الجمهورية الجزائرية الديمقراطية الشعبية

MINISTERE DE L'ENSEIGNEMENT SUPERIEUR
ET DE LA RECHERCHE SCIENTIFIQUE

UNIVERSITE DR. YAHIA FARES DE MEDEA

FACULTE DES SCIENCES ET DE LA TECHNOLOGIE

Département du Génie Électrique
Et de l'Informatique



Projet de Fin d'Études

Présenté pour l'obtention du Diplôme
De Master en Génie Électrique
Option : système de télécommunication

THEME

Conception d'une stratégie de codage de la parole pour les personnes mal-entendants

Proposé et dirigé par,
Mr. ESSAID Billel

Préparés et présentés par,

Melle. BOUMENIR Radhia
Mme. ATTA Denyazad

Année Universitaire 2015-2016

REMERCIEMENTS

En premier lieu, nous remercions ALLAH le tout puissant qui nous a guidé tout au long de ce travail et rendu possible la réalisation de notre Projet de Fin d'Etudes.

- *Nous remercions notre encadreur Mr. ESSAID Billel, à qui nous tenons à exprimer, notre profonde gratitude de nous avoir prodigué leur précieux conseils, leur aides permanentes, ses idées originales, sa compétence et ses critiques constructives, et sa patience, et qu'il trouve ici notre estimation pour sa grande personne ainsi que notre profond respect. Merci pour son encouragement dans les périodes de doute et un peu difficiles.*

  Merci Mr ESSAID Billel.

- *Mes vifs remerciements vont en suite aux membres du jury : D'avoir accepté d'évaluer notre travail, qu'ils ont faits l'honneur de juger ce travail.*
- *A Mr BOUDRAA professeur de traitement de parole.*
- *Je tiens également à exprimer ma reconnaissance à monsieur ZAGHDAR Ahmed le recteur d'université Dr. Yahia Fares de Médéa*

Nous remercions également tous les amis de notre promotion qu'ils ont aidé directement ou indirectement à la réalisation de ce travail.

… Merci à Tous et pour Tout.

Dédicace

Je rends grâce à ALLAH qui m'avoir donné la foi, le courage, la volonté ainsi que la conscience d'arriver au terme de ce premier cycle univèrsitaire.

Je dédie ce modeste travail:

A mes très chers parents qui étaient pour moi l'exemple de sacrifice et d'encouragement et qu'ils demeurant toujours gravé dans ma mémoire et j'espère qu'ils trouveront ici l'éxpression de ma profonde affection.

. A mon mari NOUMERI Mohamed

A ma chère soeure Amina qui je l'aime très profondément avec son mari et ses enfants Lilya et Neufèl, Iyad.

A mes chèrs frères les joies de ma vie :Kamel et sa femme Wissame, Sofian, Radhwane, et surtout mon petit frère Zinou.

Ama trés chèr binome :'BOUMENIR Radhia' qui j'ai partagé avec elle le travail des jours et des nuits.

A mes chérs copines : Fella, Hamida, Zohra avec ses enfants , Soraya, Zolikha, Amina BEN NAOIA, Semia, Yasmine.

.A Monsieu Othmane BAKIR TERKIqui nous a aidés dans notre recherche

En fin je dédé ce modeste travail à toutes et tous qui j'ai rencontrés de prés ou de loins et qui mon aider ou cours de la période des études.

ATTA Denyazad.

Dédicace

Merci ALLAH de m'avoir donné la capacité d'écrire et de réfléchir, la force d'y croire, la patience d'aller jusqu'au bout du mon ambition.

Avec les sentiments de fiertés je dédie ce modeste travail A la lumière de ma vie ma très chère mère pour son amour, ses prières et ses sacrifices pour mon bonheur et ma réussite.

A mon très cher père pour l'encouragement et affection, qui a été mon membre durant toutes les années d'études, à me donner l'aide et à me protéger (Que dieu les gardes et les protèges).

A ma très chère grand-mère et A ma tente Safia.

A mes deux sœurs Mounia et Nihel et à mon frère Walid et a toute ma famille qui n'a pas cessé de M'encourager dans les moments difficiles que dieu les protèges.

Ama trés chèr binome :ATTA Denyazad qui j'ai partagé avec elle le travail des jours et des nuits.

A mes très bonne copines :Soraya, Sarah, Wissem,Soumia, Amina , Assia, Siham qui je l'aiment profondément.

A mes chers frères et meilleurs amis : Nacer , Lotfi , hamza.

A Mon frére Othmane BAKIR TERKI pour ces aides

A toutes et tous qui j'ai rencontrés de prés ou de loins et qui mon aider ou cours de la période des études.

A tout promotion 2016

Radhia BOUMENIR.

ملخص:

الهدف من هذا العمل هو دراسة و محاكاة استراتيجية لتنبيه قوقعة الأذن, و التي ترتكز على استخراج معاملات MFCC من الصوت المراد سماعه, و هذا لتوليد إشارات التنبيه العصبي باستعمال MATLAB . بعد محاكاة البرامج, تحصلنا على على نتائج طيبة فيما يتعلق بإعادة تمثيل إشارة الصوت الأصلية, إلا أنه فيما يخص الأصوات فجائية الصدور, الإشارة الممثّلة تحتوي على تشويش غير مرغوب فيه.

كلمات مفتاحية : إستراتيجية تنبيه, قوقعة الأذن, ترميز, معاملات MFCC, أصوات فجائية الصدور.

Abstract :

The purpose of this work is the study and the simulation of a strategy of cochlear stimulation, based on the extraction of the coefficients MFCC of the audio signal, to generate the signals of stimulation by using the MATLAB software. After simulations, we obtained from good results point of view reconstruction of the original signal from coefficients MFCC, however, for the voiced sounds the reconstituted signal present an annoying noise

Keywords : Strategy of stimulation, cochlée, coding, coefficient MFCC, voiced sound

Résumé :

Le but de ce travail est l'étude et la simulation d'une stratégie de stimulation cochléaire, basée sur l'extraction des coefficients MFCC du signal audio, afin de générer les signaux de stimulation en utilisant le logiciel MATLAB. Après simulations, nous avons obtenus des bons résultats point de vue reconstruction du signal original à partir de coefficients MFCC, cependant, pour les sons voisés le signal reconstitué présente un bruit gênant.

Mots clés : *stratégie de stimulation, cochlée, codage, coefficient MFCC, son voisé.*

TABLE DES MATIERES

REMERCIEMENT
RESUME
TABLES DES MATIERES
LISTE DES FIGURES
LISTE DES ACRONYMES

introduction Générale……..……..…….....1

CHAPITRE I : L'OREILLE ET LES IMPLANTS COCHLEAIRES

1. Introduction 3
2. Historique 3
3. Description du systeme auditif humain 4
   3. 1. L'Oreille Externe.....5
   3. 2. L'Oreille Moyenne.....5
   3. 3. L'Oreille Interne.....5
4. L'audition 5
5. La surdite…….5
   5. 1. Surdité de Transmission...…….....5
   5. 2. Surdité Neurosensorielle.....6
   5. 3. Surdité Mixte 7
   5. 4. Surdité Rétro-cochléaire 7
6. Le son 8
   6. 1. Le spectre et le Timbre 8
      6. 1. 1. Notion de perception du spectre 8
      6. 1. 2. Notion de perception du timbre…….....9
   6. 2. L'audiogramme 9
7. Le traitement de l'information sonore à la periphérie 10
   7. 1. Le codage de l'intensité.....11
   7. 2. Le codage de la fréquence 13
      7. 2. 1. Le codage tomporel 13
      7. 2. 2. Le codage tonotopique (spatial) 14
      7. 2. 3. Un codage mixte.....15
8. Les implants cochléaires.....16

8. 1. Définition 16
8. 2. L'implant 17
8. 2. 1. Audio Processeur 17
8. 2. 2. Fonctionnement d'un implant cochléaire 18
8. 3. Les autres Différents implants auditifs 18
8. 3. 1. Implant d'oreille moyenne 18
8. 3. 2. Implant du tronc cérébral 19
8. 3. 3. Implant à ancrage osseux 19
9. Les stratégies de codage du son dans les implants cochléaires 20
9. 1. La stratégie CIS 20
9. 2. La stratégie SAS 20
9. 3. La stratégie HiRes 20
10. Conclusion 28

CHAPITRE II: CARACTRERISTIQUES DU SIGNAL DE LA PAROLE ET LES STRATEGIES DE STIMULATION COCHLEAIRES

1. Introduction 22
2. La parole 22
2. 1. Le niveau acoustique 22
2. 2. Transformée de Fourier à court terme 23
2. 3. Spectogramme 24
2. 4. Fréquence fondamentale 25
2. 5. Les caractéristiques de la parole 25
3. Les algorithmes de stimulation cochléaires et les analyseurs de parole 28
3. 1. L'algorithme de House-Urban 29
3. 2. L'algorithme Compressed Analog (CA) 30
3. 3. L'algorithme 'Continuos Interleaved Sampling 30
3. 4. Le 'Wearable speech processor' de Nucleus 31
3. 5. le Spectral Maxima Sound Processor 32
3. 6. Le 'Mini Speech Procesor' (MSP) 33
4. Conclusion 34

CHAPITRE III : RESULTATS DE SIMULATION DE LA STRATEGIE DE CODAGE BASEE SUR MFCC

1. Introduction…… ..35
2. Extraction des coefficients MFCC……35

2. 1. Données de traitement ……35
2. 2. Fenêtrage Hamming ……36
2. 3. Transformée de Fourier discrète……37
2. 4. Filtrage Mel……38
2. 5. Calcul des coefficients MFCC ……41
3. Reconstruction du spectre à partir des coefficients MFCC……42
4. Application de la methodes de reconstruction du spectre sur la totalité du signal……43
4. 1. Segmentation de Fenêtrage ……43
4. 2. Les coefficients MFCC……44
4. 3. Reconstruction du spectre……46
5. Synthèse du signal de la parole……47
6. Conclusion……49
conclusion Générale……50
Bibliographie

LISTE DES FIGURES

Figure I. 1. Anatomie de l'oreille humaine 4
Figure I. 2. La surdité de transmission 6
Figure I. 3. La surdité neurosensorielle. 6
Figure I. 4. La surdité mixte. 7
Figure I. 5. La surdité rétro-cochléaire. 7
Figure I. 6. La notion de perception du spectre. 8
Figure I. 7. Notion de perception du timbre 9
Figure I. 8. Représentation de la fréquence et l'intensité 10
Figure I. 9. Spectrogrammes et représentations temporelles pour les signaux de parole 11
Figure I. 10. Le codage de l'intensité 12
Figure I. 11. Le codage temporel. 15
Figure I. 12. Le codage mixte. 16
Figure I. 13. La partie « Implant ». 17
Figure I. 14. La partie « processeur audio ». 18
Figure I. 15. Exemple d'un implant d'oreille moyenne. 19
Figure I. 16 . Exemple d'un implant du tronc cérébral. 19
Figure I. 17. Exemple d'un implant à ancrage osseux 20
Figure I. 18. Différentes stratégies de codage. 21
Figure II. 1. Enregistrement numérique d'un signal acoustique. 23
Figure II. 2. Evolution temporelle (en haut) et transformée de Fourier discrète (en bas) du [p] de (signaux pondérés par une fenêtre de Hamming de 30 ms). 23
Figure II. 3. Spectrogrammes à large bande (en bas), à bande étroite (en haut), et évolution temporelle de la phrase anglaise 'Alice's adventures', échantillonnée à 11.25 kHz (calcul avec fenêtres de Hamming de 10 et 30 ms respectivement). 24
Figure II. 4. Evolution de la fréquence de vibration des cordes vocales dans la phrase "les techniques de traitement numérique de la parole". La fréquence est donnée sur une échelle logarithmique ; les sons non-voisés sont associés à une fréquence nulle. 25
Figure II. 5. L'appareil phonatoire. 26
Figure II. 6. Phonèmes de la langue française. 27
Figure II. 7. Spectres d'un son non-voisé et d'un son voisé. 27
Figure II. 8. Perception de tonalité en fonction de la fréquence de stimulation. 29

Figure II. 9. Schéma de principe de l'analyseur de parole de House-Urban. 29
Figure II. 10.Schéma de principe de l'analyseur de parole 'Compressed Analog'. 30
Figure II. 11.Schéma bloc de l'analyseur CIS. 31
Figure II. 12.Schéma bloc basé sur l'extraction de caractéristique de la parole. 31
Figure II. 13.Schéma du Spectral Maxima Sound Processor. 32
Figure II.14. Schéma bloc réalisé sur le MSP. 33
Figure III. 1.Schéma bloc de l'extraction des coefficients MFCC.. ……..*35*
Figure III. 2. le tracé temporel du signal audio « Ah.wav ». 36
Figure III. 3. Segment de la voyelle « A » après fenêtrage Hamming. 37
Figure III. 4. Module de spectre du segment (Haut) et son logarithme (bas) 38
Figure III. 5. Espacement linéaire des filtres jusqu'à 1 kHz et logarithmique après. 39
Figure III. 6. Banc à 22 filtres à l'échelle Mel. 41
Figure III. 7. Vingt-deux coefficients MFCC interpolés à l'aide de la fonction spline. 42
Figure III. 8. Spectre en dB (rouge) confronté au filtrage Mel interpolé par spline (en noir).. 43
Figure III. 9. Fenêtre de Hamming 44
Figure III.10. Segmentation et fenêtrage du signal « Ah.wav ». 44
Figure III. 11. Module de spectre de tous les segments du signal « Ah.wav ». 45
Figure III. 12. Logarithme du module de spectre des segments. 45
Figure III. 13. Coefficients MFCC du signal « Ah.wav ». 46
Figure III.14. Spectres reconstitués de tous les segments. 46
Figure III. 15. Signal carré bi-phasique. ……. 47
Figure III. 16. Comparaison entre le signal initial et le signal reconstitué. 47
Figure III. 17. Synthèse d'un signal représentant la 1er phrase. 48
Figure III. 18. Synthèse d'un signal représentant la 2ème phrase. 48

LISTE DES ACRONYMES

RF	Radio Fréquence
CIS	Continuous Interleaved Sampling
CI	Implant Cochléaire
MFCC	Mel-fréquence coefficients cepstraux
SAS	Simultaneous Analog Stimulation
HiRes	High Resolution
CA	Compressed Analog
WSP	wearable speech processo**r**
AGC	Contrôle Automatique du Gain
SMSP	Spectral Maxima Sound Processor
MSP	Mini Speech Processor
DSP	Digital Signal Processor
FIFO	First-In First-Out
FFT	Fast Fourier transform
ADC	Analog Degital Convertisseur
DFT	Discrete Fourier Transform
DCT	Discrete cosine transform
LOG	Logarithme
IDCT	Inverse discrete cosine transform
IFFT	Inverse fast Fourier transform

INTRODUCTION GENERALE

Introduction Générale

Les implants cochléaires sont des dispositifs qui permettent de restauré partiellement l'audition, par une stimulation électrique de l'oreille interne chez les personnes atteintes une perte auditive grave à profonde. La fonction principale de ce dispositif prothétique est un remplacement artificiel de l'oreille interne endommagée en utilisant un processeur vocal porté sur l'oreille externe et un récepteur-stimulateur interne. Un microphone capte le signal de parole ; le processeur vocal externe traite le signal de parole en utilisant diverses stratégies de traitement, et génère un ensemble de données vocales codées, qui sont reçues par le stimulateur du récepteur interne par liaison RF (fréquence radio). Le récepteur décode le signal et transmet le signal de stimulation sous une forme d'onde spécifique aux électrodes intra-cochléaire. Le récepteur-stimulateur est utilisé pour stimuler le nerf auditif via un réseau d'électrodes et qui permet la compréhension du langage humain par le cerveau [1].

L'intelligibilité de la parole ou la qualité de perception de la parole générée par les implants cochléaires a atteint seulement 70% de celle de la capacité d'audition normale, et qui est dû à un grand nombre de facteurs, dont l'un est le traitement du signal de la parole [2].

Les stratégies de traitement de la parole ont un rôle important dans la maximisation du potentiel de communication du patient. La conception des implants cochléaires a commencé avec des implants à canal unique, ces implants fournis une stimulation électrique à un seul site dans la cochlée en utilisant une seule électrode. Les implants cochléaires multi-canal sont des dispositifs largement utilisés de nos jours.

Les principales stratégies de traitement de la parole qui sont utilisés dans les implants cochléaires multicanaux sont des stratégies basées sur la forme d'onde et d'extraction des caractéristiques pertinentes de la parole.

La stratégie de stimulation analogique comprimée (CA) et de l'échantillonnage continu entrelacé (CIS) sont des stratégies basées sur la forme d'onde du signal et qui délivrent une stimulation simultanée ou non simultanée à travers les électrodes.

Dans les travaux de recherches actuels, divers modèles acoustiques de simulation des implants cochléaires sont générées, pour étudier l'effet des paramètres spécifiques du système tels que, le nombre de canaux, le type de filtres, l'exigence de la bande passante de chaque filtre dans le banc de filtres, … etc.

Dans ce travail, une stratégie basée sur d'extraction de caractéristiques du signal de la parole est étudiée. Cette stratégie consiste à l'extraction des coefficients MFCC en se basant sur un banc de filtres qui approxime la structure des bandes critiques de la cochlée.

Ce travail est organisé comme suit :

Dans le premier chapitre, nous avons décrit l'anatomie de l'oreille ainsi que les types de surdité et les implants cochléaires.

Le second chapitre décrit les caractéristiques du signal de la parole et les stratégies de stimulation cochléaire.

Dans le troisième chapitre, nous avons présenté les résultats de simulation de la stratégie de codage basé sur l'extraction des coefficients MFCC.

Nous terminerons ce travail par une conclusion générale et les perceptives.

Chapitre I :

L’Oreille et les implants cochléaire

1. Introduction :

L'implant cochléaire, dont le principe de fonctionnement reproduit le codage naturel des sons par l'oreille, est proposé aux personnes souffrant d'une surdité profonde de la périphérie du système nerveux auditif. Il est donc primordial de présenter dans un premier temps le système auditif périphérique d'un point de vue anatomique et fonctionnel pour comprendre les types de surdité impliqués et le principe de l'implant cochléaire. Nous décrirons également l'anatomie et l'organisation fonctionnelle du système auditif central afin que puissent être mieux compris par la suite les effets de la surdité et de la réhabilitation auditive sur le fonctionnement de ce système avec de présentation des propriétés de maturation et de plasticité du système nerveux auditif qui jouent un rôle prépondérant dans la réhabilitation auditive. Enfin, nous terminerons ce chapitre par les stratégies de codage du signal utilisées dans les implants cochléaires [3].

2. Historique :

On peut dire que l'histoire de l'implant cochléaire, vieille d'à peine un demi-siècle, a son avenir devant elle. Couramment pratiquée de nos jours, cette technique de réhabilitation de la surdité va connaître un développement insoupçonnable pour nous aujourd'hui. Il est bon que les chercheurs, les cliniciens et les industriels, qui vont en vivre la suite, se souviennent de ce que furent ses débuts [4].

- **1957-1985 La préhistoire de l'implant cochléaire :**

• Premières expérimentations, développement de prototypes en France, aux États Unis, en Australie et en Autriche.

• Amélioration progressive de ces prototypes, jusqu'à rendre possible la mise au point des premiers modèles commerciaux : élimination des connecteurs transcutanés, augmentation du nombre d'électrodes, mise au point de stratégies de commande des électrodes de plus en plus efficaces [5]…

- **1985-1995 Les débuts des implants cochléaires commerciaux modernes :**

• Mise sur le marché de la première génération d'implants cochléaires commerciaux, donc les processeurs sont des boitiers de la taille d'un gros paquet de cigarette. Leurs logiciels sont régulièrement améliorés, permettant ainsi aux patients de bénéficier de progrès postérieurs à leur implantation.

• Les agences de santé des pays occidentaux autorisent progressivement leur pose chez l'adulte, puis chez l'enfant. Les progrès techniques permettent d'abaisser progressivement le seuil de surdité à partir duquel l'implant cochléaire est préconisé [5].

- **À partir de 1995 :**

• Les premiers processeurs miniaturisés de type « contour d'oreille » apparaissent. Ils sont remplacés régulièrement par des modèles plus performants qui restent le plus souvent compatibles avec les implants internes plus anciens, permettant ainsi à leurs porteurs de bénéficier des progrès techniques.

• La technique a vraiment atteint sa maturité : le nombre de personnes implantées s'accroit de plus en plus, de quelques dizaines de milliers dans le monde au milieu des années 1990, on en est aujourd'hui à plusieurs centaines de milliers [5].

3. Description du système auditif humain :

Le système auditif humain est assez complexe. En effet, il est sensible aux plus fins des mouvements aléatoires des molécules d'air en contact avec le tympan. De plus, il peut supporter des bruits très forts et peut extraire des faibles sons en présence de bruits énormes. Il peut aussi distinguer et différencier plusieurs sons. Le système auditif comprend quatre parties : l'oreille externe, l'oreille moyenne, l'oreille interne et les vois nerveuses [6].

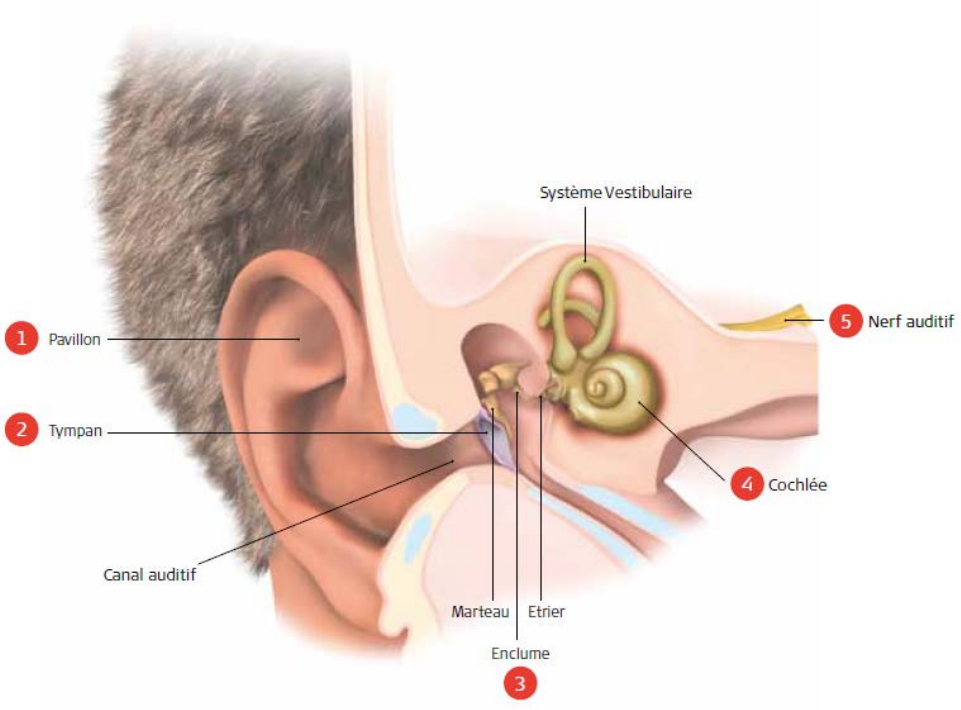


Figure I.1 : *Anatomie de l'oreille humaine.*

3.1.L'Oreille Externe :

L'oreille externe composée par le Pavillon qui recueille et transmet les sons au canal auditif et le canal auditif (conduit auditif externe) qui dirige les sons dans l'oreille. Elle amplifie les sons aigus entre 1.5 kHz et 7 kHz. Sa fonction est donc une préaccentuation du signal acoustique [7].

3.2.L'Oreille Moyenne :

L'oreille moyenne assure l'adaptation d'impédance entre le milieu aérien et les liquides de l'oreille interne, elle est composée par le Tympan (membrane tympanique) qui transforme les sons en vibrations et la chaîne des osselets (marteau, enclume, étrier) transmet les vibrations à l'oreille interne [7].

3.3.l'Oreille Interne :

L'Oreille Interne (cochlée) contient des fluides et des cellules très sensibles (cellules ciliées), avec des structures en forme de cils qui bougent grâce aux vibrations sonores. Le système vestibulaire contient des cellules qui contrôlent l'équilibre. Et le nerf auditif transmet l'information de la cochlée au cerveau [7].

4. L'audition :

Les sons captés par l'oreille externe sont transmis au conduit auditif pour atteindre le tympan qui convertit les sons en vibrations. Les vibrations sont transmises par la chaîne des osselets jusqu'à la cochlée, les fluides de la cochlée se mettent à leur tour en mouvement.

Les ondes stimulent les cellules ciliées qui génèrent des signaux électriques, transmis ensuite au nerf auditif. Les cellules ciliées de la partie apicale de la cochlée envoient l'information portant sur les fréquences graves tandis que celles de la partie basale envoient l'information portant sur les fréquences aiguës. Le nerf auditif transmet ces informations au cerveau, où elles sont interprétées en tant que sons.

5. La surdité :

5.1.Surdité de Transmission :

Tout problème concernant l'oreille externe ou l'oreille moyenne qui empêche la transmission correcte du son vers l'oreille interne est appelé surdité de transmission. Ces surdités sont en général légères ou moyennes, entraînant des pertes allant de 10 à 70 décibels.

Dans certains cas, les surdités de transmission sont temporaires. Selon l’origine du problème, le recours à la médication ou à la chirurgie peut aider. Les surdités de transmission peuvent être aussi corrigées par des aides auditives ou un implant d’oreille moyenne [7].



Figure I.2 : *La surdité de transmission.*

5.2.Surdité Neurosensorielle :

La surdité neurosensorielle est due à une perte des cellules sensorielles (cellules ciliées) dans la cochlée et elle est généralement permanente. Cette surdité peut être légère, moyenne, sévère ou profonde. Les surdités neurosensorielles légères à sévères peuvent être corrigées par des aides auditives ou un implant d’oreille moyenne. Les pertes auditives sévères à profondes peuvent être améliorées avec un implant cochléaire [7].



Figure I.3 : *La surdité neurosensorielle.*

5.3.Surdité Mixte :

Une surdité mixte est une combinaison d’une perte auditive de transmission et d’une perte auditive neurosensorielle. Elle est due à des problèmes présents au niveau de l’oreille moyenne et de l’oreille interne. Les options de traitement sont la médication, la chirurgie, les aides auditives ou l’implant d’oreille moyenne [7].



Figure I.4 : *La surdité mixte.*

5.4.Surdité Rétro-cochléaire :

L’absence de nerf auditif ou un nerf auditif endommagé peut entraîner une surdité rétro-cochléaire. Pour ces cas, la surdité rétro-cochléaire est une surdité profonde et permanente. Les aides auditives et les implants cochléaires ne sont d’aucune utilité car le nerf ne peut plus transmettre l’information auditive au cerveau. Dans certains cas, un implant du tronc cérébral peut améliorer la situation [7].



Figure I.5 : *La surdité rétro-cochléaire.*

6. Le son :

Les sons sont perçus par les oreilles et interprétés par le cerveau, Ce qui leur donnent à l'écoute un caractère très subjectif [8], il est le résultat d'un mouvement des molécules d'air qui nous entourent [8].

Il est caractérisé par les deux éléments suivant : La pression acoustique (le dB) et la fréquence (Hz). Et défini selon trois grandeurs [8] :

- L'intensité acoustique perçue (pression et dépression atmosphérique générées par l'onde de choc. (Unité : le Watt/m². Mais la grandeur la plus généralement utilisée est le décibel)
- L'intensité électrique (puissance relative perçue. Unité : le Watt)
- La fréquence et le timbre (tonalité et signature du son. Unité : le hertz).

6.1.Le Spectre et le Timbre :

6.1.1. Notion de perception du spectre :

Le son à une notion objective et subjective à la fois. L'oreille de chacun n'a pas la même sensibilité à tous les niveaux sonores et pour toutes les fréquences du spectre audible. Le spectre audible par une oreille saine est compris entre 20Hz et 20 000Hz.

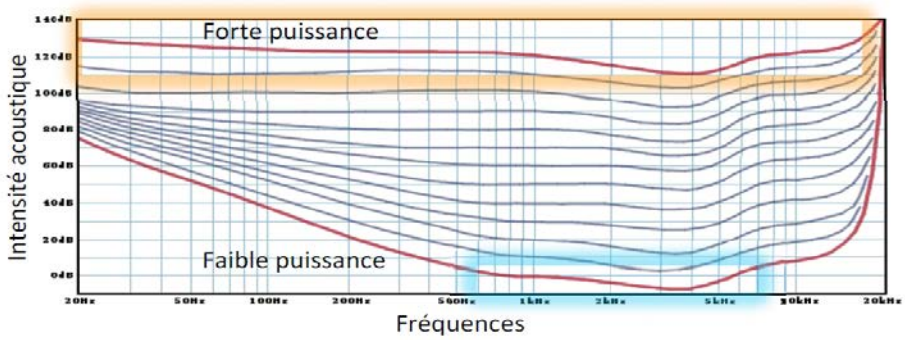


Figure I.6 : *La notion de perception du spectre.*

L'audiogramme ci-dessus illustre la variation de perception des fréquences en fonction du niveau sonore. Plus la puissance est faible, plus le spectre de fréquence perçu est petit et irrégulier (tracé rouge en bas très incurvé). Par exemple, A très faible écoute, nous n'entendons que les fréquences médiums (Encadré bleu). Phénomène constaté par exemple avec une TV à niveau très faible dans un salon. A l'inverse, plus le niveau sonore est élevé, plus le spectre de fréquence perçu est large et régulier (tracé rouge en haut, « plat ») Principe naturel de nivellement (ou compression) du signal. Lorsque nous augmentons

progressivement le niveau de cette même TV, nous entendons de plus en plus distinctement toutes les autres fréquences (très graves et très aigues). (Encadré orange) [8].

6.1.2. Notion de perception du timbre :

Le timbre est la caractéristique qui permet par exemple de différencier et reconnaitre les voix de deux individus distincts par exemple. Tout signal sonore est composé d'une fréquence ou d'un mélange de fréquences pures. Le timbre est la résultante de la sommation de toutes les fréquences pures qui compose un son [8].

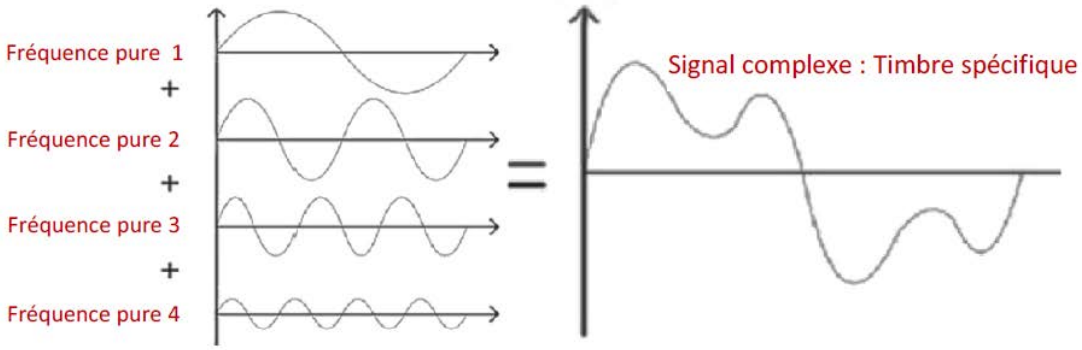


Figure I.7 : *Notion de perception du timbre.*

6.2.L'audiogramme :

Un audiogramme est un graphique qui illustre les capacités auditives d'une personne, c'est la mesure de la perte auditive pour chaque oreille. Dans l'audiogramme ci-dessous, les chiffres de la partie supérieure représentent les fréquences allant de 125 à 8000 Hz.

La fréquence est exprimée en cycles par seconde, ou Hertz. Plus la fréquence est élevée, plus le son est aigu. Le volume des sons (intensité) est mesuré en unités appelées décibel. Le zéro décibel (0 dB) ne veut pas dire « pas de son ». Il est juste très faible. Le niveau de la voix conversationnelle est autour de 55 dB. Un son à 120 dB est extrêmement fort, aussi fort que lorsque vous êtes à 25 m d'un avion en phase de décollage.

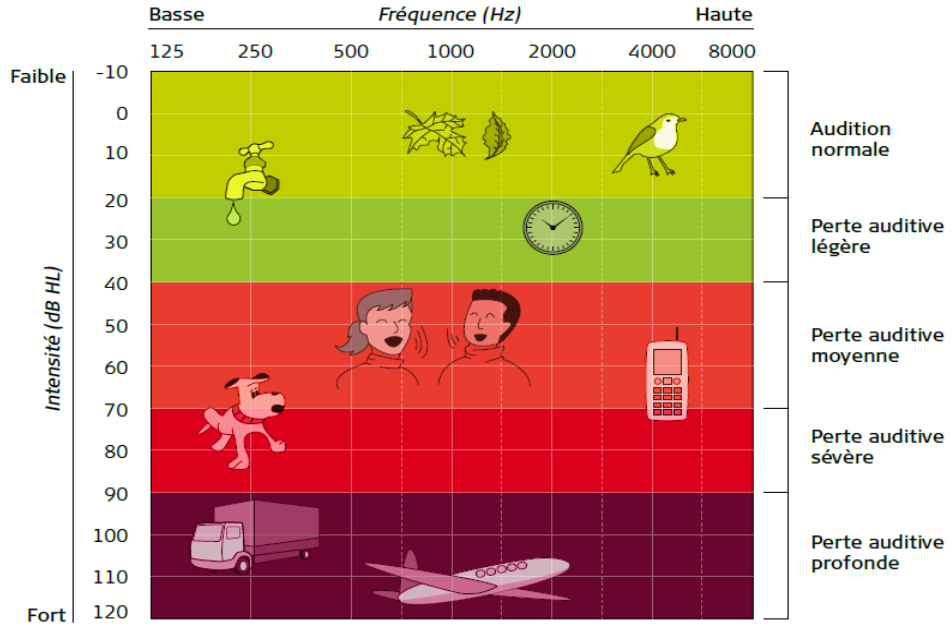


Figure I.8 : *représentation de la fréquence et l'intensité.*

Les chiffres, à la verticale de l'audiogramme, représentent des intensités en décibel. Lors d'une audiométrie (test de l'audition), l'audiologiste fait entendre au sujet des sons d'une fréquence donnée. Pour chaque fréquence testée, le volume le plus faible que la personne peut entendre est indiqué sur l'audiogramme à la fréquence et à l'intensité correspondante. On l'appelle le «Seuil Auditif ». Votre audiogramme est une «image» de votre audition. Il indique de combien votre audition varie de la norme, et s'il y a une perte auditive, son degré, son type et l'origine de cette perte [7].

7. Le traitement de l'information sonore à la périphérie :

Les sons sont des phénomènes vibratoires qui peuvent être apériodique (impulsionnel ou continu) ou périodique (simple ou complexe). Un son périodique simple peut être représenté par une courbe sinusoïdale, un son périodique complexe par une courbe complexe régulière. L'analyse fréquentielle par transformé de Fourier permet de décomposer une onde périodique complexe en ses composantes sinusoïdales, les harmoniques, caractérisés chacun par une fréquence ou hauteur (exprimée en Hz) et une amplitude ou intensité (exprimée en dB). Les fréquences des harmoniques sont des multiples entiers d'une fréquence de base : le fondamental [3].

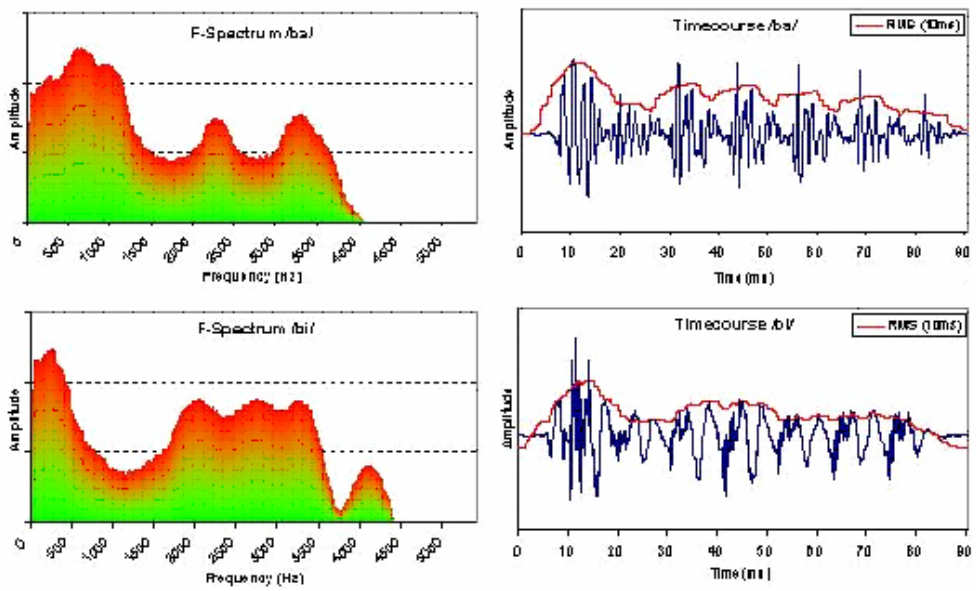


Figure I.9 : *Spectrogrammes et représentations temporelles pour les signaux de parole.*

Des études physiologiques et/ou psychophysiologiques ou encore des études impliquant des modèles ont expliqué divers aspects de l'activité du nerf auditif difficiles de coder pour les différentes caractéristiques du son. Cependant, il reste beaucoup de points à éclaircir. C'est pourquoi, seul un résumé des principales théories relatives au codage fréquentiel spatial et temporel du son est proposé ici, la façon dont ces principes sont exploités par l'implant cochléaire étant décrite plus bas [3].

7.1.Le codage de l'intensité :

Le système auditif utilise plusieurs méthodes complémentaires pour coder les variations d'intensité des sons purs. Le taux moyen de décharge d'une fibre ou d'une population de fibres coderait en partie l'intensité d'un stimulus. Lorsque pour une fibre donnée du nerf auditif on accroît l'intensité du stimulus, la fréquence de décharge des potentiels d'action augmente suivant une courbe « sigmoïde » pour toutes les hauteurs tonales (Figure I.10). La fonction d'intensité (nombre de potentiels d'action en fonction de l'intensité) est typique avec un seuil, suivi d'une augmentation progressive du nombre de potentiels d'action et une saturation (Figure I.10). Exemples schématiques de taux de décharge de neurones auditifs en fonction de l'intensité. Les courbes (a), (b), et (c) sont typiques de ce qui est observé pour les neurones avec des taux de fréquences spontanées élevés, moyens et bas respectivement [3].

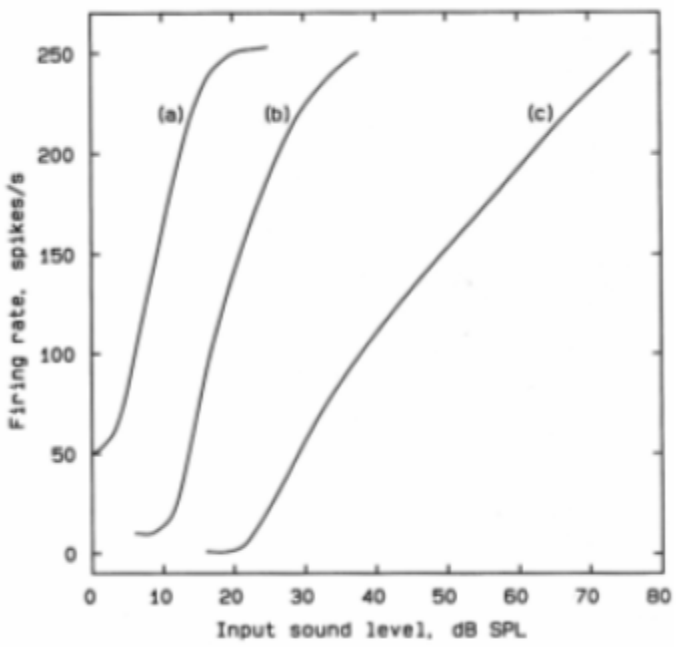


Figure I.10 : *Le codage de l'intensité.*

Le champ dynamique des fibres (l'augmentation de l'amplitude du stimulus nécessaire pour que le taux de décharge moyen augmente depuis son seuil jusqu'à saturation) est généralement compris entre 20-50 dB, alors que la dynamique acoustique est de 100 à 120 dB. Pour expliquer comment l'activité de ces fibres peut couvrir toute la dynamique acoustique, Liberman (en 1978) montre qu'il existe deux types de fibres de type I au niveau de la cochlée. Ils se différencient par leur diamètre et leur taux de décharge spontanée. D'un point de vue fonctionnel, les fibres ayant un petit diamètre et un taux de décharge spontanée élevé ont un seuil de déclenchement (défini comme étant l'intensité à laquelle le taux de décharge moyen augmente au-dessus du taux spontané) bas de 0 dB HL. Les fibres ayant un gros diamètre et un taux de décharge spontanée inférieur ont un seuil de déclenchement plus haut, vers 40-50 dB HL. Le premier type de fibres code les intensités entre 0 et 40-50 dB HL et le deuxième type de fibres code les intensités plus élevées, les fibres à seuils bas étant alors saturées. Le degré de coïncidence entre les décharges dans une population de fibres serait également utilisé pour coder l'intensité [3].

Les caractéristiques temporelles de la réponse du nerf auditif en fonction de la vibration de la membrane basilaire ne doivent pas faire oublier cependant que ces vibrations sont différentes selon la position sur la membrane, ce qui génère également un codage

tonotopique/spatial. Un modèle spatial permettrait donc également d'expliquer le codage de l'intensité.

Une théorie propose en effet que l'intensité du stimulus soit codée par la dispersion spatiale de l'excitation neurale générée par le stimulus acoustique [3].

7.2.Le codage de la fréquence :

7.2.1. Le codage temporel de la fréquence :

Même au repos (sans vibration de la membrane basilaire), les cellules ciliées internes ont une activité spontanée. Elles génèrent en moyenne 18 potentiels d'action par seconde et déchargent parfois même jusqu'à 100 fois par seconde. En présence d'une stimulation acoustique, la membrane basilaire vibre et les stéréo cils des cellules ciliées internes se penchent en avant et en arrière comme cela est décrit plus haut. Cela fait varier la quantité de neurotransmetteurs libérés de manière à ce que les potentiels d'action aient de plus grandes chances d'être générés pour une phase du stimulus plutôt que pour une autre. Ce principe de synchronisation en fonction d'une fréquence de prédilection ou fréquence caractéristique est appelé « phase-locking ». Lorsque l'amplitude des vibrations de la membrane basilaire augmente, les quantités de neurotransmetteurs libérés augmentent également. Les taux de décharge moyens et la synchronisation des décharges évoluent à peu près en fonction de l'amplitude de la vibration. Le taux de décharge cependant n'augmente pas linéairement, parce que le processus de transduction n'est pas linéaire lui-même. Dans les théories sur le codage temporel de la fréquence, il est fait l'hypothèse que les informations sur le stimulus acoustique sont codées par le rythme des réponses des fibres nerveuses. Par exemple, une théorie du codage fréquentiel temporel (pitch rate) propose que la fréquence soit codée par les intervalles inter-pics de chaque fibre ou d'une population de fibres (on parle alors de « volleying ») (par ex., Clark et coll., 1995 ; Rose et coll., 1967). Cependant, il existe des périodes réfractaires qui limitent la rapidité avec laquelle une fibre qui vient de décharger peut se trouver à nouveau dans un état qui lui permette de décharger. Le taux de décharge est ainsi limité à un maximum de 1000 pics par seconde. Il est donc généralement reconnu que ce codage est limité aux fréquences de stimulation inférieures à 4000 Hz [3].

7.2.2. Le codage tonotopique (spatial) **[3]** *:*

L'analyse fréquentielle est effectuée via des mécanismes passifs de la membrane basilaire selon la théorie de l'onde propagée décrite par von Bekesy (il reçut en 1961 le Prix Nobel de Physiologie en Médecine pour ses recherches). La distance parcourue par l'onde propagée le long de la membrane basilaire dépendrait de la fréquence du son. En effet, la structure de la membrane basilaire n'est pas identique à la base et à l'apex de la cochlée.

A la base, la membrane est moins large, d'environ un facteur 5, mais également plus rigide, d'environ un facteur 100. Ainsi, lorsqu'un son de haute fréquence atteint l'oreille, la base de la membrane vibre beaucoup, en dépensant beaucoup d'énergie et l'onde ne se propage pas très loin. Pour les basses fréquences, l'onde peut se déplacer de la base jusqu'à l'apex de la membrane basilaire, avant que toute l'énergie soit épuisée. Ce mécanisme est décrit schématiquement sur la Figure I.11. La membrane basilaire peut ainsi être présentée comme étant une série continue de filtres à bande passante, la fréquence centrale de chaque filtre diminuant de manière monotone avec la distance depuis la base. Cette relation qui lie la fréquence du son (de 20 à 20 000 Hz) et la localisation entre la base et l'apex du codage est nommée carte cochléaire fréquentielle par Liberman (1982).

Il s'agit de l'organisation tonotopique de la cochlée : les fréquences aiguës sont localisées vers la base, les fréquences graves vers l'apex. Greenwood (1990) a proposé une carte de la cochlée associant position et fréquence : la carte qu'il a décrit contient 35 bandes critiques de 1 mm chacune et est souvent utilisée comme référence dans les articles scientifiques. La figure I.10 présente l'organisation tonotopique cochléaire (A) et la réponse de la membrane basilaire aux stimulations acoustiques (B, C, D). Les sons de haute fréquence font vibrer uniquement la partie basale de la cochlée (B) alors que les sons de basse fréquence induisent une vibration jusqu'à l'apex (D).

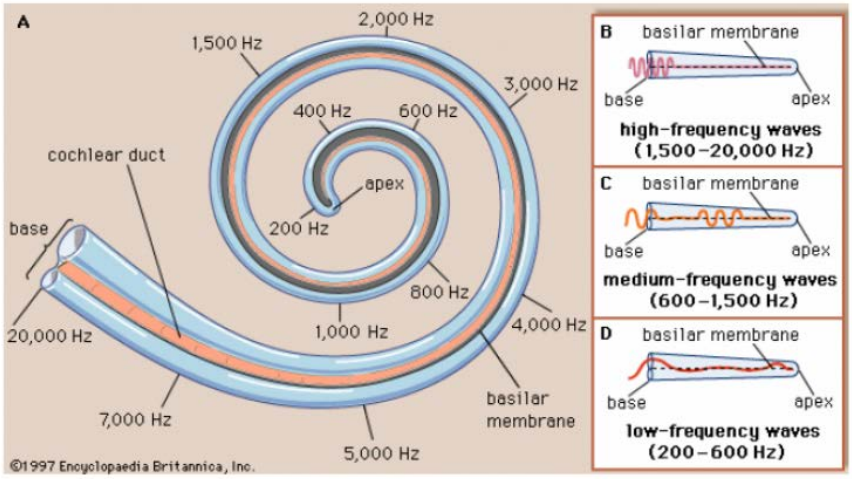


Figure I.11 : *Le codage tonotopique.*

Le taux de décharge des neurones est plus important à la fréquence caractéristique et diminue lorsque l'on s'éloigne de la zone de résonance. En réponse à des stimulations sonores, la membrane basilaire va ainsi vibrer en décrivant une courbe d'accord (« tuning curve ») (von Bekesy, 1949). Cela signifie que les cellules de la cochlée codent préférentiellement une certaine fréquence, mais également les fréquences voisines de façon moins spécifique. Les cellules ciliées externes changent également activement les vibrations de la membrane basilaire augmentant ainsi sa sensibilité et sa sélectivité fréquentielle.

7.2.3. Un codage mixte [3] :

En réalité, les théories temporelles et spatiales sont complémentaires et le codage des sons au niveau de la cochlée suit un mécanisme temps-place (Licklider, 1954). Un mécanisme temporel est particulièrement important pour le traitement des basses fréquences, alors qu'un mécanisme selon la place est essentiel pour les hautes fréquences (Wever, 1949). Il est généralement reconnu que le codage selon la place est inadapté pour les fréquences en dessous de 300-400 Hz et voire même au-delà. En effet, le seuil de discrimination de fréquence à 2000 Hz est 0.2 % (Traux, 1999), alors que cette différence de fréquence perçue est égale à la distance entre deux cellules ciliées internes voisines. Le codage par la place, selon lequel une fréquence active préférentiellement certaines cellules mais également les cellules voisines de manière moins spécifique, ne peut par conséquent expliquer à lui seul le codage de la fréquence des sons graves. D'un autre côté, le codage des fréquences supérieures à 4000-5000 Hz ne peuvent être temporel, car cela demanderait des décharges trop rapides aux neurones, par rapport à la durée de leur période réfractaire. Loeb et coll. (1983) ont proposé un modèle impliquant trois types de mécanismes pour la perception de la fréquence.

Le premier mécanisme utilise le codage spatial pour les fréquences allant de 5000 à 20 000 Hz, le deuxième utilise le « volleying » pour les fréquences entre 500 et 5000 Hz (Kiang, 1965), et le troisième le codage temporel pour les fréquences allant de 300 à 500 Hz où les nerfs sont capables de décharger à chaque cycle (Rose et coll., 1967). Une théorie, basée sur la latence de l'activité cochléaire et sur son expansion au niveau du cortex auditif, a été développée par Greenberg et coll. (1998). Elle intégrerait la notion de délai fréquence-dépendant lié au passage de l'onde acoustique dans la cochlée au codage par la place et le rythme. Les sons graves stimuleraient ainsi le tronc cérébral (Greenberg, 1980) et le cortex auditif (Woods et coll., 1993) plus tardivement que les sons aigus qui transitent très rapidement par la cochlée (le délai est de l'ordre de dizaines ou centaines de microsecondes pour des fréquences supérieures à 4 kHz). Le temps total pour aller de la base à l'apex serait de 10 ms au moins chez l'animal (Goldstein et coll., 1971 ; Ruggero & Rich, 1987). Il peut être estimé par la latence de l'activité du nerf auditif. La figure I.11 montre la latence de l'excitation des fibres nerveuses auditives en fonction de la fréquence du signal sinusoïdal [5].

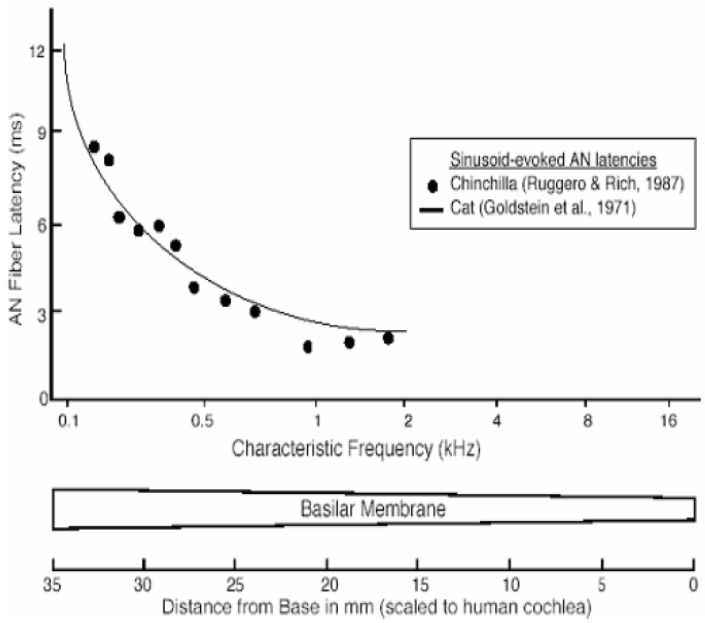


Figure I.12 *: le codage mixte.*

8. LES IMPLANTS COCHLEAIRES

8.1.Définition :

Un implant cochléaire est un dispositif électronique conçu pour permette une perception sonore optimale et la meilleure compréhension possible de la parole. Il est destiné aux enfants

et aux adultes atteints de surdité sévère a profonde auxquels les prothèses auditives conventionnelles apportent une aide insuffisante. Non parce que ces dernière ne peuvent pas amplifier suffisamment les sons, mais parce que les minuscules cellules ciliées de l’oreille interne également appelée cochlée sont devenues trop déficientes ou sont absentes et que, par conséquent, les sons, même amplifiés, ne peuvent plus être transmis au cerveau [9].

Elle est composée par deux parties : l’implant et audio processeur.

8.2.L’implant :

Placé chirurgicalement sous la peau. L’implant est constitué d’un boîtier, qui contient l’électronique, d’un porte-électrodes ainsi que d’une antenne réceptrice et d’un aimant qui maintient l’antenne en place derrière l’oreille [7].

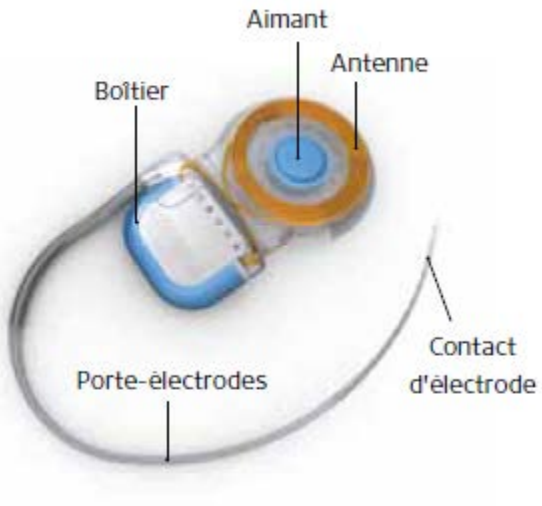


Figure I.13 : *la partie « Implant ».*

8.2.1. Audio Processeur :

Il est porté derrière l’oreille. L’audio processeur est constitué d’une unité de contrôle, d’un boîtier de piles et d’une antenne qui transmet l’information à l’implant à travers la peau [7].

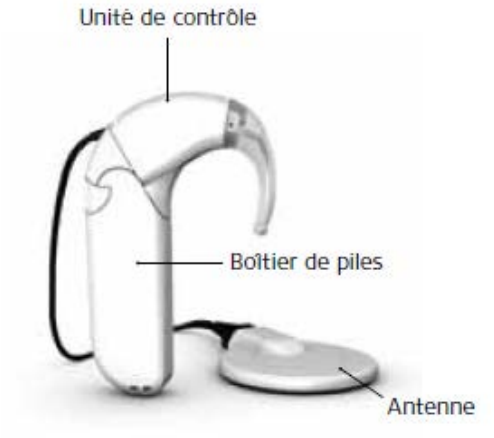


Figure I.14 *: la partie « processeur audio ».*

8.2.2. Fonctionnement d'un implant cochléaire :

Les systèmes d'implants cochléaires convertissent les sons environnants en impulsions de courant électrique codées. Ces impulsions électriques stimulent le nerf auditif et le cerveau les interprète comme des sons. Comme le cerveau reçoit les informations sonores très rapidement, les sons sont perçus au moment où ils sont produits.

- Les sons sont captés par le microphone de l'audio processeur.
- L'audio processeur analyse et code les sons en séquences spécifiques d'impulsions électriques.
- Ces impulsions sont envoyées vers l'antenne et transmises à travers la peau vers l'implant.
- L'implant envoie les impulsions aux électrodes dans la cochlée.
- Le nerf auditif achemine les signaux électriques vers les centres auditifs du cerveau qui les interprète comme des sons [7].

8.3.Les autres Différents implants auditifs [15] :

8.3.1. L'implant d'oreille moyenne :

C'est une prothèse à fonctionnement électromagnétique. Elle s'adresse à des surdités modérées, c'est-à-dire des surdités plutôt légères et moyennes peu importantes, en particulier les surdités très en pente sur les aigus qui entraînent des effets Larsen importants, aussi dans le cas de lésion de l'oreille externe, et après échec d'un appareillage conventionnel. L'implant d'oreille moyenne a une bande passante plus large et va plus loin dans les aigus.

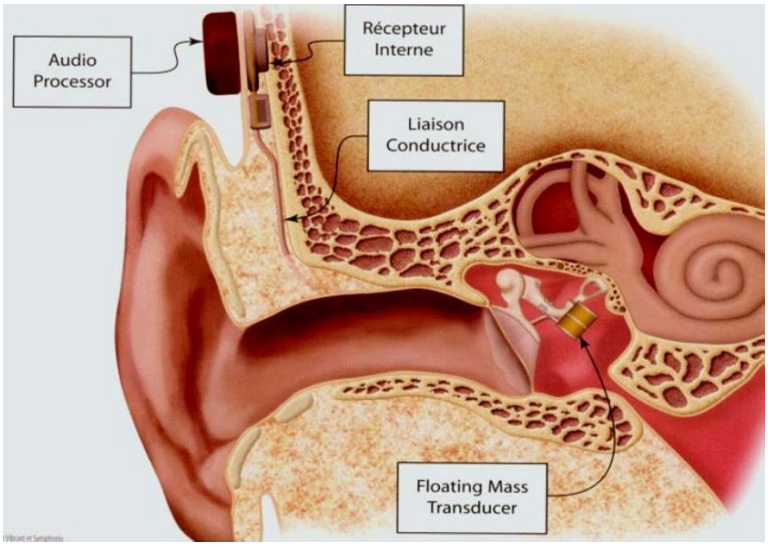


Figure I.15 : *Exemple d’un implant d’oreille moyenne.*

8.3.2. Implant du tronc cérébral :

L'implant auditif du tronc cérébral est un petit dispositif implanté, lors d'une intervention chirurgicale, dans le cerveau d'une personne atteinte de surdité dont les nerfs auditifs sont endommagés. Grâce à cet implant, une personne atteinte de surdité totale aura l'impression d'entendre.

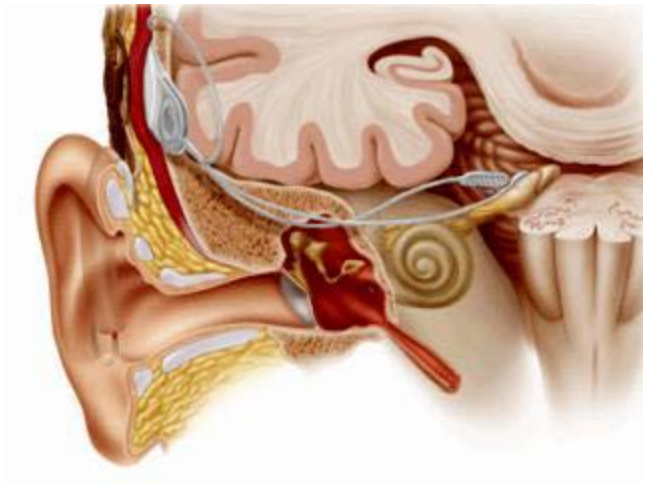


Figure I.16 : *Exemple d’un implant du tronc cérébral.*

8.3.3. Implant à ancrage osseux :

Un appareil auditif à ancrage osseux peut être utilisé dans certains cas de surdité de transmission sévère. Ceci est possible car les os, tout comme l'air peuvent transporter les vibrations du son.

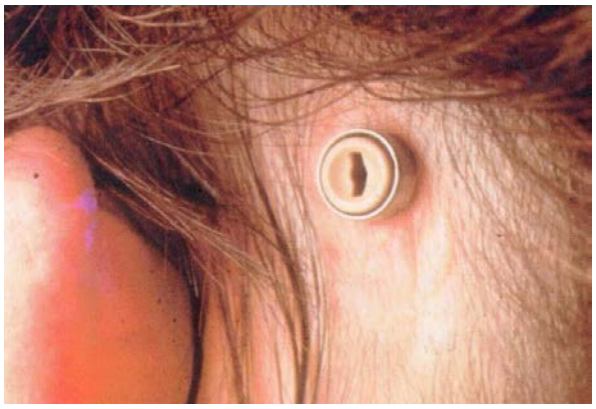


Figure I.17 : *Exemple d'un implant à ancrage osseux.*

9. Les stratégies de codage du son dans les implants cochléaires [5] :

9.1.La stratégie CIS :

Cette stratégie a été développée au Research Triangle Institute (Wilson et coll., 1991). Dans la stratégie CIS, des trains d'impulsions bi-phasiques sont délivrés avec une fréquence constante élevée dans le but de transmettre les variations rapides de l'enveloppe temporelle contenues dans le signal de parole. Les électrodes sont stimulées à des instants légèrement différents selon le principe de la stimulation séquentielle (non-simultanée) ce qui permet de minimiser les interactions inter-électrodes.

9.2.La stratégie SAS :

La stratégie SAS consiste à digitaliser les ondes analogiques puis à les reconvertir en ondes analogiques qui seront transmises simultanément aux paires d'électrodes. Elle permet une bonne retranscription des fluctuations temporelles du signal ce qui est important pour la perception de la parole mais également celle de la musique et la localisation des sons. Cette stratégie présente cependant le désavantage d'augmenter les interactions de champ électrique entre les électrodes.

9.3.La stratégie HiRes :

Avec la stratégie HiRes, la présentation de la stimulation est séquentielle et le nombre de canaux est plus élevé. Cela permet une plus grande résolution spectrale. La fréquence de stimulation est particulièrement élevée ce qui présente l'avantage de préserver la nature stochastique de l'activité neurale, d'augmenter la plage dynamique et de respecter les caractéristiques temporelles fines du signal. Il existe deux types de stratégies HiRes : HiRes-P

qui permet la stimulation simultanée d'une paire d'électrodes et HiRes-S qui stimule les électrodes en séquence.

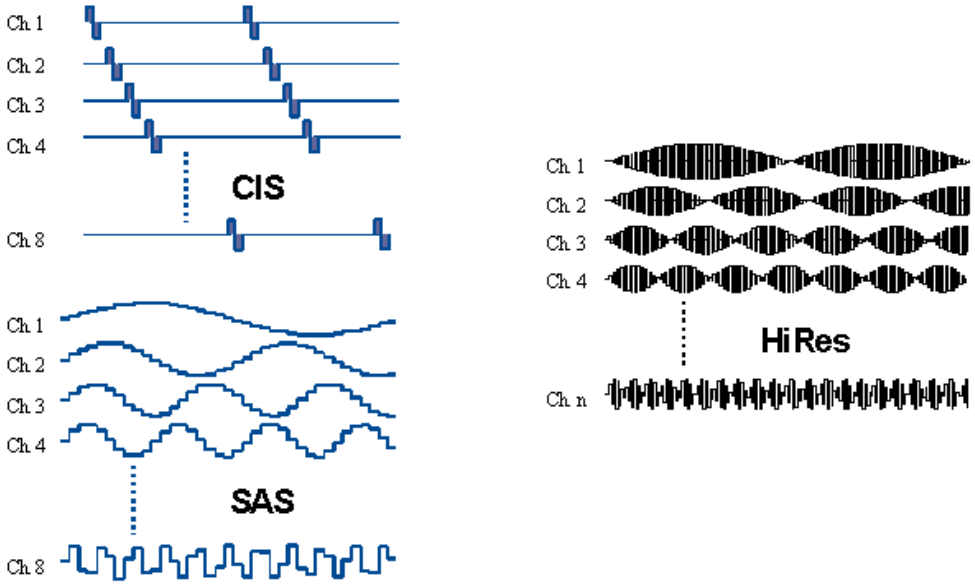


Figure I.18 : *Différentes stratégies de codage.*

10. Conclusion :

L'oreille perçoit et interprète les sons, elle permet à chacun de communiquer avec son entourage, de participer à des conversations. Les personnes atteintes de surdité sévère à profonde sont malheureusement souvent privées de cette faculté. L'implant cochléaire peut être une solution efficace. En effet, lorsque les prothèses auditives classiques n'apportent plus suffisamment de bénéfices pour la compréhension de la parole, l'implant cochléaire permet aux personnes atteintes de surdité sévère à profonde d'améliorer leur qualité de vie en leur donnant une meilleure intelligibilité dans le calme et dans le bruit, et en les aidant à affronter les défis du quotidien, de la vie professionnelle ou scolaire. Les patients implantés déclarent se sentir davantage en sécurité, ils sont aussi plus indépendants et retrouvent une vie sociale. Pouvoir utiliser le téléphone, prendre rendez-vous chez le médecin, regarder la télévision, participer à un dîner, entendre une alarme, partager des secrets, profiter du rire des enfants. Autant de sons qui ne seront plus à deviner.

Chapitre II :

Caractéristique du signal de la parole et les stratégies de stimulation cochléaires

1. INTRODUCTION :

Le traitement de la parole est aujourd'hui une composante fondamentale des sciences de l'ingénieur. Située au croisement du traitement du signal numérique et du traitement du langage, cette discipline scientifique a connu depuis les années 60 une expansion fulgurante, liée au développement des moyens et des techniques de télécommunications [10].

La parole est un moyen de communication très efficace et naturel utilisé par l'humain. Depuis longtemps, il rêve de pouvoir s'adresser par ce même moyen à des machines ce qui les rendre plus intelligentes [11].

L'importance de la compréhension de la parole en milieu bruité pour les personnes normo -entendantes, malentendantes ainsi que sourdes profondes est primordiale. En effet la compréhension de la parole permet aux humains, de pouvoir communiquer.

Dans ce chapitre, nous présentons les caractéristiques de la parole ainsi que les algorithmes de stimulation cochléaires et les analyseurs de parole.

2. La parole :

L'information portée par le signal de parole peut être analysée de bien des façons. On en distingue généralement plusieurs niveaux de description non exclusifs : acoustique, phonétique, phonologique, morphologique, syntaxique, sémantique, et pragmatique [12].

2.1.Le niveau acoustique [12] :

La parole apparaît physiquement comme une variation de la pression de l'air causée et émise par le système articulatoire. La *phonétique acoustique* étudie ce signal en le transformant dans un premier temps en signal électrique grâce au transducteur approprié : le microphone (lui-même associé à un préamplificateur).

De nos jours, le signal électrique résultant est le plus souvent numérisé. Il peut alors être soumis à un ensemble de traitements statistiques qui visent à en mettre en évidence les traits acoustiques : sa fréquence fondamentale, son énergie, et son spectre. Chaque trait acoustique est lui-même intimement lié à une grandeur perceptuelle : pitch, intensité, et timbre. L'opération de numérisation, schématisée à la figure II.1, requiert successivement : un filtrage de garde, un échantillonnage, et une quantification.

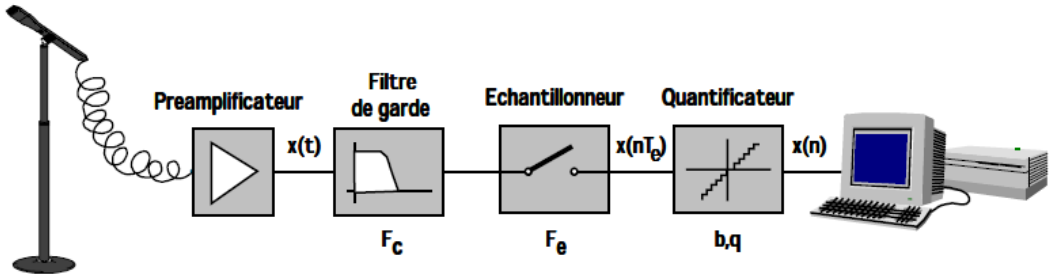


Figure II.1 : *Enregistrement numérique d'un signal acoustique.*

2.2.Transformée de Fourier à court terme :

La transformée de Fourier à court terme est obtenue en extrayant de l'audiogramme une 30aine de ms de signal vocal, en pondérant ces échantillons par une fenêtre de pondération (souvent une fenêtre de Hamming) et en effectuant une transformée de Fourier sur ces échantillons.

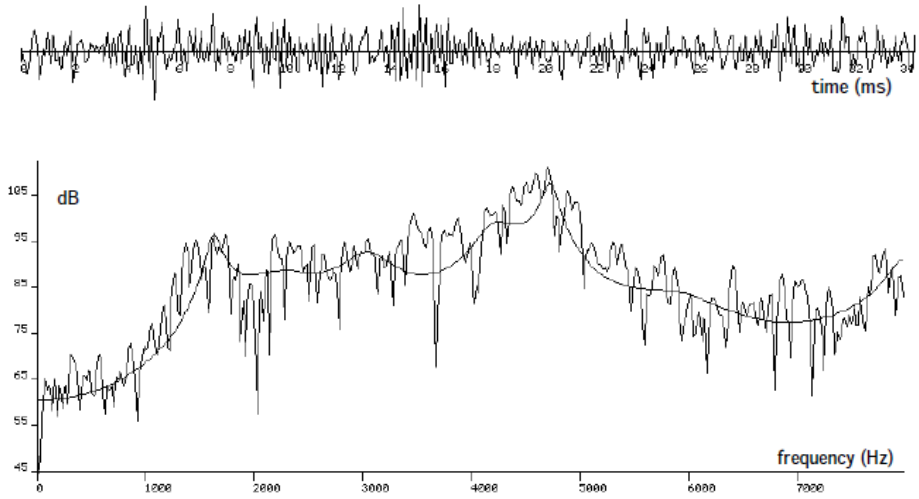


Figure II.2 : *Evolution temporelle (en haut) et transformée de Fourier discrète (en bas) du [p] de (signaux pondérés par une fenêtre de Hamming de 30 ms).*

La forme générale de ce spectre, appelée *enveloppe spectrale*, présente elle-même des pics et des creux qui correspondent aux résonances et aux antirésonances du conduit vocal et sont appelés *formants* et *anti-formants*. L'évolution temporelle de leur fréquence centrale et de leur largeur de bande détermine le timbre du son. Il apparaît en pratique que l'enveloppe spectrale des sons voisés est de type passe-bas, avec environ un formant par kHz de bande passante, et dont seuls les trois ou quatre premiers contribuent de façon importante au timbre. Par contre, les sons non-voisés présentent souvent une accentuation vers les hautes fréquences [12].

2.3.Spectrogramme [12] :

Il est souvent intéressant de représenter l'évolution temporelle du spectre à court terme d'un signal, sous la forme d'un spectrogramme. L'amplitude du spectre y apparaît sous la forme de niveaux de gris dans un diagramme en deux dimensions temps-fréquence. On parle de spectrogramme à large bande ou à bande étroite selon la durée de la fenêtre de pondération (Figure II.3). Les spectrogrammes à bande large sont obtenus avec des fenêtres de pondération de faible durée (typiquement 10 ms), ils mettent en évidence l'enveloppe spectrale du signal, et permettent par conséquent de visualiser l'évolution temporelle des formants. Les périodes voisées y apparaissent sous la forme de bandes verticales plus sombres. Les spectrogrammes à bande étroite sont moins utilisés. Ils mettent plutôt la structure fine du spectre en évidence : les harmoniques du signal dans les zones voisées y apparaissent sous la forme de bandes horizontales.

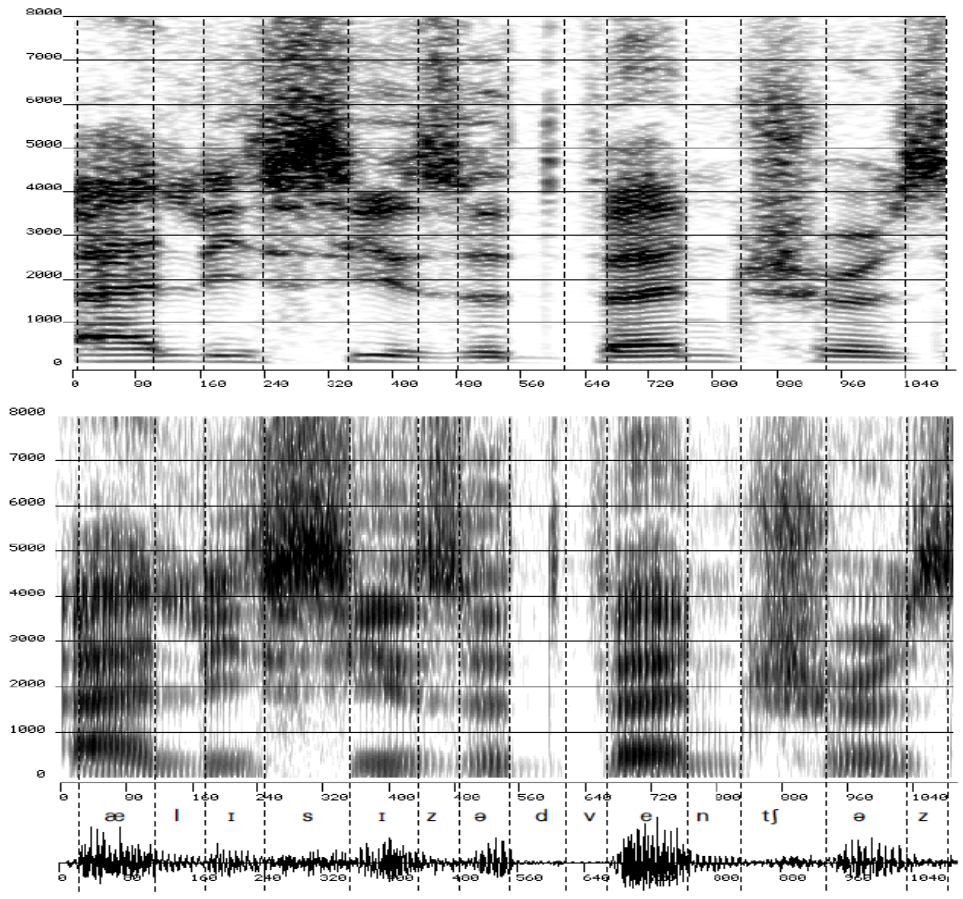


Figure II.3 : *Spectrogrammes à large bande (en bas), à bande étroite (en haut), et évolution temporelle de la phrase anglaise 'Alice's adventures', échantillonnée à 11.25 kHz (calcul avec fenêtres de Hamming de 10 et 30 ms respectivement).*

2.4.Fréquence fondamentale [12] :

Une analyse d'un signal de parole n'est pas complète tant qu'on n'a pas mesuré l'évolution temporelle de la fréquence fondamentale ou « pitch ». La figure II.4 donne l'évolution temporelle de la fréquence fondamentale de la phrase "les techniques de traitement de la parole". On constate qu'à l'intérieur des zones voisées la fréquence fondamentale évoluent lentement dans le temps.

Elle s'étend approximativement de 70 à 250 Hz chez les hommes, de 150 à 400 Hz chez les femmes, et de 200 à 600 Hz chez les enfants.

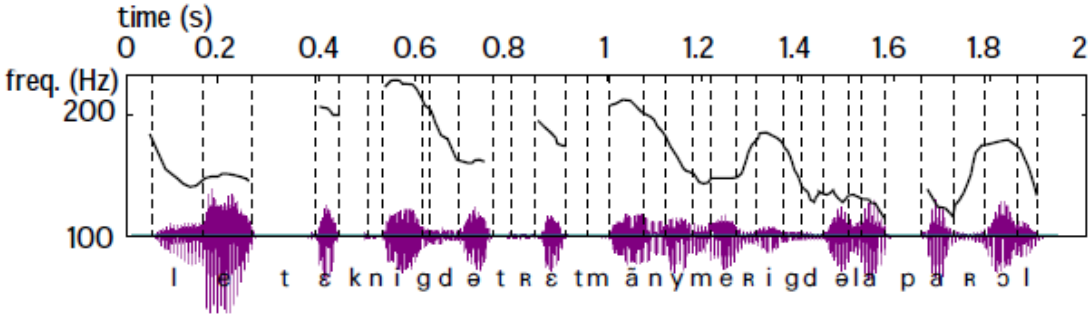


Figure II.4 : *Evolution de la fréquence de vibration des cordes vocales dans la phrase "les techniques de traitement numérique de la parole". La fréquence est donnée sur une échelle logarithmique ; les sons non-voisés sont associés à une fréquence nulle.*

2.5.Les Caractéristiques de la parole [13] :

Le signal vocal résulte des fluctuatons de la pression de l'air engendrées puis émises par l'appareil phonatoire. Ces organes sont les poumons, la trachée, le larynx, les cordes vocales, les pharynx, le nez et la bouche. Ils constituent un tube de forme complexe dont la partie supérieure s'appelle le conduit vocal et comprend le pharynx, la bouche et le nez. L'appareil respiratoire fournit l'énergie nécessaire à l'émission de la parole lorsque l'air est expiré par la trachée-atère. Au sommet de celle-ci se trouve le larynx ou la pression de l'air est modulée avant d'etre appliquée au conduit vocal qui s'étend du pharynx jusqu'aux lèvres. Les cordes vocales, placées au niveau du larynx peuvent se fermer et s'écarter en formant une ouverture triangulaire appelée glotte qui agit comme obstacle ajustable au passage de l'air. Ces cordes constituent la source d'exitation convertissant l'énergie en un bourdonnement qui se traduit par un langage perceptible.

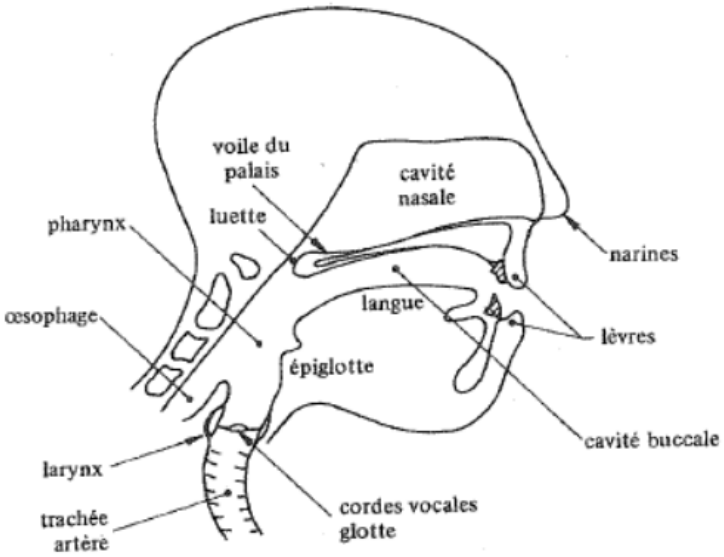


Figure II.5 : *L'appareil phonatoire.*

Selon la façon dont l'air passe à travers les cordes vocales, les sons de la parole peuvent être classés en des sons voisés et des sons non-voisés. Ces derniers sont engendrés par un libre passage de l'air via la glotte alors que les sons voisés résultent d'une vibration périodique des cordes vocales qui génère des impulsions de pression appliquées au conduit vocal. Les sons voisés sont ainsi périodiques ou semi-périodiques. Ils sont caractérisés par des résonances entre 200 Hz et 3500 Hz et des durées dépassant les 20 ms. Les sons non-voisés sont plutôt aléatoires et ils ne présentent aucune structure périodique. L'écoulement de l'air dans une constriction étroite du conduit vocal provoque un sifflement : C'est le son fricatif. Un tel son peut être non-voisé ou voisé. Son énergie est localisée dans les fréquences supérieures à 2 kHz.

L'occlusion momentanée suivie par une ouverture brusque du conduit vocal génère un son plosif (ou occlusif). Celui-ci peut être voisé ou non voisé. Enfin les sons dits sonnants résultent d'une libération de la pression de l'air dans la cavité nasale. Ceux-ci présentent une bande passante de 4 kHz. La figure II.6 illustre la répartition des phonèmes de la langue française en fonction des modes de phonation.

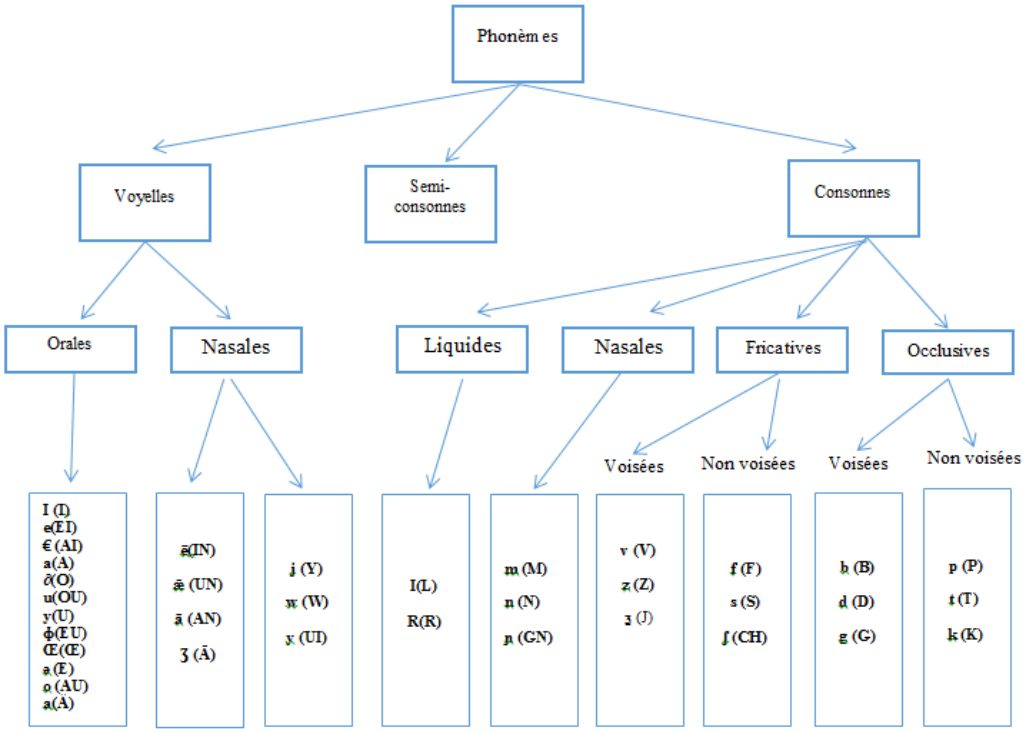


Figure II.6 : *Phonèmes de la langue française.*

De point de vue spectral, la parole présente des propriétés intéressantes pour le traitement. Le spectre du signal vocal dépend en général de la transmittance du conduit vocal, du rayonnement à partir des lèvres et du spectre de l'excitation, mais il peut être décomposé suivant deux allures de spectres correspondant aux séquences voisées et non-voisés du signal. Les spectres des sons voisés sont marqués par la présence de raies qui correspondent aux harmoniques du fondamental dont les enveloppes constituent les formants du son. Les spectres des sons non-voisés ne présentent aucune structure harmonique puisqu'ils correspondent à des séquences aléatoires.

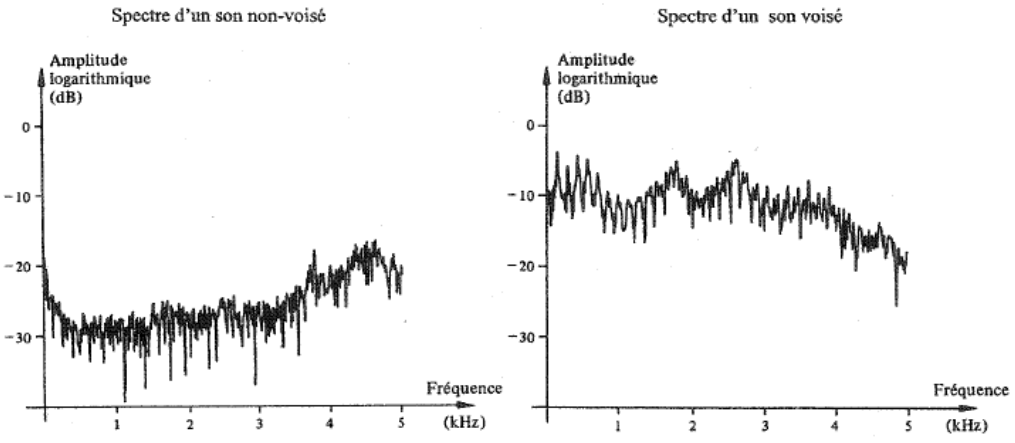


Figure II.7 : *Spectres d'un son non-voisé et d'un son voisé.*

Les propriétés spectrales de la parole révèlent un intérêt particulier pour les stratégies de stimulation. En effet, dans l'approche basée sur l'extraction des caractéristiques de la parole, on procède à une analyse de l'onde acoustique pour appliquer son allure spectrale selon l'organisation tonotopique de la cochlée. Une telle analyse devrait dégager les paramètres spectraux caractéristiques du segment de parole qui représenteront l'information à coder puis à envoyer à la cochlée par stimulation électrique [13].

3. **Les algorithmes de stimulation cochléaires et les analyseurs de parole [14] :**

Les premiers analyseurs de la parole sont intimement liés aux premiers algorithmes de stimulation ainsi qu'à leur implant respectif. Leur concept découlait directement des expériences cliniques effectuées ou du degré de compréhension du système auditif. Dès l'apparition des implants multi-électrodes, des analyseurs de paroles plus complexes ont été conçus pour réaliser des algorithmes de stimulation qui tiennent compte des avantages de ces systèmes. Malgré les progrès effectués sur l'électronique numérique, la majorité des systèmes qu'on retrouve aujourd'hui utilisent encore une approche analogique. Cependant, une évolution s'effectue lentement vers des systèmes partiellement numériques, voire vers des systèmes entièrement numériques. Jusqu'à présent, la majorité des systèmes sont conçus d'une manière trop spécifique aux algorithmes de stimulation qu'ils utilisent. Il est possible d'observer deux approches principales dans tous les algorithmes utilisés aujourd'hui, il s'agit des algorithmes basés sur l'extraction des caractéristiques de la parole, et des algorithmes basés sur l'allure temporelle des signaux large bande.

Ces deux approches se basent sur déférentes caractéristiques de la cochlée. La perception de la tonalité en fonction de la fréquence de stimulation peut être représentée par la figure II.8. Jusqu'à environ 400 Hz, l'oreille humaine moyenne peut discerner des variations de fréquences de stimulation. Au-delà de cette fréquence, il n'y a plus de discernement. Des algorithmes basés sur l'extraction des caractéristiques de la parole, se servent de la partie 0 – 400 Hz. Ils se servent de contenu fréquentiel du signal de la parole et jugent que la majorité de l'information est continue dans la fréquence fondamentale et les raies d'amplitudes maximales que sont formants.

Ainsi, pour représenter ces caractéristiques de la parole dans l'oreille interne, ils utilisent la fréquence fondamentale déterminée (<400 Hz) comme fréquence de stimulation et les formants pour déterminer la position de l'électrode qui fournira le stimulus parmi le réseau qui longe la membrane basilaire. Ces algorithmes nécessitent un système capable de détecter

les sons voisés et non voisés. Lors de l'apparition d'un son non voisé. La fréquence de stimulation n'est plus la fondamentale mais un signal aléatoire. Seulement quelques formants (F0, F1, F2) sont utilisés dans ces algorithmes. En ce qui concerne les algorithmes basés sur l'information temporelle large bande, ils travaillent à des fréquences supérieures à la limite de discrimination fréquentielle de la cochlée (>400 Hz) et tirent profit surtout de la tonotopicité de cette dernière. L'enveloppe du signal sortant des filtres varie dans le temps et module la fréquence de stimulation. Cette approche est intéressante car les filtres de basses fréquences présenteront un signal modulé se rapprochant de la fondamentale alors que les autres filtres présenteront la réponse fréquentielle du signal présent à l'entrée du système. Lors de la présence de signaux non voisés. Les filtres présent une réponse erratique ce qui fait que la stimulation s'adapte par elle-même à l'excitation présente à l'entrée.

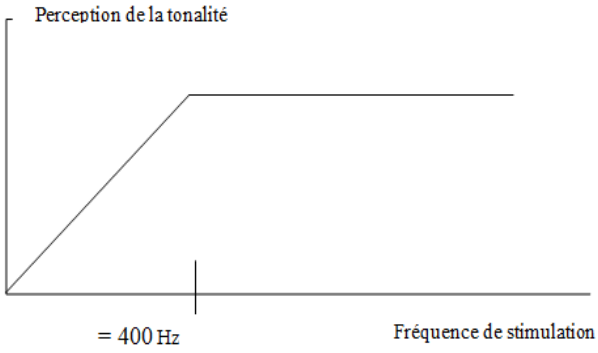


Figure II.8 : *Perception de tonalité en fonction de la fréquence de stimulation.*

3.1.L'algorithme de House-Urban :

Dans cette analyseur de parole, le son est tout simplement capté par un microphone, amplifié puis filtrés étroitement (200-400 Hz). Le signal disponible à la sortie du filtre surmodule l'amplitude d'une porteuse de 16 kHz. Le signal modulé est alors amplifié et expédié à l'implant. Cet implant a eu peu de succès étant donné sa simplicité et ses performances limitées. Cet algorithme est basé uniquement sur l'allure temporelle du signal.

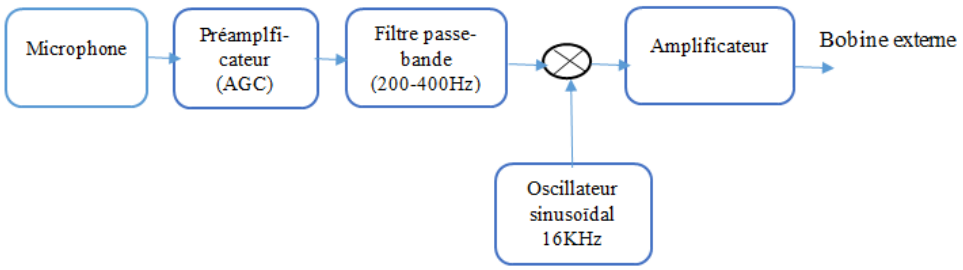


Figure II.9 : *Schéma de principe de l'analyseur de parole de House-Urban.*

3.2.L'algorithme Compressed Analog (CA) :

Cet algorithme, conçu à la fin des années 1980 consiste en une extension de l'algorithme de House-Urban. En effet, le principe de l'implant réside à multiplier le nombre de canaux de l'implant House-Urban et d'utiliser des filtres passe-bande à différentes fréquences.

L'analyseur consiste à utiliser quatre filtres passe-bande couvrant différentes plages de fréquence. Cette approche, permet d'utiliser la répartition fréquentielle de la cochlée mais le stimulus fourni est basé sur l'allure temporelle des sorties des filtres.

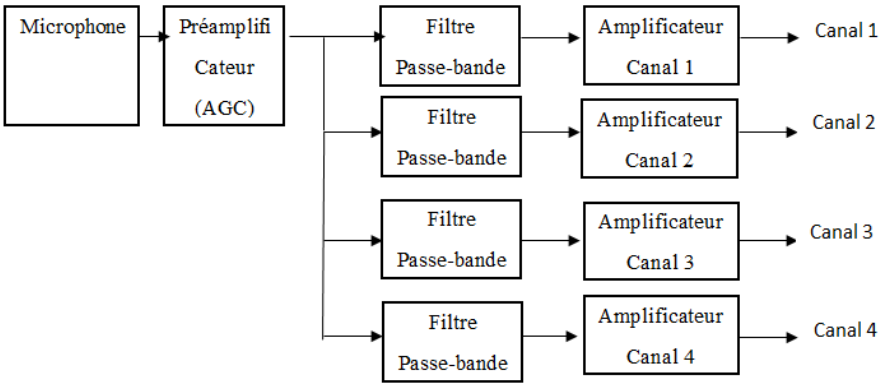


Figure II.10 : *Schéma de principe de l'analyseur de parole 'Compressed Analog'.*

3.3.L'Algorithme 'Continuous Interleaved Sampling' (CIS):

Avec l'algorithme Compressed Analog, les stimuli sont présentés simultanément sur les quatre canaux, ceci peut engendrer de l'interférence entre ces derniers lors de la stimulation. L'objectif de Continuous Interleaved Sampling (CIS) est d'éliminer les interférences en générant des stimulations non simultanées sur les canaux.

Dans cet algorithme, l'enveloppe du signal sera extraite grâce à un détecteur d'enveloppe et à un filtre passe bas au lieu d'utiliser directement le signal d'entrée comme dans le cas de House-Urban et Compressed Analog. Chaque branche comporte un filtre passe-bande qui extrait la partie du signal voulue et d'une fonction non-linéaire destinée à compresser le signal dans la plage dynamique du patient. Les canaux sont modulés par des impulsions non simultanées permettant d'éliminer toutes interférences entre canaux. Des études cliniques ont démontré la supériorité du CIS par rapport au CA.

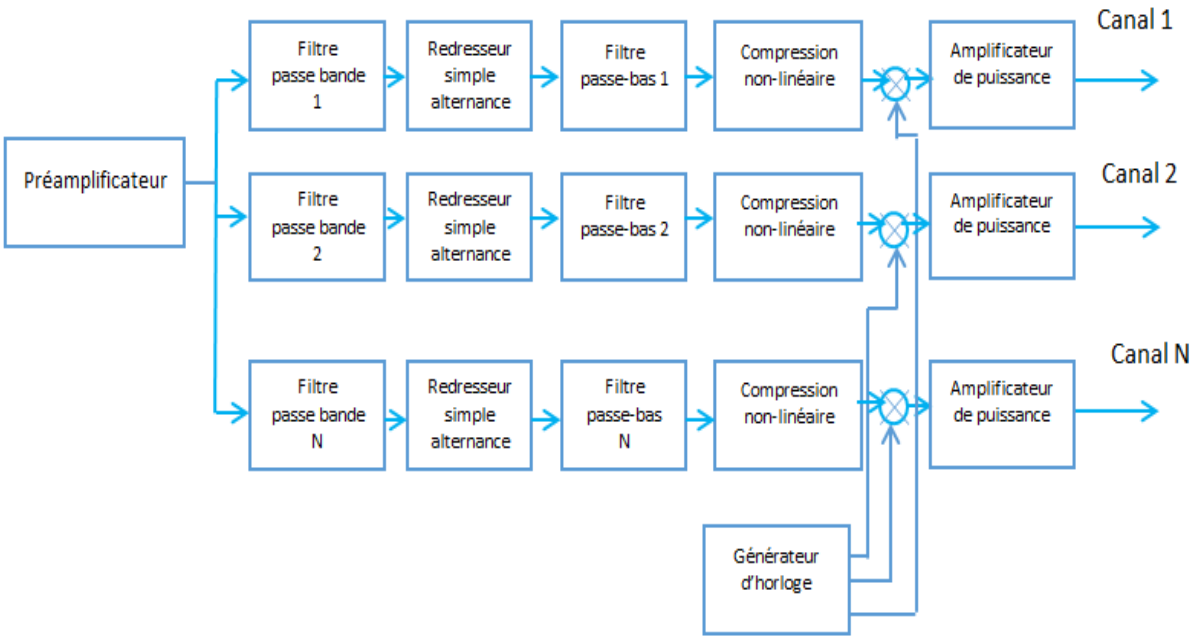


Figure II.11 : *Schéma bloc de l'analyseur CIS.*

3.4.Le 'wearable speech processor' de Nucleus:

Les algorithmes basés sur les caractéristiques de la parole vont chercher à déterminer différents paramètres du signal tel que la fréquence fondamentale (F0), la fréquence du premier et deuxième formant (F1 et F2) ainsi que leurs amplitudes (A1 et A2) respectives tel que le démontre la figure II.12.

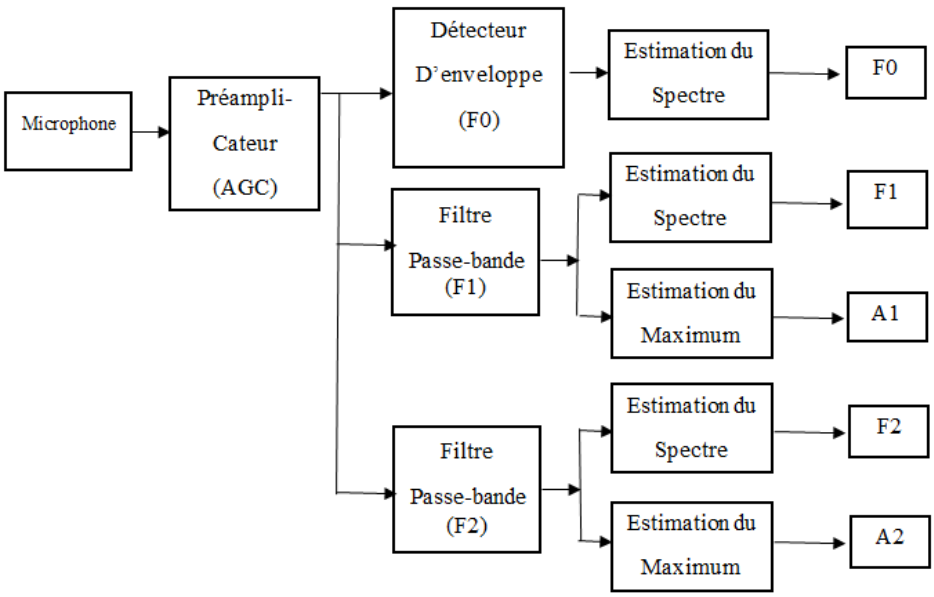


Figure II.12 : *Schéma bloc basé sur l'extraction de caractéristique de la parole.*

Le WSP est le premier analyseur conçu pour extraire les caractéristiques fréquentielles de la parole. Cet algorithme utilise un amplificateur à gain contrôlé (AGC) pour compresser dynamiquement le signal sonore. La fréquence fondamentale est extraite par un détecteur d'enveloppe composé d'une redresseuse double alternance suivit d'un filtre passe-bas à 270 Hz. Les premiers et deuxièmes formants sont extraits via des filtres passe-bande et des détecteurs de maximum. Les différents paramètres puis expédiés à l'implant qui s'occupe de stimuler la cochlée. La majeure partie de l'analyseur de parole est analogique.

3.5.Le Spectral Maxima Sound Processor :

Cet analyseur de parole de Nucleus, utilisé pour la première fois en 1989, utilise une extension de l'approche précédente. Le signal capté par un microphone, est amplifié et compressé à une banque de 16 filtres passe-bande à capacité commutée. Les 16 filtres sont répartis entre les fréquences 250 Hz et 5400 Hz. La sortie des filtres est numérisée sur 8 bits avant d'être traitée par un microprocesseur. Le rôle de ce dernier est de corriger le gain des filtres et de sélectionner les six filtres présentant les plus grandes énergies spectrales qui seront utilisés pour stimuler sur leurs électrodes correspondantes. Les données appropriées sont alors générées et expédiées à l'encodeur avant d'être présentées au transmetteur RF. Les amplitudes sont rafraichies toutes les quatre millisecondes. Ce taux est limité par la capacité de communication du lien RF.

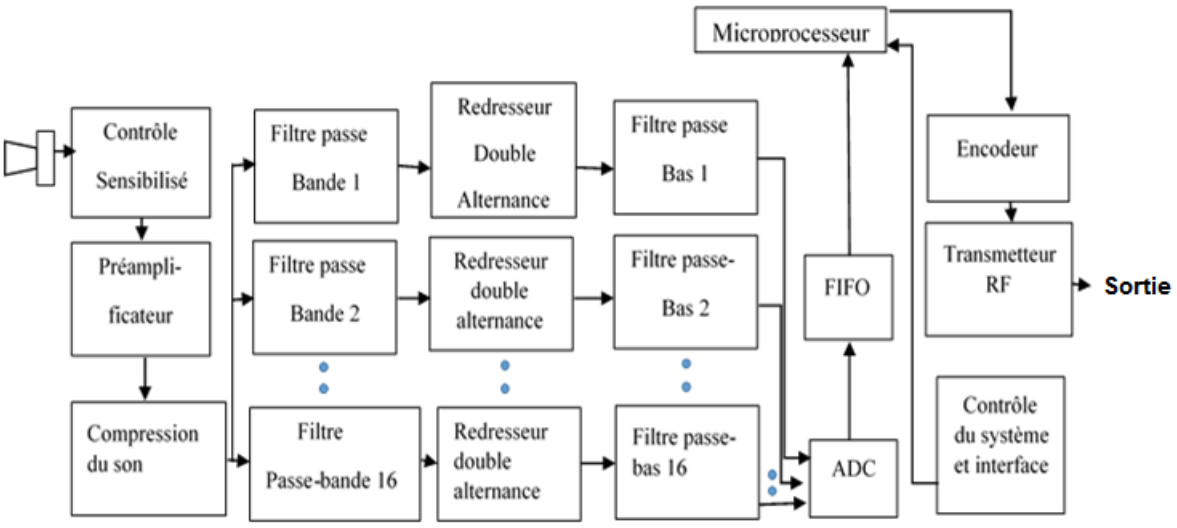


Figure II.13 : *Schéma du Spectral Maxima Sound Processor.*

3.6.Le ‘Mini Speech Processor’ (MSP):

Le ‘Mini Speech Processor’ (MSP) de Nucleus est une version numérique du WSP. Il est un des premiers analyseurs basés sur un DSP. Le MSP a été utilisé pour simuler numériquement l’algorithme du SMSP. Une transformée de Fourier est effectuée sur le signal sonore et génère 65 amplitudes également espacées de 0 à 6 KHz. Seuls les composantes impaires sont conservées et la première valeur générée par la transformée est éliminée. Des 31 bandes de fréquences restantes, 16 sont utilisés pour déterminer les électrodes qui seront activés. La partie basse du spectre estime huit bandes également espacées en regroupant les données produites par la transformée. La partie haute regroupe huit bandes de façon logarithmique afin d’être près des bandes de fréquences du SMSP. Sur les 16 bandes, les six ayant les plus hautes amplitudes seront sélectionnées pour effectuer une stimulation sur leurs électrodes correspondantes. Le rafraichissement des stimulations s’effectue à une fréquence de 400 Hz. La figure II.14 résume le traitement effectué au signal sonore.

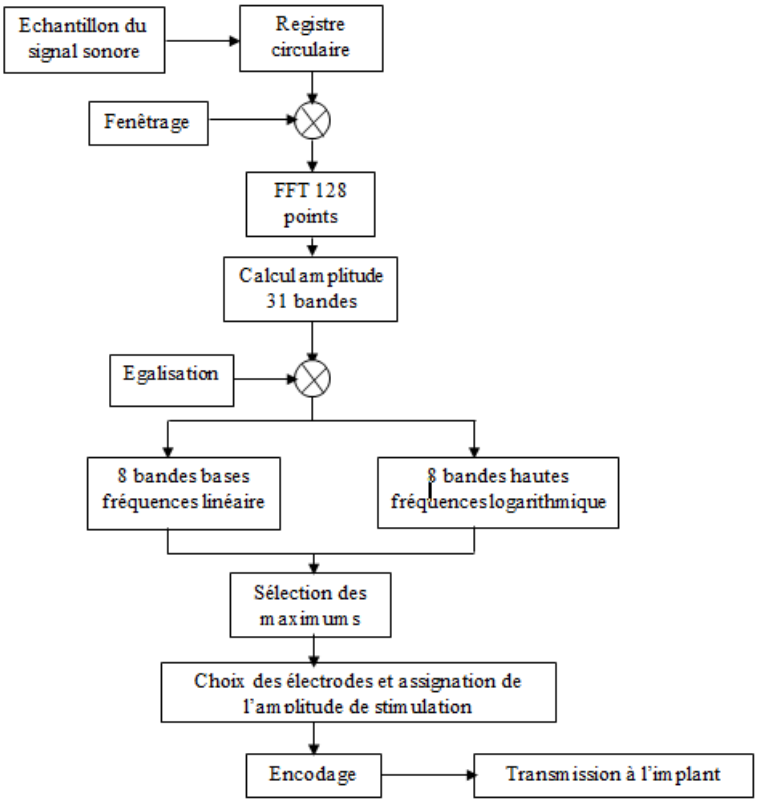


Figure II.14 : *Algorithme réalisé sur le MSP.*

Des essais cliniques ont été effectués pour comparer le SMSP et le MSP. Choisissant les six filtres ayant l’énergie la plus élevée alors que le MSP utilisait quatre électrodes

sélectionnées à chaque impulsion glottale. Dans ce cas, les résultats démontrent que SMSP est meilleure que le MSP.

4. CONCLUSION :

Dans ce chapitre, nous avons présenté les caractéristiques du signal de la parole, ainsi que les différentes stratégies de stimulation cochléaires existantes, ces dernières ont des performances remarquables dans les implants de nos jours, et restent toujours pas suffisant de point de vue identification des locuteurs, ainsi que les sons environnants.

Chapitre III :

Résultats de simulation de la stratégie de codage basée sur MFCC

1. INTRODUCTION :

Dans ce chapitre, nous présentons les résultats de simulation de la stratégie de codage de la parole étudiée pour les implants cochléaires. Dans une première partie, nous commençons à expliquer le principe d'extraction des paramètres de la parole dite « MFCC », qui est appliquée sur un seul segment du signal afin de bien comprendre son principe. Une deuxième partie sert à l'extraction des coefficients MFCC d'un signal audio, et de générer les signaux de stimulation cochléaire à partir de ces coefficients.

Les résultats obtenus pour la stratégie de codage réalisée sont présentés à la fin de ce chapitre.

2. Extraction des coefficients MFCC [16] :

Le schéma bloc suivant (figure III.1), montre les différents étapes aboutiront à l'extraction des coefficients MFCC.

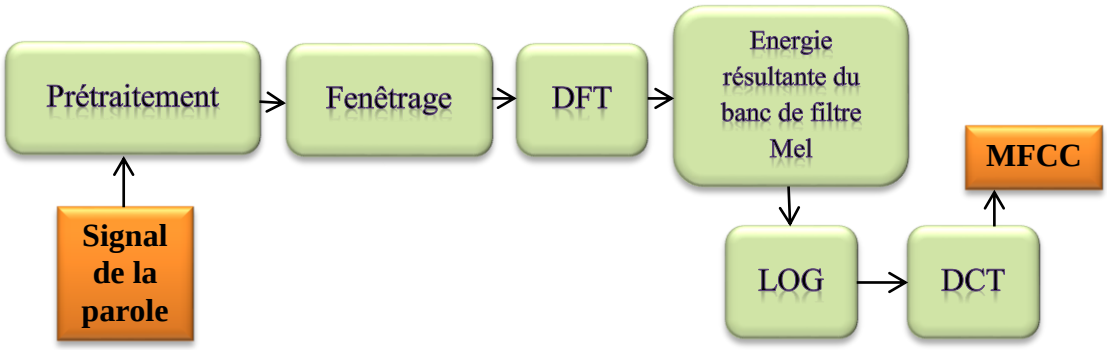


Figure III.1 : Schéma bloc de l'extraction des coefficients MFCC.

2.1.Données de traitement :

Un fichier audio contenant le phonème de la voyelle A, a été enregistré sous le nom « Ah.wav ». L'enregistrement est sous format 16 bits, à échantillonnage 8 KHz. La zone utile entière sans silence, est située entre les échantillons n°2320 et n°4000. Soit 1680 échantillons sur 210 ms.

Le script suivant permet de lire et de tracer le signal audio :

```
[y, fs, nbits] = wavread('Ah.wav');   % fs=8kHz, nbits=16bits,
plot([1:length(y)]/fs,y) % le tracé de y en fonction de temps
```

La figure III.2 montre le résultat de ce script.

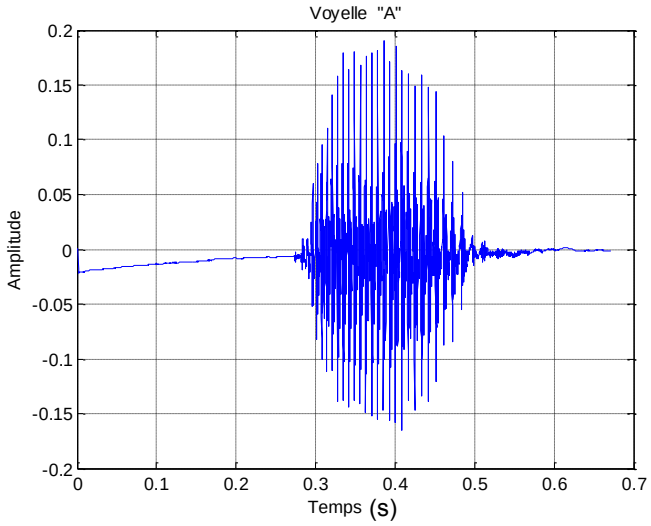


Figure III.2 : *Le tracé temporel du signal audio « Ah.wav ».*

Un segment de 2^8 = 256 échantillons (cela équivaut à une durée de 32ms), situé approximativement au centre de la zone utile du fichier « Ah.wav » est extrait.

Il est nécessaire de mettre en forme le signal de parole. Pour cela, quelques opérations sont effectuées avant tout traitement. Le signal est tous d'abord filtré puis échantillonné à une fréquence donnée. Une préaccentuation est effectuée afin de relever les hautes fréquences, qui sont moins énergétiques que les basses fréquences ; la préaccentuation $s_2(n)$ de l'échantillon $s(n)$ à l'instant α est calculée pour une valeur comprise entre 0,9 et 1 c'est comme le son articulé $s(n)$ est envoyé à un filtre passe-haut :

$$s_2(n) = s(n) - \alpha * s(n-1) \quad \text{(III.1)}$$

La transformée en Z du filtre est :

$$H(z) = 1 - \alpha z^{-1} \quad \text{(III.2)}$$

2.2.Fenêtrage Hamming :

Un prétraitement ensuite du segment consiste à son fenêtrage Hamming sur la même durée de 32 ms.

Une multiplication par une fenêtre de pondération W_n est effectuée, afin de réduire les effets de bords. Le choix se porte généralement sur les fenêtres de Hamming :

$$W(n) = 0{,}54 - 0{,}46\cos\left(2\pi\frac{n}{N-1}\right), 0 \le n \le N-1 \quad \text{(III.3)}$$

$$\text{Avec :} \quad s_3(n) = Wn.s(n) \quad \text{(III.4)}$$

Le script suivant décrit cette tâche :

```
segment = 256;
e=y(3088:3087+segment); % fenêtrage à 256 échantillons (durée 32ms)
preAccentuation = filter([1 -.97], 1, e);
Data = preAccentuation.*hamming(segment);
plot([1:length(Data)]/fs,Data)
```

Le résultat de ce script est représenté sur la figure III.3.

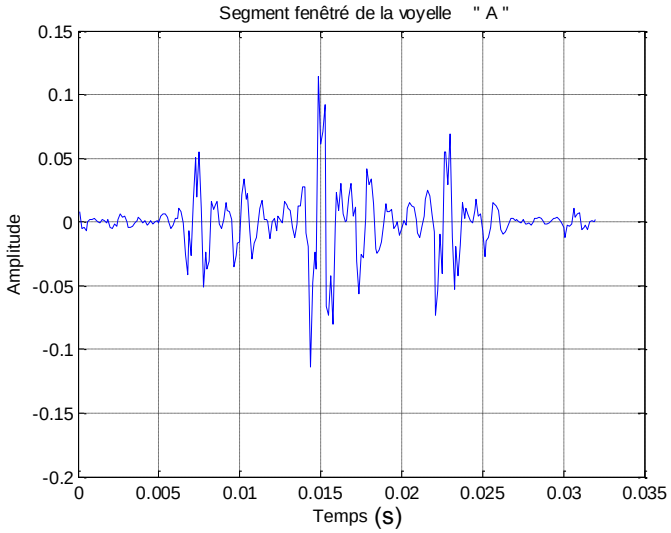


Figure III.3 : *Segment de la voyelle « A » après fenêtrage Hamming.*

2.3.Transformée de Fourier discrète :

Après cette mise en forme du signal (commune à la plupart des méthodes d'analyse de la parole), une transformée de Fourier discrète DFT en particulier FFT (Transformé de Fourier Rapide) est appliquée pour passer dans le domaine fréquentiel.

Dans ce contexte, on passe par le calcul du module du spectre et son logarithme, qui sont retrouvés en exécutant le script suivant :

```
F= [1:31:3968]';
TF=(abs(fft(Data,segment)));     % spectre par FFT
subplot(211),
plot(F,TF(1:segment/2))          % le trace du module de spectre
subplot(212),
plot(F,20*log10(TF(1:segment/2))       % le trace de Log du module de spectre
```

La figure III.4 montre le résultat d'exécution du script :

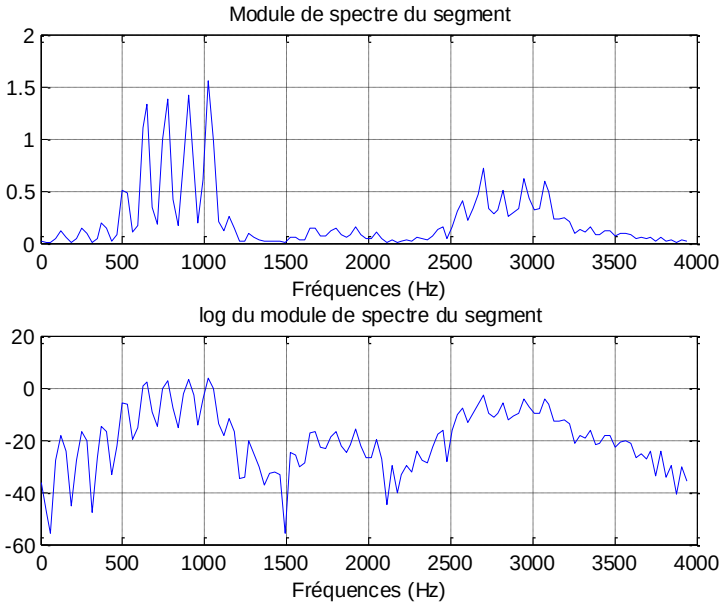


Figure III.4 : *Module de spectre du segment (Haut) et son logarithme (bas).*

2.4.Filtrage Mel :

La bande Mel est répartie suivant les fréquences :

$$F_{mel} = \frac{1000}{log_{10}(2)} log_{10}\left(1 + \frac{f}{1000}\right) \quad \text{(III.5)}$$

Le banc de filtre Mel préconisé en pratique pour une bande de 4kHz de la parole, comporte 22 ou 24 filtres opérant sur l'énergie du signal d'entrée prétraité par fenêtrage Hamming et préaccentuation. L'énergie est tirée à partir du spectre du signal (module au carré du spectre $X(k)$), alors que le filtrage Mel au rang *r*, $MF(r)$, assure l'opération suivante :

$$MF(r) = log_{10}\left(\frac{\sum_{k=L_r}^{U_r} |V(k)X(k)|^2}{\sum_{k=L_r}^{U_r} |V(k)|^2}\right) \quad \text{(III.6)}$$

Où $V(k)$ sont les paramètres de construction propre au filtre *r*, s'étendant d'une limite inférieure L_r (lower) vers une limite supérieure U_r (upper). Le dénominateur apporte une pondération du filtre qui assure une certaine normalisation en amplitude. Les coefficients cepstraux filtrés en Mel, MFCC, sont alors déterminés en prenant la transformée en cosinus du signal filtré $MF(r)$.

Ces coefficients sont une approximation du cepstre, et fournissent simplement une représentation orthogonale et compacte du log du module du spectre.

Dans le banc de filtre Mel, la détermination du nombre de filtre N_{Lin} opérant dans la partie fréquentielle linéaire (au-dessous de 1 kHz) et du nombre de filtre N_{Log} dans celle logarithmique (au-dessus de 1 kHz), s'est effectuée de la manière approximative suivante :

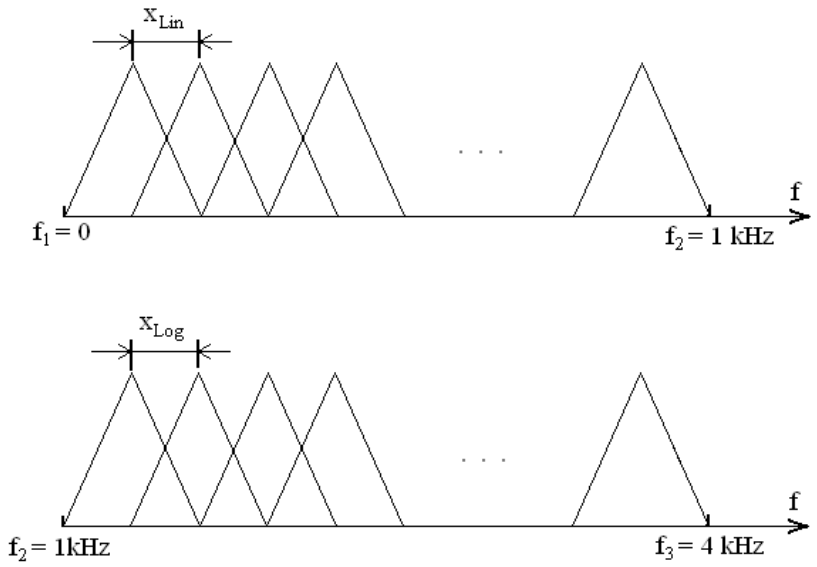


Figure III.5 : *Espacement linéaire des filtres jusqu'à 1 kHz et logarithmique après.*

Selon la Figure III.5, on peut écrire pour la partie à espacement linéaire :

$$f_1 + x_{Lin}.N_{Lin} = f_2 \quad \rightarrow \quad x_{Lin}.N_{Lin} = f_2 - f_1$$

Pour $x_{Lin} = 100\ Hz$, nous obtenons : $N_{Lin} = f_2/x_{Lin} = 1\ kHz/0.1\ kHz = 10\ filtres.$

Pour la partie à espacement logarithmique :

$$log(f_2) + N_{log} log(x_{log}) = log(f_3) \quad \rightarrow \quad log(f_2) + log({x_{log}}^{N_{log}}) = log(f_3)$$
$$\rightarrow \quad f_2.{x_{log}}^{N_{log}} = f_3$$

Pour $N_{log} \geq N_{Lin}$, et en fixant $N_{log} = 13\ filtres$, on obtient :

$$x_{log} = (f_3/f_2)^{1/N_{log}} = 4^{1/13} = 1.1125$$

Mais en recalculant tout en tenant en compte cet espacement, on aboutit à $f_3 = 4.45\ kHz$ au lieu de 4 kHz fixé au départ. Le bon ajustage donne $N_{log} = 12$, soit un nombre total des filtres du banc atteignant les 22.

Nous choisissons un banc de 22 filtres Mel occupant une bande de 4 kHz. Nous prenons 10 filtres à espacement linéaire de 100, et 12 filtres à espacement logarithmique de 1.11253. Une normalisation (<1) du poids de chaque filtre est effectuée en fonction des limites de bande, basse et haute :

$$Taille\ Triangle = 1/(haut - bas)$$

Ce qui se manifeste par la décroissance des sommets de gabarit des filtres, notamment pour ceux opérant au-dessus de 1 kHz et dont les bandes s'élargissent en allant vers la fréquence limite de 4 kHz.

Le script suivant montre comment le banc de filtres Mel est construit :

```
% paramètres du banc de filtres (calcul des poids du filtre mel) :
FreqBasse = 0; FiltresLIN = 10; EspaceLIN = 100; FiltresLOG = 12; EspaceLOG =
1.11253;
FiltresTotal = FiltresLIN + FiltresLOG;          % banc de 22 filtres Mel
freqs = FreqBasse + (0:FiltresLIN-1)*EspaceLIN; % freqs avec 1 ligne et 10 col
freqs(FiltresLIN+1:FiltresTotal+2) = ...         % ajout de 14 colonnes, de la
                                                 % 11ème à la 24ème colonne
freqs(FiltresLIN) * EspaceLOG.^(1:FiltresLOG+2);
bas = freqs(1:FiltresTotal);
centre = freqs(2:FiltresTotal+1);
haut = freqs(3:FiltresTotal+2);
TailleTriangle = 1./(haut-bas); % normalisation (<1) du poids de chaque filtre
fftFreqs = (0:duree-1)/duree*fs;
PoidsFiltres = zeros(FiltresTotal,duree); % initialisation
for num=1:FiltresTotal  % réalisation de chaque gabarit trianglulaire
  PoidsFiltres(num,:) = ...
 (fftFreqs > bas(num) & fftFreqs <= centre(num)).* ... % pente ascendante
  TailleTriangle(num).*(fftFreqs-bas(num))/(centre(num)-bas(num))+ ...
  % pondération des sommets
  (fftFreqs > centre(num) & fftFreqs < haut(num)).* ...% pente descendante
   TailleTriangle(num).*(haut(num)-fftFreqs)/(haut(num)-centre(num));
  % pondération des minimums
End
```

La figure III.6 montre le tracé du banc de filtres Mel réalisé :

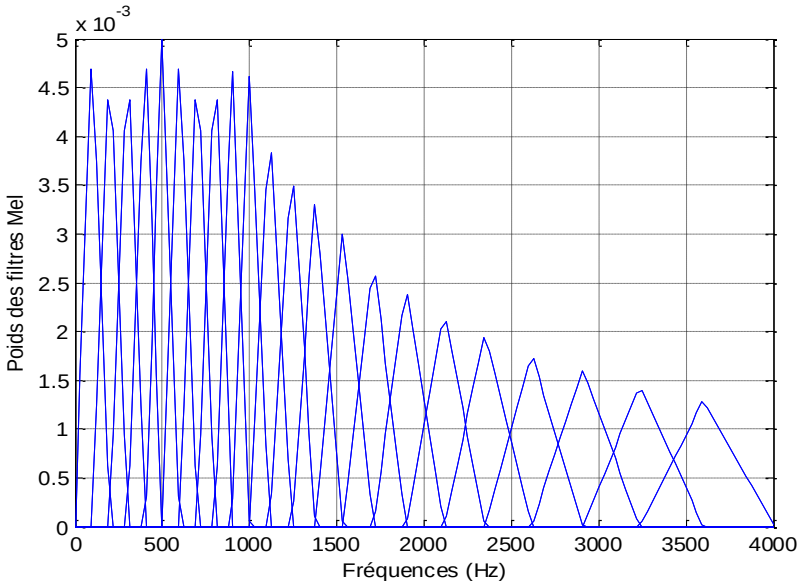


Figure III.6 : *Banc à 22 filtres à l'échelle Mel.*

2.5.Calcul des coefficients MFCC :

Une fois l'énergie spectrale du segment de la parole déterminée (carré du module du spectre), qui est calculée comme étant la somme des 256 raies de l'énergie spectrale, elle passe à travers le banc de filtres Mel.

Le script suivant montre la méthode de calcul de l'énergie spectrale :

```
AmpFFT = abs(fft(Data));
AmpFFT = AmpFFT.^2;       % énergie du segment sur la durée 32ms
Energie = 0;
for num=1:segment;        % calcul de l'énergie spectrale
    Energie = AmpFFT(num) + Energie;
end
Energie
```

Le spectre résultant est mis en échelle dB puis pondéré par l'énergie totale du segment comme il est montré dans le script suivant :

```
spectre=10*log10(Energie * PoidsFiltres * AmpFFT );    % spectre mel lissé
(pondéré par l'énergie), en dB
```

En calculant la transformée en cosinus (dct) du spectre (pondéré et mis en échelle dB), vingt-deux coefficients MFCC sont alors déduits.

La transformée en cosinus d'un vecteur x de longueur N, s'écrit :

$$y(k) = w(k).\sum_{n=1}^{N} x(n).cos\left[\frac{\pi(2n-1)(k-1)}{2N}\right] \quad , \quad k = 1,2,\dots,N \qquad \text{(III.7)}$$

Où : $w(k) = \begin{cases} 1/\sqrt{N} & k = 1 \\ \sqrt{2/N} & 2 \le k \le N \end{cases}$

Une interpolation à l'aide de la fonction Matlab « spline » rassemble les 22 coefficients sous forme d'une courbe, comme il est montré dans la Figure III.7.

```
MFCC = dct(spectre);
T= [1:0.1:length(MFCC)]';              % 128 points de résolution fréquentielle
MFCCInterpolee = spline(1:length(MFCC),MFCC,T);        % interpolation
```

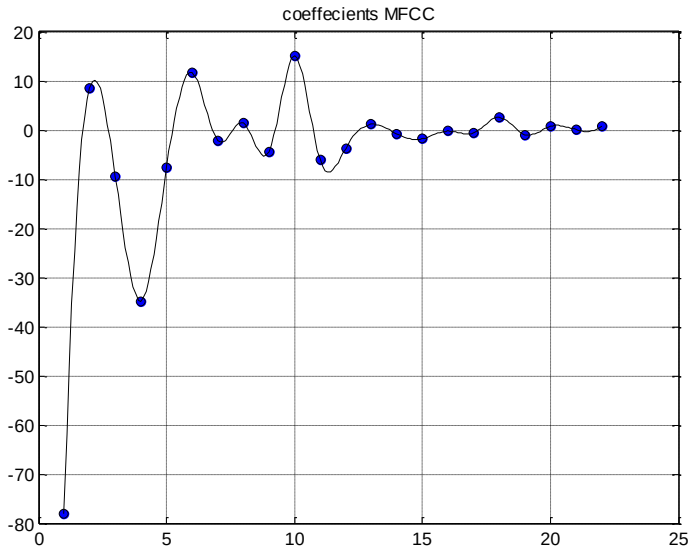


Figure III.7 : *vingt-deux coefficients MFCC interpolés à l'aide de la fonction spline.*

3. Reconstruction du spectre à partir des coefficients MFCC [16] :

Dans cette étape, nous montrons comment le spectre du signal original est reconstruit à partir des coefficients MFCC.

La méthode est décrite dans ce script :

```
% reconstruction du spectre lissé mel cepstral
melcep = idct(MFCC);    % melcep = spectre mel lissé
figure(1); plot(centre,melcep,'ok'); hold on;
MelCepInterpolee = spline(centre,melcep,F);
% interpole melcep aux valeurs centre suivant une résolution F
```

```
plot(F,MelCepInterpolee,'k');
hold on
plot(F,20*log10(TF(1:segment/2)),'r');    % |FFT|-en-dB (rouge)
```

Le spectre lissé, est retrouvé par transformée en cosinus inverse « idct » des coefficients MFCC. En réunissant les 22 points du spectre lissé par interpolation spline, la Figure III.7, montre la trace résultante (en noir) en superposition au tracé en dB du spectre du segment de la parole (en rouge). On constate que le spectre lissé épouse mieux la forme du spectre |FFT| en dB.

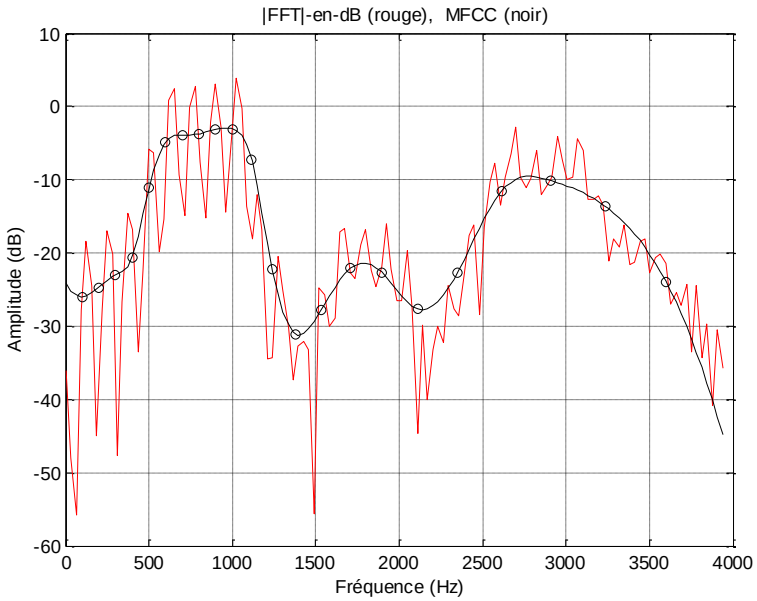


Figure III.8 : *Spectre en dB (rouge) confronté au filtrage Mel interpolé par spline (en noir).*

4. APPLICATION DE LA METHODE DE RECONSTRUCTION DU SPECTRE SUR LA TOTALITE DU SIGNAL [16] :

4.1. Segmentation et Fenêtrage :

Le signal de la parole est de nature non stationnaire, il est donc nécessaire, avant d'extraire les paramètres du signal, de la subdiviser en segments. Cette étape permet d'obtenir pour chaque segment de parole un signal quasi stationnaire. Les discontinuités aux extrémités des segments peuvent être amoindries en multipliant chaque segment par une fenêtre de Hamming (figure III.9).

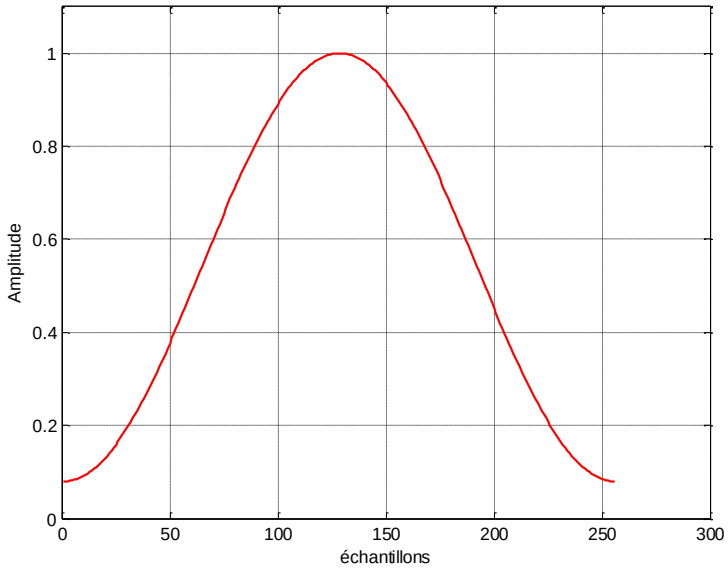


Figure III.9 : *Fenêtre de Hamming.*

Une segmentation du signal audio « Ah.wav » par des segments de 256 échantillons, nous délivre un nombre total de 21 segments.

La figure III.10, montre le tracé en superposition de tous les segments après son fenêtrage Hamming.

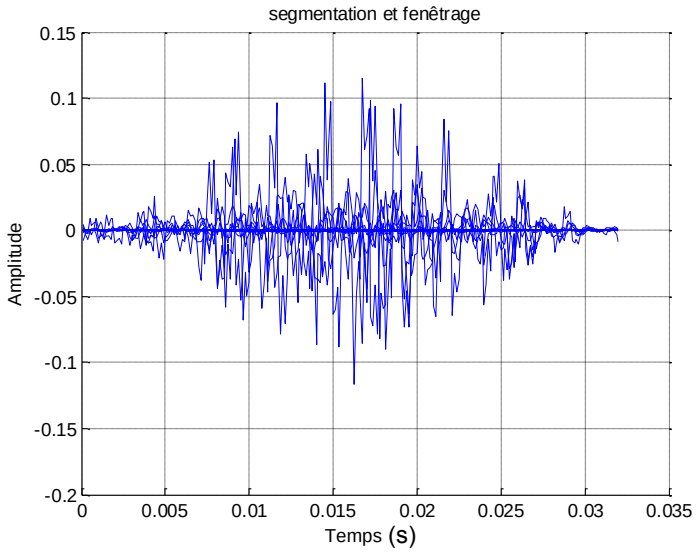


Figure III.10 : ***Segmentation et fenêtrage du signal « Ah.wav ».***

L'étape suivante consiste à calculer le module de spectre de tous les segments du signal de la parole, la figure III.11 nous présente le tracé des spectres superposés des segments.

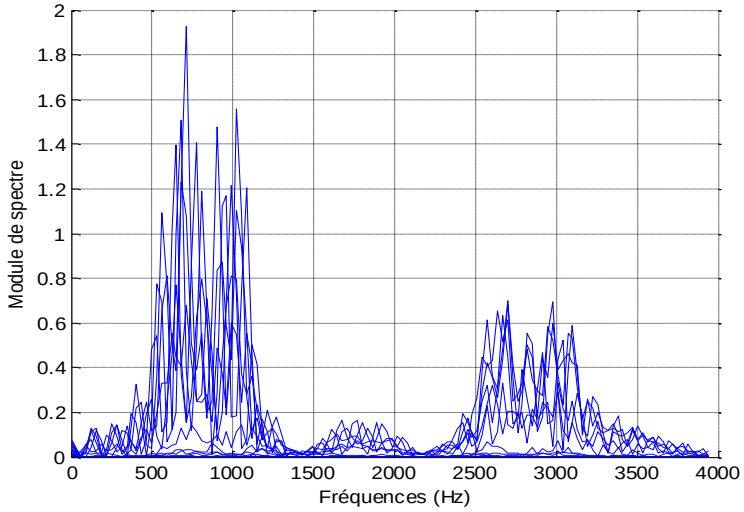


Figure III.11 : *Module de spectre de tous les segments du signal « Ah.wav ».*

Après le calcul de logarithme du module de spectre de tous les segments du signal de la parole, nous pouvons tracer le résultat comme il est montré dans la figure III.12.

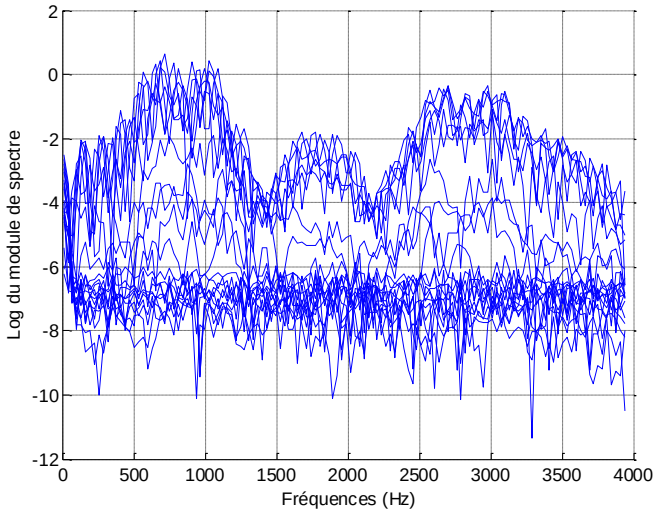


Figure III.12 : *Logarithme du module de spectre des segments.*

4.2.Les coefficients MFCC :

Afin d'extraire les coefficients MFCC des segments, nous procédant comme il est été décrit dans la partie précédente pour un seul segment, donc il faut d'abord calculer l'énergie spectrale des segments, cette dernière est pondérée par la suite par un banc de filtres Mel et est mis en échelle dB.

La dernière étape consiste à calculer la transformée en cosinus discrète (dct) du spectre résultant afin d'obtenir les coefficients MFCC de tous les segments (figure III.13).

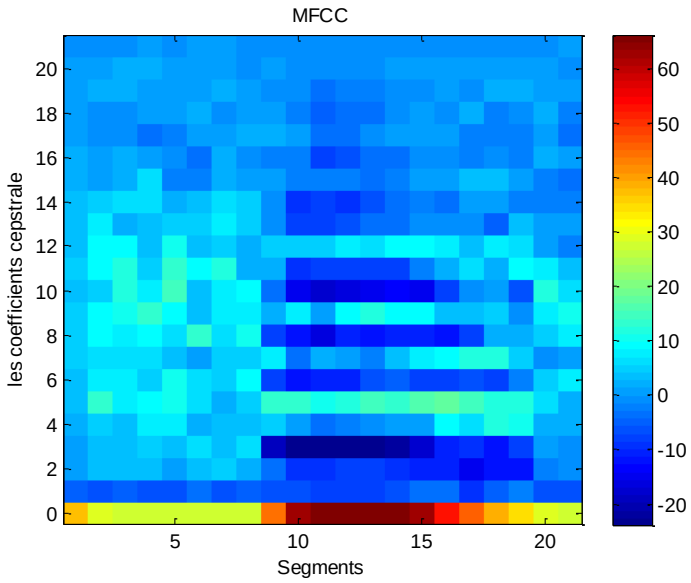


Figure III.13 : *Coefficients MFCC du signal « Ah.wav ».*

4.3.Reconstruction du spectre :

Afin de reconstruire le spectre du signal audio à partir des coefficients MFCC, nous suivons les mêmes étapes que pour un seul segment du signal, présentés dans la première partie de ce chapitre.

Après le calcul de la transformée en cosinus inverse (idct) des coefficients MFCC de chaque segment, nous utilisons la fonction Matlab (spline) afin d'interpoler ces coefficients par une courbe.

La figure III.14, présente les spectres reconstitués des 21 segments tracés en superposition.

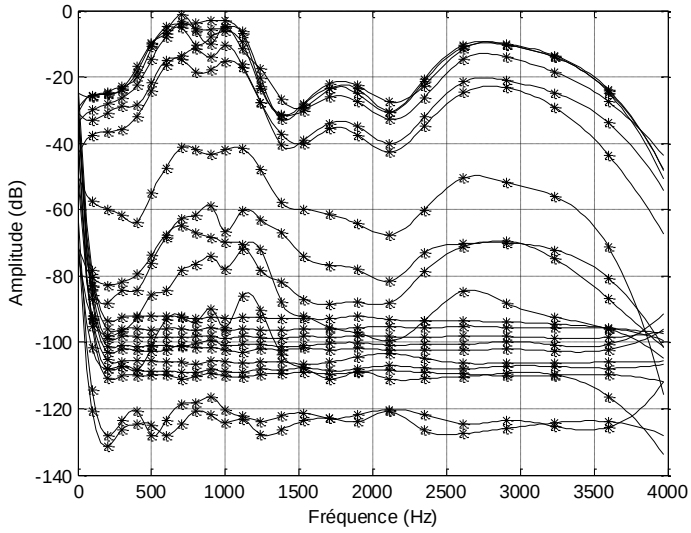


Figure III.14 : *Spectres reconstitués de tous les segments.*

5. Synthèse du signal de la parole :

Après l'analyse MFCC du signal de référence, nous passons à l'étape de synthèse qui a pour but de reconstruire le signal original à partir des coefficients MFCC.

Pour cela, nous avons généré un signal carré bi-phasique, similaire au signal utilisé pour stimuler le nerf auditif du patient en utilisant les implants cochléaires. La figure III.15 présente le signal généré.

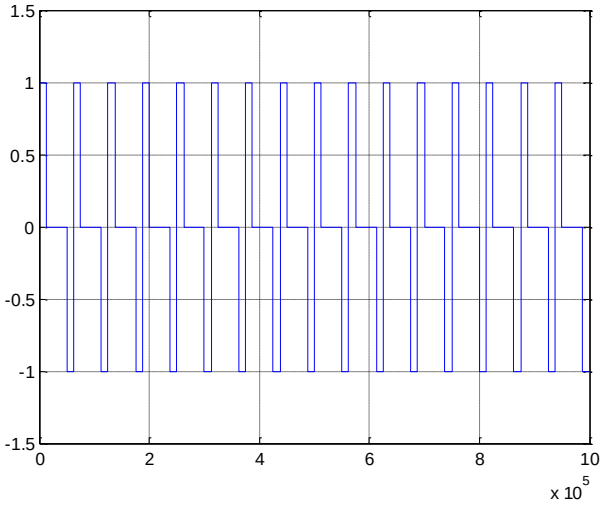


Figure III.15 : *Signal carré bi-phasique.*

En modulant le spectre des segments par le signal carré précédent, et en appliquant la transformé de Fourier inverse (IFFT) afin de retrouver le signal temporel, nous obtenons le signal synthétisé qui présente une estimation du signal initial « Ah.wav ».

La figure III.16, montre la similitude entre le signal initial et le signal synthétisé.

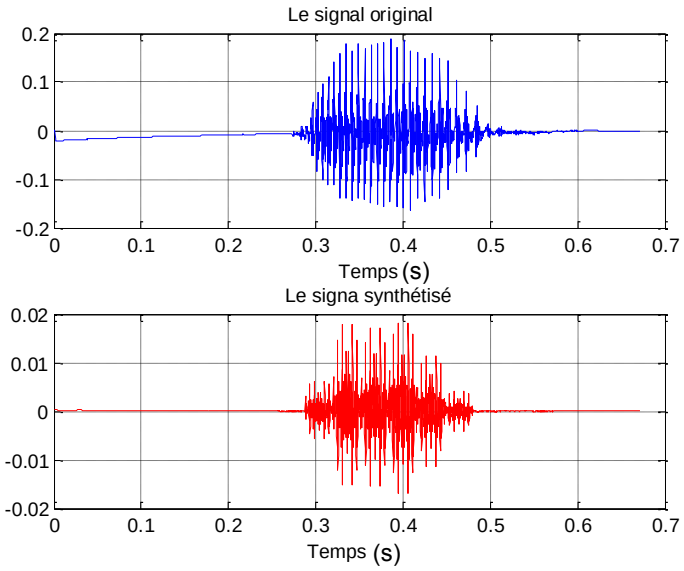


Figure III.16 : *Comparaison entre le signal initial et le signal reconstitué.*

Nous remarquons que le signal synthétisé a une amplitude inférieure à celui du signal original. En utilisant la commande Matlab « soundsc », nous pouvons entendre la voyelle « A » reconstitué avec une clarté suffisante.

Les mêmes étapes de reconstruction du signal à partir des coefficients MFCC, sont appliquées sur plusieurs mots et phrases, les figures suivantes montrent les tracés temporels des signaux originaux et les signaux reconstitués de deux essais.

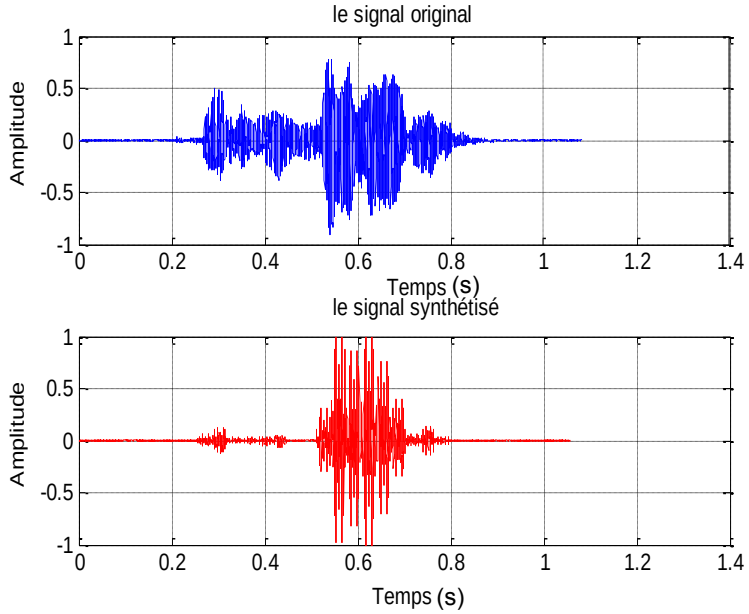


Figure III.17 : Synthèse d'un signal représentant la 1er phrase

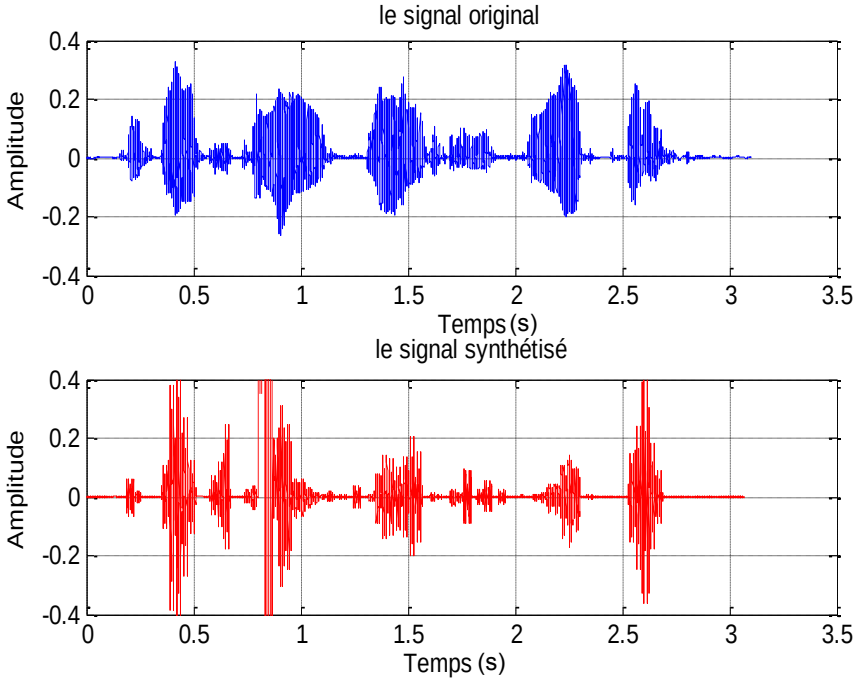


Figure III.18 : *Synthèse d'un signal représentant la 2ème phrase.*

Après plusieurs essais, nous remarquons que la reconstruction de certaines lettres donne des résultats médiocres, notamment pour les lettres « z, r, g, j, ch, v ».

6. CONCLUSION :

Dans ce chapitre nous avons démontré la méthode de reconstruction d'un signal à partir de ces coefficients MFCC. Le but de ce travail a été de construire les signaux de stimulation adéquats pour les implants cochléaires, donc le signal de synthèse correspond à une IFFT du signal résultant de la modulation du spectre retrouvé à partir des coefficients MFCC du signal original par un signal carré bi-phasique.

Ce résultat montre la possibilité d'utiliser les coefficients MFCC pour générer les signaux de stimulation allants aux électrodes placés sur la cochlée pathologique du patient.

CONCLUSION GENERALE

Conclusion Générale

Le but de notre travail c'était l'étude et la simulation d'une stratégie de codage de la parole pour les implants cochléaires, cette stratégie consiste à l'extraction des coefficients MFCC en se basant sur un banc de filtres qui approxime la structure des bandes critiques de la cochlée réelle.

Les résultats de simulations obtenus pour la stratégie de codage réalisée, ont montré la possibilité d'utiliser les coefficients MFCC pour générer les signaux de stimulation allants aux électrodes placés sur la cochlée pathologique du patient. Cependant, la reconstruction de certaines lettres donne des résultats médiocres, notamment pour les sons voisés. Ces derniers sont périodiques ou semi-périodiques, et résultent d'une vibration périodique des cordes vocales qui génère des impulsions de pression appliquées au conduit vocal.

Comme perspectives, nous proposons d'utiliser les coefficients delta-MFCC et delta-delta-MFCC pour remédier au problème des sons voisés, et d'appliquer un classificateur neuronal pour classifier les coefficients MFCC, afin de transférer les signaux de stimulation adéquats aux ses électrodes appropriés placé sur la cochlée pour stimuler le nerf auditif.

Bibliographie

[1] P. Vijayalakshmi, T. Nagarajan, Preethi Mahadevan, "*Improving Speech Intelligibility in Cochlear Implants using Acoustic Models, Department of Electronics and Communication Engineering*", Department of Information Technology, SSN College of Engineering, Chennai, Tamil Nadu, India .

[2] J. Yao and Y. T. Zhang, "*Bionic wavelet transform: a new time-frequency method*".

[3] Par Jeanne GUIRAUD, "*Effets de la surdité totale et de la réhabilitation auditive par l'implant cochléaire sur l'organisation fonctionnelle du système auditif* ", Thèse de doctorat de Sciences cognitives, Université Lumière – LYON 2, Sciences cognitives, Institut de Psychologie, Neurosciences et systèmes sensoriels, Présentée et soutenue publiquement le 1er Février 2007.

[4] C-H. Chouard, "*Histoire de l'implant cochléaire*", Annales françaises d'oto-rhino-laryngologie et de pathologie cervico-faciale 2010.

[5] L'ambition d'entendre, "*Bienvenue dans le monde sonore*", CISIC (Centre d'Information sur la Surdité et l'Implant Cochléaire).

[6] Nabil BELHAJ, "*Etude de faisabilité d'un analyseur de parole miniature pour une prothèse cochléaire*", Mémoire de maitrise ès sciences appliquées spécialité : génie électrique, Université de SHERBROOKE, Sherbrooke (Québec), Canada, Avril 2003.

[7] Comprendre les Implants Cochléaires, medel.com.

[8] PHILIAS Dominique, "*C'EST QUOI LE SON ? "*, Enseignant Cinéma-Audiovisuel au Lycée de Bellevue 97200 Fort de France. CLEMI académie de Martinique,

[9] Advanced Bionics, "*Qu'est-ce qu'un implant cochléaire*", SARL all rightsreserved. Printed in France, 2007.

[10] F. Artières, G. Geffriaud, N. Gresillon, M. Mondain, J-P. Piron, S. Romdhane, M. Sicard, M. Sillon, V. Tesseydre, A. Uziel, V. Ventosa, A. Vieu, "*Les implants cochleaires*", Document réalisé par les membres du GEORRIC, Edité dans la monographie numéro 33 du CCA Groupe, edition en 2002.

[11] Mr. DOUIB OUALID, *"Reconnaissance automatique de la parole arabe par cmu sphinx 4"*, MEMOIRE Présenté à la Faculté de Technologie Département d'ÉLECTRONIQUE Pour obtention du Diplôme de MAGISTER Option : Communication, Université ferhatabbas –sétif1– UFAS (Algérie), 18/12/2013.

[12] T. Dutoit*,* *"Introduction au Traitement Automatique de la Parole"*, Notes de cours/ DEC2, Faculté Polytechnique de Mons ,2000.

[13] Kais GADHOUMI, *"Mise en œuvre d'un algorithme de traitement de la parole basé sur la quantification vectorielle pour une prothèse cochléaire"*, Mémoire de maitrise ès sciences appliquées spécialité : génie électrique, Université de SHERBROOKE, Sherbrooke (Québec) Canada, Avril 2001.

[14] Réjean Fontaine, *"Conception et Développement d'un analyseur de parole multi algorithmes pour implant cochléaire"*, Thèse de doctorat ès sciences appliquées spécialité : génie électrique, Université de SHERBROOKE, Sherbrooke (Québec) Canada, Novembre 1999.

[15] Prothèse auditive à ancrage-osseux. [en ligne]. (Modifié le 6 février 2008) Disponible sur : **http://www.hear-it.org/fr/prothese-auditive-a-ancrage-osseux.**

[**16] P. Vijayalakshmi, T. Nagarajan, Preethi Mahadevan**. *Improving Speech Intelligibility in Cochlear Implants using Acoustic Models.* Department of Electronics and Communication Engineering et Department of Information Technology SSN College of Engineering, Chennai, Tamil Nadu, India. vijayalakshmip@ssn.edu.in, nagarajant@ssn.edu.in