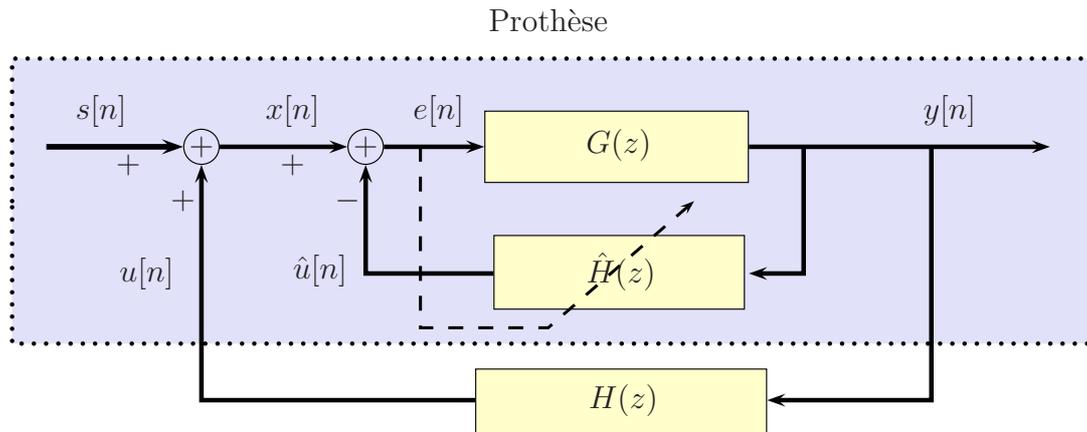


FIG. 10.2: Schéma de la prothèse auditive avec bouclage et annulation d'écho. $\mathbf{G}(\mathbf{z})$ représente la compensation des pertes auditives ; le filtre adaptatif $\widehat{H}(z)$ estime la fonction de transfert du chemin d'écho $H(z)$. D est le délai de traitement en nombre d'échantillons.

stabilité (9.11) devient :

$$|G(z) \cdot \Delta H(z)| < 1 \quad (10.5)$$

Cette nouvelle expression de la condition suffisante de stabilité montre l'intérêt d'utiliser une technique d'annulation d'écho dans le cadre des prothèses auditives.

En effet, plus l'estimation du chemin d'écho est précise, plus le terme $|\Delta H(z)|$ est faible et favorise ainsi la stabilité. L'autre remarque qui en découle directement est que plus $|\Delta H(z)|$ est faible, moins les valeurs de $|\Delta G(z)|$ sont contraintes pour assurer la stabilité et donc meilleure est la réhabilitation.

Si de plus, l'estimation est très fine et que $|\Delta H(z)|$ devient négligeable devant $|G(z)|^{-1}$ alors la *condition de moindre distorsion* (9.12) est satisfaite. En effet, cette condition se réécrit, maintenant, sous la forme :

$$|G(z) \cdot \Delta H(z)| \ll 1, \forall z \in \mathcal{C} \quad (10.6)$$

Si l'estimation est parfaite et que $|\Delta H(z)|$ tend vers 0, la fonction de transfert du système $\mathcal{T}(z)$ tend vers $G(z)$ qui est bien la fonction de transfert optimale du point de vue de la réhabilitation.

En contrepartie, la qualité d'estimation du filtre repose en grande partie sur les capacités de convergence et de poursuite du filtre adaptatif.

Remarque

Les développements qui précèdent montrent que l'utilisation d'une technique de filtrage adaptatif offre des possibilités de solutions qui n'étaient pas envisageables avec des méthodes classiques. Cet exemple illustre le bénéfice potentiel

que représente l'utilisation de techniques avancées de traitement numérique du signal dans le contexte de la réhabilitation auditive.

10.3. Annulation d'écho par filtrage adaptatif

10.3.1. Utilisation du gradient stochastique dans une prothèse auditive

Depuis quelques années, un consensus apparaît clairement au sein de la communauté scientifique sur la supériorité des solutions d'annulation d'écho adaptative employant les algorithmes de type gradient stochastique, Least Mean Square (LMS).

Quelle que soit l'application envisagée, le LMS présente les avantages suivants :

- sa complexité est extrêmement réduite
- le traitement des données est réalisé « au fil de l'eau » ce qui implique une adaptation à la non-stationnarité des sources et de l'environnement (i.e. capacités de poursuite). Parmi les difficultés de mise en œuvre, on note :
- le filtre converge en moyenne vers la solution idéale et oscille autour de celle-ci.
- le pas d'adaptation μ est difficile à régler, non seulement pour garantir la convergence mais aussi pour assurer un compromis satisfaisant entre la précision de l'estimation et les capacités de poursuite.

L'utilisation d'un annuleur d'écho adaptatif dans le cadre d'une prothèse auditive est réalisée comme dans le cadre classique d'annulation d'écho acoustique en salle ou en téléphonie. Cependant, dans le cas classique, le *signal d'écho* est issu d'un signal distant provenant d'un locuteur lointain comme le montre la figure (10.3). Le signal d'écho estimé est soustrait au *signal source* issu du locuteur proche. Dans le cas présent, au contraire, le *signal d'écho* est issu du *signal source* par l'intermédiaire de la prothèse comme le montre la figure (10.4).

Le bouclage acoustique nous met donc face à des conditions très particulières pour l'annulation d'écho. En effet, l'une des hypothèses majeures pour la convergence du gradient stochastique est la décorrélation des signaux source et écho. En annulation d'écho acoustique classique, cette hypothèse est généralement fausse même si les signaux mis en jeu sont des signaux de parole issus de locuteurs différents. Dans notre cas, cette hypothèse est encore moins réaliste car le signal d'écho est lui-même issu du signal source. La corrélation est donc importante.

En annulation d'écho classique, pour pallier ce problème, on a couramment recours à un détecteur de double parole afin de mettre à jour les coefficients du filtre uniquement lorsque seul le signal d'écho est présent. Ici, le recours à un tel système serait à la fois coûteux du point de vue calculatoire et difficile à mettre en œuvre. En effet, il n'y a, à

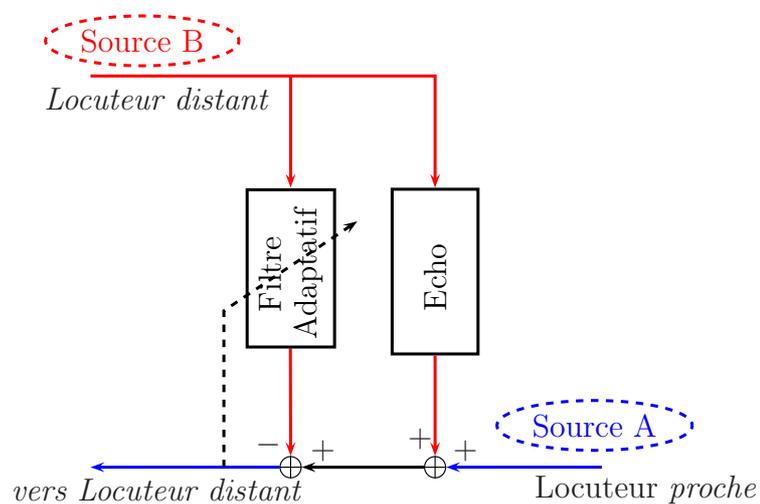


FIG. 10.3: Annulation d'écho acoustique classique

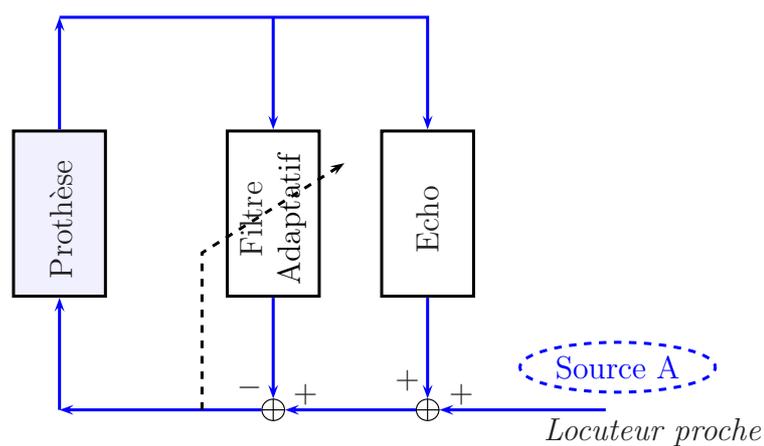


FIG. 10.4: Annulation d'écho acoustique pour une prothèse

proprement parler, de signal d'écho que s'il y a un signal source. La situation de double parole est donc quasi permanente.

Afin de réduire la corrélation des signaux, une solution couramment employée est d'ajouter un délai en sortie en amont de l'annuleur d'écho. Nous reviendrons plus tard sur ce point dans la section 10.3.4.

Si les différents travaux entrepris sur le sujet semblent tous converger vers l'utilisation de techniques d'annulation d'écho en raison de leurs performances, il n'en reste pas moins que ces algorithmes sont confrontés à des conditions particulièrement délicates de mise en œuvre.

10.3.2. Continuité de l'adaptation

En ce qui concerne la mise en œuvre du filtre adaptatif deux possibilités ont été envisagées :

- l'estimation discontinue : l'adaptation se fait uniquement pendant les moments de silence ou lorsqu'une instabilité est détectée.
- l'estimation continue : la mise à jour du filtre est effectuée à chaque trame.

Le principe de l'estimation discontinue est de profiter des moments de silence pour couper la transmission du signal reçu par le microphone et émettre une séquence de bruit aléatoire en sortie de la prothèse. Ainsi, l'entrée et la sortie étant découplées, les conditions sont conformes aux hypothèses de convergences de l'algorithme. Ceci permet alors une identification précise du chemin d'écho à un instant donné.

Cette adaptation n'étant pas continue, les variations du chemin d'écho en dehors des périodes d'identification ne sont pas prises en compte par le système. Dès lors, le système peut devenir instable si les conditions ont évolué depuis la dernière mesure. Il est donc parfois nécessaire de couper le système lorsque une instabilité est détectée et de relancer l'adaptation du filtre en injectant à nouveau une séquence de bruit. L'utilisation de ce type de méthode est justifiée par les difficultés de convergence du gradient stochastique dans ces conditions particulières d'utilisation.

Ce type de méthode présente donc les inconvénients de ne pas s'adapter continûment et d'être relativement gênant pour le malentendant du fait de l'interruption de la transmission du son et de l'émission de bruit aléatoire aussi souvent que nécessaire.

Les capacités de convergence et de poursuite des algorithmes de gradient ayant été améliorées, l'estimation continue est donc jugée préférable.

10.3.3. Émission d'une séquence de bruit

Dans certaines approches, on injecte en continu dans le signal émis un bruit à faible niveau pour améliorer la décorrélation des signaux source et écho (voir [45] et références internes).

Cette approche permet d'avoir un certain contrôle des qualités du signal excitateur vis-à-vis du processus adaptatif.

Cependant, une telle démarche compromet le rapport signal à bruit en sortie de prothèse et ce genre de solution est donc de plus en plus écartée.

Toutefois, dans quelques années, l'évolution des technologies pourrait permettre de disposer de capacités de calcul avancées. Il serait alors éventuellement possible d'utiliser les connaissances sur le phénomène de masquage pour émettre ce type de séquence de bruit sous le seuil de masquage du signal utile. En particulier dans notre cas, une grande partie du champ acoustique est disponible car elle se trouve sous le seuil audible du fait du déficit auditif.

Domaine d'implémentation du LMS

Les articles les plus récents sur le sujet emploient des algorithmes de type LMS dans le domaine fréquentiel ([68, 4, 79]).

En dehors de leur efficacité propre dans le contexte de l'annulation d'écho (complexité plus faible, rapidité, meilleure capacité de convergence), l'avantage de tels algorithmes est d'opérer dans le domaine fréquentiel comme la plupart des autres traitements de la prothèse (débruitage, compensation du recrutement de la sensation d'intensité, rehaussement spectral).

C'est vers ce type d'algorithme, ou plus généralement vers les algorithmes dans un domaine transformé que nous avons opté comme cela sera détaillé ultérieurement dans ce document.

10.3.4. Le délai

Afin de réduire la corrélation entre le signal en sortie de prothèse et le signal source, il est pertinent d'introduire un délai temporel avant la restitution du signal [69]. Ceci permet alors de réduire le biais dans l'identification du chemin d'écho. Cependant la longueur de ce délai ne doit pas excéder une certaine limite entre 10 et 20 ms afin de ne pas être gênant du point de vue perceptif.

D'autre part, on peut noter que la présence d'un délai entre l'acquisition et la restitution du signal sonore est inhérent aux techniques de traitement par blocs qui sont utilisées en filtrage adaptatif. Il est donc opportun de choisir la taille des blocs en fonction du délai de calcul et du délai maximum.

10.4. Synthèse

Les différentes sources bibliographiques étudiées montrent que le problème de retour acoustique dans les prothèses auditives est délicat. Dans ce contexte difficile, la supériorité des méthodes employant un module d'annulation d'écho par filtrage adaptatif est clairement établie. En outre, une telle approche peut permettre, en théorie, de libérer les contraintes sur le module de réhabilitation auditive.

Dans la suite de ce document, nous allons donc proposer une solution employant un algorithme de type LMS. Comme le cadre d'application est très exigeant, il nous semble nécessaire d'envisager une variante du LMS dans un domaine transformé (en restant dans un contexte proche du domaine fréquentiel). Les avantages d'une telle solution sont

1. son moindre coût calculatoire,
2. sa supériorité par rapport aux implémentations temporelles et
3. sa compatibilité éventuelle avec d'autres modules de traitement de la prothèse.

Dans le chapitre suivant nous présentons, dans un premier temps, les algorithmes rapides dérivés du LMS temporel. Puis les chapitres 12 et 13 sont consacrés plus spécifiquement à la solution à laquelle nous proposons de recourir.

En ce qui concerne le contexte de mise en oeuvre du LMS, il nous semble plus approprié de se tourner vers une solution réalisant une estimation continue du chemin d'écho sans injection de bruit dans le signal émis. Le fait de proposer une solution qui ne nuit pas au confort auditif du malentendant paraît fortement souhaitable étant donné que les malentendants sont très sensibles au bruit.

11. Annulation d'écho par filtrage adaptatif

Dans ce chapitre, nous allons présenter brièvement les méthodes de filtrage adaptatif issues de l'algorithme du gradient stochastique (LMS) et utilisées en annulation d'écho. Dans la section 11.1, nous allons d'abord décrire l'algorithme du LMS par bloc. A partir de cet algorithme de référence, plusieurs méthodes rapides ont été développées, nous les présenterons dans les deux autres sections qui constituent ce chapitre.

11.1. LMS temporel par blocs

11.1.1. Principes

Commençons par rappeler la définition des signaux mis en jeu, on note à chaque instant discret n :

- $x[n]$, le signal discret issu du microphone en entrée de prothèse (en sortie du CAN),
- $y[n]$, le signal discret transmis à l'écouteur en sortie de prothèse (en entrée du CNA),
- $\hat{u}[n]$, le signal d'écho estimé en sortie du filtre adaptatif,
- $e[n] = x[n] - \hat{u}[n]$, l'estimation du signal *source* $s[n]$ que nous avons introduit précédemment (*cf.* figure 10.2, page 147).

L'algorithme LMS est relativement facile à implémenter et repose sur un principe assez simple. Le signal $e[n]$ qui est l'estimation du signal *source* est considéré, du point de vue du LMS, comme le *signal d'erreur*.

Sous les hypothèses d'indépendance et de décorrélation des signaux $s[n]$ et $y[n]$, l'erreur quadratique moyenne (*Least Mean Square*) est minimale si le filtre adaptatif modélise parfaitement le chemin d'écho. A chaque itération de l'algorithme LMS, les coefficients du filtre adaptatif sont mis à jour selon une technique de descente de gradient afin de tendre vers la solution optimale.

L'algorithme du gradient stochastique réalise ainsi une minimisation adaptative.

11.1.2. Calcul de l'erreur par convolution

Le signal est traité par bloc de N échantillons ($N \in \mathbb{N}$), sans recouvrement entre blocs. A chaque itération de l'algorithme, on considère les échantillons de l'intervalle $[sN; sN + N - 1]$ où s est l'indice de bloc ou indice d'itération (avec $s \in \mathbb{N}$).

Soit \mathbf{h}_s le filtre adaptatif de longueur L ($L \in \mathbb{N}$) conçu pour estimer le chemin de retour acoustique et mis à jour à chaque itération s .

On considère les vecteurs de données \mathbf{x}_s , $\hat{\mathbf{u}}_s$ et \mathbf{e}_s de longueur N tels que :

$$\mathbf{x}_s = [x[sN] \quad x[sN+1] \quad \dots \quad x[sN+N-1]]^T \quad (11.1)$$

$$\hat{\mathbf{u}}_s = [\hat{u}[sN] \quad \hat{u}[sN+1] \quad \dots \quad \hat{u}[sN+N-1]]^T \quad (11.2)$$

$$\mathbf{e}_s = \mathbf{x}_s - \hat{\mathbf{u}}_s \quad (11.3)$$

ainsi que la matrice \mathbf{Y}_s de dimension $[N \times L]$ et définie par :

$$\mathbf{Y}_s = \begin{bmatrix} y[sN] & y[sN-1] & \dots & y[sN-L+1] \\ y[sN+1] & y[sN] & \dots & y[sN-L+2] \\ \vdots & \vdots & \ddots & \vdots \\ y[sN+N-1] & y[sN+N-2] & \dots & y[sN+N-L] \end{bmatrix} \quad (11.4)$$

Le calcul de l'écho estimé, $\hat{\mathbf{u}}_s$, est réalisé par convolution discrète selon le produit matriciel suivant :

$$\hat{\mathbf{u}}_s = \mathbf{Y}_s \cdot \mathbf{h}_s \quad (11.5)$$

11.1.3. Mise à jour du filtre adaptatif

Le filtre adaptatif est mis à jour par une technique de descente de gradient. La fonction de coût associée est l'*erreur quadratique instantanée* évaluée sur un bloc de données :

$$J(\mathbf{h}_s) = |\mathbf{e}_s|^2 \quad (11.6)$$

Les coefficients du filtre adaptatif \mathbf{h}_s sont alors mis à jour selon l'équation :

$$\mathbf{h}_{s+1} = \mathbf{h}_s - \mu \cdot \vec{\nabla}_{\mathbf{h}_s} J(\mathbf{h}_s), \quad (11.7)$$

où $\vec{\nabla}_{\mathbf{h}_s} J(\mathbf{h}_s)$ est le gradient de la fonction de coût J par rapport au vecteur \mathbf{h}_s défini par :

$$\vec{\nabla}_{\mathbf{h}_s} J(\mathbf{h}_s) = \frac{\partial J(\mathbf{h}_s)}{\partial \mathbf{h}_s} \quad (11.8)$$

et où μ est le *pas d'adaptation* du LMS.

Afin de calculer ce terme de gradient, on développe l'expression de la fonction de coût $J(\mathbf{h}_s)$:

$$\begin{aligned} J(\mathbf{h}_s) &= |\mathbf{e}_s|^2 \\ &= |\mathbf{x}_s - \mathbf{Y}_s \cdot \mathbf{h}_s|^2 \\ &= \mathbf{x}_s^T \cdot \mathbf{x}_s - 2\mathbf{h}_s^T \cdot \mathbf{Y}_s^T \cdot \mathbf{x}_s + \mathbf{h}_s^T \cdot \mathbf{Y}_s^T \cdot \mathbf{Y}_s \cdot \mathbf{h}_s \end{aligned} \quad (11.9)$$

Par dérivation de l'équation (11.9) et en reconnaissant les expressions respectives de $\hat{\mathbf{u}}_s$ et du vecteur d'erreur \mathbf{e}_s des équations (11.5) et (11.3), on en conclut que le gradient admet comme expression :

$$\begin{aligned} \vec{\nabla}_{\mathbf{h}_s} J(\mathbf{h}_s) &= 0 - 2\mathbf{Y}_s^T \cdot \mathbf{x}_s + 2\mathbf{Y}_s^T \cdot \mathbf{Y}_s \cdot \mathbf{h}_s \\ &= -2\mathbf{Y}_s^T \cdot (\mathbf{x}_s - \mathbf{Y}_s \cdot \mathbf{h}_s) \\ &= -2\mathbf{Y}_s^T \cdot \mathbf{e}_s \end{aligned} \quad (11.10)$$

La mise à jour des coefficients du filtre adaptatif est donc réalisée selon l'équation :

$$\mathbf{h}_{s+1} = \mathbf{h}_s + 2\mu \cdot \mathbf{Y}_s^T \cdot \mathbf{e}_s \quad (11.11)$$

L'algorithme du LMS temporel par blocs est très simple à mettre en place. Son coût de calcul est fortement conditionné par le calcul des produits matriciels des équations (11.5) et (11.10) qui sont de l'ordre de $N \times L$. L'équation (11.5) correspond à une convolution et peut donc être implémentée de manière rapide dans le domaine fréquentiel. Par incidence, il en est de même pour l'équation (11.10) qui correspond à une corrélation croisée.

11.2. LMS Rapide dans le domaine fréquentiel

11.2.1. Principes

L'algorithme LMS par bloc peut être rendu moins coûteux en temps de calcul si on choisit de réaliser la convolution temporelle dans le domaine fréquentiel. En opérant de la sorte, on tire alors avantage des propriétés de convolution circulaire de la TFD et de la rapidité de calcul de la Fast Fourier Transform (FFT). Cette méthode est appelée Fast LMS (FLMS), elle a été introduite par Ferrara dans [25] ; on la retrouve aussi dans la littérature sous l'appellation Frequency Domain Adaptive Filter (FDAF).

Outre sa rapidité, le FLMS possède des qualités de convergence supérieures au LMS temporel par blocs. En effet, les coefficients fréquentiels des signaux étant approximativement décorrélés, il est possible par une procédure de normalisation du gradient d'obtenir

des taux de convergence plus rapides et quasi-uniformes pour tous les modes du filtre adaptatif (*cf.* [27]). Le chapitre 13 est consacré à la normalisation du gradient qui est intégrée à la gestion du pas d'adaptation.

11.2.2. Convolution rapide dans le domaine de la transformée de Fourier

Les valeurs du signal d'écho $\hat{\mathbf{u}}$ sont estimées, comme on l'a vu plus haut, au moyen d'une convolution linéaire représentée matriciellement par l'équation (11.5), $\hat{\mathbf{u}}_s = \mathbf{Y}_s \cdot \mathbf{h}_s$, dont nous rappelons, ici, le détail :

$$\begin{bmatrix} \hat{u}_{[sN]} \\ \hat{u}_{[sN+1]} \\ \vdots \\ \hat{u}_{[sN+N-1]} \end{bmatrix} = \begin{bmatrix} y_{[sN]} & y_{[sN-1]} & \dots & y_{[sN-L+1]} \\ y_{[sN+1]} & y_{[sN]} & \dots & y_{[sN-L+2]} \\ \vdots & \vdots & \ddots & \vdots \\ y_{[sN+N-1]} & y_{[sN+N-2]} & \dots & y_{[sN+N-L]} \end{bmatrix} \cdot \begin{bmatrix} h_s[0] \\ h_s[1] \\ \vdots \\ h_s[L-1] \end{bmatrix}$$

Cette opération est très lourde du point de vue calculatoire. L'idée principale des algorithmes de convolution rapide est de réaliser la convolution dans le domaine de la transformée de Fourier selon le principe de dualité. En effet, une convolution circulaire dans le domaine temporel est équivalent à une multiplication terme à terme dans le domaine fréquentiel¹ de la TFD et les algorithmes de calcul rapide de la TFD permettent alors de réaliser cette opération avec une complexité réduite.

Dans le cadre de ce rapport de thèse, nous nous contentons simplement de rappeler la structure du FLMS. Pour consulter des références plus complètes sur le sujet, on peut notamment se reporter à [25], [27] et [61].

Le FLMS consiste à mettre en œuvre dans le domaine fréquentiel la convolution circulaire entre les vecteurs $\tilde{\mathbf{h}}_s$ et $\tilde{\mathbf{y}}_s$, tous deux de longueur $N + L - 1$, et définis respectivement par :

$$\begin{aligned} \tilde{\mathbf{h}}_s &= \left[h_s[0] \quad h_s[1] \quad \dots \quad h_s[L-1] \quad \vdots \quad 0 \quad \dots \quad 0 \right]^T \\ &= \left[\mathbf{h}_s^T \quad \vdots \quad \mathbf{0}_{1 \times N-1} \right]^T \end{aligned} \quad (11.12)$$

$$\tilde{\mathbf{y}}_s = \left[y_{[sN-L+1]} \quad y_{[sN-L+2]} \quad \dots \quad y_{[sN+N-1]} \right]^T \quad (11.13)$$

1. Une autre manière d'aborder cette propriété est de dire que la matrice \mathbf{Y}_s est une matrice Toëplitz à laquelle on peut associer une matrice circulante. Or, une matrice circulante est diagonalisable dans la base de Fourier.

Notons $\tilde{\mathbf{u}}_s$ le produit de la convolution circulaire entre $\tilde{\mathbf{y}}_s$ et $\tilde{\mathbf{h}}_s$:

$$\tilde{\mathbf{u}}_s = \tilde{\mathbf{y}}_s \circledast \tilde{\mathbf{h}}_s \quad (11.14)$$

où \circledast représente le produit de convolution circulaire entre deux vecteurs.

Le vecteur $\tilde{\mathbf{u}}_s$ est de longueur $N + L - 1$, ses N dernières composantes correspondent à la convolution linéaire de l'équation (11.5) c'est-à-dire aux composantes de $\hat{\mathbf{u}}_s$. Les $L - 1$ premières composantes résultent de la convolution circulaire et doivent être écartées.

De manière formelle, on écarte ces $L - 1$ premières composantes en considérant la matrice de troncature \mathbf{P}_N de dimension $N \times (N + L - 1)$ définie par :

$$\mathbf{P}_N = \left[\begin{array}{c|c} \mathbf{0}_{N \times (L-1)} & \mathbf{I}_N \end{array} \right] \quad (11.15)$$

On obtient alors :

$$\begin{aligned} \hat{\mathbf{u}}_s &= \mathbf{P}_N \cdot \tilde{\mathbf{u}}_s \\ &= \mathbf{P}_N \cdot \tilde{\mathbf{y}}_s \circledast \tilde{\mathbf{h}}_s \end{aligned} \quad (11.16)$$

Notons \mathbf{W}_F la matrice de transformation de Fourier discrète de dimension M ($M \in \mathbb{N}$) de coefficients :

$$\mathbf{W}_F[k, l] = \exp\left(-2j\pi \frac{kl}{M}\right) \quad (11.17)$$

et \mathbf{W}_F^{-1} la matrice de la transformation inverse :

$$\mathbf{W}_F^{-1}[k, l] = \frac{1}{M} \exp\left(2j\pi \frac{kl}{M}\right) \quad (11.18)$$

Remarque

Pour le calcul de la FFT, il faut choisir M comme une puissance de 2. En général, on choisit donc $N + L - 1$ si cette valeur convient ou la puissance de 2 supérieure la plus proche sinon (auquel cas il faut compléter les vecteurs par le nombre de coefficients nuls nécessaire et écarter les dernières composantes vectorielles introduites lors de la transformation inverse). Pour simplifier, on suppose, ici, que l'on choisit $M = N + L - 1$.

Pour un exemple d'implémentation avec des valeurs de M supérieures à $N + L - 1$, on peut se reporter à [61].

La réalisation de la convolution circulaire se fait dans le domaine fréquentiel à l'aide des TFD respectives de $\tilde{\mathbf{y}}_s$ et $\tilde{\mathbf{h}}_s$ notées $\tilde{\mathcal{Y}}_s$ et $\tilde{\mathcal{H}}_s$ et de longueur M :

$$\tilde{\mathcal{Y}}_s = \mathbf{W}_F \cdot \tilde{\mathbf{y}}_s \quad (11.19)$$

$$\tilde{\mathcal{H}}_s = \mathbf{W}_F \cdot \tilde{\mathbf{h}}_s \quad (11.20)$$

Le calcul de la convolution linéaire est donc réalisé selon l'expression suivante :

$$\hat{\mathbf{u}}_s = \mathbf{P}_N \cdot \mathbf{W}_F^{-1} \left(\tilde{\mathcal{Y}}_s \odot \tilde{\mathcal{H}}_s \right) \quad (11.21)$$

où \odot représente le *produit de Schür* ou produit composante à composante des vecteurs.

En reprenant l'équation (11.3) exprimant le signal d'erreur, on obtient alors :

$$\mathbf{e}_s = \mathbf{x}_s - \mathbf{P}_N \cdot \mathbf{W}_F^{-1} \left(\tilde{\mathcal{Y}}_s \odot \tilde{\mathcal{H}}_s \right) \quad (11.22)$$

Grâce à l'utilisation de FFT, le calcul de l'erreur par blocs est de cette façon rendu moins coûteux que dans le cas du LMS temporel par blocs.

11.2.3. Mise à jour du filtre adaptatif

Comme pour le calcul de l'erreur, la mise à jour du filtre $\tilde{\mathbf{h}}_s$ peut être réalisée dans le domaine fréquentiel avec un coût de calcul moindre. L'équation (11.11) de mise à jour dans le domaine temporel s'exprime, en effet, très simplement dans le domaine fréquentiel.

A cet effet, on définit la TFD, $\tilde{\mathcal{E}}_s$, de la séquence d'erreur de la manière suivante :

$$\tilde{\mathcal{E}}_s = \mathbf{W}_F \cdot \begin{bmatrix} \mathbf{0}_{L-1} \\ \dots \\ \tilde{\mathbf{e}}_s \end{bmatrix} \quad (11.23)$$

En reprenant la méthode de calcul du gradient comme on l'a exprimé, dans le domaine temporel, à la section 11.1.3, on peut mettre à jour les coefficients du filtre adaptatif directement dans le domaine de la TFD.

Le terme de mise à jour des coefficients du filtre représentant la *corrélation croisée* de l'équation (11.10) s'exprime alors par $\tilde{\mathcal{Y}}_s^* \odot \tilde{\mathcal{E}}_s$, $\tilde{\mathcal{Y}}_s^*$ étant le conjugué de $\tilde{\mathcal{Y}}_s$.

A chaque itération s , la mise à jour du filtre adaptatif est alors réalisée selon l'équation :

▷ *sans contrainte* :

$$\tilde{\mathcal{H}}_{s+1} = \tilde{\mathcal{H}}_s + \mu_s \odot (\tilde{\mathcal{Y}}_s^* \odot \tilde{\mathcal{E}}_s) \quad (11.24)$$

▷ *avec contrainte* :

$$\tilde{\mathcal{H}}_{s+1} = \tilde{\mathcal{H}}_s + \mathbf{C} \mu_s \odot (\tilde{\mathcal{Y}}_s^* \odot \tilde{\mathcal{E}}_s) \quad (11.25)$$

où

- \mathbf{C} est la matrice de contrainte qui impose aux $N - 1$ dernières composantes du filtre $\tilde{\mathbf{h}}_s$ d'être nulles,

- μ_s est le pas d'adaptation qui contrôle la vitesse de convergence de l'algorithme (cf. sa définition dans le chapitre 13).

\mathbf{C} est définie par :

$$\mathbf{C} = \mathbf{W}_F \cdot \begin{bmatrix} \mathbf{I}_{L \times L} & \mathbf{0}_{L \times (N-1)} \\ \mathbf{0}_{(N-1) \times L} & \mathbf{0}_{N-1} \end{bmatrix} \cdot \mathbf{W}_F^{-1} \quad (11.26)$$

11.3. FLMS dans le domaine de la transformée de Hartley discrète

Le FLMS peut être exactement transposé à toute autre transformée possédant la propriété de convolution circulaire. Dans [61], Prado et Moulines présentent son implantation dans le domaine de la Transformée de Hartley Discrète (THD). Cette configuration nous semble être un choix intéressant pour une utilisation du LMS dans une prothèse auditive.

11.3.1. La transformée de Hartley discrète

La THD d'un vecteur réel de dimension M , $\mathbf{x} = [x_0, \dots, x_{M-1}]$, a une définition proche de celle de la TFD et est donnée par :

$$\mathbf{x}_k^h = \sum_{l=0}^{M-1} x_l \left(\cos \left(2\pi \frac{kl}{M} \right) + \sin \left(2\pi \frac{kl}{M} \right) \right) \quad (11.27)$$

Notons \mathbf{W}_H la matrice de transformation de Hartley discrète de dimension M ($M \in \mathbb{N}$) de coefficients :

$$\mathbf{W}_H[k, l] = \cos \left(2\pi \frac{kl}{M} \right) + \sin \left(2\pi \frac{kl}{M} \right) \quad (11.28)$$

La matrice de la transformation inverse est \mathbf{W}_H^{-1} telle que :

$$\mathbf{W}_H^{-1} = \frac{1}{M} \mathbf{W}_H \quad (11.29)$$

L'utilisation de la THD dans ce contexte particulier des prothèses auditives a été choisi, ici, car la THD possède les propriétés suivantes :

- c'est une *transformée réelle*,
- elle possède la propriété d'être *auto-inversible* (au coefficient de normalisation près) comme l'indique l'équation (11.29). Ainsi, seule la programmation d'une transformée dans le code embarqué du processeur DSP de la prothèse est nécessaire.

11.3.2. Convolution rapide dans le domaine de la transformée de Hartley

Nous pouvons exprimer le calcul de $\hat{\mathbf{u}}_s$ de manière équivalente à l'équation (11.21) en utilisant, cette fois-ci, les THD respectives des vecteurs $\tilde{\mathbf{y}}_s$ et $\tilde{\mathbf{h}}_s$ en lieu et place de leurs TFD. Afin de ne pas surcharger les notations et comme nous utiliserons, à partir de ce point, uniquement la transformée de Hartley discrète, nous allons conserver les mêmes notations pour les THD des vecteurs $\tilde{\mathbf{y}}_s$ et $\tilde{\mathbf{h}}_s$ que pour leur équivalent dans le domaine de la TFD. On considère donc $\tilde{\mathcal{Y}}_s$ et $\tilde{\mathcal{H}}_s$ telles que :

$$\begin{aligned}\tilde{\mathcal{Y}}_s &= \mathbf{W}_H \cdot \tilde{\mathbf{y}}_s \\ \tilde{\mathcal{H}}_s &= \mathbf{W}_H \cdot \tilde{\mathbf{h}}_s\end{aligned}\tag{11.30}$$

La réalisation de la convolution circulaire est moins immédiate à exprimer dans le domaine de la transformée de la THD que son équivalent dans le domaine fréquentiel. La base de la TFD a la propriété de diagonaliser les matrices circulantes. Ainsi, la convolution circulaire entre $\tilde{\mathbf{y}}_s$ et $\tilde{\mathbf{h}}_s$ s'exprime, en fréquentiel, sous la simple forme du produit de Schür $\tilde{\mathcal{Y}}_s \odot \tilde{\mathcal{H}}_s$.

Dans [6], Bini et Favati montrent que la base de la THD, quant à elle, diagonalise les matrices \mathbf{Y} carrées de dimension M de la forme $\mathbf{Y} = \mathbf{A} + \mathbf{J}_M \cdot \mathbf{B}$ telles que :

- \mathbf{A} est une matrice circulante symétrique
- \mathbf{B} est une matrice circulante anti-symétrique

\mathbf{J}_M étant la matrice de permutation suivante²

$$\mathbf{J}_M = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 0 & & 1 \\ \vdots & & \ddots & \\ 0 & 1 & & 0 \end{bmatrix}\tag{11.31}$$

A partir de ces propriétés ou en s'appuyant sur la similitude entre TFD et THD, on peut alors montrer que la convolution circulaire entre $\tilde{\mathbf{y}}_s$ et $\tilde{\mathbf{h}}_s$ s'exprime, dans le domaine de la THD, sous la forme $\tilde{\mathcal{Y}}_s \bullet \tilde{\mathcal{H}}_s$, où l'opération \bullet est telle que :

$$\tilde{\mathcal{Y}}_s \bullet \tilde{\mathcal{H}}_s = \frac{1}{2} \left((\mathbf{I}_M - \mathbf{J}_M) \tilde{\mathcal{Y}}_s \odot \tilde{\mathcal{H}}_s + (\mathbf{I}_M + \mathbf{J}_M) \tilde{\mathcal{Y}}_s \odot (\mathbf{J}_M \tilde{\mathcal{H}}_s) \right)\tag{11.32}$$

Au sujet de l'utilisation de la THD pour la convolution rapide, on peut se reporter à [6] et [21]. Pour la transcription des algorithmes rapides de type LMS à la THD, une description plus détaillée est présentée dans [61] et [52].

2. La notation \mathbf{J}_M est usuellement réservée à la matrice de retournement dont les coefficients valent 1 sur l'anti-diagonale et sont nuls ailleurs. Nous avons néanmoins opté pour cette notation, ici, car il est généralement d'usage de présenter ces propriétés en écrivant la THD sur $M + 1$ points. Du fait de la périodicité de la THD, le premier et le dernier point sont égaux et la matrice est donc bien alors la matrice de retournement.

L'équation (11.16) se reformule alors de la manière suivante :

$$\hat{\mathbf{u}}_s = \mathbf{P}_N \cdot \mathbf{W}_H^{-1} \left(\tilde{\mathcal{Y}}_s \star \tilde{\mathcal{H}}_s \right) \quad (11.33)$$

Et l'expression du signal d'erreur devient :

$$\mathbf{e}_s = \mathbf{x}_s - \mathbf{P}_N \cdot \mathbf{W}_H^{-1} \left(\tilde{\mathcal{Y}}_s \star \tilde{\mathcal{H}}_s \right) \quad (11.34)$$

11.3.3. Mise à jour du filtre adaptatif

On définit la THD, $\tilde{\mathcal{E}}_s$, de la séquence d'erreur de la manière suivante :

$$\tilde{\mathcal{E}}_s = \mathbf{W}_H \cdot \begin{bmatrix} \mathbf{0}_{L-1} \\ \dots \\ \mathbf{e}_s \end{bmatrix} \quad (11.35)$$

Comme pour le calcul de la convolution circulaire, l'expression de mise à jour du filtre adaptatif dans le domaine de la THD met en jeu une expression plus complexe. En effet, la corrélation croisée est calculée dans le domaine de la THD par la formule $\tilde{\mathcal{Y}}_s \star \tilde{\mathcal{E}}_s$, où \star est l'opération telle que :

$$\tilde{\mathcal{Y}}_s \star \tilde{\mathcal{E}}_s = \frac{1}{2} \left((\mathbf{I}_M + \mathbf{J}_M) \left(\tilde{\mathcal{Y}}_s \odot \tilde{\mathcal{E}}_s \right) - (\mathbf{I}_M - \mathbf{J}_M) \left(\tilde{\mathcal{Y}}_s \odot \left(\mathbf{J}_M \tilde{\mathcal{E}}_s \right) \right) \right) \quad (11.36)$$

A chaque itération s , le filtre adaptatif est mis à jour selon l'équation :

▷ *sans contrainte* :

$$\tilde{\mathcal{H}}_{s+1} = \tilde{\mathcal{H}}_s + \boldsymbol{\mu}_s \odot \left(\tilde{\mathcal{Y}}_s \star \tilde{\mathcal{E}}_s \right) \quad (11.37)$$

▷ *avec contrainte* :

$$\tilde{\mathcal{H}}_{s+1} = \tilde{\mathcal{H}}_s + \mathbf{C} \boldsymbol{\mu}_s \odot \left(\tilde{\mathcal{Y}}_s \star \tilde{\mathcal{E}}_s \right) \quad (11.38)$$

où comme dans le cas précédent utilisant la TFD :

- \mathbf{C} est la matrice de contrainte qui impose aux $N - 1$ dernières composantes du filtre $\tilde{\mathbf{h}}_s$ d'être nulles,
- $\boldsymbol{\mu}_s$ est le pas d'adaptation qui contrôle la vitesse de convergence de l'algorithme (cf. sa définition dans le chapitre 13).

Dans la nouvelle définition de \mathbf{C} , la matrice \mathbf{W}_H intervient donc maintenant en lieu et place de \mathbf{W}_F et \mathbf{W}_F^{-1} selon l'expression :

$$\mathbf{C} = \mathbf{W}_H \cdot \begin{bmatrix} \mathbf{I}_{L \times L} & \mathbf{0}_{L \times (N-1)} \\ \mathbf{0}_{(N-1) \times L} & \mathbf{0}_{N-1} \end{bmatrix} \cdot \mathbf{W}_H \quad (11.39)$$

12. Filtre à Délais Multiples Généralisé

12.1. Principes

Initialement, une modification du FLMS a été introduite par Soo et Pang dans [71]. Cette méthode est connue sous le nom de Filtre à Délais Multiples, Multi-Delay Filter (MDF), ou également Partitioned Block FDAF (PBFDAF).

L'idée originale qui a motivé la mise au point de cette méthode est le découpage d'un filtre adaptatif en plusieurs blocs adjacents afin de réduire la taille des transformées utilisées. En effet, dans le contexte de l'annulation d'écho classique, la taille des réponses impulsionnelles des salles peut atteindre plusieurs centaines de milli-secondes ce qui entraîne pour des fréquences d'échantillonnage de 16kHz ou 32kHz des transformées de plusieurs milliers de coefficients. Ces tailles sont donc souvent inadaptées aux processeurs de signaux numériques (ou Digital Signal Processor (DSP)) sur lesquels sont implantés les algorithmes. L'autre point est le délai de traitement qui se révèle alors trop important et qui peut gêner la conversation.

Dans le cadre des prothèses auditives, la taille des réponses impulsionnelles est plus courte (environ de quelques milli-secondes à quelques dizaine de milli-secondes) ce qui implique des transformées sur plusieurs centaines de coefficients. Cependant l'architecture des DSP des prothèses n'est pas adaptée au calcul de longues transformées et on trouve également certains DSP avec des transformées sur 32 points (à la fréquence d'échantillonnage de 16kHz) préimplantées (notamment du fait de leur utilisation dans la décomposition par bancs de filtres sur 32 canaux). D'autre part, le délai de traitement dans une prothèse est un élément particulièrement sensible qui doit rester impérativement inférieur à 10-20ms.

L'idée est donc de diviser le traitement de la réponse impulsionnelle de taille L' en K sous-blocs ($K \in \mathbb{N}$) de taille L telle que L soit un sous-multiple entier de L' , i.e. $L' = K \cdot L$. Le schéma de la figure 12.1 représente une telle décomposition.

L'algorithme que nous proposons d'utiliser ici est une généralisation du MDF développée par Aït-Amrane, Grenier et Moulines [1, 56]. Cette généralisation introduit une reconstruction du signal par addition-recouvrement pondéré, WOLA. Ceci permet alors de mettre à jour le filtre adaptatif plus régulièrement ce qui améliore ses capacités de convergence et de poursuite.

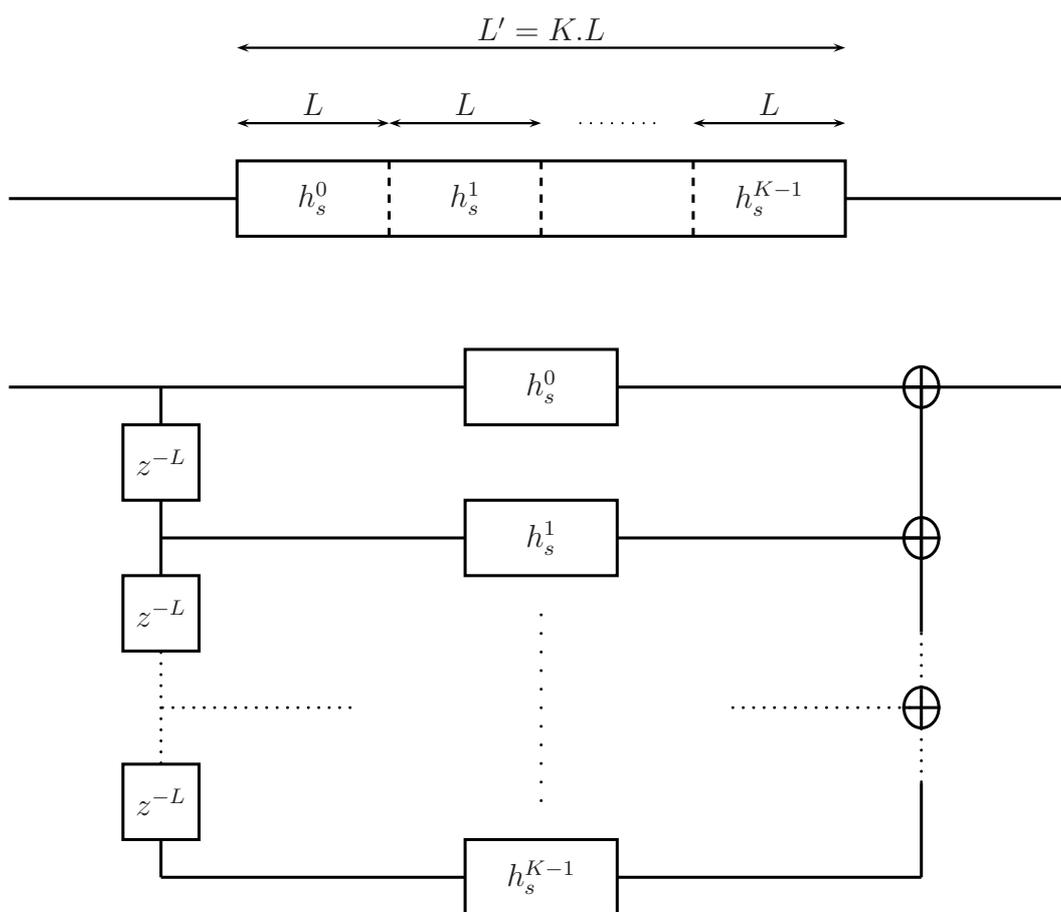


FIG. 12.1: Décomposition d'un filtre en sous-blocs pour l'implémentation du GMDF. Le filtre original de taille $L' = K \cdot L$ est décomposé en K sous blocs de taille L au moyen d'une structure à délais multiples

Cette généralisation du MDF est appelée Generalized-MDF α (GMDF α). Prado et Moulines ont explicité son implantation dans le domaine de la THD dans [61]. Dans la suite de ce document, nous désignons cette implantation particulière sous le terme de Hartley-Domain-GMDF α (HD-GMDF α).

La particularité de cette méthode, initialement développée pour un contexte pourtant très différent, nous semble, ici, être tout à fait appropriée à une utilisation dans le cadre des prothèses auditives.

12.2. Synthèse WOLA

Étant donné la structure WOLA, les blocs d'échantillons traités ont une définition différente de celle vue précédemment. Soit R le pas d'avancement, à chaque indice de bloc s le traitement s'effectue sur les échantillons de l'intervalle $[sR; sR + N - 1]$.

Soit α_R le taux de recouvrement défini par :

$$\alpha_R = \frac{N}{R}$$

On suppose que α_R est un entier positif et on le choisira comme tel.

Pour chaque indice de bloc s , on considère les signaux vectoriels \mathbf{x}_s et $\hat{\mathbf{u}}_s$, respectivement signal d'entrée et le signal d'écho estimé :

$$\mathbf{x}_s = [x_{[sR]} \quad x_{[sR+1]} \quad \dots \quad x_{[sR+N-1]}]^T \quad (12.1)$$

$$\hat{\mathbf{u}}_s = [\hat{u}_{[sR]} \quad \hat{u}_{[sR+1]} \quad \dots \quad \hat{u}_{[sR+N-1]}]^T \quad (12.2)$$

$$\mathbf{e}_s = \mathbf{x}_s - \hat{\mathbf{u}}_s \quad (12.3)$$

Dans l'approche WOLA, la synthèse du signal de sortie à l'itération s est réalisée en deux étapes :

- les échantillons de \mathbf{e}_s sont pondérés à l'aide d'une fenêtre de synthèse classique \mathbf{w} (de type hanning, blackmann, ...),
- puis on synthétise les échantillons du signal de sortie $\bar{\mathbf{e}}$ appartenant à l'intervalle $[(s-1)R + N; sR + N - 1]$ par addition-recouvrement des échantillons des α_R dernières itérations.

$$\begin{bmatrix} \bar{e}_{[(s-1)R+N]} \\ \vdots \\ \bar{e}_{[sR+N-1]} \end{bmatrix} = \sum_{j=0}^{\alpha_R-1} \begin{bmatrix} w_{[N-(j+1)R]} \\ \vdots \\ w_{[N-1-jR]} \end{bmatrix} \odot \begin{bmatrix} e_{s[N-(j+1)R]} \\ \vdots \\ e_{s[N-1-jR]} \end{bmatrix} \quad (12.4)$$

12.3. Partitionnement des données

L'originalité du Generalized-MDF (GMDF) consiste à partitionner le filtre adaptatif \mathbf{h}_s de taille L' en K sous-blocs de taille identique L telle que $L' = K \cdot L$. On définit alors K vecteurs \mathbf{h}_s^k tels que :

$$\mathbf{h}_s = \begin{bmatrix} \mathbf{h}_s^0 \\ \vdots \\ \mathbf{h}_s^{K-1} \end{bmatrix} \quad (12.5)$$

Chaque vecteur \mathbf{h}_s^k est donc défini par :

$$\forall k \in [0; K-1], \quad \mathbf{h}_s^k = [h_{s[kL]} \quad h_{s[kL+1]} \quad \dots \quad h_{s[(k+1)L-1]}]^T \quad (12.6)$$

Chaque sous-bloc \mathbf{h}_s^k est ensuite convolué dans le domaine de la THD avec $N + L - 1$ échantillons issus de $y[n]$.

De même qu'on a défini précédemment $\tilde{\mathbf{h}}_s$ par (11.12), page 156, on définit alors les vecteurs $\tilde{\mathbf{h}}_s^k$ de dimension $(N + L - 1) \times 1$ à partir des vecteurs \mathbf{h}_s^k complétés de $(N - 1)$ coefficients nuls selon l'expression :

$$\begin{aligned} \forall k \in [0; K-1], \quad \tilde{\mathbf{h}}_s^k &= [h_{s^k[0]} \quad h_{s^k[1]} \quad \dots \quad h_{s^k[L-1]} \quad \vdots \quad 0 \quad \dots \quad 0]^T \\ &= \begin{bmatrix} \mathbf{h}_s^{kT} \\ \mathbf{0}_{1 \times N-1} \end{bmatrix}^T \end{aligned} \quad (12.7)$$

On cherche à calculer le produit de convolution circulaire entre \mathbf{h}_s et le vecteur $\tilde{\mathbf{y}}_s$ de taille $N + L' - 1$ et défini par :

$$\tilde{\mathbf{y}}_s = [y_{[sR-L'+1]} \quad y_{[sR-L'+2]} \quad \dots \quad y_{[sR+N-1]}]^T \quad (12.8)$$

Comme nous avons choisi de segmenter \mathbf{h}_s en K sous-blocs \mathbf{h}_s^k , il est alors nécessaire d'effectuer K produits de convolution entre les vecteurs \mathbf{h}_s^k et les vecteurs $\tilde{\mathbf{y}}_s^k$ de longueur $(N + L - 1)$ et définis par :

$$\forall k \in [0; K-1], \quad \tilde{\mathbf{y}}_s^k = [y_{[sR-L'+1+kL]} \quad y_{[sR-L'+2+kL]} \quad \dots \quad y_{[sR+N-1-L'+(k+1)L]}]^T \quad (12.9)$$

Le schéma de la figure 12.2 explicite de manière schématique ce partitionnement de $\tilde{\mathbf{y}}_s$ en K sous-vecteurs $\tilde{\mathbf{y}}_s^k$ *non-disjoints*.

Afin d'effectuer ces K produits de convolution circulaire dans le domaine de la THD, on introduit également leurs THD respectives exactement comme c'était le cas dans la section 11.3 du chapitre précédent.

$$\begin{aligned} \tilde{\mathcal{Y}}_s^k &= \mathbf{W}_H \cdot \tilde{\mathbf{y}}_s^k \\ \tilde{\mathcal{H}}_s^k &= \mathbf{W}_H \cdot \tilde{\mathbf{h}}_s^k \end{aligned} \quad (12.10)$$

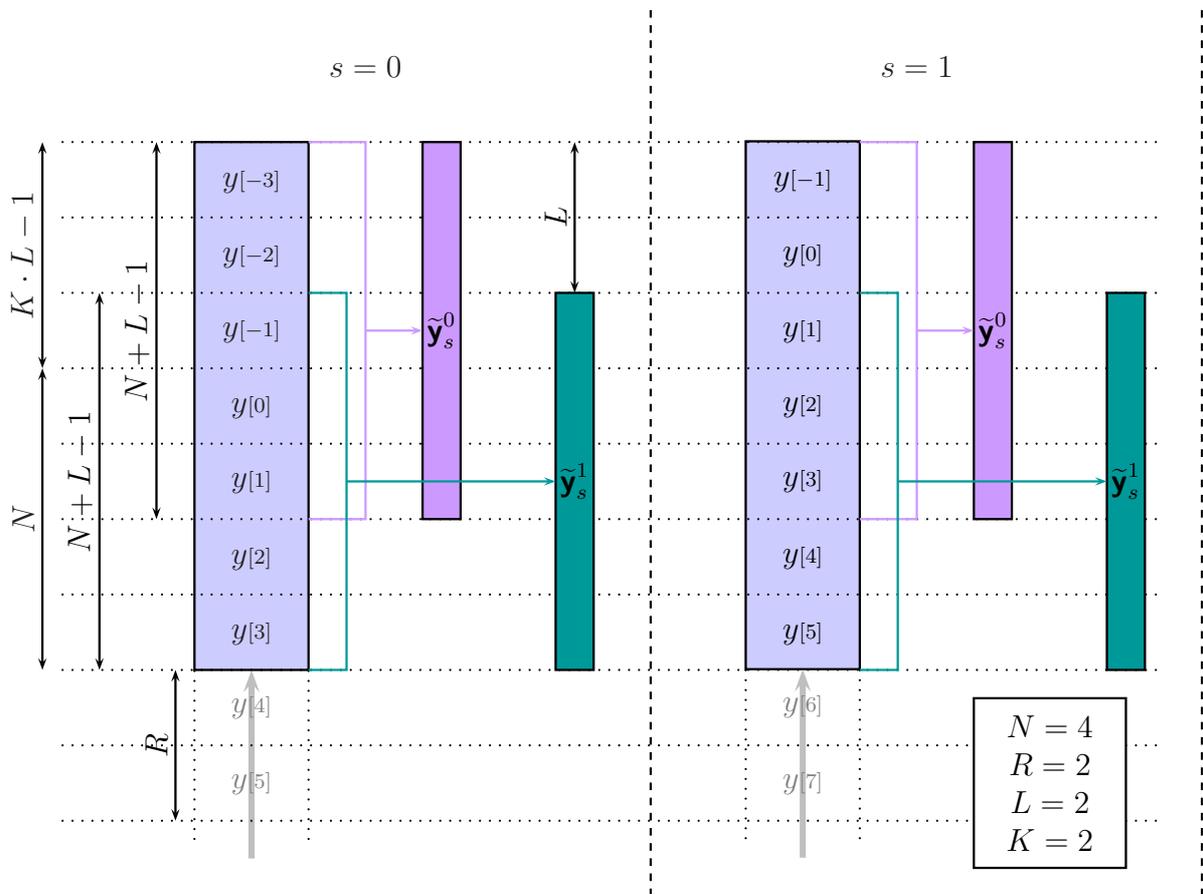


FIG. 12.2: Composition des vecteurs $\tilde{\mathbf{y}}_s^k$ pour les indices de blocs $s = 0$ et $s = 1$ pour $N = 4$, $R = 2$, $L = 2$ et $K = 4$

12.4. Implémentation du LMS

Les N échantillons de la séquence $\hat{\mathbf{u}}_s$ sont alors simplement construits à partir du produit suivant :

$$\hat{\mathbf{u}}_s = \mathbf{P}_N \cdot \mathbf{W}_H^{-1} \sum_{k=0}^{K-1} \left(\tilde{\mathcal{Y}}_s^k \bullet \tilde{\mathcal{H}}_s^k \right) \quad (12.11)$$

Comme 11.34 L'expression du signal d'erreur devient alors :

$$\mathbf{e}_s = \mathbf{x}_s - \mathbf{P}_N \cdot \mathbf{W}_H^{-1} \sum_{k=0}^{K-1} \left(\tilde{\mathcal{Y}}_s^k \bullet \tilde{\mathcal{H}}_s^k \right) \quad (12.12)$$

Le détail des calculs nécessaires pour parvenir à la formulation de la mise à jour du gradient est explicité dans l'annexe A.

Dans le domaine de la THD, l'erreur est définie comme dans le FLMS selon l'équation (11.35).

A chaque indice de bloc temporel s , la mise à jour du filtre adaptatif est réalisée *pour chaque sous-bloc k* du filtre adaptatif selon l'équation de mise à jour :

▷ *sans contrainte* :

$$\tilde{\mathcal{H}}_{s+1}^k = \tilde{\mathcal{H}}_s^k + \boldsymbol{\mu}_s \odot (\tilde{\mathcal{Y}}_s^k \star \tilde{\mathcal{E}}_s) \quad (12.13)$$

▷ *avec contrainte* :

$$\tilde{\mathcal{H}}_{s+1}^k = \tilde{\mathcal{H}}_s^k + \mathbf{C} \boldsymbol{\mu}_s \odot (\tilde{\mathcal{Y}}_s^k \star \tilde{\mathcal{E}}_s) \quad (12.14)$$

où les paramètres \mathbf{C} , $\boldsymbol{\mu}_s$ ont les mêmes définitions que celles présentées précédemment.

13. Modification du pas d'adaptation

13.1. Pas d'adaptation du LMS

13.1.1. Rôle du pas d'adaptation

Le paramètre μ_s est le pas d'adaptation qui contrôle la mise à jour du filtre adaptatif. μ_s est donc un paramètre fondamental du point de vue de la convergence du LMS.

Si le pas μ_s est petit, à chaque itération, on réalise une mise à jour fine du filtre adaptatif dans le sens du gradient. La convergence est alors assurée et la précision de la modélisation est très bonne. En contrepartie, la vitesse de convergence est faible et donc les capacités de poursuite de la réponse du chemin d'écho sont limitées.

Si le pas μ_s est grand, l'algorithme est susceptible de diverger. Si, toutefois, l'algorithme converge, l'estimation du chemin d'écho n'est pas très précise (et les valeurs des coefficients oscillent autour des valeurs de la solution idéale). En contrepartie, la vitesse de convergence est grande et les capacités de poursuite du chemin d'écho sont bonnes.

13.1.2. Normalisation du gradient

Dans la définition classique du LMS temporel par bloc que l'on a indiqué dans la section 11.1.3, le pas d'adaptation, μ , est une constante réelle. Il est cependant d'usage de faire varier μ comme l'inverse de la puissance du signal $y[n]$. Le pas d'adaptation réalise ainsi une normalisation du gradient. On parle alors de *LMS Normalisé* (cf. [27]). L'utilité d'une telle démarche est de s'adapter ainsi au caractère non-stationnaire des signaux acoustiques.

Dans le domaine fréquentiel ou dans le domaine de la THD, une même démarche est envisageable. Cependant la structure des algorithmes rapides permet d'ajouter un degré de liberté supplémentaire dans la définition du pas d'adaptation. Il est alors en effet possible de pondérer le pas d'adaptation avec un poids spécifique pour chaque coefficient fréquentiel (ou coefficient de la THD). Ainsi, dans les algorithmes adaptatifs décrits dans les deux chapitres précédents, le pas d'adaptation, μ_s , est une variable vectorielle réelle

variable telle que :

$$\boldsymbol{\mu}_s = \mu_0 \boldsymbol{\mathcal{T}}_s \quad (13.1)$$

où

- μ_0 est un paramètre constant
- $\boldsymbol{\mathcal{T}}_s$ est le vecteur de normalisation dans le domaine transformé.

Ainsi chaque composante de $\boldsymbol{\mathcal{T}}_s$ permet de normaliser de manière adaptée la composante correspondante du gradient dans le domaine fréquentiel ou dans le domaine de la THD. A cet effet, chaque coefficient de $\boldsymbol{\mathcal{T}}_s$ est défini comme l'inverse de la puissance spectrale de $\widetilde{\boldsymbol{\mathcal{Y}}}_s$:

$$\boldsymbol{\mathcal{T}}_s = \left[\frac{1}{\mathcal{P}_s^y[0]}, \dots, \frac{1}{\mathcal{P}_s^y[M-1]} \right]^T \quad (13.2)$$

où la $i^{\text{ème}}$ composante $\mathcal{P}_s^y[i]$ est obtenue par filtrage passe-bas de la séquence $\{|\widetilde{\boldsymbol{\mathcal{Y}}}_s[i]|^2\}_{s \in \mathbb{N}}$.

- Dans le domaine fréquentiel, \mathcal{P}_s^y est donc estimé selon l'équation de récurrence :

$$\mathcal{P}_{s+1}^y = \gamma \mathcal{P}_s^y + (1 - \gamma) (\widetilde{\boldsymbol{\mathcal{Y}}}_s^* \odot \widetilde{\boldsymbol{\mathcal{Y}}}_s) \quad (13.3)$$

où γ est le facteur de lissage tel que $0 < \gamma < 1$.

- Dans le domaine de la THD, étant donné les relations entre TFD et THD, le terme équivalent au terme $|\widetilde{\boldsymbol{\mathcal{Y}}}_s^k[i]|^2$ dans le domaine spectral est $\frac{1}{2} \left(|\widetilde{\boldsymbol{\mathcal{Y}}}_s^k[i]|^2 + |\widetilde{\boldsymbol{\mathcal{Y}}}_s^k[M-i]|^2 \right)$. \mathcal{P}_s^y est donc donné par la relation de récurrence suivante :

$$\mathcal{P}_{s+1}^y = \gamma \mathcal{P}_s^y + (1 - \gamma) \frac{1}{2} (\widetilde{\boldsymbol{\mathcal{Y}}}_s \odot \widetilde{\boldsymbol{\mathcal{Y}}}_s + \mathbf{J}_M (\widetilde{\boldsymbol{\mathcal{Y}}}_s \odot \widetilde{\boldsymbol{\mathcal{Y}}}_s)) \quad (13.4)$$

- Enfin, pour le HD-GMDF α , comme le signal excitateur est segmenté en K sous-vecteurs, on considère uniquement le plus récent d'entre eux pour faire la mise à jour de \mathcal{P}_s^y :

$$\mathcal{P}_{s+1}^y = \gamma \mathcal{P}_s^y + (1 - \gamma) \frac{1}{2} (\widetilde{\boldsymbol{\mathcal{Y}}}_s^0 \odot \widetilde{\boldsymbol{\mathcal{Y}}}_s^0 + \mathbf{J}_M (\widetilde{\boldsymbol{\mathcal{Y}}}_s^0 \odot \widetilde{\boldsymbol{\mathcal{Y}}}_s^0)) \quad (13.5)$$

Remarque

Les définitions respectives de la TFD et de la THD entraînent une symétrie des coefficients de \mathcal{P}_s^y . En pratique, le calcul de \mathcal{P}_s^y se fait donc uniquement sur les $M/2 + 1$ premiers coefficients. Et les valeurs suivantes sont déduites de ces $M/2 + 1$ composantes suivant la relation :

$$\forall i \in [M/2 + 2; M - 1], \mathcal{P}_s^y[M-i] = \mathcal{P}_s^y[i] \quad (13.6)$$

Pour le GMDF, d'après [56], la valeur optimale pour le pas d'adaptation μ_0 est la valeur :

$$\mu_0 = \frac{1}{K + 1} \quad (13.7)$$

La normalisation du gradient par \mathcal{T}_s est un des éléments clé des performances des algorithmes LMS rapides.

Sans cette normalisation, les algorithmes rapides sont des implémentations du LMS temporel par bloc aux performances strictement identiques.

Avec la normalisation, les performances sont alors supérieures à celle du LMS temporel par bloc car elle permet approximativement de rendre égaux les taux de convergences de chaque mode du processus d'adaptation.

13.2. Modification de la définition du pas d'adaptation

Afin d'assurer la convergence du filtre adaptatif tout en conservant un coût de calcul faible, une règle de mise à jour simple doit être mise au point.

S'il est intéressant de se focaliser sur \mathcal{T}_s , c'est que \mathcal{T}_s est un important paramètre qui contrôle le processus de convergence.

Dans ce contexte d'adaptation en système fermé, assez particulier et très exigeant, nous proposons, ici, d'adopter une stratégie de prudence dans le processus d'adaptation afin de ne pas être trop sensible à la corrélation de signaux. Nous proposons donc de définir une définition alternative de \mathcal{T}_s . Cette nouvelle règle prend le risque de ralentir la convergence comme prix à payer pour se prévenir de biais éventuels dans l'estimation du filtre optimal.

13.2.1. Seuil minimal de la puissance spectrale par sous-bande

A une itération s donnée, si $\mathcal{P}_s^y[i]$ a une valeur très faible, le signal ne contient vraisemblablement pas d'information pertinente à cette $i^{\text{ème}}$ composante de la THD et on peut considérer que le système n'est pas excité de manière significative à cette composante du signal. Cependant, si l'on se réfère à l'équation (13.2), de faibles valeurs de $\mathcal{P}_s^y[i]$ entraînent une mise à jour plus importante du filtre adaptatif qui peut être inappropriée et éloigner les coefficients du filtre des valeurs optimales. Afin de se préserver de ce genre d'artefacts tout en maintenant des capacités de poursuite convenables, nous proposons, ici, une définition alternative de \mathcal{T}_s . La nouvelle stratégie consiste à empêcher de trop faibles coefficients de normalisation en considérant un seuil minimal par sous-bandes.

Les $M/2+1$ premiers coefficients de \mathcal{P}_s^y sont regroupés en un nombre K_B de sous-bandes de même taille¹ L_B où K_B et L_B sont deux entiers naturels tels que $M/2 = K_B \cdot L_B$.

1. à l'exception d'une des bandes de taille supérieure $L_B + 1$

Soit \mathcal{B}_b l'ensemble des indices i , $i \in [0; M/2 + 1]$ appartenant à la sous-bande b , $b \in K_B$. On définit alors le vecteur $\overline{\mathcal{P}}_s^y$ dont les coefficients sont les moyennes respectives des coefficients de \mathcal{P}_s^y dans chaque sous-bande. Soit i un indice quelconque sur l'ensemble $[0; M/2 + 1]$, soit b la sous-bande \mathcal{B}_b c'est-à-dire :

$$\forall b \in [0; K_B - 1], \forall i \in \mathcal{B}_b, \overline{\mathcal{P}}_s^y[i] = \text{moyenne} \{ \mathcal{P}_s^y[j] \}_{j \in \mathcal{B}_b} \quad (13.8)$$

La figure 13.1 illustre cette définition de $\overline{\mathcal{P}}_s^y$ par rapport à \mathcal{P}_s^y .

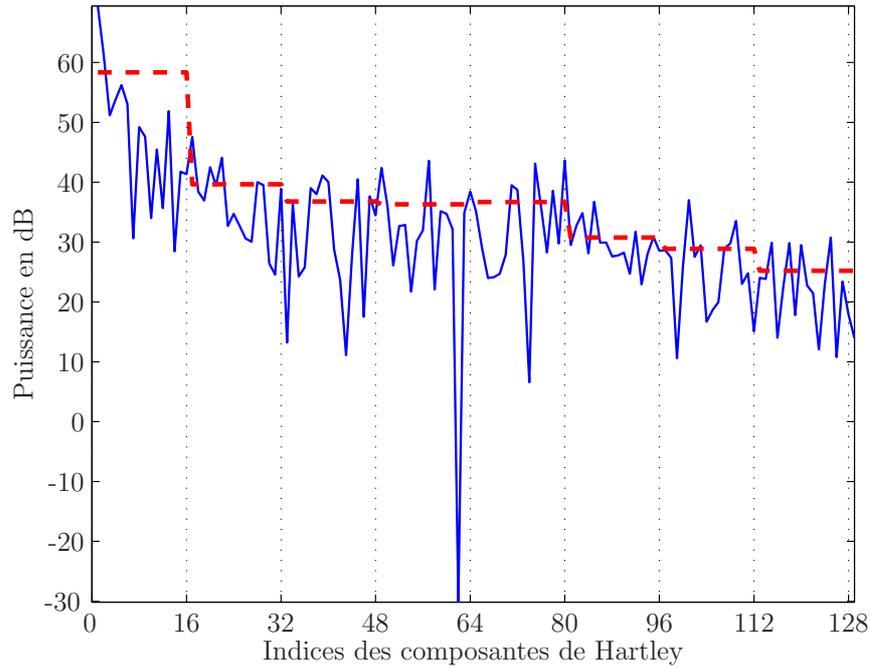


FIG. 13.1: Définition de la puissance moyenne par sous-bandes. $\overline{\mathcal{P}}_s^y$ est représenté en trait pointillé. Ses coefficients sont définis comme la moyenne des coefficients de \mathcal{P}_s^y (représenté en trait plein) pour chaque sous-bande. Du fait de la symétrie, on ne représente, ici, que les $M/2 + 1$ premiers coefficients.

A partir de cette valeur moyenne et dans le but de ne pas normaliser le gradient par une valeur de puissance trop faible, $\overline{\mathcal{P}}_s^y$ est utilisé comme seuil minimum et \mathcal{I}_s est défini par :

$$\forall i \in [0; M/2 + 1], \mathcal{I}_s[i] = \frac{1}{\max(\mathcal{P}_s^y[i], \overline{\mathcal{P}}_s^y[i])} \quad (13.9)$$

13.2.2. Prise en compte de la puissance spectrale de l'erreur

Dans les systèmes récents d'annulation d'écho, des détecteurs de double parole sont généralement utilisés afin de stopper ou de ralentir le processus d'adaptation lorsque le signal issu du locuteur éloigné est trop faible par rapport à la voix du locuteur proche.

Dans le contexte des prothèses auditives, la complexité algorithmique des détecteurs de double parole est trop importante relativement aux faibles capacités de calcul des DSP. Néanmoins, la corrélation entre le signal d'écho et le signal source étant élevée, il est nécessaire d'adopter une stratégie afin de ralentir le processus d'adaptation lorsque la puissance du signal d'erreur est trop importante.

Considérons le cas où le signal d'erreur, $\tilde{\mathcal{E}}_s$, atteint un niveau élevé en puissance. On peut alors supposer que l'on se trouve dans l'une des situations suivantes :

- (A) *le signal source $s[n]$ est bien présent* et est considéré comme une erreur par l'algorithme adaptatif. Dans ce cas il est nécessaire de ralentir le processus d'adaptation.
- (B) *le signal source n'est pas présent ou très faible*. Il apparaît donc que la forte puissance du signal d'erreur est bien due à une mauvaise estimation des coefficients du filtre. Le processus d'adaptation doit alors être favorisé.

On constate donc qu'un tel cas est très délicat puisque suivant l'interprétation de la cause, on est amené à prendre deux décisions contraires. Comme il est impossible, par ailleurs, de déterminer dans quelle situation on se trouve, nous proposons d'adopter une approche prudente pour pallier ce problème. La stratégie choisie est de pondérer le vecteur de normalisation \mathcal{T}_s à la fois par la puissance du signal de sortie $\tilde{\mathcal{Y}}_s$ et par celle du signal d'erreur $\tilde{\mathcal{E}}_s$.

Dans la situation (A), cette approche convient tout à fait puisqu'elle tend à ralentir le processus d'adaptation. Dans la situation (B), malgré le fait qu'on ne ralentisse pas l'adaptation, la normalisation choisie n'est cependant pas complètement dénuée d'intérêt. En effet, la puissance du signal source est alors très faible, et cela revient à normaliser le gradient par le signal d'erreur ce qui peut sembler cohérent.

Comme précédemment pour \mathcal{P}_s^y , nous définissons la puissance du signal d'erreur dans le domaine de la THD par :

$$\mathcal{P}_{s+1}^e = \gamma \mathcal{P}_s^e + (1 - \gamma) \left| \tilde{\mathcal{E}}_s[l] \right|^2 \quad (13.10)$$

— dans le domaine fréquentiel :

$$\mathcal{P}_{s+1}^e = \gamma \mathcal{P}_s^e + (1 - \gamma) \left(\tilde{\mathcal{E}}_s^* \odot \tilde{\mathcal{E}}_s \right) \quad (13.11)$$

— dans le domaine de la THD :

$$\mathcal{P}_{s+1}^e = \gamma \mathcal{P}_s^e + (1 - \gamma) \frac{1}{2} \left(\tilde{\mathcal{E}}_s \odot \tilde{\mathcal{E}}_s + \mathbf{J}_M \left(\tilde{\mathcal{E}}_s \odot \tilde{\mathcal{E}}_s \right) \right) \quad (13.12)$$

— enfin, pour le HD-GMDF α :

$$\mathcal{P}_{s+1}^e = \gamma \mathcal{P}_s^e + (1 - \gamma) \frac{1}{2} \left(\tilde{\mathcal{E}}_s^0 \odot \tilde{\mathcal{E}}_s^0 + \mathbf{J}_M \left(\tilde{\mathcal{E}}_s^0 \odot \tilde{\mathcal{E}}_s^0 \right) \right) \quad (13.13)$$

On définit également le vecteur de puissance moyenne par sous-bandes, $\overline{\mathcal{P}}_s^e$, tel que :

$$\forall b \in [0; K_B - 1], \forall i \in \mathcal{B}_b, \overline{\mathcal{P}}_s^e[i] = \text{moyenne} \{ \mathcal{P}_s^e[j] \}_{j \in \mathcal{B}_b} \quad (13.14)$$

Le vecteur de normalisation \mathcal{T}_s prend alors la définition alternative :

$$\mathcal{T}_{s[i]} = 1 / \left(\max \left(\mathcal{P}_{s[i]}^y, \overline{\mathcal{P}}_{s[i]}^y \right) + \rho \cdot \max \left(\mathcal{P}_{s[i]}^e, \overline{\mathcal{P}}_{s[i]}^e \right) \right) \quad (13.15)$$

où ρ est un paramètre de pondération de l'importance relative des puissances des signaux excitateur et d'erreur dans la normalisation.

14. Résultats - Simulations

14.1. Motivations

Comme nous l'avons vu dans la partie théorique, les conditions de mise en oeuvre d'un filtre adaptatif pour annuler l'écho sont particulièrement difficiles dans une prothèse auditive.

Afin de vérifier l'efficacité de la méthode proposée, il convient donc d'imposer des conditions de simulation rigoureuses et réalistes.

Pour effectuer ce travail de validation, nous avons, dans un premier temps, mené une série de simulations à l'aide du logiciel Matlab. L'ensemble des programmes écrits à cet effet permet de simuler de manière réaliste le comportement d'une prothèse auditive.

Comme nous allons le voir plus loin, ces premières expériences ont permis de vérifier le bon comportement de la méthode proposée face au problème de bouclage acoustique. Il convenait donc dans une seconde série d'expériences de confronter l'algorithme à une situation réelle.

Pour ce faire, nous disposions d'un prototype de prothèse de type *contour d'oreille* que nous avons couplé à un ordinateur personnel réalisant le traitement du signal.

14.2. Simulations Matlab

14.2.1. Description de l'environnement

Pour cette première série de simulations, nous avons mis en place un ensemble de programmes et de fonctions permettant de simuler à la fois le comportement de la prothèse et l'impact de l'environnement acoustique.

Ces fonctions permettent de simuler le système de la prothèse auquel on fournit le signal d'entrée et qui le transforme pour générer le signal de sortie. La fonction de transfert de

la prothèse pour ces expérimentations a été modélisée par un simple gain statique (c'est-à-dire sans prise en compte de la compression).

L'environnement acoustique est pris en compte à travers le chemin d'écho modélisé par un filtre à réponse impulsionnelle finie de 128 échantillons. Les simulations sont réalisées à la fréquence d'échantillonnage de 16 kHz .

On dispose de deux modèles distincts de chemin d'écho (cf. figures 14.1) issus de mesures effectuées sur des prothèses de type *intra-canal*.

Afin de simuler le système en boucle fermée, le signal de sortie est donc filtré à l'aide d'un des modèles de chemin d'écho et ajouté au signal d'entrée.

Les réponses impulsionnelles modélisant l'écho sont normalisés de telle sorte que leur gain maximum soit de 0 dB . Ainsi, lorsque la prothèse est réglée pour apporter un gain unité, le système se trouve à la limite des conditions suffisantes de stabilité (9.11).

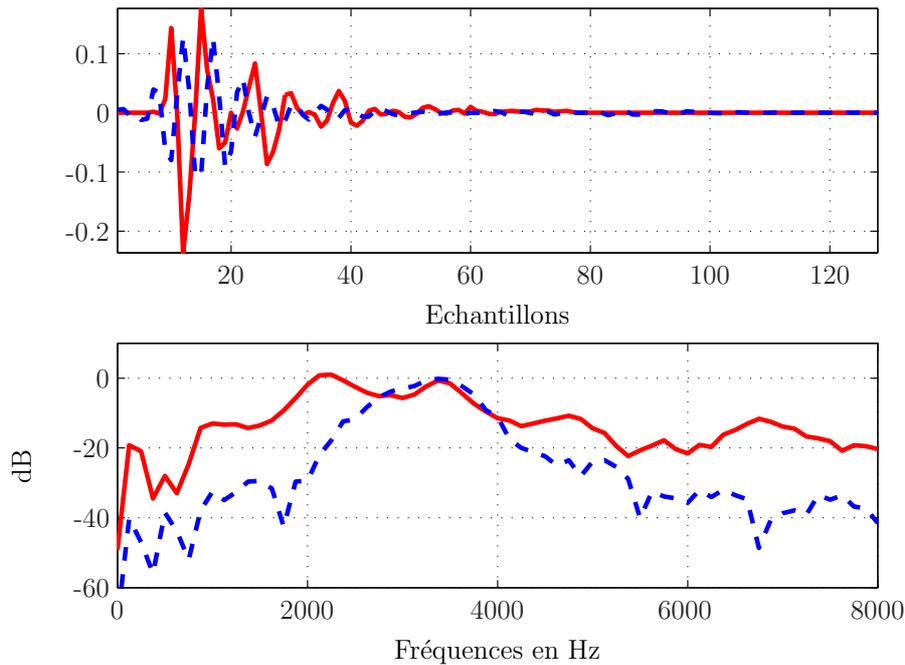


FIG. 14.1: Modèles de chemin d'écho utilisés pour les simulations. En haut, réponses impulsionnelles. En bas, leur spectre en puissance respectif.

En sortie de prothèse, le signal est écrêté afin de limiter le niveau du signal de sortie.

14.2.2. Mise en oeuvre du HD-GMDF

Lors de simulations précédentes, non présentées ici, nous avons préalablement vérifié la supériorité du HD-GMDF α par rapport aux implémentations classiques du Block-LMS

temporel.

Dans l'expérience présentée ici, nous comparons l'implémentation classique HD-GMDF α et l'implémentation proposée dans le chapitre précédent utilisant une nouvelle méthode de calcul du pas d'adaptation.

Afin de simplifier la lecture du document, nous ferons désormais référence aux deux implémentations testées lors de ces expériences sous les dénominations d'*implémentation A* et d'*implémentation B*.

14.2.3. Résultats

Le signal test utilisé est un extrait de 5s d'un fichier de parole échantillonné à 16 kHz. Pendant la simulation deux modèles de réponses impulsionnelles sont utilisés alternativement et on bascule du premier au deuxième au milieu de la simulation. Ceci nous permet de confronter l'algorithme à un changement brutal d'environnement acoustique et d'observer ses qualités d'adaptation.

Pour ces expériences, le gain statique de la prothèse a été fixé respectivement à 5 dB et à 15 dB. Pour ces deux expériences, les résultats sont retranscrits à l'aide des spectrogrammes des signaux de sortie représentés sur les figures 14.2 et 14.3.

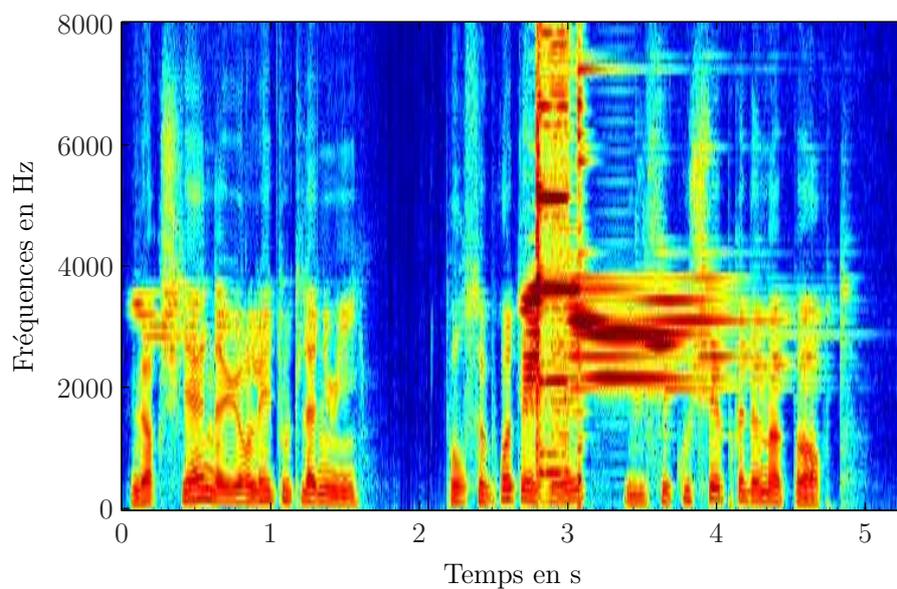
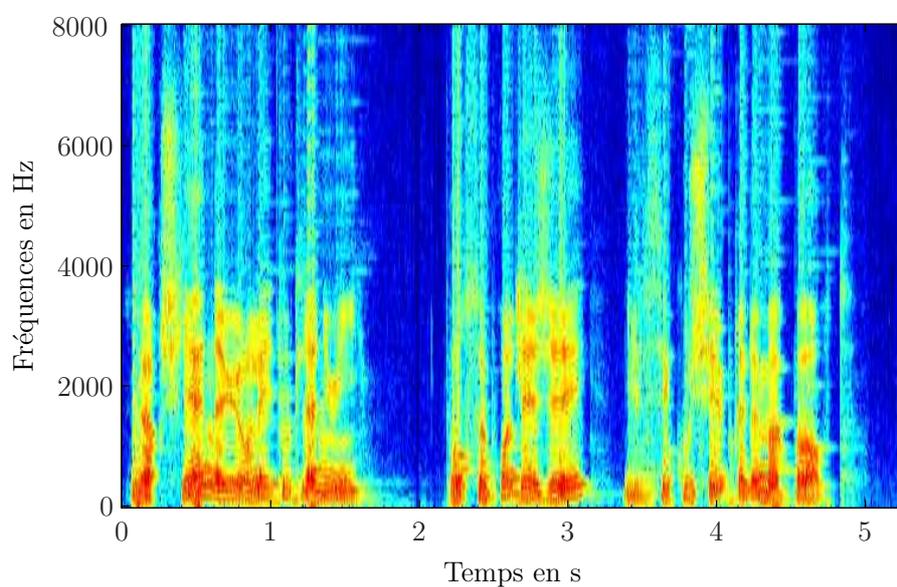
Gain statique à 5 dB

Dans cette première série de simulations, le gain de la prothèse est fixé à 5dB. La figure 14.2 présente les résultats obtenus par comparaison des spectrogrammes des signaux de sorties.

Le système fonctionne donc 5 dB au dessus des conditions de stabilité suffisante. Si le filtre adaptatif n'est pas mis en oeuvre, le système est fortement instable et l'effet Larsen se produit.

L'implémentation A parvient à rendre le système stable jusqu'au basculement de modèle de chemin d'écho. A partir de l'instant de basculement, le système devient instable et ne parvient à se stabiliser qu'au bout de 2s environ. (*cf.* figure 14.2 (a))

Pour l'implémentation B, dans les mêmes conditions de simulations, la stabilité est maintenue même à l'instant de la modification de chemin d'écho. Aucun phénomène d'auto-oscillation n'est observable comme le montre la figure 14.2 (b).

(a) $HD - GMDF\alpha$ (b) $HD - GMDF\alpha$ avec le nouveau procédé de normalisation.FIG. 14.2: Spectrogrammes des signaux en sortie de prothèse pour $|G(z)| = 5 \text{ dB}$.

Gain statique à 15 dB

Dans une seconde série de simulations, le gain de la prothèse est fixé à 15 dB.

La figure 14.3 présente les résultats obtenus par comparaison des spectrogrammes des signaux de sortie.

Avec un réglage aussi important du gain, l'implémentation A ne parvient absolument pas à compenser le phénomène de retour acoustique et à stabiliser le système. Sur la figure 14.3, on peut en effet constater que l'instabilité provoque des auto-oscillations. La présence de nombreuses harmoniques est due à l'écrêtage du signal de sortie qui induit des non-linéarités et donc de l'*aliasing*.

On peut d'ailleurs constater que la fréquence d'accrochage reste fixe, le filtre adaptatif ne parvient pas à converger vers la solution.

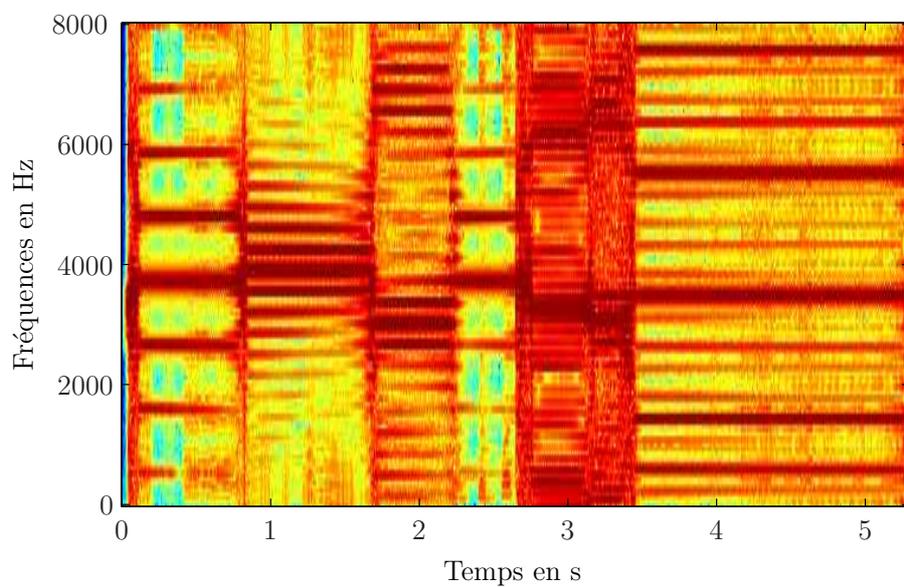
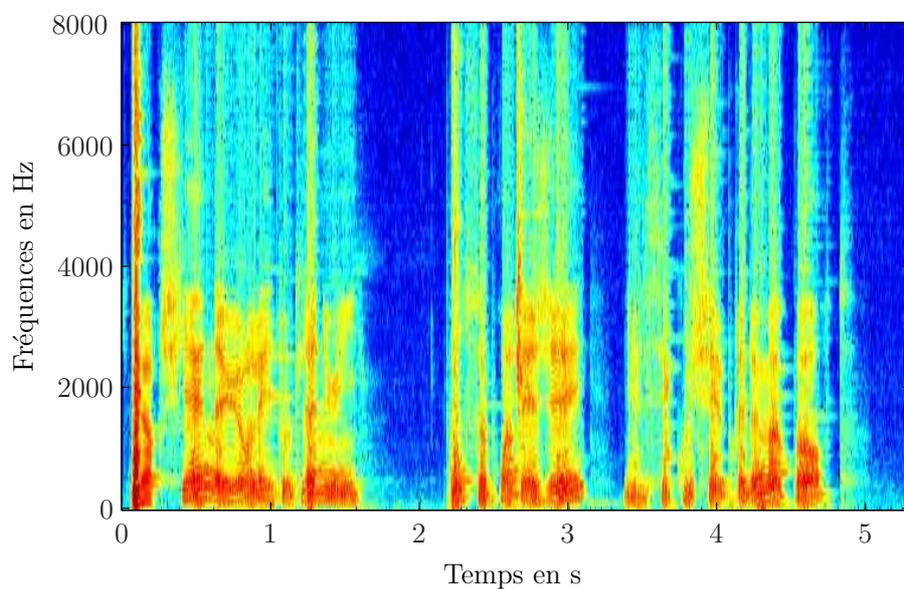
Dans les mêmes conditions de test, l'implémentation B parvient, quant à elle, à stabiliser le système. Le comportement du filtre adaptatif est comparable à celui observé avec le réglage du gain à 5 dB, aucune auto-oscillation ne se produit, même lors du basculement de modèle de chemin d'écho. La qualité sonore reste aussi du même ordre que dans les conditions précédentes.

Analyse des résultats

A la lumière de ces résultats préliminaires, la procédure de normalisation du gradient semble améliorer sensiblement les performances du HD-GMDF α . L'amélioration est constatée d'une part au niveau du temps de convergence comme le montre l'absence de problème lors du basculement de modèle de chemin d'écho et d'autre part au niveau de la stabilisation du système. D'après ce que nous avons vu dans la partie théorique, si le système est maintenu stable c'est que le filtre adaptatif a convergé vers la réponse impulsionnelle du chemin d'écho.

Des résultats similaires ont été observés avec différentes configurations. Notamment, lors de ces simulations, si on introduit un module de compression du signal, des valeurs de gain plus élevées peuvent être utilisées sans que le système soit victime de l'effet Larsen. En effet, les non-linéarités introduites alors au niveau du module de compensation des pertes auditives diminuent la corrélation entre le signal source et le signal d'écho ce qui améliore l'estimation du chemin d'écho.

Ces expériences ont permis de constater que le gain du système peut être fixé à de très hautes valeurs (relativement à la zone de stabilité du système sans compensation de l'écho) lorsque la nouvelle procédure de normalisation est utilisée sans que le système ne devienne instable. Toutefois, à partir de valeurs de gain supérieures à 15 dB, le signal en

(a) $HD - GMDF\alpha$ (b) $HD - GMDF\alpha$ avec le nouveau procédé de normalisation.FIG. 14.3: Spectrogrammes des signaux en sortie de prothèse pour $|G(z)| = 15\text{ dB}$.

sortie est de mauvaise qualité.

14.3. Implémentation en temps-réel

14.3.1. Environnement de test

Suite aux bons résultats obtenus lors des simulations sous MATLAB présentés précédemment, il convenait de vérifier sa mise en œuvre dans des conditions plus réalistes et notamment l'efficacité de poursuite pour des chemins d'échos variant dans le temps.

Un environnement de test en temps-réel a donc été mis en place.

Le système de test est composé d'un ordinateur personnel équipé d'une carte son de qualité professionnelle dont l'entrée et la sortie sont respectivement reliées au microphone et à l'écouteur d'un prototype de prothèse auditive dont le microprocesseur a été ôté. Le prototype est placé sur un moulage d'oreille en latex. Le PC réalise l'ensemble du traitement du signal de la prothèse.

L'algorithme d'annulation d'écho a été implémenté en langage C en s'appuyant sur l'environnement de programmation audio-numérique *PortAudio* [5].

Le programme fonctionne à la fréquence d'échantillonnage de 22050 kHz et les signaux sont restitués et acquis sur 16 bits. Cependant le traitement interne est effectué au format réel (flottant 32 bits).

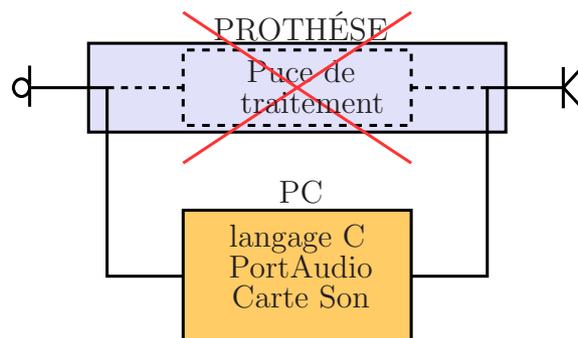


FIG. 14.4: Schéma de principe de l'implémentation temps-réel

La prothèse utilisée est un *contour d'oreille*. Afin de faciliter l'analyse des résultats, seules les fonctions d'amplification et de compression du signal acoustique sont utilisées.

Les performances de l'algorithme sont évaluées par la mesure de la *marge de gain* au dessus du *gain critique* que le système autorise, le gain critique étant la valeur limite du

gain de l'appareil au-delà de laquelle le système est instable. La différence de gain critique obtenue avec ou sans l'annulateur d'écho donne la marge de gain.

14.3.2. Résultats

Pour l'estimation des performances de l'annulateur d'écho en temps-réel, le prototype de prothèse est placé dans un environnement sans obstacle. Le signal source utilisé dans cette simulation est un signal de parole de 15 secondes émis par un haut-parleur situé à une distance de 1m de la prothèse. Afin d'avoir une référence pour estimer la précision de la convergence du filtre adaptatif, la réponse impulsionnelle du chemin d'écho est auparavant mesurée à l'aide d'une méthode d'identification utilisant des codes de Golay (cf. [26, 80]). L'annexe B décrit brièvement cette méthode de mesure. A l'instant initial, les coefficients du filtre adaptatif sont arbitrairement fixés à des valeurs nulles afin de vérifier la capacité du filtre adaptatif à converger. Les paramètres de configuration de l'algorithme sont indiqués dans la Table 14.1.

Taille du filtre adaptatif	128 échantillons
Taille des sous-blocs	64 échantillons
Nombre de sous-blocs	2
Taux de recouvrement entre fenêtres	2

TAB. 14.1: Configuration de l'algorithme pour l'expérience.

Expérience sans compression du signal

Dans une première série de tests, le seul traitement interne de la prothèse est une amplification linéaire du signal et une limitation afin de restreindre le niveau de sortie à des valeurs sur 16bits.

Dans cette configuration, le gain critique de fonctionnement sans annulation d'écho est de $24,4 \text{ dB}$ à la fréquence de 6150 Hz . Cette valeur a été déduite de la mesure de référence et confirmée lors de l'expérience.

L'utilisation de l'algorithme d'annulation d'écho a permis d'utiliser la prothèse avec une amplification de 45 dB soit une marge de gain de $20,6 \text{ dB}$.

A l'intérieur de cette marge de gain aucun phénomène de Larsen ne se produit et la qualité sonore reste acceptable malgré localement quelques distorsions audibles dues à de brèves divergences du filtre adaptatif dans les périodes de silence. Dès les premiers instants de l'expérience, aucune instabilité n'est observée ce qui signifie que l'algorithme converge très rapidement.

Pour des valeurs de gain au-delà de cette valeur, les distorsions deviennent de plus en plus fréquentes et de plus en plus gênantes, cependant jusqu'à un niveau très élevé (gain à 60 dB), on ne note toujours pas d'accrochage du système autour d'une fréquence particulière. A ces niveaux de gains, il faut cependant noter que le signal en sortie de prothèse est fortement saturé et qu'il devient donc difficile de juger de l'efficacité de l'algorithme.

La figure 14.5 montre les très bons résultats de convergence du filtre adaptatif qui parvient de manière très précise à modéliser la réponse en fréquence de référence sauf dans la région des basses fréquences du fait de la très forte atténuation du signal d'écho dans cette plage de fréquences. D'après cette figure, la condition suffisante de stabilité de l'équation (9.11) est donc préservée.

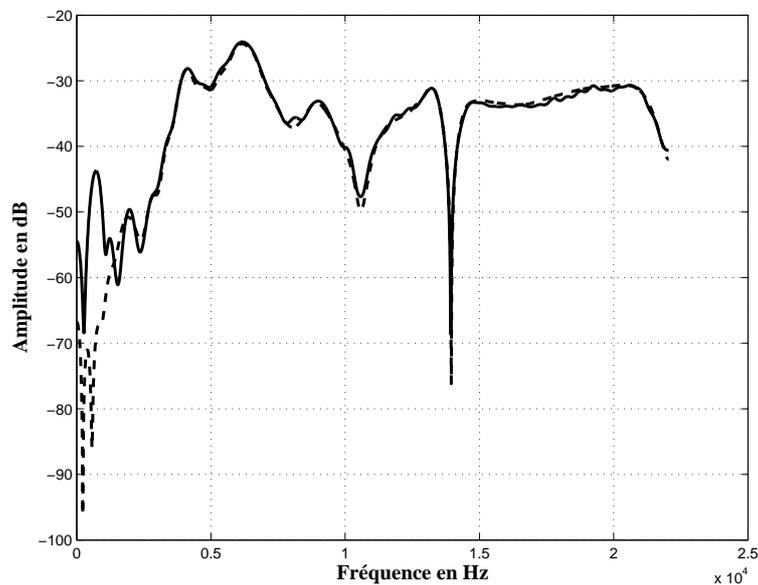


FIG. 14.5: Convergence en fréquence du filtre adaptatif. Comparaison entre la réponse en fréquence du chemin d'écho (en pointillés) et la réponse en fréquence du filtre adaptatif (en trait plein) en cours d'expérience sous la configuration sans compression.

Expérience avec compression du signal

Dans une seconde série de test, un module de compression est ajouté en sortie de l'étage d'amplification.

Dans cette configuration, le gain critique de fonctionnement sans annulation d'écho observé lors de l'expérience est de 38 dB .

L'utilisation de l'algorithme d'annulation d'écho a permis dans de telles conditions d'obtenir une marge de gain de 30 dB , cependant le signal en sortie exhibe plus de distorsions locales que dans la configuration sans compression.

A nouveau, pour cette série de tests, aucun phénomène de Larsen n'a été observé pour des niveaux de gains plus élevés, mais à ces niveaux le signal devient particulièrement corrompu par les distorsions dues aux divergences locales du filtre adaptatif. De même que précédemment, aucune instabilité n'est audible durant les premiers instants de l'expérience.

La figure 14.6 montre les bonnes performances de convergence de l'algorithme. On note que l'estimation est moins fine que dans la première configuration.

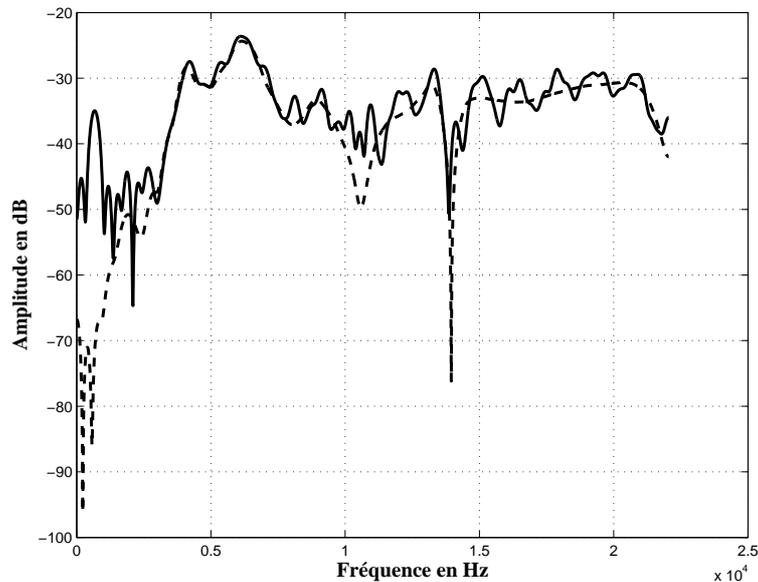


FIG. 14.6: Convergence en fréquence du filtre adaptatif. Comparaison entre la réponse en fréquence du chemin d'écho (en pointillés) et la réponse en fréquence du filtre adaptatif (en trait plein) en cours d'expérience sous la configuration avec compression.

Les distorsions locales et brèves du son causées par des divergences du filtre adaptatif notamment pendant les périodes de silence sont un phénomène courant dans cette application. Dans [45], J.M. Kates propose plusieurs méthodes pour éviter ces artefacts et restreindre les coefficients du filtre adaptatif à des valeurs cohérentes. Ces solutions n'ont pas été implémentées dans le cadre de cette étude afin, dans un premier temps, de pouvoir identifier clairement la pertinence de l'utilisation du HD-GMDF α . Néanmoins, dans de futurs travaux, des méthodes complémentaires seront envisagées pour pallier ces artefacts.

Enfin, une dernière série de tests a permis de vérifier les bonnes capacités de poursuite de l'algorithme. Pendant l'expérience, un obstacle est approché de la prothèse pour modifier le chemin d'écho. Le système utilisé opère avec l'étage de compression et le gain est réglé à 10 dB au dessus de la marge de gain. Sans annulation d'écho, on peut observer une modification de la fréquence d'accrochage et un renforcement de l'amplitude des oscillations. Ces résultats sont présentés sur le spectrogramme de la figure 14.7. Dans la même expérience effectuée avec l'annuleur d'écho, la transition entre les deux situations se fait sans perturbations audibles comme le montre le spectrogramme de la figure 14.8.

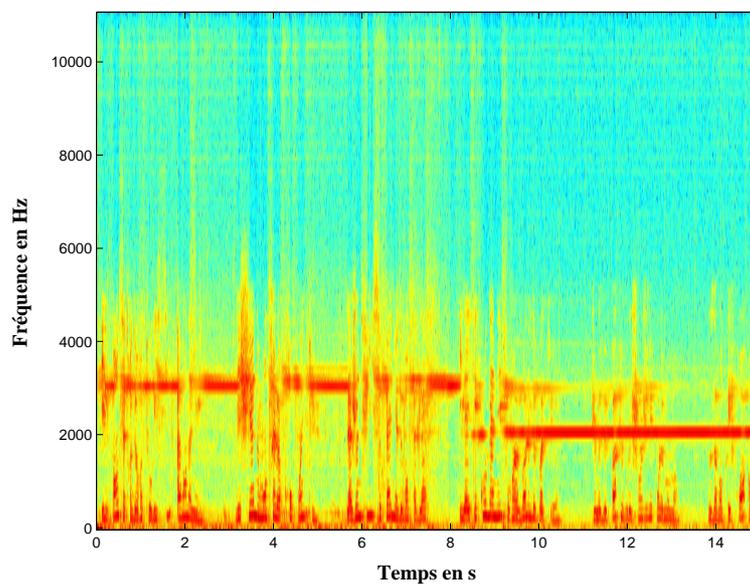


FIG. 14.7: Spectrogramme du signal en sortie de prothèse sans annulation d'écho. Un obstacle est approché de la prothèse au bout de 7s provoquant un changement de la fréquence d'accrochage du Larsen.

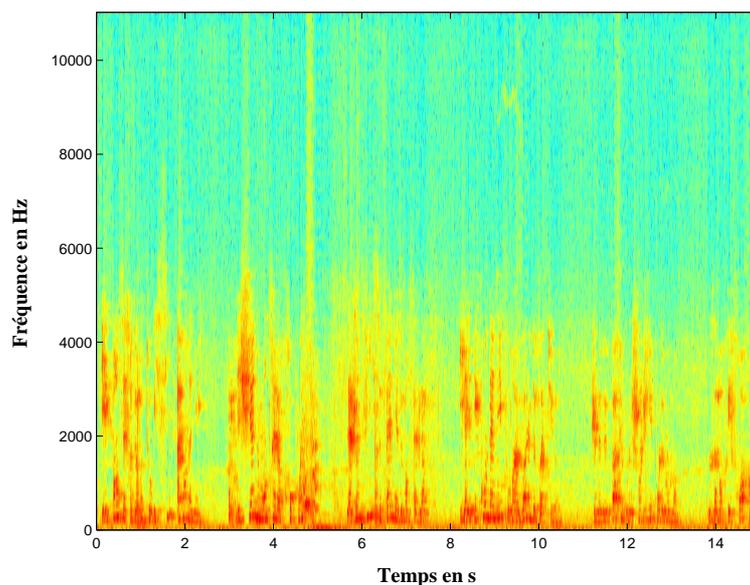


FIG. 14.8: Spectrogramme du signal en sortie de prothèse avec annulation d'écho. Un obstacle est approché de la prothèse au bout de 7s, cependant aucun phénomène de Larsen n'est observé.

14.3.3. Analyse des résultats

Dans les deux configurations étudiées, l'algorithme proposé a permis d'apporter une marge de gain supplémentaire importante. Dans les deux cas, le gain critique était donné par la valeur au delà de laquelle la qualité sonore est gravement corrompue par de brèves divergences du filtre adaptatif sans toutefois qu'un phénomène de Larsen soit observé. La solution adoptée semble donc bien adaptée à la situation particulière de l'annulation d'écho au sein d'une prothèse auditive. En outre, la modification du pas d'adaptation proposée dans le chapitre 13 améliore sensiblement les capacités de convergence du HD-GMDF α .

Par ailleurs, les conditions réalistes dans lesquelles ces simulations ont été menées, ont permis de faire une analyse précise et rigoureuse du problème qui pourra éventuellement mener à plusieurs améliorations de la méthode, notamment du point de vue de la qualité sonore afin de se prévenir des divergences locales du filtre adaptatif.

Néanmoins quelques améliorations restent à apporter du point de vue de la qualité sonore. On observe notamment un problème de divergence locale du filtre adaptatif dans les instants qui suivent les périodes de silence. De même un problème d'estimation semble se poser dans les basses fréquences. A l'origine de ces problèmes, se trouve probablement une même cause. En effet, dans les deux cas, ces divergences se produisent à des instants ou dans des zones du spectre où le signal d'écho est faible. Le signal de retour capté au niveau du microphone est sans doute trop faible et se trouve fortement corrompu par le bruit ambiant ou même le bruit de quantification sur 16 bits. Donc, cela biaise vraisemblablement l'estimation du chemin d'écho. Il faudrait prendre en compte ce phénomène dans la mise à jour.

Conclusion de la troisième partie

Conclusion

Comme cela a été plusieurs fois souligné au cours de cette troisième et dernière partie, le bouclage acoustique dans les prothèses auditives est un problème très gênant et très difficile à résoudre. Dans la continuité des travaux scientifiques récents sur le sujet, nous avons proposé d'utiliser un algorithme de filtrage adaptatif. En particulier, nous avons opté pour le HD-GMDF α qui nous semble particulièrement adapté à la situation même s'il a été conçu à l'origine pour une application *a priori* très différente. Afin d'améliorer les capacités de convergence de l'algorithme, nous avons proposé une nouvelle méthode de gestion du pas d'adaptation.

Les simulations ont permis de montrer que la méthode proposée semblait convenir et les expériences en temps réel ont confirmé cette première observation. Les contraintes d'utilisation réelle étant très fortes, nous nous sommes attachés à réaliser ces expériences dans des conditions tout à fait réalistes. L'utilisation du HD-GMDF α avec le nouveau pas d'adaptation réalise un compromis équilibré entre précision de l'estimation et capacité de poursuite.

Les résultats sont donc très satisfaisants et très encourageants. Avec l'algorithme proposé, il est possible de faire fonctionner les prothèses auditives avec un gain d'amplification largement supérieur au gain critique.

Perspectives

Problème des signaux faibles

Si les résultats obtenus sont très encourageants, le sujet reste néanmoins largement ouvert à d'autres améliorations. Notamment, un travail de recherche doit être mené pour limiter la divergence du filtre adaptatif après les périodes de silence ou lorsque le signal est

d'un niveau inférieur à celui du bruit de quantification. La difficulté consiste à parvenir à ce résultat tout en maintenant les capacités de poursuite.

Pas d'adaptation

Du point de vue de la gestion du pas d'adaptation, une autre possibilité de perfectionnement est possible. Nous avons précédemment vu que le passage du LMS par bloc temporel au LMS rapide permet d'ajouter un degré de liberté supplémentaire pour la définition du pas d'adaptation. La normalisation dans le domaine transformé est alors possible. De même, l'utilisation d'un filtre à délais multiples dans le HD-GMDF α offre un degré de liberté supplémentaire. Il devient en effet possible de spécifier un comportement différent pour chaque sous-bloc du filtre adaptatif. Il est envisageable d'améliorer le processus d'adaptation si on autorise un réglage différent entre le début et la fin de la réponse impulsionnelle. Ainsi pourrait-on optimiser, pour chaque sous-bloc, le compromis entre précision de l'estimation et capacité de poursuite. Il est plausible, par exemple, que les premiers échantillons de la réponse impulsionnelle du chemin d'écho varient moins que les derniers qui seraient plus dépendants de l'environnement acoustique extérieur à la prothèse. D'autre part, les premiers échantillons sont plus énergétiques. A ce titre, ils influent de manière prépondérante sur les conditions de mise en oeuvre de l'effet Larsen. Il semble alors judicieux de les estimer avec une plus grande précision. Si ces hypothèses sont valables, il convient d'opter pour un pas d'adaptation plus faible pour les premiers sous-blocs que pour les derniers. Tout ceci reste, bien sûr, à vérifier en menant une étude précise de la variabilité des réponses impulsionnelles, mais il nous semble que ce sont des pistes très intéressantes.

Imposition de contraintes sur les coefficients du filtre

Dans ces travaux, J.M. Kates propose de restreindre les coefficients du filtre adaptatif à une certaine plage de valeurs (*cf.* [45]). Notamment, les coefficients ne peuvent pas atteindre de trop grandes valeurs. Ceci permet d'éviter que le filtre ne diverge lorsque le signal de référence et le signal excitateur sont corrélés ou lors des périodes de silence. Ce choix est très pertinent et mérite d'être pris en considération dans de prochains travaux sur le HD-GMDF α . Par contre, il présente l'inconvénient d'introduire un *a priori* sur les valeurs des coefficients du filtre. L'utilisation de cette technique requiert de définir avec justesse la plage de valeurs autorisée et pour cela, une étude précise sur la dynamique des valeurs des coefficients de la réponse impulsionnelle du chemin d'écho est nécessaire.

Prise en compte d'un filtre fixe

Les perspectives de recherche des deux paragraphes précédents requièrent que la plage de variation du chemin d'écho soit étudiée afin de pouvoir ajuster les hypothèses de travail au plus près. Si les réponses impulsionnelles qui modélisent le bouclage acoustique d'une prothèse auditive varient largement, il est possible par contre qu'elles partagent des caractéristiques communes, comme, par exemple, certains modes de résonance de l'appareil ou de l'événement. Dans l'hypothèse que de telles caractéristiques existent, il serait alors opportun qu'elles soient clairement identifiées et que le processus adaptatif ne s'y consacre pas. Il est alors peut être pertinent de considérer un filtre invariant (un filtre à réponse impulsionnelle infinie par exemple) en complément du filtre adaptatif.

Compatibilité avec le rehaussement de la parole

Dans notre contexte, les algorithmes adaptatifs dans un domaine transformé possèdent également l'avantage d'être compatibles avec les méthodes de rehaussement de la parole de type soustraction spectrale comme l'algorithme d'Ephraïm et Malah.

Il est en effet très simple de transcrire les méthodes de soustraction spectrale dans le domaine de la transformée de Hartley. En outre, les variables d'estimation de la puissance du signal bruité regroupées au sein du vecteur \mathcal{P}_s^y dans la deuxième partie correspondent au vecteur \mathcal{P}_s^e introduit dans la gestion du pas d'adaptation. Il est donc possible de bénéficier des calculs effectués dans le module d'annulation du bouclage si, à la suite, le module de réduction du bruit est monté en série.

Il serait d'autre part très intéressant de mesurer l'impact de l'utilisation de l'algorithme d'Ephraïm et Malah en sortie de l'annuleur d'écho. On peut en particulier se demander comment l'EMSR se comporte face à l'écho résiduel ou face à une auto-oscillation.

Il semble donc fortement souhaitable de mettre en série ces deux modules et de tester leur efficacité conjointe.

Implémentation en banc de filtres

Il est possible de faire un parallèle très intéressant entre le GMDF et le filtrage adaptatif par banc de filtres [22, 52]. L'implémentation à délais multiples est, en effet, assez proche en pratique d'un banc de filtres.

Comme nous avons déjà eu l'occasion de le dire dans ce document, un traitement par bancs de filtres est tout à fait justifié dans le cadre de la réhabilitation auditive.

Nous avons souligné que l'EMSR s'adapte très bien à ce type d'analyse. À la vue des résultats encourageants du HD-GMDF α , il semble possible qu'une implémentation par banc de filtres de l'annulation d'écho se révèle également probante.

Des travaux de recherche sont actuellement en cours à ce sujet au sein du laboratoire TSI de Télécom-Paris.

Injection de séquences masquées

Une autre perspective préalablement abordée est la possibilité d'injecter des séquences de bruit dans le signal émis afin de faciliter l'identification du chemin d'écho. Comme la présence de bruit diminue le rapport signal à bruit en sortie et se révèle gênante pour le malentendant, nous avons préféré écarter ce type de méthode.

Cependant la perspective de pouvoir disposer dans l'avenir de capacités de calcul accrues permet d'envisager à nouveau l'utilisation de telles séquences. Il s'agirait de les mettre en forme spectralement avant de les ajouter au signal de sortie afin qu'elles se situent sous le seuil de masquage du signal utile.

Dans un même ordre d'idée, on peut également placer ces séquences sous les seuils auditifs des malentendants. Ainsi, une zone relativement large du champ acoustique, pourrait être utilisée à cet effet.

Conclusion générale

Ce rapport de thèse a présenté notre travail de recherche.

Outre la compréhension des mécanismes auditifs normaux et pathologiques et des différentes techniques de réhabilitation possibles, nous avons abordé deux problèmes spécifiques : le rehaussement de la parole et l'annulation du retour acoustique.

Pour le rehaussement de la parole, nous avons proposé d'utiliser la règle de suppression d'Ephraïm et Malah (EMSR) qui est une méthode très efficace et dont la structure d'analyse fréquentielle peut se décliner en analyse en banc de filtres. Une telle structure est très appréciable dans le contexte des prothèses. A l'issue de ce travail, nous avons présenté une perspective d'utilisation future de l'EMSR avec un banc de filtres en échelle ERB.

Notre travail sur l'annulation du retour acoustique a abouti à la conception d'une solution originale largement basée sur le filtrage multi-délai généralisé issu du travail de thèse d'Omar Aït Amrane. L'originalité de notre solution tient dans l'utilisation de la transformée de Hartley pour une utilisation dans les prothèses auditives et dans la définition d'une nouvelle méthode de gestion du pas d'adaptation.

Les perspectives de recherche futures ont déjà été proposées en conclusion des trois parties de ce document.

A l'issue de ce rapport, nous soulignerons simplement qu'un travail important reste à faire pour s'assurer de la compatibilité des algorithmes proposés en prenant également en compte les modules de traitement du signal standards d'une prothèse auditive. Les interactions entre les différentes étapes du traitement peuvent produire des artefacts. D'autre part, il faut s'assurer que, du point de vue du malentendant, la qualité sonore finale est acceptable et exempte de tout défaut gênant.

La validation de toute nouvelle technique de traitement introduite dans une prothèse auditive nécessite la réalisation de tests d'écoute auprès de malentendants. La spécificité des déficiences auditives est telle que des normo-entendants ne peuvent pas juger de la qualité sonore d'un traitement. En effet, certaines différences de qualité perçues par un normo-entendant peuvent ne pas l'être par un malentendant. Inversement, certains artefacts qui ne seraient pas gênants pour les normo-entendants pourraient être désagréables

pour les malentendants et être perçues comme une source de fatigue par exemple.

Annexes

A. Calculs du gradient dans le domaine de Hartley

Dans cette annexe, nous présentons le détail des calculs qui permettent d'aboutir aux équations de mise à jour du HD-GMDF α . Nous établissons ces équations directement dans le domaine de la THD sans les obtenir à partir des équations de mise à jour dans le domaine temporel.

En adoptant une même procédure, il est possible d'arriver aux équations équivalentes dans le domaine fréquentiel.

Le HD-GMDF α étant une généralisation des algorithmes LMS rapides, il est possible de se ramener à des implémentations standards en considérant $K = 1$ ou $\alpha = 1$.

A.1. Expression de la fonction de coût dans le domaine de la THD

A.1.1. Expression de l'erreur dans le domaine de la THD

Dans le domaine temporel, l'expression du vecteur d'erreur à l'indice de blocs s est donné par (12.12) :

$$\mathbf{e}_s = \mathbf{x}_s - \mathbf{P}_N \cdot \mathbf{W}_H^{-1} \sum_{k=0}^{K-1} \left(\tilde{\mathcal{Y}}_s^k \bullet \tilde{\mathcal{H}}_s^k \right)$$

La THD, $\tilde{\mathcal{E}}_s$, de la séquence d'erreur est définie selon (11.35) par :

$$\tilde{\mathcal{E}}_s = \mathbf{W}_H \cdot \begin{bmatrix} \mathbf{0}_{L-1} \\ \dots\dots\dots \\ \mathbf{e}_s \end{bmatrix}$$

Par similitude avec $\tilde{\boldsymbol{\epsilon}}_s$, on construit le vecteur $\tilde{\boldsymbol{x}}_s$ tel que :

$$\tilde{\boldsymbol{x}}_s = \mathbf{W}_H \cdot \begin{bmatrix} \mathbf{0}_{L-1} \\ \vdots \\ \mathbf{x}_s \end{bmatrix} \quad (\text{A.1})$$

A partir de la matrice \mathbf{P}_N , on définit la matrice de projection \mathbf{P}'_N de dimension $(N + L - 1) \times (N + L - 1)$ telle que :

$$\mathbf{P}'_N = \begin{bmatrix} \mathbf{0}_{(L-1) \times (N+L-1)} \\ \vdots \\ \mathbf{P}_N \end{bmatrix} = \begin{bmatrix} \mathbf{0}_{L-1} & \mathbf{0}_{(L-1) \times N} \\ \vdots & \vdots \\ \mathbf{0}_{N \times (L-1)} & \mathbf{I}_{N \times N} \end{bmatrix} \quad (\text{A.2})$$

On peut alors exprimer l'erreur de blocs de manière homogène dans le domaine de la THD :

$$\tilde{\boldsymbol{\epsilon}}_s = \mathbf{W}_H \cdot \mathbf{P}'_N \cdot \mathbf{W}_H^{-1} \cdot \left(\tilde{\boldsymbol{x}}_s - \sum_{k=0}^{K-1} \tilde{\boldsymbol{y}}_s^k \bullet \tilde{\boldsymbol{H}}_s \right) \quad (\text{A.3})$$

Pour simplifier l'expression, on définit la matrice de projection \mathbf{P} telle que :

$$\mathbf{P} = \mathbf{W}_H \cdot \mathbf{P}'_N \cdot \mathbf{W}_H^{-1} \quad (\text{A.4})$$

On obtient finalement l'expression de $\tilde{\boldsymbol{\epsilon}}_s$ suivante :

$$\tilde{\boldsymbol{\epsilon}}_s = \mathbf{P} \cdot \left(\tilde{\boldsymbol{x}}_s - \sum_{k=0}^{K-1} \left(\tilde{\boldsymbol{y}}_s^k \bullet \tilde{\boldsymbol{H}}_s^k \right) \right) \quad (\text{A.5})$$

A.1.2. Expression de la fonction de coût sans contrainte

Dans le domaine temporel, nous avons défini la fonction de coût :

$$J(\mathbf{h}_s) = |\mathbf{e}_s|^2$$

De la même manière, il est possible de définir dans le domaine de la THD la fonction de coût, $\mathcal{J} \left(\{\tilde{\boldsymbol{H}}_s^k\}_{k \in [0; K-1]} \right)$. Pour être plus concis, notons plus simplement \mathcal{J}_s cette fonction de coût. Elle est définie par :

$$\mathcal{J}_s = |\tilde{\boldsymbol{\epsilon}}_s|^2 \quad (\text{A.6})$$

En développant cette expression, on obtient :

$$\mathcal{J}_s = \left| \mathbf{P} \cdot \left(\tilde{\boldsymbol{x}}_s - \sum_{k=0}^{K-1} \left(\tilde{\boldsymbol{y}}_s^k \bullet \tilde{\boldsymbol{H}}_s^k \right) \right) \right|^2 \quad (\text{A.7})$$

Cette fonction de coût est définie pour des vecteurs $\widetilde{\mathcal{H}}_s^k$ quelconques sur l'espace vectoriel $\mathbb{R}^{M \times 1}$. Elle n'est donc pas rigoureusement adaptée à notre problème puisque l'ensemble des vecteurs $\widetilde{\mathcal{H}}_s^k$ auquel nous nous intéressons est l'ensemble $\underline{\mathbf{H}}$ tel que :

$$\forall \widetilde{\mathcal{H}}_s^k \in \underline{\mathbf{H}}, \exists \mathbf{h}^k \in \mathbb{R}^L, \text{ tel que } \widetilde{\mathcal{H}}_s^k = \mathbf{W}_H \cdot \begin{bmatrix} \mathbf{h}^k \\ \dots\dots\dots \\ \mathbf{0}_{(N-1) \times 1} \end{bmatrix} \quad (\text{A.8})$$

Il s'agit de l'ensemble des vecteurs qui sont les THD des sous-blocs du filtre adaptatif.

A.1.3. Expression de la fonction de coût avec contrainte

On peut facilement construire la matrice de projection de $\mathbb{R}^{M \times 1}$ sur $\underline{\mathbf{H}}$, notée \mathbf{C} et définie par :

$$\mathbf{C} = \mathbf{W}_H \cdot \begin{bmatrix} \mathbf{I}_{L \times L} & \mathbf{0}_{L \times (N-1)} \\ \mathbf{0}_{(N-1) \times L} & \mathbf{0}_{N-1} \end{bmatrix} \cdot \mathbf{W}_H^{-1} \quad (\text{A.9})$$

A partir de cette matrice de projection, on peut alors proposer une nouvelle expression de l'erreur $\widetilde{\mathcal{E}}_s^C$ évaluée uniquement pour la projection de $\{\widetilde{\mathcal{H}}_s^k\}_{k \in [0; K-1]}$ sur $\underline{\mathbf{H}}$. La nouvelle fonction de coût associée, \mathcal{J}_s^C , a pour expression :

$$\widetilde{\mathcal{E}}_s^C = \mathbf{P} \cdot \left(\widetilde{\mathcal{X}}_s - \sum_{k=0}^{K-1} \left(\widetilde{\mathcal{Y}}_s^k \bullet \mathbf{C} \cdot \widetilde{\mathcal{H}}_s^k \right) \right) \quad (\text{A.10})$$

$$\mathcal{J}_s^C = \left| \widetilde{\mathcal{E}}_s^C \right|^2 \quad (\text{A.11})$$

Pour $\{\widetilde{\mathcal{H}}_s^k\}_{k \in [0; K-1]} \in \underline{\mathbf{H}}$, on a $\widetilde{\mathcal{E}}_s^C = \widetilde{\mathcal{E}}_s$ et

$$\mathcal{J}_s^C = J(\mathbf{h}_s)$$

A.2. Gradient de la fonction de coût sans contrainte

A.2.1. Expression de la fonction de coût

On montre facilement la propriété suivante :

$$\mathbf{P} = \mathbf{P}^T \cdot \mathbf{P} = \mathbf{P} \cdot \mathbf{P}^T$$

A partir de cette propriété, on peut exprimer développer l'expression de la fonction de coût \mathcal{J}_s .

$$\begin{aligned} \mathcal{J}_s &= \left| \mathbf{P} \cdot \left(\tilde{\mathbf{x}}_s - \sum_{k=0}^{K-1} \left(\tilde{\mathbf{y}}_s^k \bullet \tilde{\mathcal{H}}_s^k \right) \right) \right|^2 \\ &= \tilde{\mathbf{x}}_s^T \mathbf{P}^T \mathbf{P} \tilde{\mathbf{x}}_s - \tilde{\mathbf{x}}_s^T \mathbf{P}^T \mathbf{P} \sum_{k=0}^{K-1} \left(\tilde{\mathbf{y}}_s^k \bullet \tilde{\mathcal{H}}_s^k \right) - \sum_{k=0}^{K-1} \left(\tilde{\mathbf{y}}_s^k \bullet \tilde{\mathcal{H}}_s^k \right)^T \mathbf{P}^T \mathbf{P} \tilde{\mathbf{x}}_s \\ &\quad + \sum_{k=0}^{K-1} \left(\tilde{\mathbf{y}}_s^k \bullet \tilde{\mathcal{H}}_s^k \right)^T \mathbf{P}^T \mathbf{P} \sum_{k=0}^{K-1} \left(\tilde{\mathbf{y}}_s^k \bullet \tilde{\mathcal{H}}_s^k \right) \\ &= \tilde{\mathbf{x}}_s^T \mathbf{P} \tilde{\mathbf{x}}_s - 2 \sum_{k=0}^{K-1} \left(\tilde{\mathbf{y}}_s^k \bullet \tilde{\mathcal{H}}_s^k \right)^T \mathbf{P} \tilde{\mathbf{x}}_s + \sum_{k=0}^{K-1} \left(\tilde{\mathbf{y}}_s^k \bullet \tilde{\mathcal{H}}_s^k \right)^T \mathbf{P} \sum_{k=0}^{K-1} \left(\tilde{\mathbf{y}}_s^k \bullet \tilde{\mathcal{H}}_s^k \right) \end{aligned}$$

On en déduit alors l'expression du gradient $\vec{\nabla}_{\tilde{\mathcal{H}}_s^k} \mathcal{J}_s$.

$$\begin{aligned} \vec{\nabla}_{\tilde{\mathcal{H}}_s^k} \mathcal{J}_s &= -2 \left(\frac{\partial}{\partial \tilde{\mathcal{H}}_s^k} \left(\tilde{\mathbf{y}}_s^k \bullet \tilde{\mathcal{H}}_s^k \right) \right)^T \mathbf{P} \tilde{\mathbf{x}}_s + 2 \left(\frac{\partial}{\partial \tilde{\mathcal{H}}_s^k} \left(\tilde{\mathbf{y}}_s^k \bullet \tilde{\mathcal{H}}_s^k \right) \right)^T \mathbf{P} \sum_{k=0}^{K-1} \left(\tilde{\mathbf{y}}_s^k \bullet \tilde{\mathcal{H}}_s^k \right) \\ &= -2 \left(\frac{\partial}{\partial \tilde{\mathcal{H}}_s^k} \left(\tilde{\mathbf{y}}_s^k \bullet \tilde{\mathcal{H}}_s^k \right) \right)^T \mathbf{P} \left(\tilde{\mathbf{x}}_s^T - \sum_{k=0}^{K-1} \left(\tilde{\mathbf{y}}_s^k \bullet \tilde{\mathcal{H}}_s^k \right)^T \right) \\ &= -2 \left(\frac{\partial}{\partial \tilde{\mathcal{H}}_s^k} \left(\tilde{\mathbf{y}}_s^k \bullet \tilde{\mathcal{H}}_s^k \right) \right)^T \tilde{\boldsymbol{\varepsilon}}_s \end{aligned}$$

A.2.2. Calcul de la dérivée

$$\text{Calculons } \frac{\partial}{\partial \tilde{\mathcal{H}}_s^k} \left(\tilde{\mathbf{y}}_s^k \bullet \tilde{\mathcal{H}}_s^k \right)$$

Rappelons l'expression (11.32) qui définit $\widetilde{\mathcal{Y}}_s^k \bullet \widetilde{\mathcal{H}}_s^k$:

$$\widetilde{\mathcal{Y}}_s^k \bullet \widetilde{\mathcal{H}}_s^k = \frac{1}{2} \left[(\mathbf{I}_M - \mathbf{J}_M) \widetilde{\mathcal{Y}}_s^k \odot \widetilde{\mathcal{H}}_s^k + (\mathbf{I}_M + \mathbf{J}_M) \widetilde{\mathcal{Y}}_s^k \odot (\mathbf{J}_M \widetilde{\mathcal{H}}_s^k) \right]$$

Afin de calculer le gradient de manière plus directe, nous introduisons la matrice diagonale de dimension M $\mathbf{D}_{\widetilde{\mathcal{Y}}_s^k}$ qui est la matrice diagonale dont les éléments diagonaux sont les coefficients de $\widetilde{\mathcal{Y}}_s^k$. Ceci nous permet de transformer le produit de Schür en produit matriciel. On obtient alors :

$$\widetilde{\mathcal{Y}}_s^k \bullet \widetilde{\mathcal{H}}_s^k = \frac{1}{2} \left[(\mathbf{I}_M - \mathbf{J}_M) \mathbf{D}_{\widetilde{\mathcal{Y}}_s^k} \widetilde{\mathcal{H}}_s^k + (\mathbf{I}_M + \mathbf{J}_M) \mathbf{D}_{\widetilde{\mathcal{Y}}_s^k} \mathbf{J}_M \widetilde{\mathcal{H}}_s^k \right] \quad (\text{A.12})$$

On en déduit donc la dérivée :

$$\frac{\partial}{\partial \widetilde{\mathcal{H}}_s^k} \left(\widetilde{\mathcal{Y}}_s^k \bullet \widetilde{\mathcal{H}}_s^k \right) = \frac{1}{2} \frac{\partial}{\partial \widetilde{\mathcal{H}}_s^k} \left[(\mathbf{I}_M - \mathbf{J}_M) \mathbf{D}_{\widetilde{\mathcal{Y}}_s^k} \widetilde{\mathcal{H}}_s^k + (\mathbf{I}_M + \mathbf{J}_M) \mathbf{D}_{\widetilde{\mathcal{Y}}_s^k} \mathbf{J}_M \widetilde{\mathcal{H}}_s^k \right] \quad (\text{A.13})$$

$$= \frac{1}{2} \left[(\mathbf{I}_M - \mathbf{J}_M) \mathbf{D}_{\widetilde{\mathcal{Y}}_s^k} + (\mathbf{I}_M + \mathbf{J}_M) \mathbf{D}_{\widetilde{\mathcal{Y}}_s^k} \mathbf{J}_M \right] \quad (\text{A.14})$$

$$\vec{\nabla}_{\widetilde{\mathcal{H}}_s^k} \mathcal{J}_s = - \left[(\mathbf{I}_M - \mathbf{J}_M) \mathbf{D}_{\widetilde{\mathcal{Y}}_s^k} + (\mathbf{I}_M + \mathbf{J}_M) \mathbf{D}_{\widetilde{\mathcal{Y}}_s^k} \mathbf{J}_M \right]^T \widetilde{\mathcal{E}}_s \quad (\text{A.15})$$

$$= - \left[\mathbf{D}_{\widetilde{\mathcal{Y}}_s^k} (\mathbf{I}_M - \mathbf{J}_M) + \mathbf{J}_M \mathbf{D}_{\widetilde{\mathcal{Y}}_s^k} (\mathbf{I}_M + \mathbf{J}_M) \right] \widetilde{\mathcal{E}}_s \quad (\text{A.16})$$

$$= - \left[\mathbf{D}_{\widetilde{\mathcal{Y}}_s^k} - \mathbf{D}_{\widetilde{\mathcal{Y}}_s^k} \mathbf{J}_M + \mathbf{J}_M \mathbf{D}_{\widetilde{\mathcal{Y}}_s^k} + \mathbf{J}_M \mathbf{D}_{\widetilde{\mathcal{Y}}_s^k} \mathbf{J}_M \right] \widetilde{\mathcal{E}}_s \quad (\text{A.17})$$

$$= - \left[(\mathbf{I}_M + \mathbf{J}_M) \mathbf{D}_{\widetilde{\mathcal{Y}}_s^k} - (\mathbf{I}_M - \mathbf{J}_M) \mathbf{D}_{\widetilde{\mathcal{Y}}_s^k} \mathbf{J}_M \right] \widetilde{\mathcal{E}}_s \quad (\text{A.18})$$

D'où finalement en réintroduisant le produit de Schür :

$$\vec{\nabla}_{\widetilde{\mathcal{H}}_s^k} \mathcal{J}_s = - \left[(\mathbf{I}_M + \mathbf{J}_M) \widetilde{\mathcal{Y}}_s^k \odot \widetilde{\mathcal{E}}_s - (\mathbf{I}_M - \mathbf{J}_M) \widetilde{\mathcal{Y}}_s^k \odot \mathbf{J}_M \widetilde{\mathcal{E}}_s \right] \quad (\text{A.19})$$

On retrouve ici l'équivalent dans le domaine de la THD de la corrélation croisée défini par (11.36).

$$\widetilde{\mathcal{Y}}_s^k \star \widetilde{\mathcal{E}}_s = \frac{1}{2} \left[(\mathbf{I}_M + \mathbf{J}_M) \widetilde{\mathcal{Y}}_s^k \odot \widetilde{\mathcal{E}}_s - (\mathbf{I}_M - \mathbf{J}_M) \widetilde{\mathcal{Y}}_s^k \odot (\mathbf{J}_M \widetilde{\mathcal{E}}_s) \right]$$

D'où finalement :

$$\vec{\nabla}_{\widetilde{\mathcal{H}}_s^k} \mathcal{J}_s = -2 \widetilde{\mathcal{Y}}_s^k \star \widetilde{\mathcal{E}}_s \quad (\text{A.20})$$

On retrouve ici, l'expression du gradient *sans contrainte* (12.13) :

$$\widetilde{\mathcal{H}}_{s+1}^k = \widetilde{\mathcal{H}}_s^k + \mu_s \cdot \widetilde{\mathcal{Y}}_s^k \star \widetilde{\mathcal{E}}_s \quad (\text{A.21})$$

A.3. Gradient de la fonction de coût avec contrainte

A.3.1. Expression de la fonction de coût

On calcule le gradient de la fonction de coût \mathcal{J}_s^C définie précédemment dans (A.11) par :

$$\mathcal{J}_s^C = \left| \mathbf{P} \cdot \left(\tilde{\boldsymbol{x}}_s - \sum_{k=0}^{K-1} \left(\tilde{\boldsymbol{y}}_s^k \bullet \mathbf{C} \cdot \tilde{\boldsymbol{H}}_s^k \right) \right) \right|^2$$

On montre facilement la propriété suivante :

$$\mathbf{C} = \mathbf{C}^T \cdot \mathbf{C} = \mathbf{C} \cdot \mathbf{C}^T$$

A partir de ces propriétés et en développant l'expression de \mathcal{J}_s^C , on obtient de la même manière que pour le gradient sans contrainte :

$$\begin{aligned} \vec{\nabla}_{\tilde{\boldsymbol{H}}_s} J_C(\tilde{\boldsymbol{H}}_s) &= -2 \frac{\partial}{\partial \tilde{\boldsymbol{H}}_s^k} \left(\tilde{\boldsymbol{y}}_s^k \bullet \mathbf{C} \cdot \tilde{\boldsymbol{H}}_s^k \right)^T \mathbf{P} \left(\tilde{\boldsymbol{x}}_s - \sum_{k=0}^{K-1} \left(\tilde{\boldsymbol{y}}_s^k \bullet \mathbf{C} \cdot \tilde{\boldsymbol{H}}_s^k \right) \right) \\ &= -2 \frac{\partial}{\partial \tilde{\boldsymbol{H}}_s^k} \left(\tilde{\boldsymbol{y}}_s^k \bullet \mathbf{C} \cdot \tilde{\boldsymbol{H}}_s^k \right)^T \tilde{\boldsymbol{\varepsilon}}_s^C \end{aligned}$$

A.3.2. Calcul de la dérivée

Calculons $\frac{\partial}{\partial \tilde{\boldsymbol{H}}_s^k} \left(\tilde{\boldsymbol{y}}_s^k \bullet \mathbf{C} \cdot \tilde{\boldsymbol{H}}_s^k \right)$:

$$\begin{aligned} \frac{\partial}{\partial \tilde{\boldsymbol{H}}_s^k} \left(\tilde{\boldsymbol{y}}_s^k \bullet \mathbf{C} \cdot \tilde{\boldsymbol{H}}_s^k \right) &= \frac{1}{2} \frac{\partial}{\partial \tilde{\boldsymbol{H}}_s^k} \left((\mathbf{I}_M - \mathbf{J}_M) \mathbf{D}_{\tilde{\boldsymbol{y}}_s^k} \mathbf{C} \tilde{\boldsymbol{H}}_s + (\mathbf{I}_M + \mathbf{J}_M) \mathbf{D}_{\tilde{\boldsymbol{y}}_s^k} \mathbf{J}_M \mathbf{C} \tilde{\boldsymbol{H}}_s \right) \\ &= \frac{1}{2} \left((\mathbf{I}_M - \mathbf{J}_M) \mathbf{D}_{\tilde{\boldsymbol{y}}_s^k} \mathbf{C} + (\mathbf{I}_M + \mathbf{J}_M) \mathbf{D}_{\tilde{\boldsymbol{y}}_s^k} \mathbf{J}_M \mathbf{C} \right) \end{aligned}$$

L'expression transposée vaut alors :

$$\left(\frac{\partial}{\partial \tilde{\boldsymbol{H}}_s^k} \left(\tilde{\boldsymbol{y}}_s^k \bullet \mathbf{C} \cdot \tilde{\boldsymbol{H}}_s^k \right) \right)^T = -\frac{1}{2} \mathbf{C} \left((\mathbf{I}_M + \mathbf{J}_M) \mathbf{D}_{\tilde{\boldsymbol{y}}_s^k} - (\mathbf{I}_M - \mathbf{J}_M) \mathbf{D}_{\tilde{\boldsymbol{y}}_s^k} \right) \quad (\text{A.22})$$

D'où finalement :

$$\vec{\nabla}_{\widetilde{\mathcal{H}}_s^k} J(\widetilde{\mathcal{H}}_s^k) = -2\mathbf{C} \cdot \widetilde{\mathcal{Y}}_s^k \star \widetilde{\mathcal{E}}_s^C \quad (\text{A.23})$$

D'où la relation de mise à jour du filtre adaptatif directement dans le domaine de la THD :

$$\widetilde{\mathcal{H}}_{s+1}^k = \widetilde{\mathcal{H}}_s^k + \mu \cdot \mathbf{C} \cdot \widetilde{\mathcal{Y}}_s^k \star \widetilde{\mathcal{E}}_s^C$$

Cette relation fait intervenir le terme d'erreur $\widetilde{\mathcal{E}}_s^C$ défini dans (A.10). On a vu que $\widetilde{\mathcal{E}}_s^C$ et $\widetilde{\mathcal{E}}_s$ sont égaux si $\widetilde{\mathcal{H}}_s^k$ appartient à l'espace $\underline{\mathbf{H}}$. D'après la relation de mise à jour (A.3.2), il apparaît que si on initialise correctement $\widetilde{\mathcal{H}}_s^k$ pour le premier bloc $s = 0$ avec $\widetilde{\mathcal{H}}_0^k \in \underline{\mathbf{H}}$ alors par récurrence simple sur s , on obtient $\widetilde{\mathcal{H}}_s^k \in \underline{\mathbf{H}}$ pour tout indice de blocs $s \in \mathbb{N}$.

On retrouve alors bien l'expression de l'algorithme du gradient *contraint* dans le domaine de la THD (12.14) :

$$\widetilde{\mathcal{H}}_{s+1}^k = \widetilde{\mathcal{H}}_s^k + \mu_s \cdot \mathbf{C} \cdot \widetilde{\mathcal{Y}}_s^k \star \widetilde{\mathcal{E}}_s \quad (\text{A.24})$$

B. Développement d'une plate-forme de test

B.1. Motivations

A l'issue du travail de recherche préliminaire et théorique sur l'annulation du retour acoustique dans les prothèses, nous avons proposé d'utiliser le HD-GMDF α et introduit une nouvelle méthode de gestion du pas d'adaptation.

La série de simulations numériques présentées dans la section 14.2 a permis de vérifier la supériorité de la solution proposée sur les solutions classiques. Toutefois, le contexte des prothèses auditives étant un contexte particulièrement difficile, il est nécessaire, pour juger des performances de l'algorithme, de le faire fonctionner dans des conditions réelles.

A cet effet, nous avons donc mis au point une plate-forme de test permettant d'évaluer notre méthode directement sur une prothèse auditive afin de se confronter à un environnement acoustique approprié et selon des contraintes de fonctionnement en temps-réel.

B.2. Prototype de prothèse

Pour réaliser cette plate-forme, nous disposons d'un prototype de prothèse réalisé à partir d'un contour d'oreille. L'appareil est simplement constitué du corps en plastique et des deux transducteurs (microphone et écouteur). Tous les autres éléments qui constituent une prothèse ont été enlevés et un connecteur permet d'accéder aux transducteurs et de les relier à la carte-son d'un PC.

La carte son utilisée est une carte son de qualité professionnelle pouvant réaliser la quantification du signal sur 24 bits à la fréquence d'échantillonnage de 96 kHz. Dans notre cas, nous avons limité ses capacités et opté pour un fonctionnement sur 16 bits/32 kHz pour opérer de manière comparable à celui d'une prothèse.

Celle-ci dispose de deux entrées et deux sorties (droite et gauche) qui nous permettent de contrôler deux transducteurs supplémentaires en plus de ceux du prototype de pro-

thèse. Cela nous permet de pouvoir également utiliser un microphone de référence ou le microphone d'une oreille artificielle¹. Le haut-parleur permet de simuler quant à lui la présence d'une source extérieure de manière reproductible. La carte-son est reliée aux appareils de la manière suivante :

- *Sortie gauche* → écouteur de la prothèse
- *Sortie droite* → haut-parleur (source extérieur)
- *Entrée gauche* → microphone de la prothèse
- *Entrée droite* → microphone de référence ou oreille artificielle

Le schéma correspondant est illustré par la figure B.1.

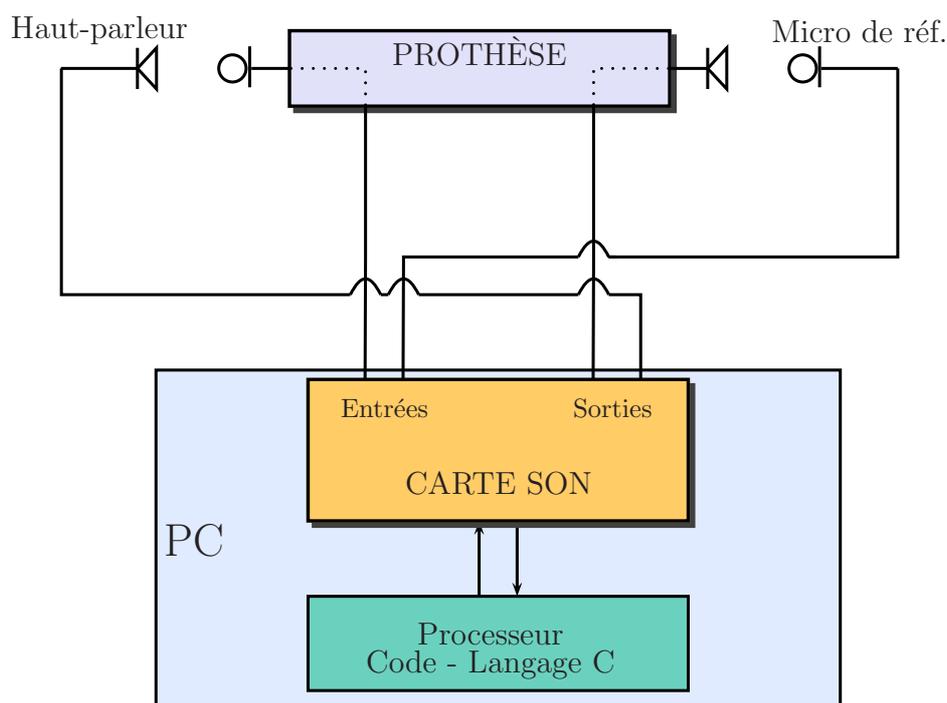


FIG. B.1: Schéma de la plate-forme de test

B.3. Calibration du matériel

La calibration du matériel est nécessaire pour assurer les équivalences de niveau acoustique entre l'émission et l'acquisition pour les différents transducteurs. De même, elle permet de connaître pour un niveau numérique donné le niveau acoustique correspondant. Avant de calibrer le prototype de prothèse, il faut d'abord calibrer le matériel de

1. Une oreille artificielle est un système composé d'un microphone et de différents embouts qui permettent de reproduire l'impédance acoustique de l'oreille. Ceci permet de tester un écouteur ou une prothèse auditive en conditions réelles. Chaque embout simule un positionnement de l'écouteur plus ou moins profondément dans le conduit auditif. Le microphone capte alors le son comme s'il arrivait au tympan.

référence. Il s'agit d'équipements de mesure acoustique de grande qualité de marque *Bruël & Kjaër*. Outre le microphone de référence et l'oreille artificielle, nous disposons également d'un caisson anéchoïque muni d'un haut-parleur. L'utilisation de ces équipements nous a permis de réaliser les mesures et les expérimentations de manière précise et rigoureuse.

La calibration du prototype de prothèse consiste à régler les appareils de conversion de niveau entre la carte-son et la prothèse.

B.4. Mesure de réponses impulsionnelles

Pour effectuer la mesure de la réponse impulsionnelle du chemin de bouclage acoustique, nous avons eu recours à une technique utilisant les *codes de Golay* (cf. [26, 80]). Ce type de technique est comparable à l'utilisation de séquences de longueur maximale MLS [72].

Les codes de Golay sont des paires de séquences binaires complémentaires. Ils ont été introduits par Golay dans [30]. Soit $(a; b)$ une paire de codes de Golay de longueur L (où L est une puissance de deux). Notons C_{aa} et C_{bb} les autocorrélations respectives des séquences a et b .

Les paires de codes de Golay ont la propriété suivante : la somme de leurs autocorrélations est exactement un Dirac.

$$C_{aa} + C_{bb} = 2L\delta_0$$

Supposons qu'on cherche à mesurer la réponse impulsionnelle h entre l'écouteur et le microphone de la prothèse.

Le système est alors successivement excité par la séquence a puis par la séquence b . On note \tilde{a} et \tilde{b} les signaux correspondant aux réponses respectives du système et captés au microphone.

Soit $C_{\tilde{a}a}$ et $C_{\tilde{b}b}$ les intercorrélations respectives entre a et \tilde{a} d'une part et entre b et \tilde{b} d'autre part.

La réponse impulsionnelle du système peut alors être obtenue en sommant les intercorrélations $C_{\tilde{a}a}$ et $C_{\tilde{b}b}$:

$$C_{\tilde{a}a} + C_{\tilde{b}b} = 2L \cdot h$$

Nous avons utilisé ce principe pour effectuer les mesures du chemin de bouclage entre les

transducteurs de la prothèse. Nous avons effectué ces mesures dans le domaine fréquentiel, c'est-à-dire que nous avons calculés les autocorrélations dans le domaine fréquentiel. La longueur des séquences a été fixée à une valeur assez grande pour se prévenir de l'aliasing temporel inhérent à ce choix de méthode. En effet, en travaillant sur la FFT nous identifions la réponse impulsionnelle à partir de la convolution circulaire. Enfin, les réponses impulsionnelles identifiées ont été définies à partir de la moyenne des résultats pour un grand nombre de mesures.

Bibliographie

- [1] Omar AIT AMRANE : *Identification de systèmes à réponse impulsionnelle longue par filtrage adaptatif en fréquence : application à l'annulation d'écho acoustique*. Thèse de doctorat, Ecole Nationale Supérieure des Télécommunications, 1992.
 - [2] Frédéric APOUX, Olivier CROUZET et Christian LORENZI : Temporal envelope expansion of speech in noise for normal-hearing and hearing-impaired listeners : effects on identification performance and response times. *Hearing Research*, 153(1-2):123–131, mars 2001.
 - [3] J.M. ARAN, A. DANCER, J.M. DOLMAZON et R. PUJOL : *Physiologie de la Cochlée*. INSERM/SFA, 1988.
 - [4] Françoise BEAUFAYS : Transform-domain adaptive filters : An analytical approach. *IEEE Transactions on Signal Processing*, 43(2):pages 422–431, février 1995.
 - [5] Ross BENCINA et Phil BURK : Portaudio, an open source cross platform audio api. *In Proceeding of the ICMC 2001*, 2001. <http://www.portaudio.com>.
 - [6] Dario BINI et Paola FAVATI : On a matrix algebra related to the discrete Hartley transform. *SIAM J. Matrix Anal. Appl.*, 14(2):pages 500–507, avril 1993.
 - [7] Gérard BLANCHET et Jacques PRADO : *Elements d'automatique*. Ellipses, 1995.
 - [8] Jens BLAUERT : *Spatial Hearing*. MIT Press, Cambridge MA, 1983.
 - [9] M.C. BOTTE, G. CANEVET, L. DEMANY et C. SORIN : *Psychoacoustique et Perception Auditive*. INSERM/SFA/CNET, 1988.
 - [10] Diane K. BUSTAMANTE, Thomas L. WORALL et Malcom J. WILLIAMSON : Measurement and adaptative suppression of acoustic feedback in hearing aids. *In Proc. IEEE ICASSP*, pages 2017–2020, 1989.
 - [11] Georg von BÉKÉSY : *Experiments in Hearing*. McGraw-Hill Book company, 1960.
 - [12] CALLIOPE : *La parole et son traitement automatique*. Collection technique et scientifique des télécommunications. Masson, 1989.
 - [13] Olivier CAPPÉ : *Techniques de réduction de bruit pour la restauration d'enregistrements musicaux*. Thèse de doctorat, Ecole Nationale Supérieur des Télécommunications, 1993.
 - [14] Olivier CAPPÉ : Elimination of the musical noise phenomenon with the Ephraïm and Malah noise suppressor. *IEEE Trans. on Speech and Audio Processing*, vol. 2(2):pp. 345–349, avril 1994.
-

-
- [15] Alain CASENAVE, Michel MONDAIN, Bruno FRACHET, Christian HAMANN et Olivier STERKERS : *Les Surdités - De la prothèse à l'implant*. Numéro 33 de Les monographies du CCA groupe. CCA, 2002.
- [16] Israel COHEN : Enhancement of speech using bark-scaled wavelet packet decomposition. *In Proc. 7th European Conf. Speech, Communication and Technology, EUROSPEECH-2001, Aalborg, Denmark, 3-7 September*, pages pp. 1933–1936, 2001.
- [17] Israel COHEN : On speech enhancement under signal presence uncertainty. *In Proc. 26th IEEE Internat. Conf. Acoust. Speech Signal Process., ICASSP-2001, Salt Lake City, Utah, 7-11 May 2001*, pages 167–170, 2001.
- [18] Israel COHEN et Baruch BERDUGO : Spectral enhancement by tracking speech presence probability in subbands. *In Proc. IEEE Workshop on Hands Free Speech Communication, HSC'01, Kyoto, Japan, 9-11 April 2001*, pages 95–98, avril 2001.
- [19] Eric J. DIETHORN : *Acoustic Signal Processing For Telecommunication*. Kluwer Academic Publishers, 2000.
- [20] D.L. DONOHO : De-noising by soft-thresholding. *IEEE Transactions on Information Theory*, 41(3):pp 613 – 627, mai 1995.
- [21] Pierre DUHAMEL et Martin VETTERLI : Improved fourier and hartley transform algorithms : Application to cyclic convolution of real data. *IEEE Trans. on Acoustics, Speech, and Signal Processing*, ASSP-35(6):pp. 818–824, juin 1987.
- [22] Koen ENEMAN et Marc MOONEN : Hybrid subband/frequency-domain adaptive systems. *Signal Processing*, Volume 81(1):pages 117–136, janvier 2001.
- [23] Yariv EPHRAÏM et David MALAH : Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator. *IEEE Trans. on Acoustics, Speech and Signal Processing*, vol. ASSP-32(6):pp. 1109–1121, dec 1984.
- [24] Yariv EPHRAÏM et David MALAH : Speech enhancement using a minimum mean-square error log-spectral amplitude estimator. *IEEE Trans. on Acoustics, Speech and Signal Processing*, vol. ASSP-33(2):pp. 443–445, april 1985.
- [25] E.R. FERRARA : Fast implementaton of LMS adaptive filter. *IEEE Trans. on Acoustic, Speech and Signal Processing*, 28:pages 474–475, août 1980.
- [26] Scott FOSTER : Impulse response measurement using golay codes. *In proc. International Conference on Acoustics, Speech, and Signal Processing (ICASSP86 Tokyo)*, pages pages 929–932, 1986.
- [27] André GILLOIRE, Eric MOULINES, Dirk SLOCK et Pierre DUHAMEL : *Digital Signal Processing in Telecommunications*, chapitre State of the art in acoustic echo cancellation, pages 45–91. Springer, 1996.
- [28] B.R. GLASBERG et Brian C. J. MOORE : Derivation of auditory filter shapes from notched-noise data. *Hearing Research*, 47:p 103–138, août 1990.
- [29] Brian R. GLASBERG et Brian C. J. MOORE : Frequency selectivity as a function of level and frequency measured with uniformly exciting notched noise. *J. Acoust. Soc. Am.*, 108:p 2318–2328, 2000.
- [30] Marcel J.E. GOLAY : Complementary series. *IRE Transactions on Information Theory*, pages 82–87, avril 1961.
-

-
- [31] M. GOODWIN : Nonuniform filterbank design for audio signal modelling. *Conference Record of the Thirtieth Asilomar Conference on Signals, Systems and Computers, 1996*, vol. 2:pp. 1229–1233, 1997.
- [32] Alain GOYÉ : *Le réhaussement fréquentiel dans l'aide auditive*. Thèse de doctorat, Ecole Nationale Supérieure des Télécommunications, 2000.
- [33] Yves GRANJON : *Automatique. Systèmes linéaires, non linéaires, à temps continu, à temps discret, représentation d'état*. Dunod, 2001.
- [34] T. GÜLZOW, A. ENGELSBERG et U. HEUTE : Comparison of a discrete wavelet transformation and a nonuniform polyphase filterbank applied to spectral-subtraction speech enhancement. *Signal Processing*, vol. 64:pp. 5–19, 1998.
- [35] Joseph L. HALL : *The Digital Signal Processing Handbook*, chapitre Auditory Psychophysics for Coding Applications, pages 39.1–39.24. CRC Press LLC, 1998.
- [36] Johan HELLGREN, Thomas LUNNER et Stig ARLINGER : Variations in the feedback of hearing aids. *J.Acoust.Soc.Am.*, 106(5):2821–2833, 1999.
- [37] Hermann L.F. HELMHOLTZ : *On the Sensations of Tone*. Dover Publications, 1954.
- [38] Aki HÄRMA, Matti KARJALAINEN, Lauri SAVIOJA, Vesa VÄLIMÄKI, Unto K. LAINE et Jyri HUOPANIEMI : Frequency-warped signal processing for audio applications. *Journal of the Audio Engineering Society*, 48(no. 11):pp. 1011–1030, November 2000.
- [39] Toshio IRINO et Roy D. PATTERSON : A time-domain, level-dependent auditory filter : the gammachirp. *J.Acoust.Soc.Am.*, 101(1):412–419, janvier 1997.
- [40] Toshio IRINO et Roy D. PATTERSON : A compressive gammachirp auditory filter for both physiological and psychophysical data. *J.Acoust.Soc.Am.*, 109(5):2008–2022, mai 2001.
- [41] B.M. JOHNSTONE, R. PATUZZI et G.K. YATES : Basilar membrane measurements and the travelling wave. *Hearing Research*, 22(1-3):147–153, 1986.
- [42] Jean-Marc JOT, Véronique LARCHER et Olivier WARUSFEL : Digital signal processing issues in the context of binaural and transaural stereophony. *In Proceedings of the 98th Convention of the Audio Engineering Society, Paris, France*, 1995.
- [43] James M. KATES : Feedback cancellation in hearing aids : Results from a computer simulation. *IEEE Trans. on Signal Processing*, 39(3):pages 553–562, mars 1991.
- [44] James M. KATES : *Applications of Digital Signal Processing to Audio and Acoustics*, chapitre Signal Processing for Hearing Aids, pages 235–276. Kluwer Academic Publishers, 1998.
- [45] James M. KATES : *Adaptive Signal Processing, Applications to Real-World Problems*, chapitre Adaptive Feedback Cancellation in Hearing Aids, pages 23–57. Springer, 2003.
- [46] Mead C. KILLION : *Encyclopedia of Acoustics*, volume 4, chapitre Hearing Aid Transducers, pages 1979–1991. John Wiley & Sons, 1997.
- [47] Ian Doré LANDAU : *Commande des systèmes - Conception, identification et mise en œuvre*. Hermès Science, 2002.
-

-
- [48] F. LEGENT, L. PERMUTER et C. VANDENBROUCK : *Cahier d'anatomie O.R.L.* Masson, 3ème édition, 1979.
- [49] Christian LORENZI, Frédéric BERTHOMMIER, Frédéric APOUX et Nicole BACRI : Effects of envelope expansion on speech recognition. *Hearing Research*, 136(1-2):131–138, octobre 1999.
- [50] David MALAH, Richard V. COX et Anthony J. ACCARDI : Tracking speech-presence uncertainty to improve speech enhancement in non-stationary noise environments. In *ICASSP*, pages 789–792, 1999.
- [51] Joseph A. MAXWELL et Patrick M. ZUREK : Reducing acoustic feedback in hearing aids. *IEEE Trans. on Speech and Audio Processing*, 3(4):pages 304–313, juillet 1995.
- [52] Ricardo MERCHED et Ali H. SAYED : An embedding approach to frequency-domain and subband adaptive filtering. *IEEE Trans. on Signal Processing*, 48(9):2607–2619, septembre 2000.
- [53] Brian C.J. MOORE : *Perceptual Consequences of Cochlear Damage*. Oxford University Press Inc., 1995.
- [54] Brian C.J. MOORE : *Introduction to the psychology of hearing*. Academic Press, 5ème édition édition, 2003.
- [55] Brian C.J. MOORE et B.R. GLASBERG : Suggested formulae for calculating auditory-filter bandwidths and excitation pattern. *J.Acoust.Soc.Am.*, 74:pp. 750–753, 1983.
- [56] Eric MOULINES, Omar AIT AMRANE et Yves GRENIER : The generalized multi-delay adaptive filter : Structure and convergence analysis. *IEEE Trans. on Signal Processing*, 43(1):pages 14–28, janvier 1995.
- [57] Ted PAINTER et Andreas SPANIAS : Perceptual coding of digital audio. *Proceeding of the IEEE*, 88(4):451–513, avril 2000.
- [58] Roy D. PATTERSON : Auditory filter shapes derived with noise stimuli. *J.Acoust.Soc.Am.*, 59:p 640–654, 1976.
- [59] James O. PICKLES : *An Introduction to the Physiology of Hearing*. Academic Press, 1982.
- [60] Jacques PRADO : *Traité d'électronique*, volume E3 162, chapitre Filtres numériques, Conversion de fréquences et bancs de filtres, pages 1–33. Techniques de l'ingénieur, 2000.
- [61] Jacques PRADO et Eric MOULINES : Frequency-domain adaptive filtering with applications to acoustic echo cancellation. *Annales des Télécommunications*, 49(7-8):pages 414–428, 1994.
- [62] R. PUJOL : Promenade autour de la cochlée. <http://www.iurc.montp.inserm.fr/cric/audition/index.htm>, 2004.
- [63] Lawrence R. RABINER : *Fundamentals of Speech Processing*. Prentice Hall Signal Processing Series. PTR Prentice-Hall, Inc., 1993.
- [64] RICE UNIVERSITY : Signal processing information base (SPIB). <http://spib.rice.edu/>, 2004.
-

-
- [65] A. ROBIER : Otologie. <http://www.med.univ-tours.fr/enseign/orl/Otol/index.html>, 2004. Site ORL de l'université de Médecine de Tours.
- [66] P.M. SELICK, R PATUZZI et B.M. JOHNSTONE : Measurement of basilar membrane motion in the guinea pig using the mossbauer technique. *J.Acoust.Soc.Am.*, 72:pages 131–141, 1982.
- [67] E.A.G. SHAW : *Handbook of Sensory Physiology*, volume 5/1, chapitre The External Ear, pages pages 455–490. Springer, 1974.
- [68] John J. SHYMK : Frequency-domain and multirate adaptive filtering. *IEEE Signal Processing Magazine*, 9(1):14–37, janvier 1992.
- [69] Marcio G. SIQUEIRA et Abeer ALWAN : Steady-state analysis of continuous adaptation in acoustic feedback reduction systems for hearing-aids. *IEEE Trans. on Speech and Audio Processing*, 8(4):pages 443–453, juillet 2000.
- [70] Julius O. SMITH et Jonathan S. ABEL : Bark and ERB bilinear transforms. *IEEE Transactions on Speech and Audio Processing*, pages 697–708, novembre 1999.
- [71] Jia-sien SOO et Khee K. PANG : Multidelay block frequency domain adaptive filter. *IEEE Trans. on Acoustics, Speech and Signal Processing*, 38(2):pages 373–376, février 1990.
- [72] G.-B. STAN, J.-J. EMBRECHTS et D. ARCHAMBEAU : Comparison of different impulse response measurement techniques. *Journal of the Audio Engineering Society*, 50(4): 249–262, avril 2002.
- [73] Michael R. STINSON et Gilles A. DAIGLE : Effect of handset proximity on hearing aid feedback. *J.Acoust.Soc.Am.*, 115(3):1147–1156, mars 2004.
- [74] P. P. VAIDYANATHAN : *Multirate systems and filter banks*. Prentice Hall, 1993.
- [75] Nathalie VIRAG : Single channel speech enhancement based on masking properties of the human auditory system. *IEEE Transactions on Speech and Audio Processing*, 7(2):126 – 137, mars 1999.
- [76] Beth WILSON : Digital signal processing applications for hearing accessibility. *IEEE Signal Processing Magazine*, 20(5):14–18, septembre 2003.
- [77] Patrick J. WOLFE et Simon J. GODSILL : Towards a perceptually optimal spectral amplitude estimator for audio signal enhancement. In *Proc. of IEEE ICASSP 2001, Istanbul, Turkey*, volume 2, pages 821–824, 2001.
- [78] B.D. WOODRUFF et D.A. PREVES : Fixed filter implementation of feedback cancellation for in-the-ear hearing aids. In *IEEE ASSP Workshop on Applications of Signal Processing to Audio and Acoustics*, pages 22 –23, octobre 1995.
- [79] Sigisbert WYRSCH et August KAELIN : Performance comparison of pbfdaf algorithms. In *Proc. IEEE Inter. Conf. on Electronics, Circuits and Systems*, volume 2, pages 831–834, septembre 1999.
- [80] B. ZHOU, D.M. GREEN et J.C MIDDLEBROOKS : Characterization of external ear impulse responses using golay codes. *Journal of the Acoustical Society of America*, 92(2) Pt. 1:pages 1169–1171, août 1992.
-

- [81] E. ZWICKER et R. FELDTKELLER : *Psychoacoustique - L'oreille récepteur d'information*. Collection Technique et scientifique des télécommunications. Masson, 1981.
- [82] E. ZWICKER et E. TERHARDT : Analytical expressions for critical-band rate and critical bandwidth as a function of frequency. *J.Acoust.Soc.Am.*, 68:p 1523–1525, 1980.
-