

UNIVERSITÉ DE SHERBROOKE  
Faculté de génie  
Département de génie électrique et de génie informatique

**MISE EN ŒUVRE D'UN ALGORITHME DE TRAITEMENT  
DE LA PAROLE BASÉ SUR LA QUANTIFICATION VECTORIELLE  
POUR UNE PROTHÈSE COCHLÉAIRE**

Mémoire de maîtrise ès sciences appliquées  
Spécialité : génie électrique

---

Kais GADHOUMI

Sherbrooke (Québec), Canada

Avril 2001

À la mémoire de mon père

À ma mère

À mes sœurs

# SOMMAIRE

Ce travail porte sur la mise en œuvre d'un algorithme de codage de parole dédié aux prothèses cochléaires. La technique proposée repose sur la quantification vectorielle statistique.

L'introduction qui suit survole l'état de l'art des prothèses cochléaires.

Dans le deuxième chapitre, on décrit d'abord, d'une façon sommaire, le système auditif et son fonctionnement. Ensuite, on fait un survol des algorithmes de stimulation et des systèmes de prothèses cochléaires qu'on retrouve actuellement.

Le troisième chapitre traite de la méthode d'analyse et de codage de la parole basée sur la quantification vectorielle. Dans une première partie, on présente les caractéristiques de la parole et les modèles de son analyse. Dans la deuxième partie, on décrit la théorie de la quantification vectorielle et les différentes étapes qui définiront la technique proposée.

Le quatrième chapitre décrit les étapes de la mise en œuvre de l'algorithme de traitement de parole. Dans un premier volet, on décrit sommairement l'analyseur de parole de Sherbrooke [15] qui servira de plate-forme technologique pour la mise en œuvre de l'algorithme. Par la suite, on présente les étapes d'implémentation logicielle de l'algorithme et la validation de son fonctionnement.

Dans le chapitre 5 quelques stratégies de stimulation qui peuvent être utilisées sont proposées. Celles-ci pourraient servir avec la technique de codage basée sur la quantification vectorielle pour compléter les algorithmes de stimulation et mettre à profit l'originalité de ce travail.

## **REMERCIEMENTS**

Je tiens tout d'abord à exprimer ma gratitude et ma profonde reconnaissance à mon directeur de recherche, Monsieur Jaouhar Mouine pour l'aide et conseils précieux qu'il n'a cessé de me prodiguer tout au long de ce travail. Qu'il trouve ici l'expression de mes sincères respects et mes vifs remerciements.

Je remercie également mes collègues du GRAMS pour leur support technique, leurs conseils judicieux et leurs commentaires constructifs qui m'ont été d'un grand secours.

Mes remerciements vont aussi à tous mes collègues et amis, et à toutes les personnes qui ont contribué de près ou de loin, à l'accomplissement de ce travail.

Enfin, qu'il me soit permis d'adresser mes infinis remerciements au Gouvernement Tunisien représenté par la mission universitaire de Tunisie qui a contribué en grande partie au financement de mes études de maîtrise.

# TABLE DES MATIÈRES

1.	<b>INTRODUCTION.....</b>	<b>1</b>
2.	<b>LA PHYSIOLOGIE DE L'OREILLE ET LES STIMULATIONS ÉLECTRIQUES DU SYSTÈME AUDITIF.....</b>	<b>4</b>
2.1	La physiologie du système auditif.....	5
2.1.1	L'oreille externe.....	5
2.1.2	L'oreille moyenne.....	6
2.1.3	L'oreille interne.....	6
2.1.4	Les voies nerveuses auditives.....	10
2.2	Principe des prothèses cochléaires.....	11
2.3	Les stratégies de stimulation.....	15
2.3.1	Stratégie $F_0F_2$ .....	15
2.3.2	Stratégie $F_0F_1F_2$ .....	16
2.3.3	Stratégie multi-pics.....	17
2.3.4	Stratégie d'extraction des maxima spectraux.....	18
2.3.5	Stratégie analogique compressée.....	19
2.3.6	Stratégie d'échantillonnage entrelacé continu.....	21
2.3.7	Stratégie des impulsions entrelacées.....	22
2.4	Le support matériel des stratégies de stimulation.....	23
2.4.1	Prothèse Ineraid.....	23
2.4.2	Prothèse Nucleus.....	24
2.4.3	Prothèse Clarion.....	25
2.4.4	Prothèse Combi 40.....	25
2.4.5	Prothèse Laura.....	26
2.4.6	Prothèse Digisonic.....	26
2.5	Conclusion.....	28
3	<b>ALGORITHME DE STIMULATION BASÉ SUR LA QUANTIFICATION VECTORIELLE.....</b>	<b>29</b>
3.1	Les caractéristiques de la parole.....	30
3.2	Les modèles d'analyse de la parole.....	33
3.2.1	Analyse par transformée de Fourier.....	33
3.2.2	Analyse par prédiction linéaire.....	34
3.3	Codage de la parole par quantification vectorielle statistique.....	38
3.3.1	Représentation de la parole.....	38
3.3.2	Quantification vectorielle.....	39
3.3.3	Construction du dictionnaire.....	41
3.3.4	Codage LPC par quantification vectorielle statistique.....	42
3.4	Conclusion.....	47

<b>4</b>	<b>MISE EN ŒUVRE DE L'ALGORITHME DE STIMULATION BASÉ SUR LA QUANTIFICATION VECTORIALE.....</b>	<b>48</b>
4.1	Limitation des systèmes actuels.....	49
4.2	La prothèse cochléaire de Sherbrooke.....	50
	4.2.1 Le stimulateur.....	50
	4.2.2 L'analyseur de parole.....	51
4.3	Mise en œuvre de l'algorithme de quantification vectorielle.....	53
	4.3.1 Traitement de la parole par quantification vectorielle.....	54
	4.3.2 Pré-accentuation du signal.....	55
	4.3.3 Estimation de la fréquence fondamentale.....	56
	4.3.4 Identification des paramètres du modèle.....	57
	4.3.5 Quantification.....	59
4.4	Validation et discussion de l'algorithme.....	60
4.5	Conclusion.....	61
<b>5</b>	<b>STRATÉGIES DE STIMULATION PROPOSÉES.....</b>	<b>63</b>
5.1	Fonctionnement du neurostomuateur de Sherbrooke.....	64
	5.1.1 Principe.....	64
	5.1.2 Dictionnaire des séquences d'impulsions.....	66
5.2	Survol des stratégies de stimulation utilisées par les systèmes actuels.....	67
5.3	Stratégies de stimulation proposées.....	69
	5.3.1 L'approche d'extraction des caractéristiques de la parole.....	70
	5.3.2 L'approche du traitement large bande.....	70
	5.3.3 L'approche mixte.....	71
5.4	Conclusion.....	72
	<b>CONCLUSION.....</b>	<b>73</b>
	<b>BIBLIOGRAPHIE.....</b>	<b>76</b>

# LISTE DES FIGURES

FIGURE 2.1:	STRUCTURES DE L'OREILLE HUMAINE [42].....	6
FIGURE 2.2:	(a)- COUPE TRANSVERSALE DE LA COCHLÉE (b)- DÉTAIL D'UNE SPIRE DE LA COCHLÉE [42].....	8
FIGURE 2.3:	COUPE TRANSVERSALE DE L'ORGANE DE CORTI [42].....	9
FIGURE 2.4:	TRANSMISSION DES ONDES DE PRESSION DANS LA COCHLÉE [42].....	10
FIGURE 2.5:	TRAJET DES VOIES NERVEUSES AUDITIVES [4].....	12
FIGURE 2.6:	SCHÉMA BLOC D'UNE PROTHÈSE COCHLÉAIRE.....	14
FIGURE 2.7:	SCHÉMA BLOC DE LA STRATÉGIE $F_0F_1F_2$ .....	16
FIGURE 2.8:	SCHÉMA BLOC DE LA STRATÉGIE MULTI-PICS.....	17
FIGURE 2.9:	SCHÉMA BLOC DE LA STRATÉGIE SMSP.....	19
FIGURE 2.10:	SCHÉMA BLOC DE LA STRATÉGIE CA POUR UN IMPLANT À QUATRE ÉLECTRODES.....	20
FIGURE 2.11:	SCHÉMA BLOC DE LA STRATÉGIE CIS DANS LE CAS D'UN IMPLANT À QUATRE ÉLECTRODES.....	22
FIGURE 2.12:	SCHÉMA BLOC DE LA STRATÉGIE IP.....	23
FIGURE 3.1:	L'APPAREIL PHONATOIRE [6].....	30
FIGURE 3.2:	PHONÈMES DE LA LANGUE FRANÇAISE [6].....	31
FIGURE 3.3:	SPECTRES D'UN SON NON-VOISÉ ET D'UN SON VOISÉ [6].....	32
FIGURE 3.4:	MODÈLE LINÉAIRE DE PRODUCTION DE LA PAROLE.....	35
FIGURE 3.5:	L'ERREUR SPECTRALE D'UN SEGMENT DE PAROLE VOISÉ ESTIMÉ PAR UN MODÈLE PRÉDICTIF.....	37
FIGURE 3.6:	PROCESSUS DE TRAITEMENT DE LA PAROLE BASÉ SUR LA QUANTIFICATION VECTORIELLE.....	45
FIGURE 4.1:	SCHÉMA BLOC DE L'ANALYSEUR DE PAROLE DE SHERBROOKE.....	52
FIGURE 4.2:	DIAGRAMME D'IMPLÉMENTATION DU PROCESSUS DE TRAITEMENT.....	54
FIGURE 4.3:	SPECTRES D'UN SEGMENT DE PAROLE VOISÉ SANS ET AVEC PRÉ- ACCENTUATION.....	55
FIGURE 5.1:	FORMAT DU MOT DE COMMANDE UTILISÉ PAR L'IMPLANT À 8 CANAUX....	65

FIGURE 5.2:    PROCESSUS DE GÉNÉRATION DES IMPULSIONS À L'AIDE D'UN  
                          DICTIONNAIRE DE SÉQUENCES DE STIMULATION..... 66

# 1. INTRODUCTION

La surdité est une des déficiences neurophysiologiques les plus courantes qui peut atteindre la population de n'importe quelle catégorie d'âge. Les progrès importants qui ont été accomplis durant les trois dernières décennies en matière de techniques chirurgicales et de développement électronique ont permis de venir en aide à plusieurs patients souffrant de cette déficience. Bien que les moyens proposés étaient assez efficaces pour la restitution des sensations sonores naturelles dans le cas de la surdité légère, cela n'était pas le cas pour la surdité profonde. Cette dernière est en effet une forme plus complexe de surdité qui origine au niveau de l'oreille interne. Le but recherché derrière les premières tentatives aussi bien que les systèmes les plus récents est de recréer chez les sourds profonds une sensation d'audition se rapprochant le mieux possible des sensations naturelles.

Les systèmes qui ont été développés pour la surdité profonde ou totale, connus sous le nom de prothèses cochléaires, reposent sur l'approche de stimulation électrique du nerf auditif. Des impulsions électriques sur le nerf auditif engendrent une activité nerveuse qui se transmet jusqu'aux parties concernées du cerveau pour être interprétée, et ainsi créer une sensation d'audition chez le patient.

Il a été bien établi que les prothèses cochléaires constituent le seul moyen de restauration de l'audition chez les sourds profonds. Les performances relatives à cette restauration,

obtenues jusqu'à nos jours, demeurent plus ou moins acceptables et dépendent très subjectivement des patients. Les travaux entrepris au cours des dernières décennies ont montré la nécessité d'impliquer de plus en plus de disciplines pour faire avancer la recherche dans le domaine des prothèses cochléaires. Chacune de ces disciplines a son rôle à jouer dans l'amélioration des performances de ces appareils. Ainsi, ingénieurs, linguistes, psychologues, psychoacousticiens, physiologistes, chirurgiens et beaucoup d'autres intervenants contribuent ensemble à l'élaboration de prothèses plus sophistiquées conduisant à de meilleurs résultats.

L'un des aspects les plus étudiés dans la recherche et développement de nouvelles prothèses cochléaires est la stimulation de la cochlée ou la façon de coder les sons dans l'oreille interne afin d'approcher le traitement naturel qui s'y effectue. Les méthodes ou stratégies de stimulation développées sont inspirées des diverses approches de modélisation du mécanisme d'audition. Ces approches sont simplifiées pour pouvoir les mettre en œuvre et restent encore loin du processus naturel. Ainsi, beaucoup de travaux doivent être effectués par les chercheurs en se basant sur des expériences empiriques et des perceptions subjectives dans le but de développer des stratégies de stimulation assurant des meilleures performances.

À cet égard, on retrouve aujourd'hui des propositions de plusieurs algorithmes de stimulation. Plusieurs d'entre eux sont déjà opérationnels sur des prothèses cochléaires existant sur le marché. Les résultats publiés de l'évaluation de ces stratégies sur certains patients, reflètent dans plusieurs cas des tendances publicitaires quant aux performances du système commercialisé. Dans la majeure partie des publications, il s'agit de reprises

d'anciens travaux avec proposition de nouvelles variantes plutôt que de nouvelles approches.

La limitation majeure des systèmes actuels est qu'ils ont été conçus d'une façon trop spécifique. Ainsi, leur conception matérielle est très dépendante de l'algorithme de stimulation qu'ils préconisent. C'est d'ailleurs la raison pour laquelle ils ne peuvent prétendre offrir un nouvel algorithme sur une nouvelle version de l'appareil utilisant le même concept matériel. Dès lors, une nouvelle approche exigera de reprendre la conception de ces prothèses. En fait, le patient se trouve obligé à s'adapter à l'appareil plutôt que de profiter d'un système qui s'adapte à son cas. La solution semble alors de disposer d'un système versatile, très flexible et complémentent programmable pour s'adapter aux différentes pathologies et offrir plus qu'un algorithme à la fois.

Ce travail constitue une contribution au développement d'un système de ce type. Il concerne l'implémentation d'un nouvel algorithme de stimulation dédié à la prothèse cochléaire de Sherbrooke. S'inspirant de techniques récentes développées en matière du traitement de la parole, un nouvel algorithme de stimulation basé sur la quantification vectorielle fera partie de l'ensemble des algorithmes proposés pour ce système. L'approche conduite vise l'amélioration de certains aspects des performances des prothèses cochléaires relatives à la réhabilitation du patient et à la qualité de reconnaissance de la parole.

## **2. LA PHYSIOLOGIE DE L'OREILLE ET LES STIMULATIONS ELECTRIQUES DU SYSTEME AUDITIF**

Le système nerveux est le centre de régulation et le réseau de communication du corps humain. Il remplit trois fonctions de base. La première fonction est sensorielle. Elle consiste à détecter tous les événements qui se produisent à l'intérieur ou à l'extérieur du corps et qui affectent le comportement de l'être humain. La deuxième fonction s'occupe d'intégrer ou d'interpréter ces événements alors que la troisième est une fonction motrice qui permet de réagir selon l'interprétation en déclenchant une action sous forme de contractions musculaires ou de sécrétions glandulaires. Les récepteurs sensoriels du corps traduisent l'information captée en impulsions électriques ou influx nerveux qui empruntent les voies nerveuses jusqu'au système nerveux central qui s'occupe de leur traitement.

Parmi les organes sensoriels, l'oreille est celui qui s'occupe de l'ouïe. Cet organe permet de collecter les sons environnant, de localiser leurs sources et d'en distinguer les nuances. De nos jours, plusieurs personnes souffrent de troubles d'audition ou de surdité à différents degrés. Ceci a un impact direct sur leur intégration sociale. En effet, les personnes atteintes de surdité se retrouvent isolées et se renferment souvent sur elles-mêmes à cause des difficultés de communication qu'elles éprouvent.

On distingue deux types de surdité : la surdité de conduction ou surdité légère, due principalement à des problèmes au niveau de l'oreille externe ou de l'oreille moyenne et

la surdité de perception ou profonde, due à une atteinte des structures de l'oreille interne. Les problèmes de la surdité de conduction peuvent généralement être résolus par l'utilisation des amplificateurs de sons ou par des chirurgies mineures, alors que la surdité profonde ou totale nécessite des moyens et des chirurgies beaucoup plus complexes. Parmi ces moyens on retrouve l'utilisation des prothèses cochléaires. Ces appareils servent à stimuler directement l'oreille interne par des impulsions électriques, conformément au son capté, afin de reproduire une sensation auditive chez le patient. Ce chapitre présente un survol de la physiologie du système auditif suivi d'une description des prothèses cochléaires. Ensuite, il sera question des différentes approches utilisées par les systèmes disponibles aujourd'hui, pour stimuler électriquement l'oreille interne.

## **2.1 La physiologie du système auditif**

Le système auditif se divise anatomiquement en quatre parties : L'oreille externe, l'oreille moyenne, l'oreille interne et les voies nerveuses. La figure 2.1 illustre les structures de l'oreille. Chaque partie de l'oreille joue un rôle spécifique dans la transformation du signal acoustique en impulsions électriques interprétables par le cerveau.

### **2.1.1 L'oreille externe**

L'oreille externe est composée du pavillon, du conduit auditif externe et de la membrane tympanique. Le pavillon s'occupe de collecter les sons pour les acheminer vers le tympan à travers le conduit auditif. Les ondes sonores qui se propagent dans l'air provoquent alors des vibrations au tympan.

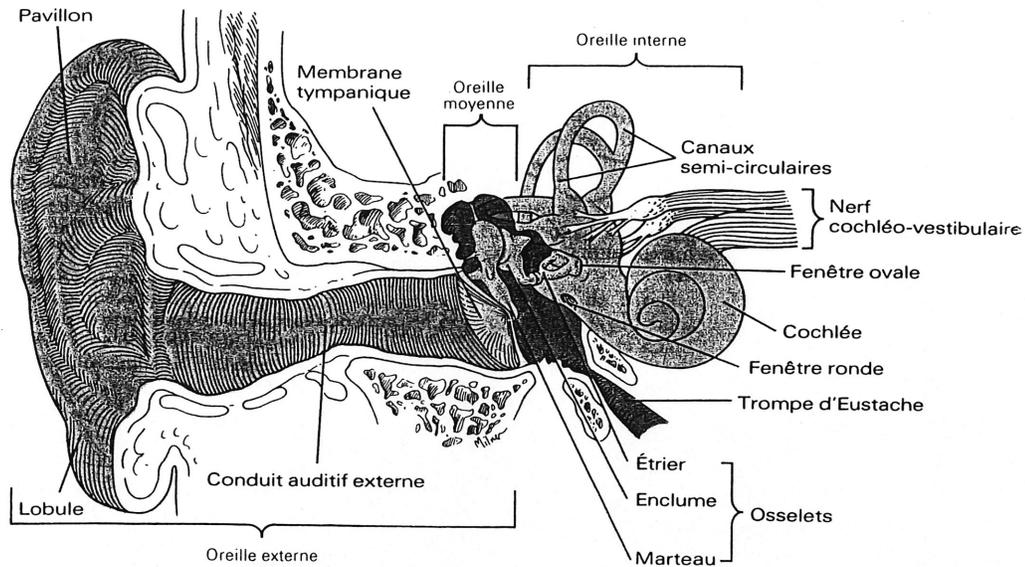


Figure 2.1 Structures de l'oreille humaine [42]

### 2.1.1 L'oreille moyenne

L'oreille moyenne est une cavité qui contient une chaîne d'osselets appelés respectivement le marteau, l'enclume et l'étrier. Le marteau, solidaire au tympan, communique les vibrations de ce dernier au reste de la chaîne qui les achemine à l'oreille interne. L'étrier, le dernier osselet de la chaîne, est fixé sur un orifice recouvert d'une membrane flexible, appelée la fenêtré ovale, qui constitue le point d'entrée de l'oreille interne.

### 2.1.2 L'oreille interne

L'oreille interne se compose d'un système de canaux creusés dans l'os et remplis de liquide. Ces canaux se divisent en trois parties : le vestibule, les canaux semi-circulaires et la cochlée. Les deux premières parties s'occupent de la position de la tête et du

balancement du corps alors que la cochlée constitue la partie responsable de l'ouïe. La cochlée est constituée d'un tube d'environ 2 mm de diamètre et 32 mm de long qui s'enroule sur lui-même en spirale en forme d'une coquille d'escargot (figure 2.2). Deux membranes appelées membrane vestibulaire et membrane basilaire la divisent longitudinalement en trois compartiments. Le compartiment central, appelé la rampe médiane ou canal cochléaire, est délimité par les deux membranes et il est rempli d'un liquide appelé l'endolymphe. Le compartiment formé par la membrane vestibulaire et les parois internes de la cochlée est appelé la rampe vestibulaire. Le dernier compartiment séparé du canal cochléaire par la membrane basilaire est appelé la rampe tympanique. Les rampes vestibulaire et tympanique sont remplies d'un liquide appelé le périlymphe et communiquent à l'apex de la cochlée par un orifice appelé l'hélicotrème. À la base de la cochlée, la rampe vestibulaire commence à la fenêtre ovale et la rampe tympanique finit à la fenêtre ronde.

Du côté du canal cochléaire, la membrane basilaire supporte une structure appelée l'organe de Corti. Comme illustré sur la figure 2.3, cet organe se compose essentiellement de cellules de soutien et de cellules ciliées reliées aux terminaisons nerveuses du nerf auditif. On distingue les cellules ciliées internes qui sont organisées en une rangée d'environ 3500 cellules et les cellules ciliées externes dénombrées à environ 12000 et disposées en trois rangées. Comme leur nom l'indique, les cellules ciliées sont munies de stéréocils se projetant dans une membrane qui les surplombe appelée membrane tectoriale.

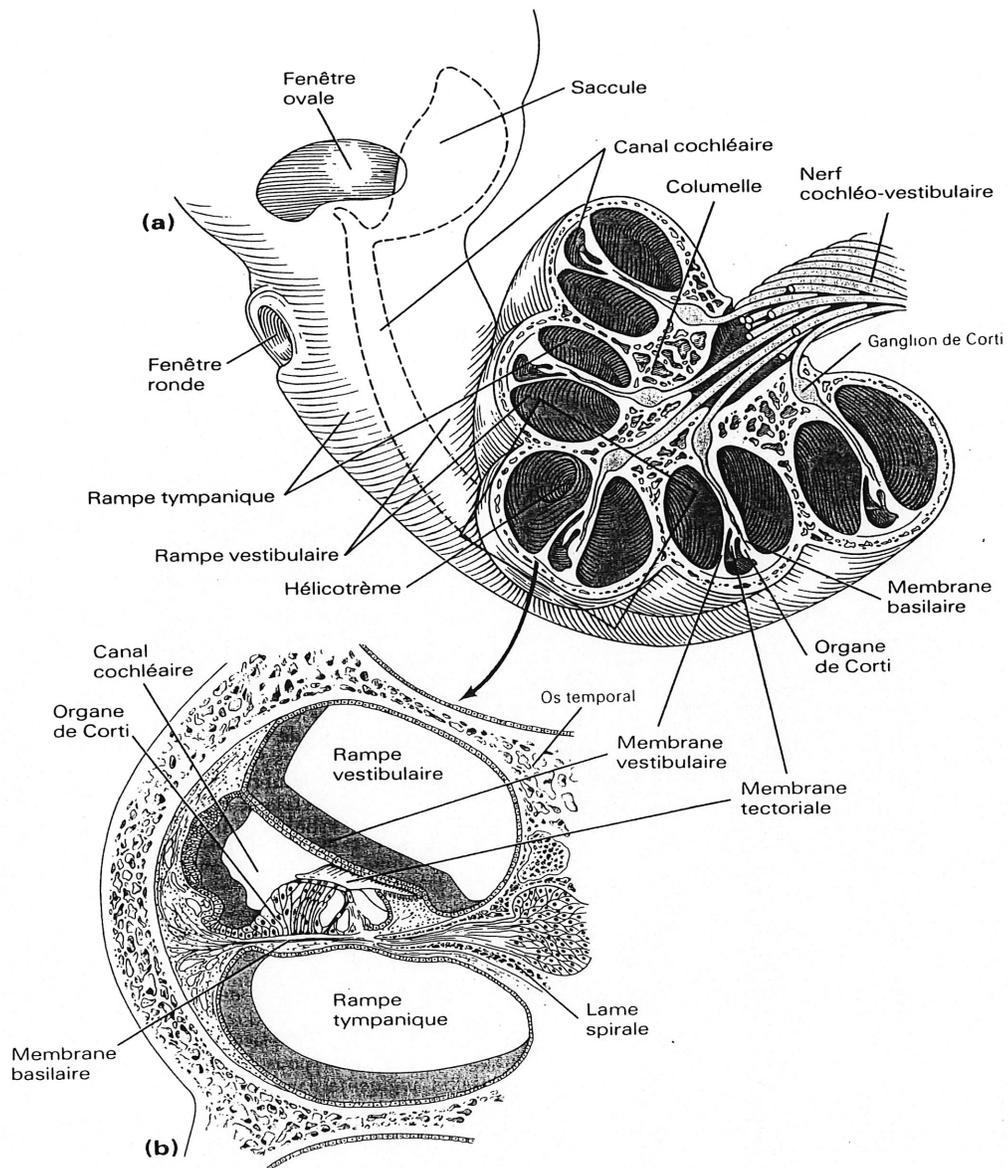


Figure 2.2 (a)- Coupe transversale de la cochlée (b)- Détail d'une spire de la cochlée [42]

Suite à la réception d'un son, les vibrations mécaniques transmises à la fenêtre ovale par les structures de l'oreille moyenne se transforment en ondes de pression hydraulique dans le liquide de la rampe vestibulaire et la rampe tympanique (figure 2.4). Pour les sons de basses fréquences, ces ondes voyagent le long de la rampe vestibulaire et communiquent à la rampe tympanique via l'hélicotrème, imprimant ainsi les mouvements mécaniques

sur la membrane de la fenêtre ronde. Dans le cas des hautes fréquences, les mouvements se produisent seulement au niveau de la base de la cochlée et se transmettent à la rampe tympanique en exerçant une pression directement sur le liquide du canal cochléaire.

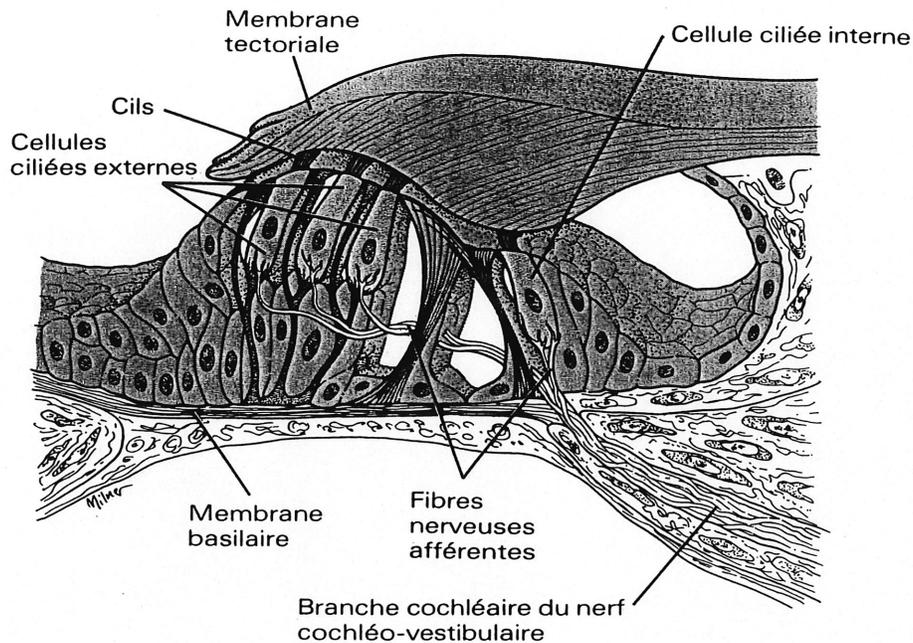


Figure 2.3 Coupe transversale de l'organe de Corti [42]

Les ondes de pression hydrauliques générées dans les rampes vestibulaires et tympanique provoquent des mouvements dans le liquide du canal cochléaire produisant ainsi des ondulations à la membrane basilaire. Ces mouvements combinés produisent un effet de cisaillement entre la membrane basilaire et la membrane tectoriale qui s'exerce sur les stéréocils. Ces derniers excitent alors leurs cellules respectives, ce qui génère un influx nerveux qui voyage le long du nerf auditif.

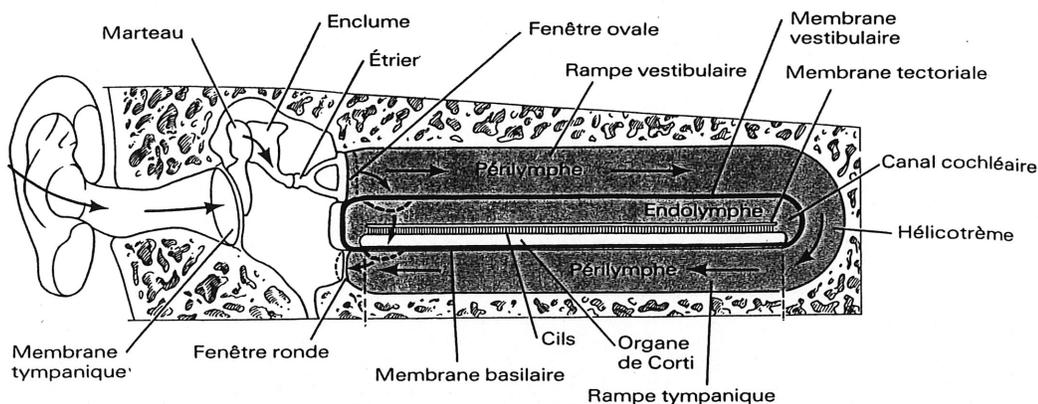


Figure 2.4 Transmission des ondes de pression dans la cochlée [42]

La membrane basilaire présente une géométrie non uniforme. Sa largeur augmente de la base vers l'apex alors que son épaisseur diminue. Cet aspect lui confère une sensibilité relative aux caractéristiques du son capté. En effet, l'amplitude de vibration maximale de la membrane se produit à un endroit spécifique pour chaque fréquence du signal acoustique reçu. Cependant, cette organisation dite tonotopique n'explique pas à elle seule le mécanisme de la perception auditive et la sélectivité fréquentielle de la cochlée. En fait, il a été établi que chaque type de cellules ciliées joue un rôle différent dans la perception des sons. Les cellules ciliées internes reliées à 95% des fibres du nerf auditif constituent les véritables récepteurs sensoriels, alors que les cellules ciliées externes en contact avec 5% des fibres du nerf auditif sont plutôt des transducteurs du son qui amplifient son effet sur la membrane basilaire [4][11].

### 2.1.3 Les voies nerveuses auditives

Les voies nerveuses auditives, illustrées sur la figure 2.5, constituent la dernière partie du système auditif et s'occupent d'acheminer l'influx nerveux généré au niveau de l'oreille interne jusqu'au cerveau. Ces voies renferment plusieurs relais. Le premier relais est constitué du ganglion de Corti (ou ganglion spiral) situé au niveau de la cochlée. Il

contient à peu près 31 000 fibres reliées aux cellules ciliées. Ces fibres se divisent en une partie efférente dite spirale et une partie afférente dite radiale. Les fibres de la partie spirale véhiculent l'innervation des cellules ciliées externes alors que celles de la partie radiale s'occupent de communiquer l'information des cellules ciliées internes. Le deuxième et le troisième relais des voies nerveuses se situent dans le cerveau moyen et le cortex cérébral, lieux d'intégration des actes réflexes et des cellules nerveuses respectivement [4]. Quoique certaines structures de ces dernières parties soient peu comprises, il semble néanmoins qu'il s'agit d'un centre de coordination et d'intégration des sensations auditives avec les autres informations sensibles susceptibles de produire une contre-réaction [32]. C'est également à cet endroit que l'influx nerveux est interprété. Il s'agit définitivement de la partie la plus importante de tout le système auditif qui doit être intacte afin d'espérer la restitution de l'ouïe aux personnes totalement sourdes.

## **2.2 Principe des prothèses cochléaires**

Les prothèses cochléaires demeurent aujourd'hui le moyen le plus répandu pour résoudre, du moins en partie, le problème de la surdité profonde ou totale. Plusieurs systèmes de prothèses cochléaires ont été développés à travers le monde, avec des caractéristiques et des performances différentes, mais tous basés sur le même principe. Ce principe consiste à traduire le son capté en une suite d'impulsions électriques transmises à l'oreille interne via un réseau d'électrodes fixé au niveau de la cochlée. Il existe toutefois certains critères qui sont considérés dans la sélection des patients qui peuvent en bénéficier. Ces critères peuvent être reliés aussi bien à la pathologie du patient qu'à son état physique et psychologique ou encore à sa motivation, son âge, sa période de surdité et autres [32][39].

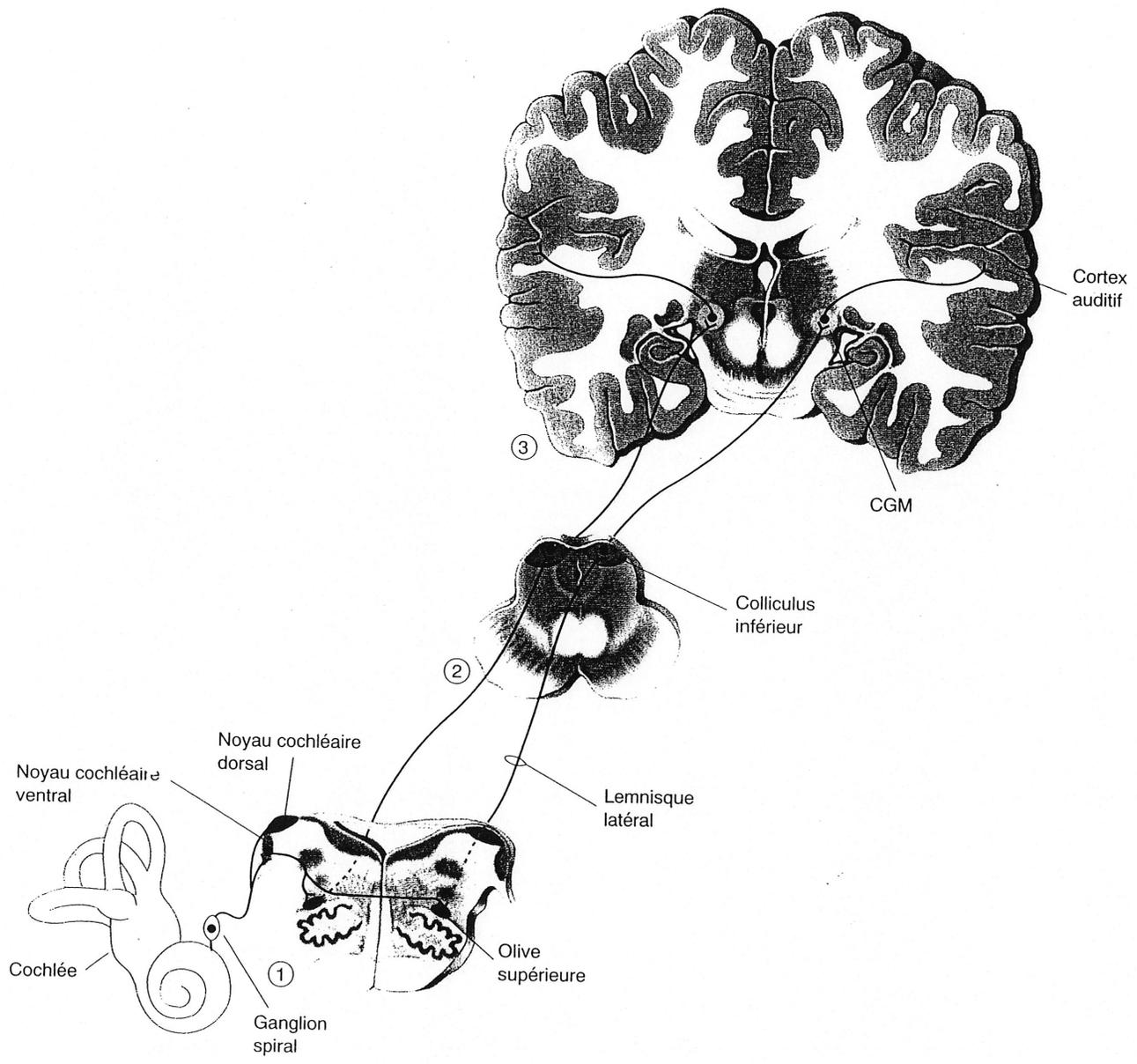


Figure 2.5 Trajet des voies nerveuses auditives [4]

Les constituants de base des différents systèmes de prothèses cochléaires qui existent sont tous les mêmes (figure 2.6). Il s'agit d'une partie externe qui s'occupe de l'acquisition et du traitement de sons et d'une partie implantable munie d'un réseau d'électrodes qui livre les stimuli électriques à la cochlée. Les deux parties sont reliées par un lien de communication qui véhicule l'information délivrée par l'unité externe.

Le nombre d'électrodes varie d'un système à l'autre. Les systèmes monoélectrodes qui se distinguaient par leur simplicité de fabrication et de chirurgie ne peuvent plus se justifier face aux nouvelles technologies et aux techniques de chirurgies modernes, surtout que la nécessité d'un système multiélectrodes a été bien établie pour espérer atteindre un degré raisonnable de compréhension de la parole. En ce qui concerne la disposition des électrodes, on distingue deux alternatives préconisées selon l'état de la cochlée. Généralement, le réseau est introduit dans la rampe tympanique via la fenêtre ronde afin d'atteindre les différents sites de stimulation le long de la membrane basilaire. Cependant, dans le cas où la cochlée serait obstruée ou ossifiée, les différentes électrodes sont placées une à une sur la paroi externe du limaçon.

Par ailleurs, deux modes de stimulations sont préconisés par les différents systèmes. Le premier mode dit monopolaire se caractérise par un pôle commun de retour de charges, placé relativement distant des électrodes. Ce mode se justifie surtout dans le cas d'un nombre petit de fibres nerveuses résiduelles. Le deuxième mode, appelé bipolaire, se caractérise par l'utilisation de deux pôles rapprochés, permettant de générer un stimulus localisé dans une région spécifique de la membrane basilaire.

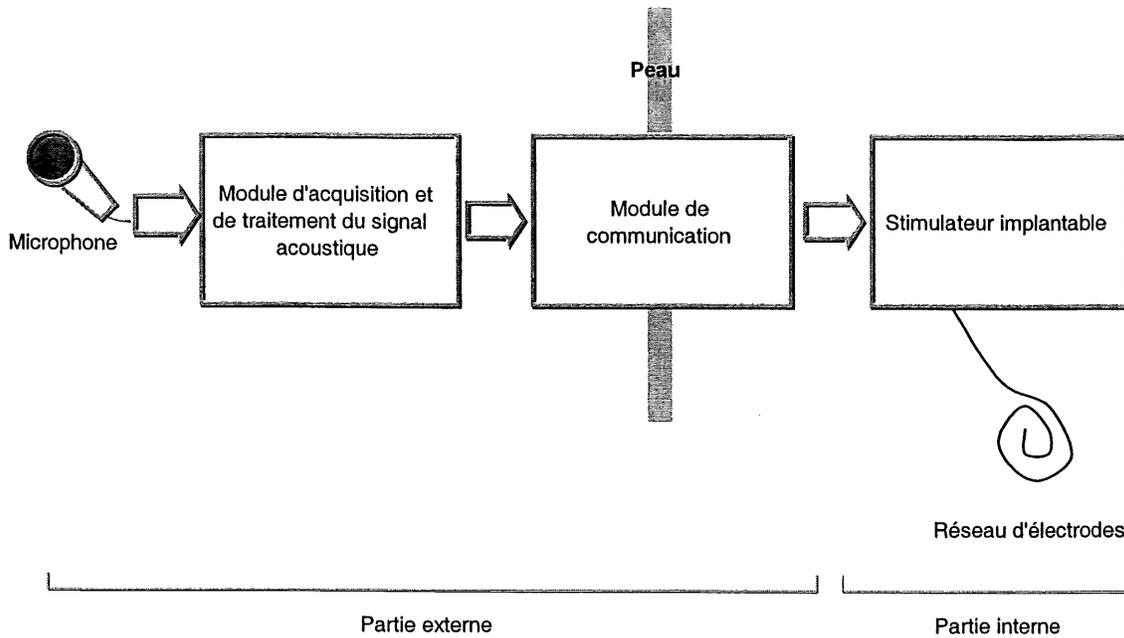


Figure 2.6 Schéma bloc d'une prothèse cochléaire

Du côté du lien de communication entre l'implant et l'unité externe, la majorité des systèmes actuels utilisent un lien inductif transcutané. Celui-ci utilise les ondes électromagnétiques générées par une bobine et un émetteur RF externes accordés à un récepteur muni d'une deuxième bobine insérée avec l'implant sous la peau.

Les derniers facteurs qui distinguent les prothèses cochléaires sont l'algorithme de traitement de signal utilisé et la stratégie de stimulation adoptée pour représenter le son dans l'oreille interne. Ces facteurs demeurent les uns des plus importants autour desquels se concentrent les travaux de recherche actuels. Ces travaux visent principalement à élaborer des stratégies de stimulation basées sur des algorithmes de traitement de parole qui offrent une meilleure représentation de l'information dans l'oreille interne. En fait, la façon exacte de représenter un signal sonore par des impulsions électriques dans la cochlée est encore inconnue. Ainsi, les seuls critères qui

sont pris en considération de nos jours sont d'une part que tous les calculs doivent s'exécuter en un laps de temps non apercevable par le patient et d'autre part que l'algorithme doit toujours être implémentable sur un analyseur de parole portable. Quoique nombreuses, les stratégies actuellement développées sont toutes basées sur deux approches principales. La première repose sur l'extraction des caractéristiques de la parole et consiste à présenter à l'oreille interne un stimuli qui représente les paramètres jugés essentiels à la compréhension de la parole selon la tonotopie de la membrane basilaire. La deuxième approche est axée sur le traitement large bande du signal de parole et consiste à le traiter et à le transformer pour pouvoir le présenter à l'oreille interne directement sous forme brute toujours selon la tonotopie de la membrane basilaire. Un survol des plus importantes stratégies de stimulation est présenté dans ce qui suit.

## **2.3 Les stratégies de stimulation**

### **2.3.1 Stratégie $F_0F_2$**

La stratégie  $F_0F_2$  est la première stratégie développée en se basant sur l'extraction des caractéristiques de la parole [27]. Expérimentée pour la première fois en 1978 par l'équipe de Clark à l'université de Melbourne, elle repose sur l'extraction de la fréquence fondamentale du son  $F_0$  et celle du deuxième formant  $F_2$ . Le choix du deuxième formant est justifié par la concentration d'énergie dans la plage de fréquences entre 800 Hz et 2300 Hz d'où son rôle important dans la compréhension de la parole. La fréquence du deuxième formant est utilisée pour désigner la paire d'électrodes active alors que l'amplitude estimée du deuxième formant servira à fixer le niveau du courant de

stimulation. La fréquence fondamentale détermine le rythme de stimulation. Typiquement, cette fréquence se situe entre 50 et 300 Hz dans le cas des segments voisés de parole. Dans le cas des segments non voisés, les stimulations sont effectuées à des intervalles pseudo-aléatoires à un rythme moyen fixe de l'ordre de 100 pulsations par seconde.

### 2.3.2 Stratégie $F_0F_1F_2$

Il s'agit d'une amélioration de la stratégie  $F_0F_2$  [27][38] où l'on a ajouté l'information sur le formant  $F_1$  correspondant à une autre concentration d'énergie entre 300 Hz et 800 Hz. La figure 2.7 montre le principe de cette stratégie. Le formant  $F_1$  et son amplitude sont utilisés pour activer séquentiellement une deuxième paire d'électrodes.

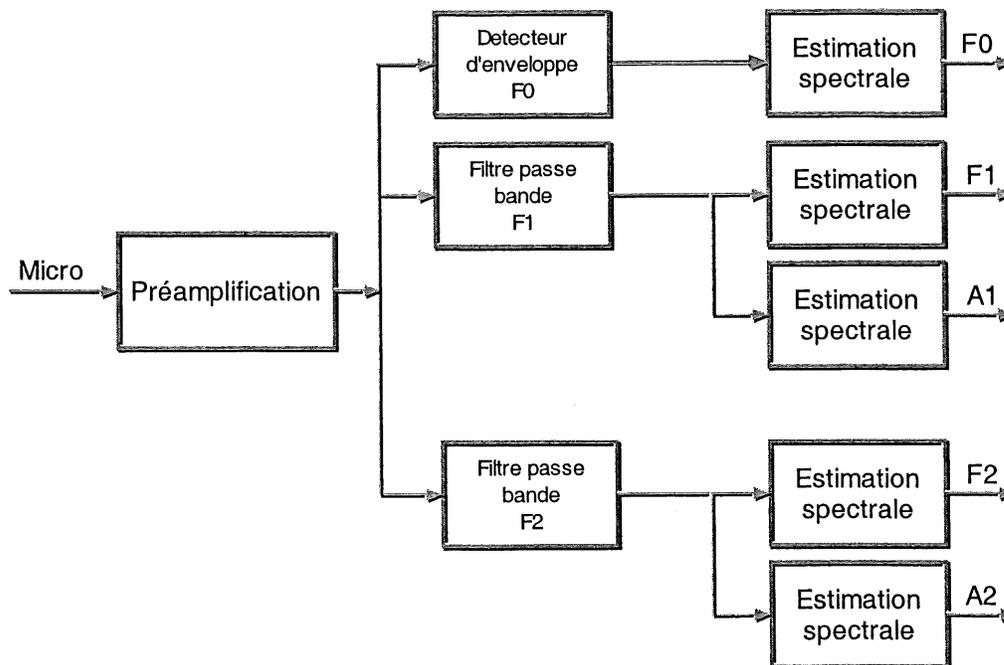


Figure 2.7 Schéma bloc de la stratégie  $F_0F_1F_2$

### 2.3.3 Stratégie multi-pics

La stratégie multi-pics (MPEAK) se présente comme une extension de la stratégie  $F_0F_1F_2$  où, en plus d'extraire les formants  $F_1$  et  $F_2$ , on procède à l'évaluation de 3 bandes d'information spectrales pour déterminer l'énergie du signal dans les plages hautes fréquences de 2 - 2.8 KHz, 2.8 - 4 KHz et 4 - 6 KHz [27][38] (figure 2.8).

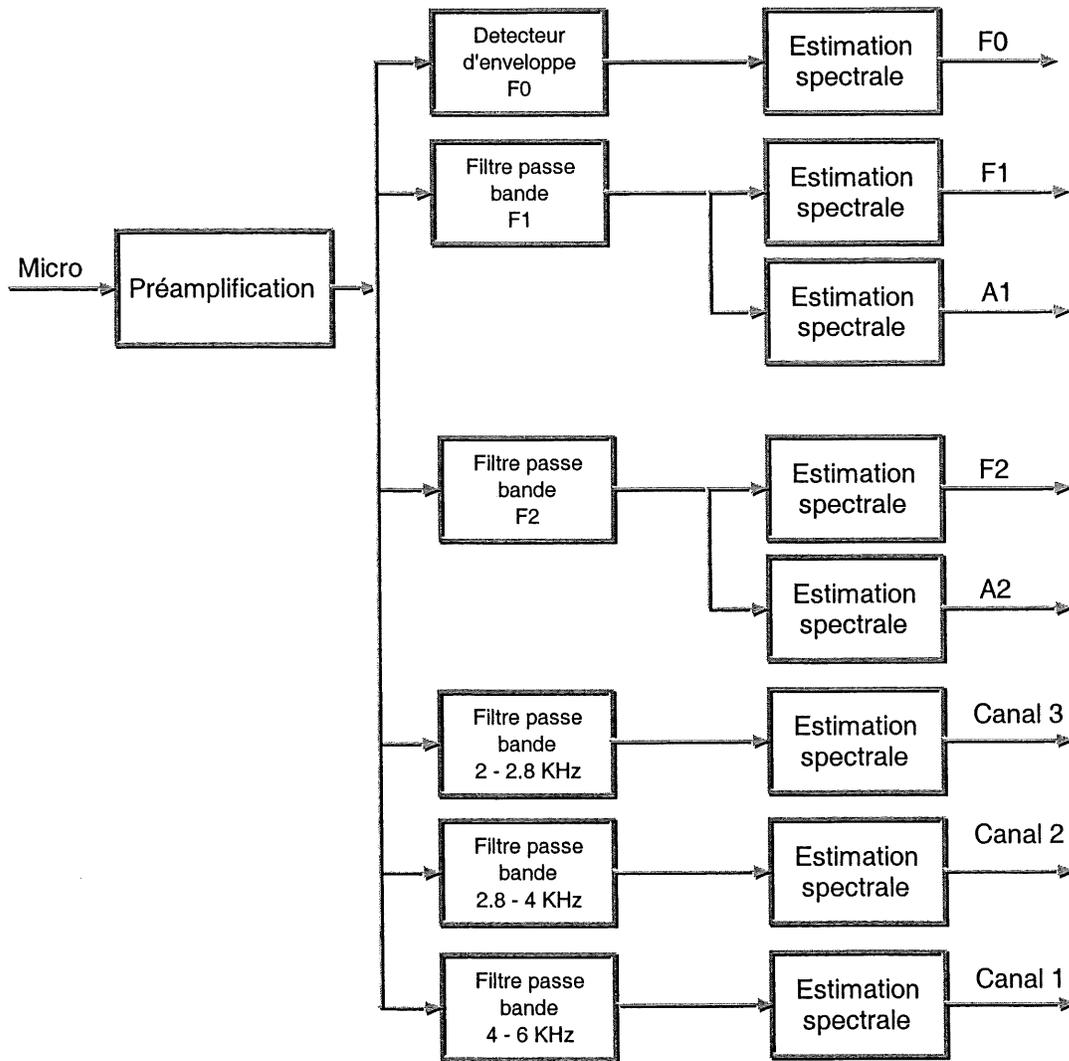


Figure 2.8 Schéma bloc de la stratégie multi-pics

Ainsi 3 stimuli additionnels peuvent être envoyés sur les électrodes positionnées dans la région basale de la cochlée. Le stimuli est alors composé d'une sélection de 4 parmi les 5 composantes fréquentielles extraites. La sélection est basée sur la nature du signal de la parole. Si le son est voisé, les canaux représentant les formants  $F_1$  et  $F_2$  et les canaux hautes fréquences 2 et 3 seront stimulés au rythme du voisement. Le canal 1 n'est pas activé puisqu'il y a généralement peu d'énergie dans le spectre des sons voisés au-delà de 4 KHz. Pour les sons non-voisés les stimulations se font sur le canal déterminé par  $F_2$  et sur les canaux 1, 2 et 3 et ce à des intervalles quasi-aléatoires, à un taux moyen de 250 impulsions par seconde. Le canal correspondant au formant  $F_1$  n'est pas activé puisqu'il y a généralement peu d'énergie en dessous de 1 KHz dans le spectre des sons non-voisés.

#### 2.3.4 Stratégie d'extraction des maxima spectraux

Cette stratégie est désignée par SMSP [30] et a été développée au début des années 1990 pour la prothèse Nucleus de la compagnie Cochlear [34]. Elle repose sur l'utilisation d'un banc de 16 filtres pour évaluer le contenu de tout le spectre du signal de la parole. Après sa pré-amplification, le signal est subdivisé en 16 bandes à l'aide de 16 filtres passe-bande centrés sur des fréquences variant de 250 à 5400 Hz (figure 2.9). La sortie de chaque filtre est alors redressée puis filtrée sur des basses fréquences par des filtres passe-bas de fréquence de coupure de 200 Hz pour déterminer l'amplitude maximale. L'information retenue est déterminée par la sélection des 6 sorties, parmi celles des 16 filtres, ayant les plus hautes amplitudes. Ces amplitudes sont transmises vers les 6 canaux de stimulation correspondants à leurs bandes de fréquence sous forme d'impulsions biphasiques à un taux constant de 250 impulsions par seconde dans le cas d'un son non voisé, et à la fréquence fondamentale si le son est voisé.

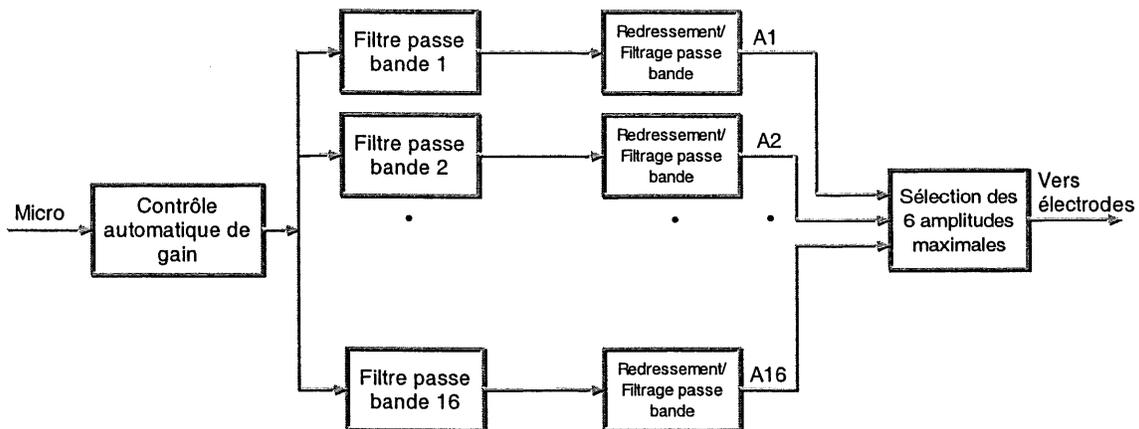


Figure 2.9 Schéma bloc de la stratégie SMSP

Une variante de la stratégie SMSP est la stratégie des pics spectraux, désignée par SPEAK [46]. Elle utilise un banc de 20 filtres centrés sur des fréquences variant de 250 Hz à 10 KHz. Les sorties de filtres les plus énergétiques, pouvant aller de 1 à 10 sont alors sélectionnées pour la stimulation. Le rythme de la stimulation varie entre 180 et 300 Hz dépendamment du nombre d'électrodes activées.

### 2.3.5 Stratégie analogique compressée

Cette stratégie, complètement analogique, a été premièrement développée par Eddington [12] pour la prothèse Ineraid de la compagnie Symbion Inc puis elle a été améliorée par Wilson [46][48]. En première étape, le signal subit une atténuation de 6 dB/octave sur la plage des fréquences en dessous du 1.2 KHz. Cette atténuation a pour effet d'aplatir le spectre du signal ce qui affaiblit les composantes basses fréquences qui peuvent dominer la perception. Le signal est par la suite compressé en utilisant un contrôle automatique de

gain puis filtré à l'aide de 3 filtres passe-bande centrés sur 0.5, 1.1 et 1.8 KHz et un filtre passe-haut de fréquence de coupure 2.2 KHz. Ces bandes sont particulièrement choisies de sorte que le premier formant soit représenté sur la premier canal et le troisième et le reste des formants supérieurs le soient sur le 4<sup>ème</sup> canal. Le deuxième formant est divisé entre les 2<sup>ème</sup> et 3<sup>ème</sup> canaux afin d'améliorer la discrimination de cette composante essentielle de la parole. Un filtre passe haut, utilisé à la sortie du filtre passe-bande du premier canal, procure une égalisation initiale du son pour les fréquences en bas de 300 Hz. Des circuits à gain ajustables sont mis en aval des filtres de sortie de façon à améliorer la perception des composantes correspondantes de la parole. Les signaux issus de chacun des filtres sont alors utilisés pour la stimulation simultanée analogique sur les 4 canaux où l'information sur le signal de parole est transmise dans l'amplitude relative des stimuli et les détails temporels des formes d'onde.

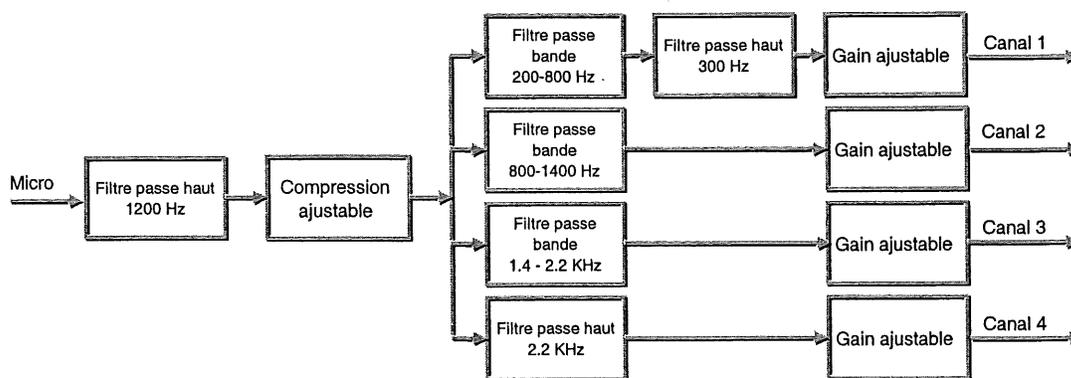


Figure 2.10 Schéma bloc de la stratégie CA

### 2.3.6 Stratégie d'échantillonnage entrelacé continu

Développée au RTI (Research Triangle Institute), la stratégie d'échantillonnage entrelacé continu ou CIS (Continuous Interleaved Sampling) repose sur la stimulation non simultanée en utilisant des impulsions entrelacées [47]. Des trains d'impulsions biphasiques de rapport cyclique égale à  $\frac{1}{2}$  sont délivrés aux électrodes avec un décalage temporel de façon à omettre tout chevauchement entre les canaux. Les amplitudes des impulsions sont déduites des enveloppes des formes d'ondes issues de filtres passe-bande. Contrairement à la stratégie CA, le signal est pré-amplifié et filtré sur 5 ou 6 bandes de fréquence pour tirer profit d'un plus grand nombre d'électrodes implantées et de moins d'interactions entre les paires d'électrodes. Les enveloppes de sortie de ces filtres sont alors extraites par redressement et filtrage passe-bas (fréquence de coupure de 200 ou 400 Hz). Les sorties des enveloppes sont compressées pour servir à moduler les pulsations biphasiques. Cette compression suit une loi non linéaire (de type logarithmique, en général) afin que les sorties des enveloppes correspondent aux plages dynamiques relatives du patient. La stimulation se déroule à un taux constant indépendamment de la nature voisée ou non-voisée du segment sonore. Ce taux dépasse souvent les 800 impulsions par seconde sur chaque canal. La figure 2.11 présente le diagramme de la stratégie CIS dans le cas de N canaux.

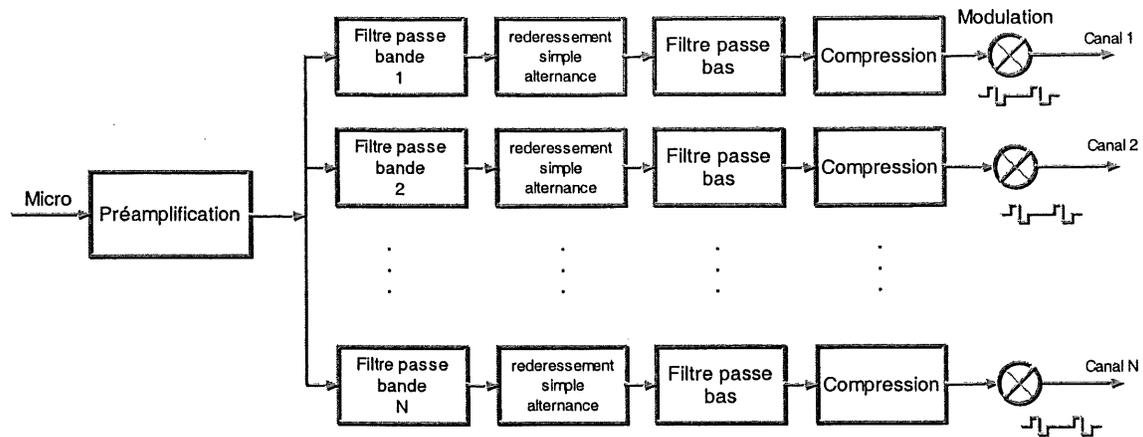


Figure 2.11 Schéma bloc de la stratégie CIS

### 2.3.7 Stratégie des impulsions entrelacées

La stratégie des impulsions entrelacées (Interleaved Pulses, IP) effectue au départ un contrôle automatique de gain sur le signal d'entrée dans le but de l'ajuster à un niveau constant [48]. Le signal subit alors un filtrage passe-haut qui réduit l'amplitude des composantes en dessous de 1.2 KHz pour se subdiviser par la suite en 6 bandes de fréquence en passant par un banc de 6 filtres passe-bande. Les fréquences centrales des filtres sont réparties selon une échelle logarithmique, sur les plages combinées des 3 premiers formant de la parole. La valeur efficace de l'énergie de chacune des bandes est détectée et envoyée à un post-processeur qui choisira périodiquement les  $n$  valeurs les plus hautes. Ce dernier contrôle également le rythme de stimulation en recevant l'information sur la nature du signal d'un détecteur de voisement. Dans le cas des segments voisés, la stimulation se fait au rythme fixe de la fondamentale, et dans le cas des segments non-voisés, les stimuli sont présentés à un rythme plus élevé ou aléatoire.

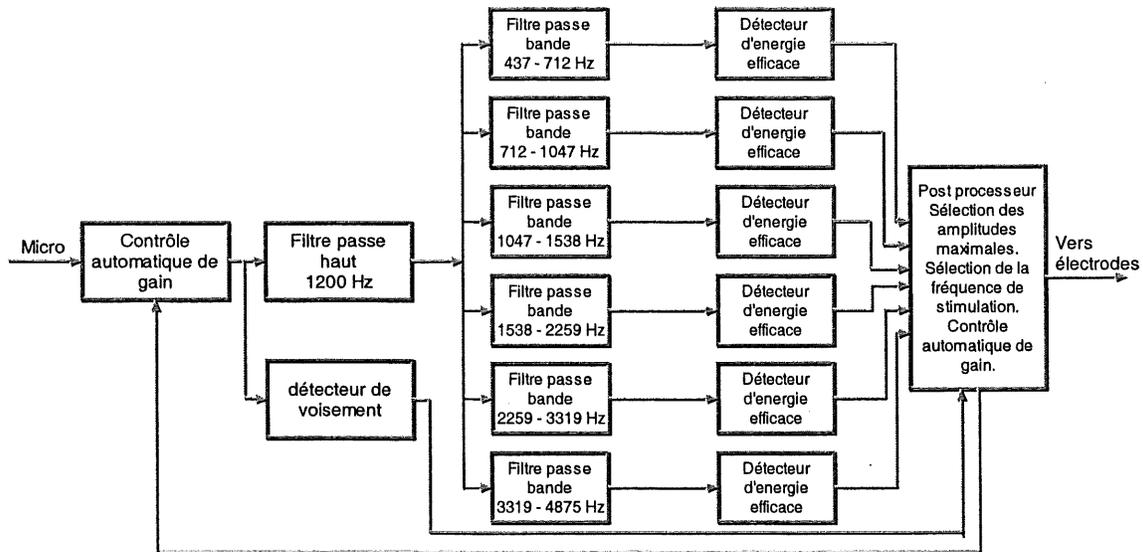


Figure 2.12 Schéma bloc de la stratégie IP

## 2.4 Le support matériel des stratégies de stimulation

Chacune des stratégies de stimulation qu'on vient de décrire, doit être implémentée sur un système portable qui sera livré au patient. Ce système sera garant de la bonne exécution de la stratégie que ce soit du point de vue de l'algorithme de traitement de signal ou du point de vue de la génération du stimuli dans l'oreille interne. Ci-après une description sommaire des différents systèmes qui offrent les différentes stratégies décrites.

### 2.4.1 Prothèse Ineraid

Développée à l'université de l'Utah durant les années 1970, la prothèse Ineraid utilise 6 électrodes intracochléaires activées en mode monopolaire et deux électrodes extracochléaires pour le retour de charge [31]. L'analyseur de parole se base sur une

stratégie CA utilisant 4 parmi les 6 électrodes. Il s'agit du premier système qui a expérimenté cette stratégie.

#### 2.4.2 Prothèse Nucleus

Le système Nucleus est le système le plus commercialisé actuellement. Il a été développé à l'université de Melbourne, et expérimenté pour la première fois en 1978. Depuis 1982, la compagnie Cochlear Pty a acquis ses droits de commercialisation [34].

La première génération de cette prothèse était complètement analogique. Elle se composait d'un stimulateur à une seule source de courant muni de 10 électrodes et d'un analyseur de parole désigné WSP-III (Wearable Speech Processor) exécutant la stratégie  $F_0F_2$  en estimant les fréquences à l'aide des détecteurs de passage par zéro. La deuxième génération de la prothèse utilisait un réseau de 22 électrodes et adoptait la stratégie  $F_0F_1F_2$ , avant que la compagnie ne développe un nouveau analyseur analogique/numérique appelé le MSP (Mini Speech Processor) qui utilise la stratégie MPEAK. Sur ce dernier système, l'estimation des fréquences a été améliorée en remplaçant les détecteurs de passage par zéro par une fonction numérique de détection. La troisième génération des systèmes de cette compagnie est le Spectra 22. Celui-ci est un système complètement digital. Il utilise un nouveau stimulateur muni de 3 sources de courant et qui est capable de générer une stimulation quasi-simultanée sur 3 paires d'électrodes. L'analyseur de parole est conçu autour du microprocesseur DSP56001 de Motorola et exécute la stratégie SMSP.

La dernière génération de la prothèse Nucleus, lancée sur le marché en 1998, est munie de 24 électrodes et utilise la variante SPEAK de la stratégie SMSP.

#### 2.4.3 Prothèse Clarion

Cette prothèse est produite par la compagnie ABC (Advanced Bionics Corporation) [10]. Elle est la deuxième la plus commercialisée après Nucleus. Le dernier système de cette compagnie utilise 16 électrodes et permet des stimulations simultanées, séquentielles ou combinées. Deux stratégies sont utilisées avec ce système: la CIS et la SAS (Simultaneous Analog Stimulation). Cette dernière est similaire à la stratégie CA où l'on filtre le signal acoustique sur 8 bandes de fréquence. Les sorties analogiques des filtres sont alors compressées et transmises simultanément aux 8 paires d'électrodes à un taux pouvant atteindre 13000 impulsions par seconde sur chaque paire. Dans le cas de la stratégie CIS, la stimulation se déroule de façon non simultanée à un rythme de 833 impulsions par seconde sur chaque paire.

#### 2.4.4 Prothèse Combi 40

Le Combi 40 est une prothèse cochléaire à 24 électrodes regroupées en 12 paires, qui est produite par la compagnie Med-El [9]. L'analyseur de parole est conçu autour du processeur, DSP56002 de Motorola. Il comporte un étage analogique qui effectue la préamplification, le contrôle de gain, le filtrage anti-recouvrement et la préaccentuation du signal, suivi d'un étage de conversion analogique/numérique. Le processeur est programmé pour exécuter la stratégie CIS et une de ses variantes appelée "*n of m*". Cette version consiste à ne conserver parmi les  $m$  bandes du signal, que les  $n$  bandes les plus

énergétiques à chaque cycle de stimulation [27]. Les sorties des filtres correspondants sont alors combinées avec la plage dynamique du patient et une fonction non-linéaire de compression pour générer l'amplitude des stimuli. Ces derniers étant délivrés à un rythme maximal de 3030 impulsions par seconde.

#### 2.4.5 Prothèse Laura

Le système Laura a été développé au début des années 1980, à l'université de l'Antwerp en Belgique. Il a été fabriqué et commercialisé par la compagnie ABS (Antwerp Bionic Systems) avant que la compagnie PHI (Philips Hearing Implants) n'acquière le droit de sa fabrication et sa commercialisation [36]. Le dernier produit de la compagnie, appelé Laura Flex, est une prothèse à 8 paires d'électrodes bipolaires. L'analyseur de parole est conçu autour de deux processeurs : un DSP, le ADS2173 de Analog Devices, pour le traitement de la parole et un microcontrôleur dédié à l'interface usager et à la mise au point des paramètres de l'analyseur. Deux stratégies de stimulation sont offertes : la stratégie CIS, et une version modifiée de celle-ci, la PLCIS (Phase Locked CIS). Dans la PLCIS, les impulsions sont envoyées à un rythme fixe de 1250 impulsions par seconde sur chaque paire d'électrodes. Ces dernières sont activées en phase avec le formant contenu dans la bande de fréquence correspondante.

#### 2.4.6 Prothèse Digisonic

Il s'agit de la deuxième version de l'ancien système appelé Chorimac comptant 12 électrodes et conçu par deux équipes de chercheurs français en 1984. La nouvelle prothèse, produite par la compagnie MXM, utilise un réseau de 15 électrodes. Elle est

munie d'un analyseur qui incorpore un algorithme basé sur la FFT pour le traitement de la parole [7]. Cette analyse fréquentielle repose sur l'extraction de 64 valeurs d'énergies sur des bandes de fréquence de 122 Hz dans une fenêtre de 100 à 7800 Hz. Ces valeurs sont par la suite distribuées sur 15 bandes de fréquence définies par le clinicien et qui représenteront l'information à encoder en des trains d'impulsions séquentielles et à envoyer sur les 15 électrodes en commençant par celles qui représentent les hautes fréquences. Les impulsions sont de largeur et d'amplitude ajustables suivant les niveaux d'énergie. Quatre valeurs d'amplitude du stimulus communes à toutes les électrodes sont offertes. S'il y a de l'interférence entre les électrodes voisines, la valeur de l'amplitude de l'impulsion peut être changée pour l'un des autres niveaux. La stimulation est effectuée à la fréquence du voisement ou à une fréquence programmable entre 125 à 400 Hz.

L'équipe de Chouard utilisant le système MXM implante en 1997 un nouveau algorithme appelé le AIS (Asynchronous Interleaved Stimulation) [35]. Cet algorithme évalue plusieurs FFT de différentes largeurs sur le même segment de parole. Les propriétés statistiques de ces transformées vont permettre de décider de la stationnarité du segment de parole ainsi que du choix d'une élévation du rythme de stimulation. Ce dernier est fixé à 600 Hz si le segment est non stationnaire ou est abaissé à 400 Hz dans le cas contraire. Les tests préliminaires ont montré une amélioration tangible de la discrimination de consonnes mais pas au niveau des voyelles sauf pour des patients ayant déjà fait preuve de bonne performance avec l'ancien algorithme.

## 2.5 Conclusion

Dans ce chapitre, on a exposé dans une première partie les principes physiologiques du système auditif et on a expliqué les mécanismes de l'audition. On a ensuite présenté le principe des prothèses cochléaires ainsi que les stratégies de stimulations utilisées et les systèmes existants.

La diversité des stratégies évoquées provient du fait qu'on n'est pas encore parvenu à une modélisation complète du fonctionnement de l'oreille. Cela est dû principalement à l'ignorance de la façon dont les informations fréquentielles et temporelles sont représentées sur le nerf auditif. Les implants sont ainsi conçus selon la perception de chaque groupe de recherche et en fonction des stratégies de stimulation qui seront adoptées. Cette représentation partielle de l'information a pour effet d'étaler la période de réhabilitation, notamment pour les patients ayant déjà certain acquis linguistique, mais aussi de nuire à la qualité de perception si le niveau de bruit environnant n'est pas judicieusement contrôlé.

En vue de résoudre certains de ces problèmes, on propose une technique de traitement de parole basé sur la quantification vectorielle statistique. L'approche adoptée vise surtout à réduire le nombre de séquences de stimulation auxquelles le patient sera exposé, d'améliorer les performances en présence du bruit, et de réduire la durée de réhabilitation.

### **3. ALGORITHME DE STIMULATION BASÉ SUR LA QUANTIFICATION VECTORIELLE**

Comme il a été évoqué dans le chapitre précédent, tous les algorithmes de stimulation préconisés par les différents systèmes tentent de créer chez le patient une sensation auditive par le biais d'une activité électrique dans sa cochlée traduisant l'information extraite du signal acoustique.

Dans le cas de la parole, ce traitement se dote d'une certaine particularité. En effet, le signal vocal présente des caractéristiques qui lui sont propres. Il est caractérisé par une très grande redondance, une stationnarité à court terme et des particularités spectrales. Ainsi, toute modélisation, analyse ou codage de la parole doit tenir compte de ces caractéristiques. L'une des méthodes prédominantes dans le traitement de parole est la prédiction linéaire. Cette approche se base sur un modèle simplifié du mécanisme de la phonation pour estimer les paramètres de base de la parole et les coder en vue d'une transmission à bas débit et à faible stockage. Il vient alors que la prédiction linéaire se prête efficacement aux applications de traitement de parole où l'on a à analyser, coder et transmettre le signal vocal. Dans ce chapitre on exposera en premier lieu les caractéristiques et les modèles d'analyse de la parole. On présentera par la suite notre approche de traitement de parole basée sur la quantification vectorielle et un modèle prédictif linéaire, et on montrera la validité de cette technique pour les prothèses cochléaires ainsi que les avantages qu'elle apporte par rapport à d'autres techniques.

### 3.1 Les caractéristiques de la parole

Le signal vocal résulte des fluctuations de la pression de l'air engendrées puis émises par l'appareil phonatoire. Les principaux organes responsables du mécanisme de la phonation sont illustrés sur la figure 3.1. Ces organes sont les poumons, la trachée, le larynx, les cordes vocales, le 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-artère. Au sommet de celle-ci se trouve le larynx où la pression de l'air est modulée avant d'être 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'excitation convertissant l'énergie en un bourdonnement qui se traduit par un langage perceptible.

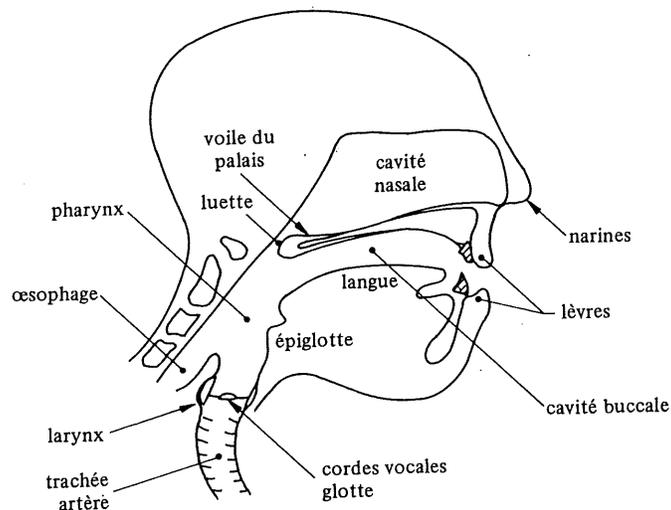


Figure 3.1 L'appareil phonatoire [6]

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 3.2 illustre la répartition des phonèmes de la langue française en fonction des modes de phonation [6].

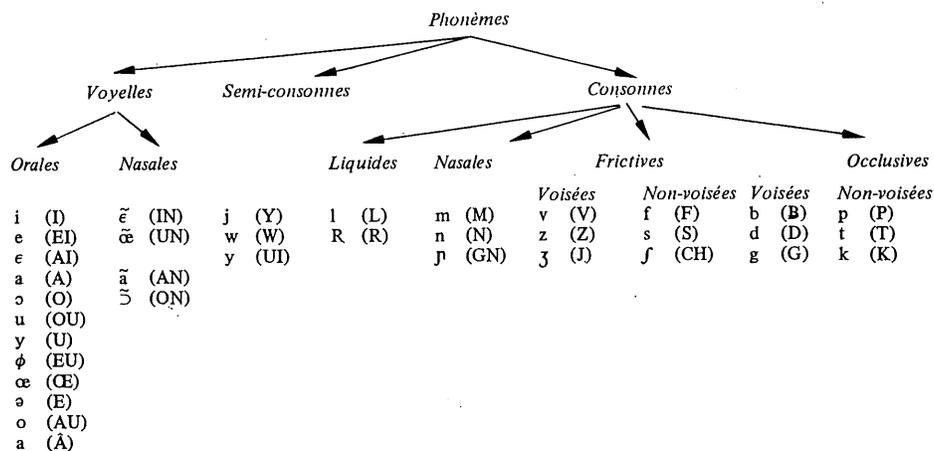


Figure 3.2 Phonèmes de la langue française [6]

De point de vue spectral, la parole présente des propriétés fort 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 [6], mais il peut être décomposé suivant deux allures de spectres (figure 3.3) correspondant aux séquences voisés 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.

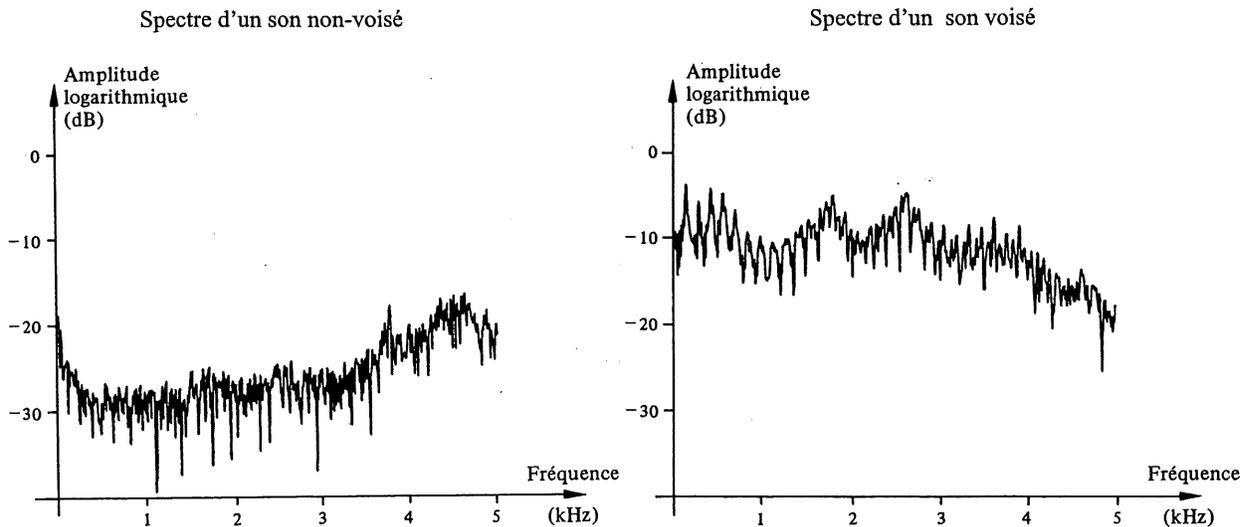


Figure 3.3 Spectres d'un son non-voisé et d'un son voisé [6]

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. On exposera dans ce qui suit quelques modèles et méthodes utilisées en traitement de parole. Par la suite on présentera l'approche de la quantification vectorielle qui sera utilisée dans ce travail.

### **3.2 Les modèles d'analyse de la parole**

Bien que nombreuses, les méthodes d'analyse de la parole peuvent être regroupées en deux grandes catégories : l'analyse par transformée de Fourier et l'analyse prédictive linéaire. La première sert à une analyse spectrale continue du signal vocal. Il en découle la méthode d'analyse par bancs de filtres qui a été dans le passé la méthode la plus utilisée (les vocodeurs à canaux). La deuxième méthode résulte d'une simple modélisation du conduit vocal et d'une modélisation autorégressive de la parole. Elle se base sur une caractérisation du spectre vocal par un ensemble de paramètres.

#### **3.2.1 Analyse par transformée de Fourier**

La transformée de Fourier est une méthode d'analyse spectrale qui ne se réfère pas à un modèle de production de la parole. Le calcul du spectre d'un signal de parole par la transformée de Fourier rapide (FFT) est souvent contraint par l'intermodulation de la source et du conduit ce qui rend difficile l'estimation des formants et de la fréquence fondamentale. Il est cependant possible de séparer la contribution de ces derniers par déconvolution grâce au lissage spectral.

La réalisation de bancs de filtres basée sur un calcul de FFT pour l'estimation des spectres représentait l'une des méthodes les plus utilisées pour la réalisation des systèmes

d'analyse de la parole. Dans les vocodeurs à canaux, les premiers systèmes d'analyse/synthèse de la parole, le signal est filtré sur plusieurs bandes de fréquences. L'enveloppe spectrale du conduit vocal est alors représentée par l'énergie moyenne dans chaque bande.

La transformée de Fourier se trouve cependant limitée par l'effet de Gibbs ou l'effet de bord qui est souvent présent dans les signaux non-périodiques. Cette limitation rend difficile l'analyse des segments de parole non-voisés et requiert l'utilisation des méthodes d'analyse locales comme la transformée de Gabor. Celle-ci suggère de rendre locale la transformée de Fourier de sorte à considérer un segment non périodique comme une succession de segments périodiques sur lesquelles la FFT est applicable sans problèmes d'effets de bord.

### 3.2.2 Analyse par prédiction linéaire

L'analyse par prédiction linéaire est devenue une technique prédominante pour l'estimation et la représentation des paramètres de base de la parole. Le modèle prédictif linéaire repose sur une modélisation simple de la production de parole qui consiste à décorrélérer la source du conduit vocal (figure 3.4). Le signal vocal est alors produit par excitation de ce modèle soit avec une suite d'impulsions quasi-périodiques dans le cas des sons voisés soit par un bruit blanc dans le cas des sons non-voisés.

Ce modèle est en réalité trop simplifié. En effet, il ne tient pas compte de l'interaction entre la source et le conduit, de la forme réelle des impulsions glottales et de la radiançe des lèvres. Cependant, il présente l'avantage de réduire la quantité d'information

nécessaire pour représenter le message (signal de parole), sans pour autant encourir une perte d'information.

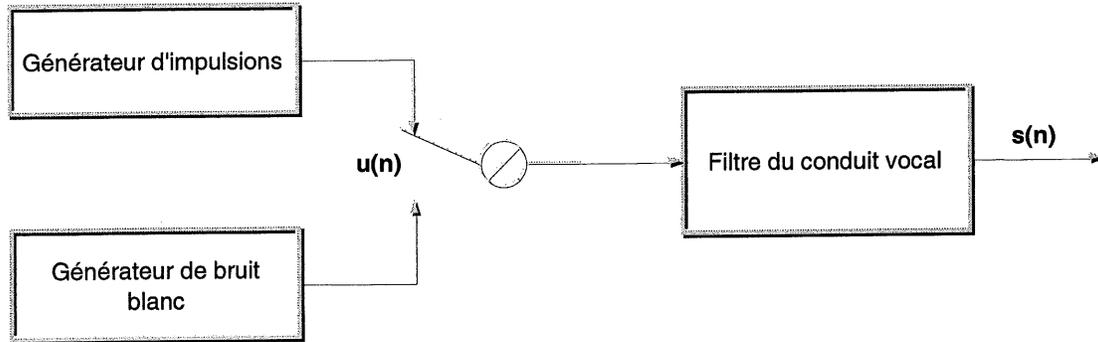


Figure 3.4 Modèle linéaire de production de la parole

La prédiction linéaire est un modèle autorégressif. La lente évolution à court terme du spectre de la parole permet de considérer ce signal stationnaire sur des intervalles de temps de l'ordre de 10 à 20 ms. Chaque échantillon du signal peut être prédit par une combinaison linéaire des  $p$  échantillons précédents. On peut alors écrire :

$$\hat{s}(n) = -\sum_{k=1}^p a(k)s(n-k) \quad (3.1)$$

où  $a(k)$  est le coefficient de prédiction,  $s(n)$  est l'échantillon du signal et  $\hat{s}(n)$  l'échantillon prédit.

On peut alors définir l'erreur  $e(n)$  produite sur la prédiction de l'échantillon  $s(n)$  comme suit :

$$e(n) = s(n) - \hat{s}(n) = s(n) + \sum_{k=1}^p a(k)s(n-k) \quad (3.2)$$

L'équation 3.2 peut être considérée comme la réponse d'un filtre inverse  $A(z)$  défini par :

$$A(z) = 1 + \sum_{k=1}^p a(k)z^{-k} \quad (3.3)$$

Ainsi, le filtre qui tient compte des caractéristiques du conduit vocal (figure 3.4) peut être défini par :

$$H(z) = G/A(z) \quad (3.4)$$

Ceci correspond aussi à l'estimation de l'enveloppe spectrale du signal  $s(n)$  où l'interférence avec la fréquence fondamentale est inhibée. D'un autre côté, la fonction de transfert  $H(z)$  est celle d'un modèle autorégressif (AR) tout pôle. Si le segment de parole contient des zéros à large bande, son spectre sera bien approximé avec le même modèle AR en augmentant l'ordre du modèle  $p$ . L'analyse par prédiction linéaire revient ainsi à l'étude du filtre  $H(z)$  pour dégager les caractéristiques spectrales du segment de parole analysé. On peut définir l'erreur spectrale  $E(z)$  en appliquant la transformée en  $z$  à l'équation 3.2. Ceci permet d'écrire :

$$E(z) = S(z) + \sum_{k=1}^p a(k)z^{-k} S(z) = \left( 1 + \sum_{k=1}^p a(k)z^{-k} \right) S(z) = A(z)S(z) \quad (3.5)$$

Pour avoir une meilleure estimation du signal de sortie  $S(z)$  par  $H(z)$ , il faut que l'erreur spectrale  $E(z)$  soit à spectre uniforme (figure 3.5). Ce sont les méthodes des autocorrélations et celle de Levinson-Durbin qui déterminent les coefficients du filtre  $A(z)$  correspondant à un tel spectre [6].

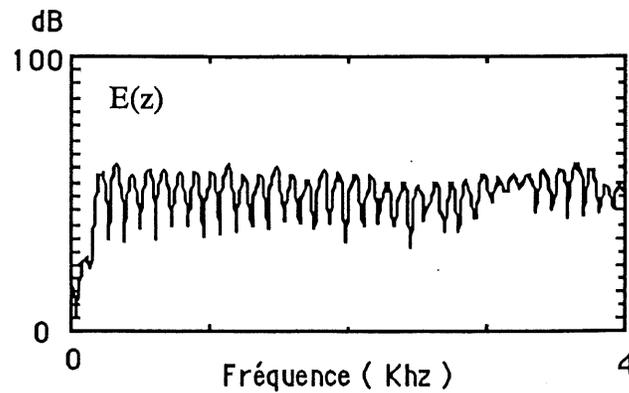


Figure 3.5 L'erreur spectrale d'un segment de parole voisé estimé par un modèle prédictif

L'estimation des formants des régions voisées de la parole peut dès lors se faire par deux alternatives. La première est directe et est basée sur la résolution des racines de l'équation caractéristique  $A(z)$  donnée par l'équation 3.3. La deuxième repose sur la recherche des pics spectraux de  $H(z)$ . La première technique permet de connaître les fréquences de résonances et les bandes passantes mais elle présente le problème d'établir une correspondance entre les pôles et les formants. Néanmoins, il est souvent possible de détecter les pôles qui ne font pas partie des formants puisqu'ils sont caractérisés par des bandes passantes très larges. Par ailleurs, le modèle tout pôle présente certains désavantages quant à l'interprétation des pôles. En effet, la signification physique des pôles dans le cas des sons nasaux demeure indéterminée. On ignore si ces pôles correspondent à des zéros de nasale, des pôles de nasale ou des résonances du conduit vocal. Une autre difficulté réside en la sensibilité des bandes passantes à la durée de la fenêtre d'analyse, de sa position et à la méthode de calcul [6].

### 3.3 Codage de la parole par quantification vectorielle

#### 3.3.1 Représentation de la parole

Le codage du signal vocal est l'étape succédant son analyse. Cette étape est généralement utilisée pour compresser le signal et ainsi réduire la quantité d'information à transmettre, accroître la capacité de stockage et avoir une meilleure robustesse (immunité au bruit, qualité de transmission). La redondance naturelle du signal vocal permet de réduire le débit binaire dans une très large mesure, au prix d'un traitement plus ou moins complexe et au risque d'une certaine dégradation de la qualité de la représentation.

Diverses techniques ont été proposées pour atteindre cet objectif; en fait le choix d'un système de représentation résulte d'un compromis entre la qualité de la représentation, la complexité du traitement, le débit binaire à transmettre et enfin la robustesse. Il existe deux grands modes de représentation : la représentation directe (ou le codage des formes d'onde) basée sur le traitement large bande du signal et la représentation paramétrique basée sur le codage des caractéristiques du signal [6]. Ces modes de codages sont en fait à la base des stratégies de stimulation évoquées dans le chapitre précédent. Le codage direct a pour objectif une représentation du signal ou de son spectre qui soit aussi fidèle que possible. Il consiste à représenter chaque échantillon du signal ou de son spectre indépendamment de tous les autres. Dans le codage paramétrique, les paramètres estimés du modèle de production sont quantifiés pour caractériser un spectre court-terme du signal.

La représentation de la parole, pour un mode ou pour l'autre, implique par ailleurs la quantification des échantillons (ou des paramètres) selon un nombre fini de valeurs

discrètes. Cette quantification peut alors agir sur des valeurs scalaires dans le cas d'une quantification scalaire comme elle peut agir sur des valeurs vectorielles dans le cas d'une quantification vectorielle. La théorie de taux de distorsion de Shannon [5] a montré que la quantification de vecteurs procure toujours des meilleurs résultats que la quantification des scalaires. Ce résultat a été en effet confirmé par les recherches récentes qui ne cessent de montrer les vertus de la quantification vectorielle en matière de codage de la parole [1] [20][28].

Dans ce qui suit, on expose le principe de la quantification vectorielle et on présentera l'algorithme de traitement de parole mettant au profit cette technique pour donner lieu à de nouveaux algorithmes de stimulation de la cochlée.

### 3.3.2 Quantification vectorielle

Dans le sens mathématique, la quantification vectorielle est la représentation d'un vecteur  $x$  dont les  $k$  composants sont à valeurs réelles continues, par un vecteur  $y$  appartenant à un ensemble fini  $\{y_i \in \mathfrak{R}^k, i = 1, 2, \dots, M\}$ . Les éléments de cet ensemble sont des vecteurs particuliers des partitions  $V_i$  de  $\mathfrak{R}^k$  (dits régions de Voronoï) appelés centroïdes ou mots de codes. L'ensemble des centroïdes constitue un dictionnaire ou vocabulaire. Dans le cas de transmission ou de stockage, un code, souvent binaire, est attribué à chaque vecteur  $y_i$ .

La substitution de  $x$  par le vecteur  $y$  provoque une distorsion  $d(x,y)$ . Cette distorsion est dans le cas général une distance de  $\mathfrak{R}^k$  [3]. Dans le cas du codage des formes d'onde, on utilise le plus souvent l'erreur quadratique :

$$d^2 = \frac{1}{k} \sum_{i=1}^k (x_i - y_i)^2 \quad (3.6)$$

La distorsion moyenne  $D$  due à la quantification est alors :

$$D = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=0}^{N-1} d[x(n), y(n)] \quad (3.7)$$

Le but poursuivi dans l'élaboration d'un système de codage est de minimiser la distorsion moyenne pour un débit de transmission donné ou réciproquement de minimiser le débit pour une distorsion acceptée.

La procédure de la quantification vectorielle peut être algébrique ou statistique. Dans le premier cas, l'ensemble des points quantifiés dans  $\mathfrak{R}^k$  présente une certaine structure mathématique. Ces points sont engendrés indépendamment des données. Alors, on admet implicitement pour celles-ci une répartition uniforme. Le problème se ramène ainsi à celui de l'empilement de sphères dans  $\mathfrak{R}^k$  où la plupart des solutions connues sont des réseaux ou treillis [1]. Dans la procédure statistique, les paramètres du système de quantification sont simplement obtenus par entraînement à partir d'un ensemble suffisamment représentatif des données. A faible débit, dans le cas de petites dimensions et de dictionnaire de tailles relativement faibles, la quantification vectorielle statistique engendre des opérations relativement moins complexes et un temps de calcul plus réduit qu'avec l'approche algébrique. Elle devient en outre intéressante dès que la distribution des données n'est pas uniforme. L'approche statistique est au fait utilisée couramment dans le codage de la parole [6]. On envisage dans notre cas d'implémenter cette approche vu la nature des données et de l'application en plus de la contrainte temps-réel. En effet,

cette dernière est un paramètre important dans la mise au point d'un algorithme de stimulation, ce qui suppose un temps de calcul réduit et donc des opérations pas trop complexes à effectuer par le processeur.

### 3.3.3 Construction du dictionnaire

La construction du dictionnaire de mots de code demeure l'étape la plus importante pour tout quantificateur. Pour établir un dictionnaire, il faut déterminer les régions de Voronoï et leurs centroïdes respectifs de façon à minimiser la distorsion de quantification. Dans le cas de la quantification statistique, cette construction se fait par apprentissage sur un nombre significatif de données. Quoique plusieurs méthodes sont préconisées, la méthode dite de LLOYD généralisée ou des K-moyennes reste la plus utilisée [6]. Elle consiste à transformer les centroïdes d'une partition pour obtenir une nouvelle partition dont la distorsion est inférieure ou égale à celle de la partition originale. Le processus est itéré jusqu'à l'obtention d'un minimum local de la distorsion. L'algorithme des K-moyennes se résume en 4 étapes :

- création d'un dictionnaire initiale formé de  $M$  centroïdes quelconques de l'ensemble d'apprentissage,
- création des  $M$  classes en cherchant pour chaque centroïde lesquels des vecteurs d'apprentissage sont les plus près, dans le sens de la minimisation d'une distorsion,
- calcul des nouveaux centroïdes. Si la distorsion utilisée est une distance euclidienne ou une distance cepstrale, les centroïdes sont les barycentres (centres

de gravité) des classes établies. On peut définir le centroïde d'une classe  $V_i$  contenant les éléments  $x_j, j = 1, 2, \dots, L$  par un vecteur donné par :

$$c_i = \frac{1}{L} \sum_{j=1}^L x_j \quad (3.8)$$

- itération du processus jusqu'à ce que la distorsion ne varie plus significativement d'une itération à l'autre.

Il est à préciser que l'algorithme des K-moyennes n'assure cependant pas un dictionnaire optimal. En effet, il peut converger vers une situation sous-optimale fonction du choix initial des centroïdes. Il y a alors intérêt à recommencer la procédure plusieurs fois avec des choix initiaux différents pour finalement retenir la solution qui fournit la distorsion la plus faible.

### 3.3.4 Codage LPC par quantification vectorielle statistique

La théorie du codage prédictif linéaire par quantification vectorielle a fait preuve de plusieurs succès notamment dans les applications de codage à faible débit de la parole. Le principe de cette approche consiste à coder les paramètres du modèle LPC par quantification vectorielle dans un dictionnaire de vecteurs de paramètres de sorte à minimiser l'erreur de prédiction.

Les paramètres d'un modèle LPC étant les coefficients de prédiction, il s'agit d'établir un quantificateur vectoriel dont le dictionnaire est formé de  $M$  vecteurs de coefficients LPC préétablis et dont le critère de distorsion est de minimiser l'erreur de prédiction donnée par l'équation (3.2). Dans le cas de la prothèse cochléaire de Sherbrooke, le dictionnaire comporte les coefficients de filtres décrivant les 31 à 36 phonèmes de la langue française.

Les  $M$  spectres ainsi décrits dans le dictionnaire serviront de base pour la réhabilitation du patient à l'identification des sons.

Soit  $\alpha$  l'énergie résiduelle résultant de l'estimation du segment de parole  $s(n)$  par  $\hat{s}(n)$ . Minimiser l'erreur de prédiction  $e(n)$  implique une minimisation de  $\alpha$ . D'après le théorème de Parseval l'énergie résiduelle peut s'écrire comme suit:

$$\alpha = \int_{-\pi}^{\pi} |S(e^{j\theta})|^2 |A(e^{j\theta})|^2 \frac{d\theta}{2\pi} \quad (3.9)$$

L'équation 3.9 peut s'écrire sous forme numérique de la façon suivante [21] :

$$\alpha = \sum_{k=-p}^p r_s(k)r_a(k) \quad (3.10)$$

où  $p$  est l'ordre du modèle,  $r_a(n) = \sum_{k=0}^p a(k)a(k+n)$  est la fonction d'autocorrélation des coefficients de prédiction et  $r_s(n) = \sum_{k=0}^{N-1} s(k)s(k+n)$  est la fonction d'autocorrélation du signal  $s(n)$ .

En se basant sur le fait que la fonction d'autocorrélation est paire. L'équation 3.10 peut se réécrire comme suit :

$$\alpha = r_s(0)r_a(0) + 2 \sum_{k=1}^p r_s(k)r_a(k) \quad (3.11)$$

Le calcul de cette quantité peut être effectué sur plusieurs domaines. Celui auquel on s'intéresse est basé sur le spectre fréquentiel vu que le traitement proposé repose sur les caractéristiques fréquentielles de la parole. La distance d'Itakura-Saito modifiée sous la

forme du rapport de vraisemblance permet alors d'établir un critère de distorsion [24].

Cette distance est donnée par:

$$d_{LR}(A_p, A) = \int_{-\pi}^{\pi} \left| A(e^{j\theta}) / A_p(e^{j\theta}) \right|^2 \frac{d\theta}{2\pi} - 1 \quad (3.12)$$

Cette expression représente la distorsion calculée en estimant le spectre  $S(z)$  par un filtre inverse quelconque  $A(z)$  sachant que  $A_p(z)$  est le filtre inverse optimal du modèle LPC d'ordre  $p$ . Si  $\alpha_p$  désigne l'énergie résiduelle optimale résultant de l'estimation de  $S(z)$  par le filtre  $A_p(z)$ , alors le rapport de vraisemblance peut s'écrire sous la forme :

$$d_{LR} = \frac{\alpha}{\alpha_p} - 1 \quad (3.13)$$

L'équation 3.13 permet donc d'établir une équivalence entre minimiser le rapport de vraisemblance  $d_{LR}$  et minimiser l'énergie résiduelle  $\alpha$  puisque  $\alpha_p$  ne dépend que du signal d'entrée.

En utilisant les équations 3.11 et 3.13, on peut écrire :

$$d_{LR}(A_p, A) = \frac{r_s(0)}{\alpha_p} r_a(0) + 2 \sum_{i=1}^p \frac{r_s(i)}{\alpha_p} r_a(i) - 1 \quad (3.14)$$

La minimisation du rapport de vraisemblance donné par l'équation (3.14) permettra de dégager le filtre du dictionnaire correspondant le mieux au spectre  $S(z)$  de la séquence  $s(n)$ . Notons que l'utilisation du rapport de vraisemblance présente un avantage remarquable qui consiste à ne pas tenir compte du facteur gain du filtre LPC ce qui

signifie la possibilité de le quantifier séparément. Cette approche offre l'avantage de contourner les contraintes de devoir représenter les variations de gain des séquences de parole dans le dictionnaire. Ce dernier gardera donc une taille moins grande et sera plus facile à aménager.

L'opération de traitement de la parole proposée se divise dès lors en deux étapes de traitement : l'identification et la quantification. Dans la première étape, les caractéristiques du signal sont identifiées en calculant ses  $p$  coefficients d'autocorrélation. Dans la seconde étape, le signal est quantifié en choisissant parmi les  $M$  filtres du dictionnaire celui qui minimise la distorsion donnée par l'équation 3.14. La figure 3.6 résume ce processus.

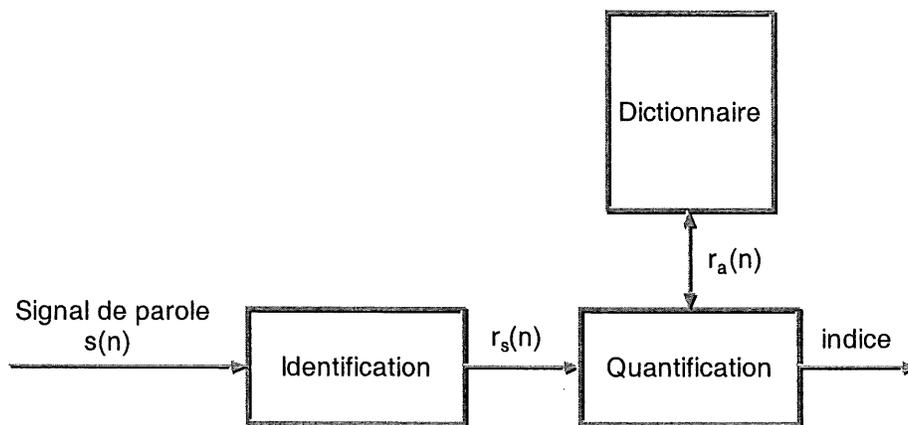


Figure 3.6 Processus de traitement de la parole basé sur la quantification vectorielle

Les avantages que procure le traitement par quantification vectorielle comme stratégie de stimulation de la cochlée sont nombreux. Par ces avantages on vise l'amélioration de plusieurs aspects relatifs au processus d'identification des sons ainsi qu'à la période de réhabilitation des patients. Le premier avantage réside dans la limitation du nombre de

sons ou phonèmes que le patient doit identifier, au contenu du dictionnaire. Le patient trouverait ainsi une plus grande facilité dans la période d'entraînement et d'habilitation à l'identification des sons [18]. Le contenu et la taille du dictionnaire étant fixés au préalable, on modifiera les sons stockés ainsi que le nombre en fonction de l'évolution du patient dans le processus de réhabilitation. Un deuxième avantage provient de la qualité de l'information transmise à la cochlée. Cette information se trouve "lissée" en ce sens que les spectres correspondants à chaque phonème sont issus d'une moyenne statistique obtenue des mêmes mots prononcés par plusieurs locuteurs différents. Le troisième avantage du traitement proposé se trouve dans l'amélioration de la qualité de perception par rapport au bruit environnant. En effet, il est connu que dans la majorité des stratégies de stimulation, les impulsions électriques délivrées à la cochlée sont synchrones des sons et des bruits de la parole comme ceux du monde environnant. Ceci fait que l'identification des phonèmes se révèle moins systématique qu'avec le traitement par quantification vectorielle qui opère avec un nombre défini et limité de spectres de fréquences représentant les sons et les bruits. Ainsi, si l'analyseur de parole peut effectuer une identification précise des phonèmes dans le dictionnaire des spectres, le rapport signal à bruit du système peut en être considérablement amélioré. Enfin, un quatrième avantage vient de la flexibilité du dictionnaire. Ce dernier pouvant être adapté à la langue maternelle ou même aux particularités linguistiques régionales de chaque patient. Cette caractéristique pourrait alors pallier la difficulté d'adapter aux autres langues un algorithme de stimulation développé dans une langue donnée, les spectres des divers phonèmes n'étant pas identiques d'une langue à l'autre, ou même d'une région à l'autre.

### 3.4 Conclusion

Ce chapitre a fait l'objet d'une présentation générale des caractéristiques de la parole ainsi que de quelques modèles préconisés pour son codage. La technique de codage prédictif linéaire par quantification vectorielle qu'on a explicitée, présente un intérêt particulier pour le traitement de parole destiné à un algorithme de stimulation de la cochlée. En effet, elle permettra de réduire la période de réhabilitation des patients qui utilisent des prothèses cochléaires et d'améliorer les performances de ces appareils en milieu bruité. Ces avantages découlent du principe de base de la technique proposée qui consiste à utiliser un dictionnaire d'un nombre limité d'éléments permettant de décrire complètement tous les phonèmes ou les sons à percevoir. Dans le prochain chapitre, on exposera la mise en œuvre de cette technique. On décrira au préalable la prothèse de Sherbrooke et en particulier l'analyseur de parole sur lequel l'algorithme a été implémenté.

## **4. MISE EN ŒUVRE DE L'ALGORITHME DE STIMULATION BASÉ SUR LA QUANTIFICATION VECTORIELLE**

La mise en œuvre de tout algorithme de stimulation destiné aux prothèses cochléaires est étroitement liée au côté matériel du système et aux données physiologiques du patient. Les performances du système relèvent alors de la pathologie et de l'état du patient, ainsi que de divers champs de compétences tels que l'ingénierie, l'audiologie et la psychoacoustique. Tous les intervenants contribuent ensemble à la mise au point de solutions susceptibles d'amener aux meilleurs résultats. Il est ainsi impératif de considérer tout les critères tant techniques que cliniques lors de la mise en œuvre de l'algorithme de stimulation, et ce aussi bien pour l'aspect traitement de signal que pour l'aspect stratégie de stimulation de l'algorithme. Le traitement de signal doit ainsi considérer les contraintes matérielles de la plate-forme d'implémentation alors que la stratégie de stimulation doit se conformer aux critères cliniques de la pathologie du patient. Vu le nombre et la diversité de ces considérations, la mise en oeuvre de l'algorithme de stimulation suppose la disposition d'un système matériel qui offre une marge de flexibilité suffisante pour supporter la complexité des algorithmes visés ainsi que la diversité de ses paramètres.

Dans ce chapitre on exposera la mise en oeuvre de l'algorithme de stimulation basé sur la quantification vectorielle en décrivant principalement les étapes de mise en œuvre de la

technique de traitement de signal proposée au chapitre précédent. La première partie sera dédiée à la prothèse cochléaire de Sherbrooke qui constitue notre plate-forme technologique. On passera en revue les systèmes développés au sein du Groupe de Recherche en Appareillage Médical de Sherbrooke (GRAMS) pour ensuite décrire les fonctionnalités de la prothèse actuellement en cours de développement. La deuxième partie sera consacrée à la description fonctionnelle de la technique proposée.

#### **4.1 Limitation des systèmes actuels**

La majorité des systèmes de prothèses cochléaires décrites dans le chapitre 2 présente beaucoup de limitations quant aux stratégies de stimulation qu'elle peut supporter. Les analyseurs de parole qu'on retrouve sont généralement conçus d'une façon trop spécifique et ne supportent qu'un algorithme de stimulation ou des variantes de ce même algorithme. Ainsi, du côté clinique, le patient ne dispose pas d'un système qui peut être adapté à sa pathologie ni de l'avantage d'avoir plusieurs algorithmes dont l'efficacité peut dépendre des conditions externes. Ceci veut dire que c'est plutôt lui qui devrait s'adapter au système en subissant une réhabilitation "forcée" qui peut ne pas répondre à ses besoins. Du côté de la mise à jour du système, un système conçu selon un algorithme de stimulation spécifique impliquera une modification de sa conception matérielle si un autre algorithme est à implémenter.

Pour remédier à ce problème, un système flexible à structure indépendante des algorithmes de stimulation serait une solution. Ce système peut alors être individualisé et adapté facilement au patient.

Depuis plusieurs années, le GRAMS poursuit des recherches pour développer une prothèse cochléaire complètement programmable et dotée d'une flexibilité illimitée. L'objectif visé étant de mettre au point un analyseur de parole transparent aux algorithmes de stimulation ainsi qu'un stimulateur à faible consommation offrant une grande versatilité de stratégies de stimulation.

## **4.2 La prothèse cochléaire de Sherbrooke**

### **4.2.1 Le stimulateur**

Le premier stimulateur a été mis au point par le groupe en 1985 [33]. Il s'agissait d'un microstimulateur à 8 électrodes conçu sur un prédiffusé ISO-CMOS à 4  $\mu\text{m}$  et monté sur un circuit hybride de 2 cm de diamètre. Il peut fonctionner en mode monopolaire ou bipolaire et peut fournir jusqu'à quatre canaux de stimulation simultanés. Les électrodes sont adressables selon toutes les combinaisons possibles et leurs stimuli peuvent être synchronisés ou générés indépendamment. En 1993, la compagnie COSEM a acquit les droits de fabrication et de commercialisation de ce neurostimulateur pour l'utiliser avec un analyseur de parole qu'elle a conçu et qui intègre un algorithme de stimulation similaire à la stratégie IP. Les travaux du GRAMS se sont poursuivis pour développer la deuxième génération de son implant ainsi que son propre analyseur de parole. Avec les nouvelles technologies d'intégration à très grande échelle qui sont apparues, et les nouveaux outils de conception assistée par ordinateur, le groupe a mis au point son implant à 16 électrodes en 1992 [32]. Ce stimulateur est présenté en deux versions. La première de taille semblable à celui à 8 électrodes, et la deuxième de taille encore plus réduite (1 cm de diamètre). Ce stimulateur est plus rapide, plus puissant et offre beaucoup

plus de combinaisons d'électrodes que son prédécesseur. L'implant ainsi réalisé permet une stimulation à un rythme plus élevé même dans le cas de 16 stimulations simultanées. Pendant ce temps, les travaux sur l'analyseur de son ont permis de réaliser une première version qui utilisait un circuit intégré dédié et par conséquent ne répondait pas nécessairement au critère de flexibilité désirée. Une deuxième génération d'analyseur de parole a été alors entreprise. En 1998, le GRAMS présente son nouvel analyseur complètement programmable conçu autour d'un puissant DSP et doté d'une grande capacité mémoire capable d'intégrer plusieurs algorithmes simultanément [15].

#### 4.2.2 L'analyseur de parole

L'analyseur de parole développé par le GRAMS est doté d'une grande flexibilité et d'une transparence sans précédent aux algorithmes de stimulation [13]. En fait, cet appareil peut être considéré comme un outil clinique très puissant qui permettra aux psycho-acousticiens et aux audiologistes d'effectuer des travaux de recherche et de développement pour découvrir les meilleurs algorithmes de stimulation menant à une meilleure compréhension de la parole.

Le système est conçu autour d'un puissant processeur de traitement de signaux numériques (DSP) de la compagnie Texas Instruments (TMS320C50). Ce dernier est un processeur à point fixe sur 16 bits. Il est doté de 10 KOctets de mémoire vive, ce qui est suffisant pour charger et exécuter des algorithmes de traitement de signal assez complexes.

À l'entrée, l'analyseur comprend une chaîne d'amplification et un filtre anti-recouvrement qui communiquent les signaux à un convertisseur analogique-numérique. Ce dernier numérise ces signaux pour les présenter sur un des deux ports sériels du processeur TMS320C50. Ce même port sert à transmettre les mots de commande encodés vers le modulateur en direction du stimulateur interne. Le deuxième port sériel est utilisé pour une communication bidirectionnel avec le PC qui sert à sa programmation et à l'usage clinique. Le système est doté d'une mémoire Flash externe de 16 Mbits utilisée pour la sauvegarde des différents algorithmes. Ces derniers peuvent être sélectionnés par le patient, selon sa convenance, à l'aide d'un bouton externe prévu à cette fin [15]. La figure 4.1 illustre un schéma bloc du système.

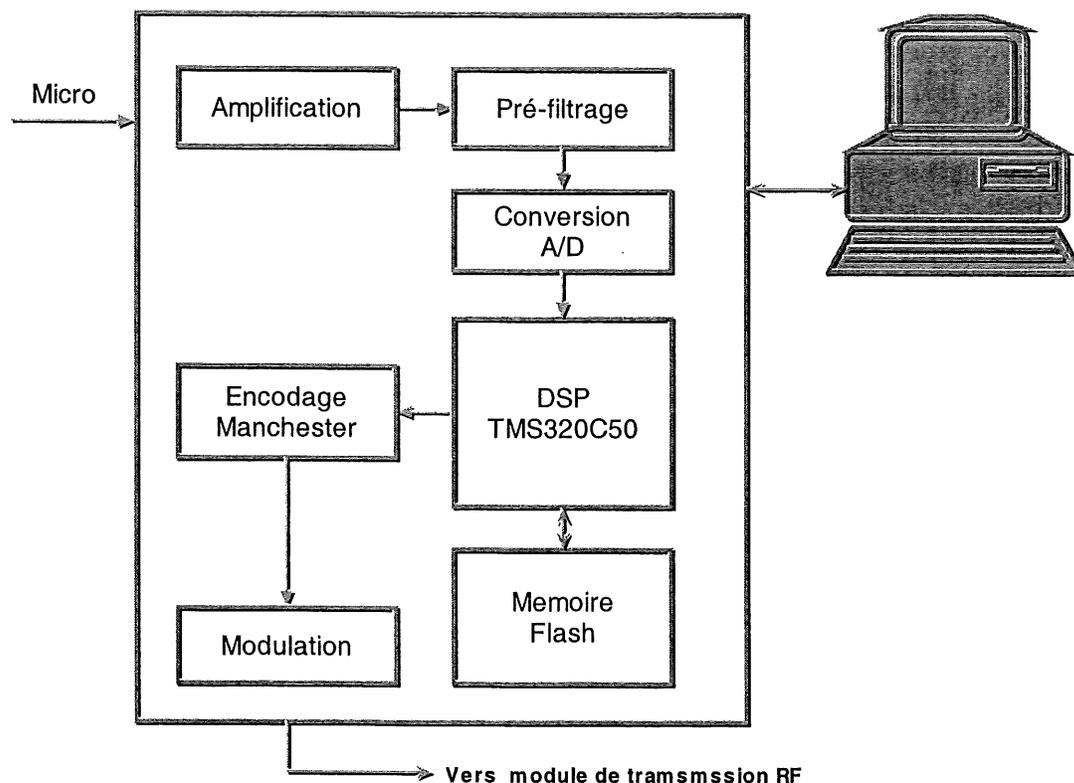


Figure 4.1 Schéma bloc de l'analyseur de parole de Sherbrooke

Afin de faciliter la tâche des cliniciens et de simplifier la procédure de programmation du système, un logiciel clinique appelé PASTIS (Programme de tests d'Audiologie et de Stimulations pour l'Implant cochléaire de Sherbrooke) est développé pour effectuer les tests de calibrage et l'adaptation de l'appareil au patient. Ce même logiciel peut ainsi servir pour tester les différents algorithmes de traitement de signal et les différentes stratégies de stimulation lors des séances de rééducation [8].

#### **4.3 Mise en oeuvre de l'algorithme de quantification vectorielle**

On a exposé dans le chapitre 3 le concept théorique du traitement de signal basé sur la quantification vectorielle et on a expliqué l'approche proposée pour appliquer cette technique comme algorithme de stimulation destiné à une prothèse cochléaire.

Comme pour tout algorithme de stimulation, la technique proposée se compose de deux processus complémentaires. Le premier s'occupe du traitement de signal pour en extraire l'information pertinente et le deuxième élabore la stratégie de stimulation qui exploite cette information en la représentant dans la cochlée sous forme d'impulsions électriques. Dans ce qui suit, on se concentrera sur la mise en oeuvre de la technique de traitement de signal basé sur la quantification vectorielle en tant que première étape dans l'implémentation d'un algorithme de stimulation. On verra dans le prochain chapitre, la deuxième étape qui consiste à lui associer des stratégies de stimulation qui peuvent être utilisées pour générer les impulsions de courant dans l'oreille interne.

### 4.3.1 Traitement de la parole par quantification vectorielle

Le traitement de la parole basé sur la quantification vectorielle est une opération de codage LPC. Ce processus est mis en œuvre en trois étapes. La première consiste en un pré-conditionnement du signal. La deuxième étape englobe les opérations d'analyse et d'identification du signal, et la troisième son codage par quantification vectorielle (figure 4.2). L'analyse spectrale de la séquence du signal reçue consiste à dégager les paramètres à coder du modèle LPC adopté, ce qui revient à calculer les coefficients d'autocorrélation du signal sur la fenêtre d'analyse. L'opération de codage consiste à quantifier ces paramètres par quantification vectorielle statistique moyennant un dictionnaire de spectres préétablis représentés par  $M$  vecteurs d'autocorrélation sur les coefficients du modèle LPC utilisé. Le spectre qui approcherait le mieux les caractéristiques de la séquence reçue sera ainsi sélectionné.

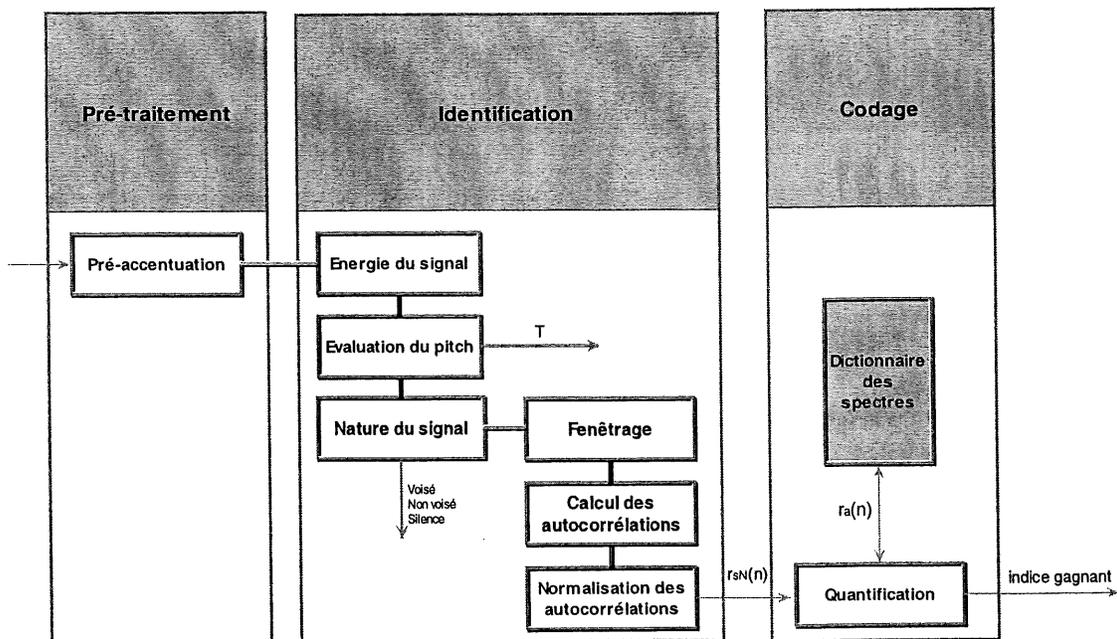


Figure 4.2 Diagramme d'implémentation du processus de traitement

### 4.3.2 Pré-accentuation du signal

Pour les sons voisés de la parole, la puissance spectrale se concentre plutôt dans les basses fréquences. Ceci a pour effet d'accentuer l'allure du spectre dans la région de ces fréquences, ce qui provoque une perte de la résolution spectrale lors de l'analyse du signal. Ceci induit, dans une analyse LPC, un problème de précision surtout apparent dans une implémentation sur un processeur un point fixe, ce qui est le cas dans notre application. Afin de remédier à ce problème, le signal est pré-accentué juste après son échantillonnage pour compenser le faible niveau des composantes spectrales de plus hautes fréquences. Cette pré-accentuation est réalisée en filtrant le signal par un filtre passe-haut de premier ordre dont la fonction de transfert est donnée par :

$$P(z) = 1 - \alpha z^{-1} \quad (4.1)$$

où  $0.9 \leq \alpha < 1$ . On peut alors écrire :

$$y(n) = s(n) - \alpha s(n-1) \quad (4.2)$$

où  $s(n)$  est la séquence d'entrée et  $y(n)$  et la séquence de sortie.

La figure 4.3 illustre l'allure spectrale d'un segment voisé avec et sans pré-accentuation.

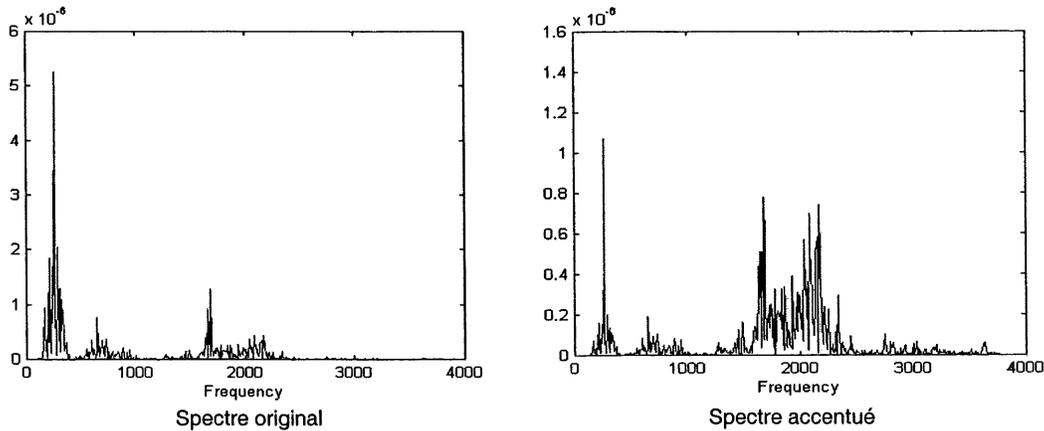


Figure 4.3 Spectres d'un segment de parole voisé sans et avec pré-accentuation

#### 4.3.3 Estimation de la fréquence fondamentale

L'identification du signal commence par l'estimation de sa fréquence fondamentale  $F_0$  ainsi que la détermination de sa nature, voisée ou non-voisée. Ces paramètres serviront pour établir les stratégies de stimulation qui seront préconisées. La mesure de  $F_0$  revient à une estimation de la fréquence des vibrations laryngées. Elle peut se faire dans le domaine temporel aussi bien que dans le domaine spectral. Différentes méthodes existent pour la "détection" de la fréquence fondamentale. Plusieurs d'entre elles font intervenir des techniques plus ou moins complexes qui nécessitent des quantités de calculs relativement élevées. Pour une application temps réel tel que la nôtre, on ne peut pas se permettre d'adopter de telles méthodes. Le calcul de  $F_0$  dans notre cas doit être relativement rapide et dans la limite du temps alloué. Généralement, la méthode dite d'autocorrélation, semble satisfaire ces contraintes temps réel [41]. Elle donne des résultats acceptables lorsque le signal ne varie pas trop d'une période à une autre (quasi-stationnaire).

La méthode d'autocorrélation est une méthode temporelle qui consiste à chercher le maximum de la fonction d'autocorrélation du signal sur une plage de temps déterminée. Compte tenu de la quasi-stabilité du pitch à long terme, la plage de temps sur laquelle il est évalué peut être réduite, ce qui diminuera la quantité de calcul. Ainsi, pour chaque séquence de parole  $s(n)$ , on calculera la période du pitch  $T$  sur une plage  $[t_0, t_1]$ . La période  $T$  sera l'argument qui maximise la fonction d'autocorrélation  $r_s(n)$  donnée par :

$$r_s(n) = \sum_{k=0}^{N-1} s(k)s(k+n) \quad (4.3)$$

L'énergie du signal ainsi que le gain du pitch sont deux autres grandeurs qui permettent d'évaluer la nature voisée ou non-voisée du signal. L'énergie du signal est estimée par la valeur de l'autocorrélation à zéro  $r_s(0)$ . En général, l'énergie d'un son voisé est beaucoup plus grande que celle d'un son non-voisé. Si l'énergie évaluée est remarquablement trop faible, la région peut être considérée comme un silence. Le gain du pitch  $\beta$  est défini comme le rapport du maximum d'autocorrélation  $r_s(T)$  sur l'énergie du signal  $r_s(0)$ . Il peut être écrit comme suit :

$$\beta = r_s(T) / r_s(0) \quad (4.4)$$

Pour un son voisé, le gain donné par l'équation 4.4 s'approche de l'unité.

#### 4.3.4 Identification des paramètres du modèle

On a mentionné précédemment que l'analyse spectrale suppose que le segment de parole traité est stationnaire. Or la lente évolution à court terme du spectre du signal de parole permet de le considérer comme stationnaire sur des intervalles de temps de l'ordre de 10

à 20 ms. ON considère alors une durée de fenêtre d'analyse de cet ordre. Autrement dit, le choix de la longueur de la fenêtre d'analyse  $N$  doit tenir compte de la stationnarité du signal. D'autre part, la valeur de  $N$  ne doit pas être trop grande pour minimiser la charge de calcul, ni trop petite afin de garantir une durée suffisante qui permettra de déterminer la fréquence fondamentale. Cette dernière ne dépasse généralement pas 500 Hz.

Le choix de l'ordre  $p$  du modèle LPC dépend de la fréquence d'échantillonnage  $F_s$  du signal. Le spectre de la parole peut en général être représenté par des pôles complexes ayant des effets dans les fréquences de l'ordre de quelques kilohertz, ce qui représente la contribution du conduit vocal [41]. En considérant le signal sur une bande de 4 KHz, on aura besoin de 4 paires de pôles complexes, ce qui correspond à un modèle LPC d'ordre  $p = 8$ . Ainsi, le signal sera échantillonné à 8 KHz et on aura 9 coefficients d'autocorrélation qui caractériseront les séquences de parole reçues. Ces coefficients sont calculés après fenêtrage de signal par une fenêtre de *Hamming* de longueur 20 ms (128 points pour des séquences de 8 KHz).

Afin de décorrélérer le spectre du signal de son énergie, il doit être normalisé par rapport à la valeur de celle-ci. Les coefficients d'autocorrélation sont alors normalisés par rapport à la valeur de l'énergie du signal, définie par l'autocorrélation à zéro. On peut alors écrire :

$$r_{sN}(n) = r_s(n) / r_s(0) \quad (4.5)$$

où  $r_s(n)$  et  $r_s(0)$  désignent respectivement le coefficient d'autocorrélation d'ordre  $n$  et l'énergie de la séquence de parole  $s(n)$ . Ils sont donnés par :

$$r_s(n) = \sum_{k=0}^{N-1} s(k)s(k+n) \quad (4.6)$$

$$r_s(0) = \sum_{k=0}^p s^2(k) \quad (4.7)$$

#### 4.3.5 Quantification

La dernière étape dans le processus de traitement de la séquence de parole est le codage des paramètres identifiés par quantification vectorielle. Dans le chapitre 2, on a montré que cette opération consiste à chercher dans le dictionnaire de spectres pré-établis celui qui correspond le mieux au spectre de la séquence traitée en terme de minimisation du rapport de vraisemblance (équation 3.14).

Pour une séquence de parole donnée, l'énergie résiduelle optimale  $\alpha_p$  est constante. La minimisation du rapport de vraisemblance revient alors à minimiser la quantité :

$$E = \sum_{k=1}^p r_s(k)r_a(k) \quad (4.8)$$

Les  $M$  spectres du dictionnaire ( $M=128$ ) sont scrutés en calculant pour chacun la quantité  $E$  donnée par l'équation 4.8. Le spectre qui sera sélectionné est alors celui qui correspond au minimum de cette quantité. Si l'on indexe les spectres du dictionnaire, l'indice  $j$  du spectre sélectionné sera celui qui correspond à :

$$\underset{j}{\text{Arg min}} E = \sum_{k=1}^p r_s(k)r_{a_j}(k) \quad (4.9)$$

où  $r_{a_j}(k)$  est l'autocorrélation d'ordre  $k$  du  $j^{\text{ème}}$  vecteur  $r_a$  du dictionnaire.

#### 4.4 Validation et discussion de l'algorithme

La validation du bon fonctionnement de l'algorithme qu'on vient de décrire, devrait se faire sur l'analyseur de son développé par le groupe. Malheureusement, on disposait seulement de deux prototypes qui étaient utilisés intensivement par deux membres du groupe pour finaliser leurs projets de doctorat. On s'est alors contenté d'effectuer quelques tests sommaires sur un module EVM (*Evaluation Module*) qui comprend principalement le même DSP que l'analyseur de son (TMC320C50), une mémoire externe SRAM de 64 Koctets et un lien bidirectionnel de communication avec le PC.

La validation de l'algorithme consiste à vérifier le bon fonctionnement du processus de codage. En considérant une séquence de parole, on vérifiera si le module de quantification fournira comme résultat l'indice du spectre approchant le mieux la séquence d'origine. Une façon simple de le faire consiste à calculer pour une séquence test  $s(n)$  le filtre correspondant  $A_s(z)$ . Ce même filtre est alors stocké dans le dictionnaire d'origine. Le programme est ensuite exécuté et devrait alors fournir l'indice de ce filtre comme résultat.

On a effectué l'expérience sur quelques séquences de parole de 128 échantillons. Les filtres inverses correspondant ont été déterminés en calculant les vecteurs des 8 autocorrélations sur les coefficients LPC d'ordre 8. Ces vecteurs ont été ensuite ajoutés au dictionnaire. Le résultat a démontré que la sortie de l'algorithme était toujours l'indice du filtre dont la séquence de parole correspondante a été présentée à l'entrée. En changeant les indices des filtres tests dans le dictionnaire, le programme ressort toujours le bon indice du filtre qui devrait être sélectionné par l'opération de codage.

Implémentée sur l'analyseur de Sherbrooke, et associée à de stratégies de stimulation adéquates, la technique proposée devrait faire preuve de ses avantages. Ces derniers étant principalement la minimisation de la période de réhabilitation du patient en permettant une identification systématique des sons et ainsi d'améliorer la reconnaissance de la parole particulièrement en milieu bruyé [18]. Il faut souligner que la taille du dictionnaire de 128 spectres n'est pas une limitation pour le patient. Le dictionnaire peut être sujet à des extensions futures selon l'état et l'évolution du patient pendant la période de réhabilitation. Le clinicien peut alors modifier le dictionnaire ou l'étendre en lui ajoutant d'autres éléments.

Il faut tout de même noter que l'extension du dictionnaire n'est pas une opération gratuite. Elle exige davantage de complexité de calcul et donc de temps de traitement, ce qui pourrait ne plus satisfaire les contraintes de temps réel exigées. Une solution à ce problème pourrait être la substitution de quantification vectorielle statistique par une quantification vectorielle algébrique [3]. La structuration mathématique des dictionnaires algébriques rend plus simple l'opération de recherche d'indice. Néanmoins, elle implique une construction plus complexe du dictionnaire. En effet, celui-ci renfermera des phonèmes ayant une certaine structure définie dans un espace donné dont l'obtention est plus complexe.

#### **4.5 Conclusion**

On a décrit dans ce chapitre la mise en œuvre de l'algorithme de traitement de parole basé sur la quantification vectorielle statistique. On a présenté brièvement le système qui

lui servira de support matériel ainsi que les différentes étapes qui permettent de le mettre en œuvre. Même si les tests de validation du programme ont été effectués sur un module d'évaluation du DSP, le code demeure facilement transférable sur l'analyseur de son.

La deuxième partie de l'algorithme de stimulation consiste à associer à l'algorithme de traitement de parole des stratégies de stimulation qui conviennent à cette technique. On propose alors dans le chapitre suivant, des approches de stratégies de stimulation qui peuvent servir à traduire les résultats du traitement de parole en impulsions électriques qui seront générées dans la cochlée.

## 5. STRATÉGIES DE STIMULATION

Dans le chapitre précédent, on a décrit les étapes essentielles nécessaires à l'implémentation d'un algorithme de traitement de la parole basé sur la quantification vectorielle. L'utilisation de cette technique pour la mise en œuvre des algorithmes de stimulation d'une prothèse cochléaire démontre une avenue très prometteuse pour ce genre d'application. En effet, la technique est généralement utilisée pour la compression de la parole, et par conséquent pour identifier l'information pertinente qui la caractérise. C'est effectivement ce qui est recherché dans le cas des prothèses cochléaires. Cependant, la synthèse du signal à la réception est remplacée par la stratégie de stimulation. C'est à dire, on ne cherche pas à reproduire le signal transmis, mais plutôt à représenter l'information identifiée à l'aide d'impulsions électriques dans l'oreille interne afin que le système nerveux du patient se charge de la reproduire par voie naturelle. Cela revient à dire que les stimulations électriques doivent provoquer une sensation auditive qui sera interprétée par le patient conformément au son transmis.

Quoique la dynamique cochléaire ait été bien définie, l'influx nerveux que cette dynamique provoque sur les voies nerveuses demeure encore difficile à déchiffrer. De plus, sa reproduction exacte à l'aide des stimulations électriques semble être impossible, puisqu'il a été démontré qu'il est le résultat d'une combinaison d'activités nerveuses ascendantes à travers le nerf auditif. De là, l'objectif des recherches effectuées sur les prothèses cochléaires n'est pas d'offrir un appareil qui reproduit la fonction naturelle de

l'oreille mais plutôt un système qui permet une compréhension maximale de la parole selon la pathologie rencontrée tout en facilitant le plus possible la période de rééducation du patient. C'est en effet, ce qu'on estime que la technique de quantification vectorielle offrirait.

En se basant sur ces constatations, le fait de mieux identifier l'information utile à coder par des impulsions électriques va permettre de trouver la meilleure façon pour la représenter. Ceci exige des expériences cliniques avec des patients et un système assez flexible pour aboutir à la solution. Le système du GRAMS répond à ce besoin, cependant on ne peut procéder à des expériences cliniques dans le moment puisqu'on ne dispose pas de porteurs de cet appareil. Pour cette raison on va proposer dans ce chapitre des stratégies de stimulation basées sur les résultats cliniques publiés dans la littérature. Néanmoins, on peut affirmer qu'une meilleure extraction de l'information contenue dans le signal donnera nécessairement de meilleurs résultats avec les mêmes stratégies de stimulation. Avant de décrire les stratégies qui pourraient être jumelées à notre algorithme de traitement de signal, on va décrire la façon de générer les stratégies de stimulation à l'aide de notre système.

## **5.1 Fonctionnement du neurostimulateur de Sherbrooke**

### **5.1.1 Principe**

L'implant opère à la manière d'un microprocesseur. Il exécute des mots de commande qu'il reçoit en série de l'analyseur de parole à travers le lien inductif. Ces mots sont de 24 bits pour la version à 8 canaux et de 32 bits pour la version à 16 canaux. La figure 5.1 représente le format du mot de commande utilisé par l'implant à 8 canaux [33]. Ce format

se compose de six champs. Il y a 3 bits pour la clef de reconnaissance de la commande, 3 bits pour le code d'opération, 6 bits pour le niveau de courant de stimulation, 4 bits pour l'adresse de l'électrode active, 4 bits pour l'adresse de l'électrode de retour de charge et 4 bits servant à adresser l'électrode avec laquelle il faut synchroniser le début des stimuli dans le cas de stimulations simultanées.

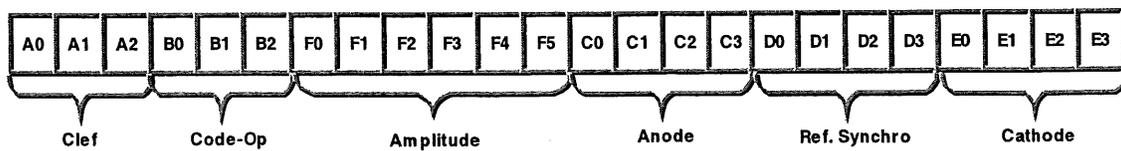


Figure 5.1 Format du mot de commande utilisé par l'implant à 8 canaux

L'implant dispose de 9 mots de commande qui permettent un contrôle total sur la forme des impulsions électriques, leur durée, leur phase active...etc. Ceci donne une souplesse dans la représentation de l'information extraite du signal à l'aide d'une séquence d'excitation tout en s'adaptant à la pathologie du patient. Le neurostimulateur offre également un grand nombre de possibilités de combinaisons de paires d'électrodes pour former des canaux de stimulation qui ne seront nécessairement pas tous utilisables. En effet, cela dépendra de la pathologie du patient ainsi que du degré d'insertion du réseau d'électrodes dans la cochlée ce qui décidera des coordonnées spatiales de chaque électrode.

### 5.1.2 Dictionnaire des séquences d'impulsions

Le concept de l'implant de Sherbrooke permet de générer diverses formes d'onde d'impulsions électriques à l'aide d'une suite de mots de commande ayant des arguments et une succession conformes aux paramètres déterminés par l'analyse du signal. Puisque les résultats de l'analyse dans le cas de la quantification vectorielle sont confinés dans le contenu du dictionnaire, il sera avantageux, de point de vue exécution, d'utiliser un dictionnaire de séquences de stimulation à l'image de ce dernier (figure 5.2). Ainsi, chaque séquence de ce dictionnaire correspondrait à un élément du dictionnaire des filtres ayant servis à l'analyse du signal. Ce dictionnaire de séquences sera alors stocké dans une mémoire morte programmable et sera éventuellement modifié selon tout changement apporté au dictionnaire des filtres.

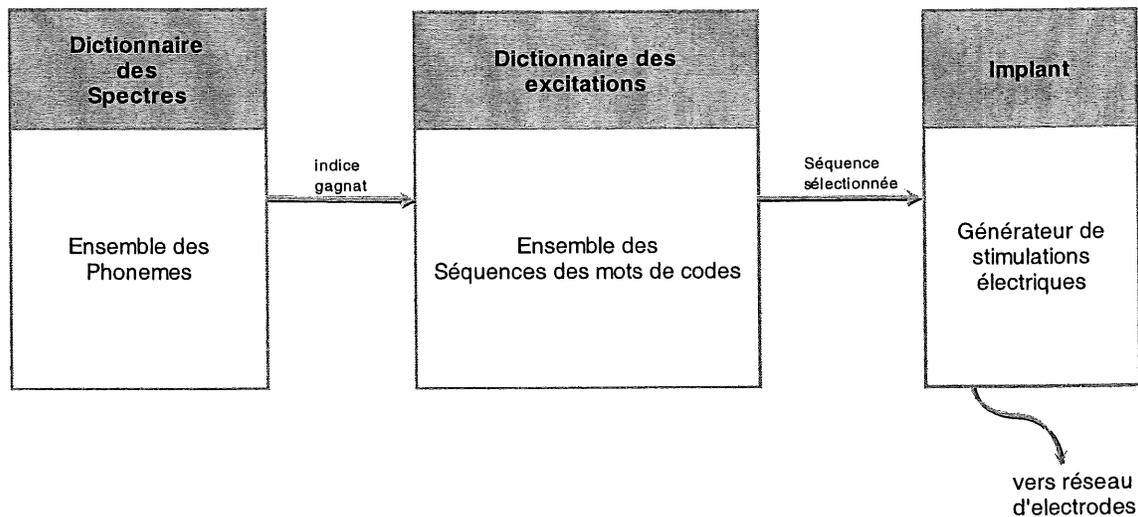


Figure 5.2 Processus de génération des impulsions à l'aide d'un dictionnaire de séquences de stimulation

Il est à noter que les séquences d'impulsions enregistrées en mémoire ne doivent pas nécessairement se conformer à un modèle psychoacoustique quelconque. Elles pourraient, à la limite, provenir de tests empiriques et relever de la préférence personnelle du patient. D'une façon plus réaliste, on peut facilement implémenter dans la mémoire morte des séquences d'impulsions conformes à des modèles psychoacoustiques différents, qui pourraient être sélectionnées selon la préférence du patient et les conditions externes où il se retrouve.

L'utilisation de dictionnaires permettra d'adapter facilement l'appareil à différentes langues et même aux particularités linguistiques régionales de chaque patient. Il suffira simplement de changer leurs contenus selon les besoins.

## **5.2 Survol des stratégies de stimulation utilisées par les systèmes actuels**

Tel que présenté au chapitre 2, les algorithmes de stimulation utilisés par les systèmes actuels sont généralement basés sur deux approches principales établies depuis les premières expériences effectuées dans le domaine. La première approche consiste à extraire les caractéristiques de la parole considérées essentielles à sa compréhension (fréquence fondamentale, un ou deux formants) et à les représenter selon la tonotopie de la membrane basilaire. Cette approche met l'accent sur l'aspect fréquentiel du signal. La deuxième approche est un traitement large-bande du signal de la parole et consiste à le décomposer en différents signaux, en subdivisant sa bande à l'aide d'une série de filtres passe-bande, puis en présentant chacun des signaux obtenus aux régions respectives sur la membrane basilaire. Cette approche met l'accent sur les détails temporels du signal.

L'utilisation de l'une ou l'autre de ces approches a permis d'atteindre un certain niveau de perception et chacune d'elles démontre des avantages et des inconvénients. La technique d'extraction des caractéristiques a démontré des meilleures performances dans l'identification des voyelles alors que le traitement large-bande a donné des meilleurs résultats dans la discrimination des consonnes. D'un autre côté, beaucoup de spécialistes s'accordent sur le fait que la technique d'extraction des caractéristiques enlève l'aspect naturel du signal acoustique et démontre une faible immunité aux bruits environnants. En plus, il a été démontré que les résultats obtenus avec cette technique sont très sensibles aux variations de l'emplacement des électrodes le long de la membrane basilaire. Du côté de la technique du traitement large-bande, celle-ci a démontré une corrélation directe entre l'augmentation du taux de stimulation et ses performances de discrimination.

De nos jours, beaucoup d'efforts sont investis pour améliorer les systèmes et étendre leur utilisation, cependant chaque système conserve son approche d'implémentation de l'algorithme de stimulation en essayant de se justifier par les avantages relatifs à l'approche. Ainsi, ceux qui préconisent la technique d'extraction des caractéristiques insistent sur l'importance de la résolution fréquentielle dans la compréhension de la parole et vantent l'avantage d'utiliser des rythmes de stimulation bas, ce qui permet de diminuer la consommation énergétique de l'appareil et de réduire l'interférence entre les canaux en donnant accès à un plus grand nombre de canaux de stimulation. Les travaux actuels de ces équipes sont généralement concentrés sur la façon d'améliorer l'immunité de leurs systèmes au bruit puisque c'est le facteur essentiel qui affecte significativement leurs performances. Quant à ceux qui utilisent le traitement large-bande, ils mettent l'accent sur l'importance de la résolution temporelle et alors leurs travaux de recherche se

concentrent sur la façon d'augmenter le taux de stimulation pour avoir une plus grande résolution.

Le concept du système de Sherbrooke n'est pas basé sur l'une ou l'autre de ces approches. Il a été conçu d'une façon modulaire complètement indépendante de l'algorithme de stimulation. Ainsi, il est capable d'émuler n'importe quel système en adoptant l'une ou l'autre des approches et il donne accès à des nouvelles approches grâce à sa flexibilité illimitée. De plus, l'algorithme de traitement de signal qui servira à extraire l'information à représenter dans l'oreille interne est complètement indépendant de la stratégie de stimulation qui servira à la représenter.

### **5.3 Stratégies de stimulation proposées**

On a déjà mentionné que les performances modestes obtenues par les systèmes actuels sont dues à leurs concepts étroitement liés à un algorithme de stimulation spécifique. Les travaux poursuivis par notre groupe visent le développement d'un système très flexible et complètement programmable qui sera transparent aux algorithmes de stimulation. L'idée du concept modulaire du système qui permet de rendre indépendantes les deux parties de l'algorithme de stimulation donnera des atouts de plus aux cliniciens pour trouver la meilleure façon d'extraire l'information pertinente du signal et la façon adéquate de la représenter dans l'oreille interne selon la pathologie du patient et les conditions externes où il peut se trouver. En fin de compte, on vise à offrir aux cliniciens l'outil de recherche puissant qui permettra d'améliorer les performances de ce type d'appareil qui semble stagner depuis longtemps. Dans ce mémoire, on a présenté une nouvelle technique de traitement de signal qui se rajoutera à ceux qui ont déjà été développées pour le système.

Cependant, tel qu'on a déjà mentionné, on ne peut affirmer l'efficacité des stratégies de stimulation sans expériences cliniques. Par conséquent, on proposera dans ce qui suit des stratégies qui peuvent être utilisées selon les deux approches préconisées par les systèmes actuels et on ajoutera une qui ne peut être mise en oeuvre que par notre système pour démontrer son potentiel. Il faut cependant rappeler que l'avantage de la technique développée dans ce projet demeure au niveau de traitement de signal en permettant une meilleure extraction de l'information à coder.

### 5.3.1 L'approche d'extraction des caractéristiques de la parole

Dans cette approche, les paramètres pertinents à extraire du signal sont les fréquences caractéristiques de la parole à savoir la fréquence du pitch  $F_0$  et les formants. Pour une meilleure description des voyelles, l'extraction de 4 formants  $F_1, F_2, F_3$  et  $F_4$  est suggérée. Quatre canaux de stimulation représenteront alors ces formants et déclencheront des trains d'impulsions à une fréquence de stimulation égale à la fréquence du pitch, en utilisant les amplitudes des formants pour fixer l'amplitude du courant. Les canaux de stimulation qui seront activés sont ceux dont l'emplacement sur la membrane basilaire correspondrait aux fréquences des formants selon la répartition tonotopique de la membrane. D'autres variantes peuvent être préconisées en utilisant des filtres pour identifier les formants et utiliser l'énergie de sortie de chaque filtre pour l'amplitude de stimulation.

### 5.3.2 L'approche du traitement large bande

Dans cette approche, une fois que le son a été identifié, sa bande fréquentielle sera divisée en quatre sous-bandes par exemple, couvrant chacune l'emplacement d'un

formant. Le signal récupéré de chaque sous-bande peut alors servir pour moduler l'amplitude d'un train d'impulsions de fréquence égale à celle du pitch ou proportionnelle à la fréquence centrale de la sous-bande. D'autres variantes peuvent être préconisées en utilisant  $N$  bandes. Les stimulations peuvent être entrelacées pour éviter l'interférence entre les canaux et la fréquence de stimulation peut dépendre de la nature voisée ou non-voisée du segment de parole.

### 5.3.3 L'approche mixte

L'approche mixte est une combinaison des deux approches précédentes. Elle est inspirée du fait que l'encodage de l'information au niveau de la cochlée peut être différent selon sa bande de fréquence. Les signaux en bas de 400 Hz et ceux de plus que 4 KHz seraient encodés à l'aide d'un traitement large bande, et ceux entre 400 Hz et 4 KHz seraient encodés par les caractéristiques du segment (formants). L'approche mixte propose alors de subdiviser l'information transmise en trois bandes. Chacune sera transmise selon le mécanisme d'encodage correspondant. Les signaux aux fréquences inférieures à 400 Hz seront dès lors synchronisés à cette fréquence puisque l'information que porte le pitch sera codée temporellement. Pour les sons voisés, les informations portées par tout formant situé dans la bande 400 Hz – 4 KHz seront transmises à la cochlée via des électrodes placées aux sites correspondants aux fréquences des formants sur la cochlée. Les stimulations sur ces électrodes se dérouleront à la fréquence fondamentale ou à des fréquences proportionnelles à celles des formants respectifs. Quant à la bande de plus de 4 KHz, la fréquence de stimulation peut être égale à la fréquence du pitch ou à une fréquence plus élevée selon la dynamique du patient.

## 5.4 Conclusion

Les stratégies de stimulations proposées dans ce chapitre sont loin d'être une présentation exhaustive de ce que le neurostimulateur de Sherbrooke peut offrir. On a proposé trois approches différentes permettant de développer des stratégies de stimulation afin de mettre en œuvre au complet un algorithme de stimulation de la cochlée basé sur le traitement de parole par quantification vectorielle. Il est évident que plusieurs autres stratégies qui seraient issues d'autres approches peuvent être envisagées. La proposition d'un dictionnaire de stimulation comme interface entre le traitement de signal par quantification vectorielle et les stratégies de stimulation pourrait s'étendre à d'autres algorithmes de traitement de signal. La mémoire morte programmable pourrait contenir plusieurs dictionnaires de séquences d'impulsions correspondants aux traductions diverses des résultats des algorithmes de traitement de signal. Il deviendra ainsi facile de sélectionner les stratégies de stimulation par simple choix de la partie de la mémoire de masse dédiée à tel ou tel algorithme. Cette extension ajouterait davantage de flexibilité et de versatilité.

## CONCLUSION

Ce travail a fait l'objet d'un nouvel algorithme de stimulation destiné aux prothèses cochléaires. L'algorithme se base sur la technique de quantification vectorielle pour le traitement des sons et tire profit de ses avantages pour apporter des améliorations aux performances des prothèses cochléaires.

Dans une première partie du travail, on a survolé l'état de l'art des systèmes présents et des algorithmes développés. On a montré que la difficulté d'arriver à une modélisation intégrale et complète du mécanisme d'audition a limité les stratégies à une représentation partielle de l'information sonore. La réhabilitation des patients est par conséquent plus longue et la qualité de perception est limitée et souffre d'une mauvaise immunité au bruit.

La méthode proposée dans ce mémoire vise, entre autres, à apporter des solutions à ces problèmes. En confinant les résultats d'analyse de la parole dans un nombre limité de solutions, on réduit d'une part le nombre de séquences de stimulation auxquelles le patient est exposé, et par conséquent la période de réhabilitation, et on améliore, d'autre part la perception en milieu bruité. L'utilisation d'un dictionnaire de spectres comme base de connaissance confère au système davantage de flexibilité. En fait, l'information sur les phonèmes qu'il contient est facilement modifiable, particulièrement selon l'évolution de l'apprentissage du patient. Le psychoacousticien pourra changer cette information suivant le progrès de son patient. Ainsi, le dictionnaire pourrait contenir des

sons élémentaires de l'environnement dans les premières phases de réhabilitation pour contenir de plus en plus de phonèmes de la langue dans des stades avancés.

En deuxième partie, on a décrit les étapes de développement et de mise en œuvre du traitement de la parole par quantification vectorielle. On a tout d'abord présenté le fondement théorique du codage prédictif linéaire par quantification vectorielle, ensuite on a exposé le processus de sa mise en œuvre pour enfin proposer des stratégies de stimulation pouvant s'apparier à cette technique afin de développer des algorithmes de stimulation complets. La génération des stratégies de stimulation est elle aussi basée sur un dictionnaire de séquences d'impulsions. De cette façon on peut étendre la flexibilité du système. Les séquences enregistrées sont au même titre que les spectres, modifiables et peuvent relever de modèles psychoacoustiques déterminés, ou de modèles empiriques et des résultats d'expérimentation déterminés selon la préférence du patient et ses particularités linguistiques.

L'approche de traitement par quantification vectorielle s'avère prometteuse. On pense que l'essai clinique sur un patient d'un algorithme de stimulation basé sur cette approche fera preuve des avantages énumérés. La puissance, la flexibilité et la transparence de l'analyseur de parole de Sherbrooke définiront une plate-forme idéale pour un tel essai. Avec la collaboration des linguistes, des phonéticiens et des psychoacousticiens, on doit être en mesure de mettre au point les paramètres de l'algorithme pour assurer une meilleure adaptation aux préférences des patients. On pourra en plus analyser les rapports entre les résultats phonétiques et les paramètres acoustiques essentiels à la stimulation électrique de la cochlée.

Des améliorations futures peuvent être envisagées. Les nouvelles recherches sur la quantification vectorielle et le codage de la parole révèlent de nouvelles techniques plus performantes en matière de débit de transmission de données, d'économie en temps de calcul et en capacité mémoire. Ces techniques ouvrent la porte à des nouvelles investigations pour aboutir à des nouveaux algorithmes. La quantification vectorielle algébrique, en particulier, s'avère très prometteuse notamment dans le cas de dictionnaire de grande taille. On pourra, dès lors, penser à l'élaboration d'une variante de l'algorithme proposé qui sera dédiée aux patients déjà réhabilités. La nouvelle variante utilisera ainsi un dictionnaire plus riche exposant le patient à de plus larges particularités linguistiques.

## BIBLIOGRAPHIE

- [1] ADOUL, J. -P. (1987) *Speech coding algorithms and vector quantization in advanced digital communications and signal processing*, K. Feher, Ed. Englewood Cliffs: NJ, Prentice Hall, pp. 133-181.
- [2] AY, S. U., ZENG, F., SHEU, B. J. (May 1997) *Hearing with Bionic Ears, Speech Processing Strategies for Cochlear Implant Devices*, IEEE Circuits & Devices.
- [3] ADOUL, J. -P. (1986) *La quantification vectorielle des signaux: approche algébrique*, *Annales des Télécommunications*, tome 41, no. 3-4, pp. 158-177.
- [4] BEAR, M.F., CONNORS, B.W., PARADISO, M.A. (1997) *Neurosciences: À la découverte du cerveau*, Editions Pradel, Paris, pp 272-307.
- [5] BERGER, T. (1971) *Rate-distorsion theory*, Prentice-Hall, New-Jersey.
- [6] BOITE, R., KUNT, M. (1987), *Traitement de la parole*, presses polytechniques romandes, Lausanne.
- [7] CHOUARD, C. H., OUAYOUN, M., MEYER, B., FUGAIN, C. (1995) *Speech Coding Strategies of Digisonic Fully Digitized Cochlear Implant*, ISSN 0001-6489, Scandinavian University Press.
- [8] CHTOUROU, Z. (1999) *Conception et développement de stratégies multi-rythmes de simulation basées sur un traitement multi-résolution du son dédiées à une prothèse cochléaire*, Thèse de doctorat, Université de Sherbrooke.
- [9] COMBI 40/40+ Cochlear Implant, [www.medel.com/international/index\\_int.html](http://www.medel.com/international/index_int.html).
- [10] CLARION *Multi-Strategy Cochlear Implant System*, DALLAS Otolaryngology, Cochlear Implant Program, [www.dallascochlear.com/clarion.htm](http://www.dallascochlear.com/clarion.htm).
- [11] DELLER, J. R., HANSON, J. H. L., PROAKIS, J. G. (1993) *Discrete-time processing of speech signals*, IEEE Press Classic Reissue.
- [12] EDDINGTON D.K., TIERNEY, J., ZISSMAN, M.A. (1994) *Digital Signal Processing Applications in Cochlear Implant Research*, The Lincoln Laboratory Journal, Vol. 7, No. 1.

- [13] FONTAINE, R., ALARY, A., MOUINE, J., DUVAL, F. (1995) *A Programmable Speech Processor For Profoundly Deaf People*, Engineering in Medicine and Biology Society, IEEE 17<sup>th</sup> Annual Conference, Vol. 2, pp. 1617-1618.
- [14] FONTAINE, R., POURMEHDI, S., MOUINE, J., DUVAL, F., WEI, M. (1995) *Cochlear Implant Stimulation Based On Vector Quantization*, IEEE trans.
- [15] FONTAINE, R., (1999) *Conception et développement d'un analyseur de parole multialgorithmes pour implant cochléaire*, Thèse de doctorat, Université de Sherbrooke.
- [16] FURUI, S. (1989) *Digital speech processing, synthesis, and recognition*, Electrical engineering and electronics, Marcel Dekker, INC, New York and Basel.
- [17] FURUI, S., SONDI, M.M. (1992) *Advances in Speech Signal Processing*, Marcel Dekker inc, New York, Basel, Hong Kong, pp 269-305.
- [18] GADHOUMI, K., MOUINE, J., CHTOUROU, Z. (July 2000) *A Cochlear Prosthesis Stimulation Algorithm Targeting To Minimize The Rehabilitation Phase and Enhance Speech Comprehension*, Engineering in Medicine and Biology Society, Proceedings of the 22<sup>nd</sup> Annual International Conference of the IEEE, Vol. 1, pp. 798-800.
- [19] GERSHO, A., GRAY, R. M. (1992) *Vector quantization and signal compression*, Kluwer Academic Publishers, Massachussets.
- [20] GRAY, R. M. (April 1984), Vector quantization, IEEE ASSP Mag., vol. 1, pp 4-29.
- [21] GRAY, A.H. JR., MARKEL, J.D. (October 1976) *Distance Measures For Speech Processing*, IEEE transactions on ASSP, vol. ASSP-24, No5.
- [22] HOUSE, W. F. (1995) *AllHear Cochlear Implants, chapter 1*, David House Editions, [www.serve.com/AllHear/Monographs/m-96-h01.html](http://www.serve.com/AllHear/Monographs/m-96-h01.html).
- [23] HOUSE, W. F. (1995) *Cochlear implants: My perspective*, AllHear inc.
- [24] ITAKURA, F. (February 1975) *Minimum Prediction Residual Principle Applied to Speech Recognition*, IEEE trans. Acoust., Speech, Signal Processing, vol. ASSP-23, pp 67-72.
- [25] JAVEL, E., MOTT, J.B. (April 1988) *Physiological and Psychophysiological correlates of temporal processes in hearing*, Elsevier Science Publishers B.V., pp. 275-294.
- [26] LAZORTHE, G.(1983) *Le système nerveux central*, Masson, Paris, 3<sup>ème</sup> édition, pp 32-54.

- [27] LOIZOU, P.C. (1997), *Signal Processing for cochlear prosthesis: A Tutorial Review*, IEEE MWSCAS.
- [28] LOOKABAUGH, T., GRAY, R. M. (April 1987) *High-resolution quantization theory and vector quantization advantage*, IEEE Trans. Inform. Theory, vol. IT-35, pp. 1020-1033.
- [29] MARKEL, J. D, GRAY, A. H. JR., WAKITA, H. (1973) *Linear prediction of speech theory and practice*, speech communications research laboratory, SCRT monograph, no. 10.
- [30] McKAY, C. M., VANDALI, A. E., McDermott, H. Clark, G.M. (1994) *Speech processing for multichannel cochlear implants: variations of the Spectral Maxima Sound Processor Strategy*, Acta Otolary., vol. 114, pp. 52-58.
- [31] MOUNTANDON, P., PELLIZONE, M., KASPER, A. (1991) *L'implant intracochléaire multicanal "Ineraid"*, Revue de Laryngol., vol. 112, no. 4, pp. 339-342.
- [32] MOUINE, J. (1992) *Conception et développement d'un neurostimulateur implantable dédié à une prothèse cochléaire à 16 canaux*, Thèse de doctorat, Université de Sherbrooke.
- [33] NADEAU-DOSTIE, B. (1985) *Design et évaluation d'un stimulateur nerveux multi-canal destiné à une prothèse auditive*, Thèse de doctorat, Université de Sherbrooke.
- [34] NUCLEUS 24 Channel Cochlear Implant System, DALLAS Otolaryngology, Cochlear Implant Program, [www.dallascochlear.com/nucleus.htm](http://www.dallascochlear.com/nucleus.htm).
- [35] OUAYOUN, M., PEAN, V., GENIN, J., BACHELOT, G., FUGAIN, C., MEYER, B., CHOUARD, C. (1997) *Asynchronous interleaved stimulation (AIS): A new speech coding strategy for cochlear implants*, Acta Otolaryngol., vol. 117, pp. 182-186.
- [36] PEETERS, S., MARQUET, J., OFFECIERS, F.E., BOSIERS, W., KINSBERGEN, J., VAN DURME, M. (1989) *Cochlear Implants: The Laura prothesis*, Journal of Medical Engineering & Technology, vol. 13, no. 1/2, pp 76 - 80.
- [37] PEPIN, G. (1984) *Codage numérique de la parole utilisant les TDHS et la quantification vectorielle sphérique*, Thèse de doctorat, Université de Sherbrooke.
- [38] PIJL, S. (1994) *Audiological results with the MSP/MPEAK and WSP/FOF1F2 processors and coding strategies for the Nucleus Cochlear Implant*, The Journal of Otolaryngol., vol. 23, no. 4, pp. 286-291.
- [39] POIRIER, D. (1991) *Neurophysiologie applicable aux prothèses auditives implantables*, Rapport interne, Université de Sherbrooke.

- [40] POURMEHDI, S. (Mars 1989) *Codage acoustique destiné à une prothèse auditive intra-cochléaire*, Mémoire de maîtrise, Université de Sherbrooke.
- [41] POURMEHDI, S. (Mai 1993) *Conception et développement d'un système de traitement de la parole basé sur la quantification vectorielle dédié à une prothèse auditive*, Thèse de doctorat, Université de Sherbrooke.
- [42] SPENCE, A. P., MASON, E.B. (1983) *Anatomie et physiologie, une approche intégrée*, Éditions du renouveau pédagogique inc., Ottawa, 855 pages.
- [43] SUMMERS, I.R., GRATTON, D.A. (March 1995) *Choice of Speech Features for Tactical Presentation to the Profoundly Deaf*, IEEE trans. on rehabilitation engineering, Vol. 3, No. 1.
- [44] TORTORA, G.J., ANAGNOSTAKES, N.P. (1988) *Principes d'anatomie et de physiologie*, Centre Éducatif et Culturel inc, Montréal, pp 270-347.
- [45] VAN DER BORGHT, G., PEETERS, S., BOUCHATAOUI, I., SNELLINGS, J., *LAURA FLEX : a new speech processor for the Laura cochlear implant*, [www.phi.be](http://www.phi.be).
- [46] WILSON, B.S., LAWSON, D.T., ZERBI, M., FINLEY, C.C., WOLFORD, R.D. (September 1995) *New Processing Strategies in Cochlear Implantation*, The American Journal Of Otology, Vol. 16, No. 5. pp. 669-675.
- [47] WILSON, B.S., LAWSON, D.T., ZERBI, M. (1994) *Recent development with CIS strategies*, Advances in cochlear implants, edited by Hochmair-Desoyer, I. J. and Hochmair, E. S., International Inter-Science Series, Wein, pp. 103-112.
- [48] WILSON, B.S., LAWSON, D.T., FINLEY, C.C. (1993) *New Processing Strategies for Multi-channel Cochlear Protheses*, Progress in Brain Research, Vol. 97, Elsevier Science Publishers B.V.
- [49] WILSON, B.S., LAWSON, D.T., FINLEY, C.C., WOLFORD, R.D. (September 1988) *Speech Processors for Cochlear Implants*, IEEE proc., Vol. 76, No 9.
- [50] WONG, D. Y., JUANG, B. -H., GRAY, A. H. JR. (October 1982) *An 800 bit/s Vector quantization LPC vocoder*, IEEE trans. Acoust., Speech, Signal Processing, vol. ASSP-30, pp 770-780.