République Algérienne Démocratique Et Populaire

Ministère De L'enseignement Supérieur Et De La Recherche Scientifique

Université Yahia Fares de Medea

Faculté De La Technologie

Département De Génie Mécanique

MEMOIRE DE FIN D'ETUDES

EN VUE DE L'OBTENTION DE DIPLÔME DE MASTER EN GÉNIE MÉCANIQUE

Option : MECATRONIQUE

Thème

Traitement vocale d'un robot manipulateur

Préparé et présenté par
- MADANE Fahima.
- KHELIFA Khadidja.

Proposé et dirigé par
- . ABUDURA Salem.
- Dr. BELHADJ Abd-El-Moneim.

Année Universitaire 2016-2017

REMERCIMENT

Merci Allah (mon dieu) de nous avoir donné la capacité d'écrire et de réfléchir, la force d'y croire, la patience d'aller jusqu'au bout du rêve et le bonheur de lever nos mains vers le ciel et de dire " Ya Kayoum ".

Tout d'abord, ce travail ne serait pas aussi riche et n'aurait pas pu avoir le jour sans l'aide et l'encadrement de Monsieur Abudhura Salem , on le remercie pour la qualité de son encadrement exceptionnel, pour sa patience, sa rigueur et sa disponibilité durant notre préparation de ce projet.

Nos remerciements s'adresse aussi aux membres du jury Belhadj, Ghemana et Bachen qui on accepter d'examiner notre travail.

A tous nos enseignants et les membres du section mécatronique de L'université YAHIA FARES.

Nos profonds remerciements vont également à toutes les personnes qui nous ont aidés et soutenue de prés ou de loin.

Dédicaces

Je dédie ce modeste travail :

A l'homme de ma vie, mon exemple éternel, mon soutien moral et source de joie et de bonheur, celui qui s'est toujours sacrifié pour me voir réussir, que dieu vous garde et te protège, à toi mon ***père****.*

A la lumière de mes jours, la source de mes efforts, la flamme de mon cœur, ma vie et mon bonheur ***maman*** *que j'adore.*

Mon ange gardien et mon fidèle accompagnant dans les moments les plus délicats de cette vie mystérieus mes frère : ***Sidali,Yahia ,Hacen et Hocin.***

En témoignage de l'attachement, de l'amour et de l'affection que je porte pour vous mes sœurs : ***Khadidja, Nour el houda et Hadjer et Abir***

A Mes chers amis qui sans leur encouragement ce travail

A mes collègues ***Nabil, Moussa, Youcef*** *et* ***Ahmed***

A toute ma famille ***MADEN et SOUMATI***

A tous ceux que j'aime.

FAHIMA

Dédicaces

Je dédie ce modeste travail :

A l'homme de ma vie, mon exemple éternel, mon soutien moral et source de joie et de bonheur, celui qui s'est toujours sacrifié pour me voir réussir, que dieu vous garde et te protège, à toi mon ***père****.*

A la lumière de mes jours, la source de mes efforts, la flamme de mon cœur, ma vie et mon bonheur ***maman*** *que j'adore.*

Mon ange gardien et mon fidèle accompagnant dans les moments les plus délicats de cette vie mystérieus mes frère : ***Morad,Amin*** *et* ***Wassim****.*

En témoignage de l'attachement, de l'amour et de l'affection que je porte pour vous mes sœurs : ***FAHIMA*** *et* ***NOUR EL HOUDA****.*

A Mes chers amis qui sans leur encouragement ce travail

A mes collègues ***Nabil, Moussa, Youcef*** *et* ***Ahmed***

A toute ma famille ***KHELIFA***

A tous ceux que j'aime.

KHADIDJA

Liste des Abréviations et Nomenclatures

R.U.R	**Rossum's Universal Robot**
PUMA	**Programmable Universal Machine for Assembly**
SMA	**Structure Mécanique Articulé**
TTS	**Text-To Speech**
OCR	**Optical Character Reconition**
LPC	**Linear Predictive Coding**
MFCC	**Mel Frequency Castrum Coefficient**
MMC	**Modèle de Markov Cachés**
HMM	**Hidden markov models**

Table des matières

LISTE DES FIGURES……I
LISTE DE TABLEAU……...II
LISTE DES ABREVIATION……III
Introduction générale……...1
CHAPITRE I : généralité sur les robots
I.1. Introduction 3
I.2. Définition générale 3
I.2.1. Le robot industriel 3
I.2.2. La robotique 3
I.3. L'histoire de la robotique industrielle 4
I.4. Types de robots 5
I.4.1. Robots mobiles 5
I.4.2. Robots manipulateurs 6
I.5. Les éléments constitutifs d'un robot 7
I.5.1. Unité informationnelle 7
I.5.2. Unité opérationnelle 7
I.6. La structure mécanique articulée 8
I.6.1. Structure mécanique articulée à chaîne cinématique simple 8
I.6.2. Structure mécanique articulée à chaîne cinématique fermée 9
I.7. Architecture des robots 10
I.8. Classification des robots 13
I.8.1. Classification fonctionnelle 13
I.8.1.a. Manipulateurs à commande manuelle ou télécommande 13
I.8.1.b. Manipulateurs automatiques à cycles préréglés 14
I.8.1.c. Robots programmables 14
I.8.1.d. Robots intelligents 15
I.8.2. Classification géométrique 16
I.9. Caractéristiques d'un robot 19
I.10. Domaines d'applications 20
I .11. Avenir de la robotique 22
I .12. Conclusion 22
Chapitre II :**Le traitement et la synthèse de Parole**
II.1. Introduction 24

II.2.Le traitement de la parole 24
II.2.1.Mécanisme de la phonation 26
II.2.1.1. Phonation de sons voisés 26
II.2.1.2. Phonation de sons non voisés 28
II.2.2. Mécanisme de l'audition 29
II.2.2. Mécanisme de l'audition 29
II.3. Le niveau acoustique de la parole 33
II.3.1. Audiogramme 33
II.3.2.Transformée de Fourier à court terme 36
II.3.3. Spectrogramme 37
II.3.4. Fréquence fondamentale 39
II.4. Synthèse de parole 39
II.4.1 Applications 40
II.4.2. Le modèle prédictif Linéaire du signal vocal 42
II.5. Codage LPC 44
II.6. Conclusion 46
Chapitre III : **La Reconnaissance De La Parole**
III.1. Introduction 47
III.2. Le traitement automatique de la parole 47
III.2.1. Niveaux de complexité 48
III.2.2. Extraction des paramètres 49
III.2.2.1. Acquisition et préaccentuation du signal 50
III.2.2.2. Segmentation et Fenêtrage 51
III.2.2.3. Transformée de Fourier discrète et l'énergie du signal 53
III.2.2.4. l'échelle Mel 54
III.2.3. La reconnaissance 55
III.3. Modélisation acoustique à base de HMM 58
III.4. Application des HMM à la reconnaissance des mots isolés 59
III.4.1. Reconnaissance de mots isolés en nombre limité (<100 mots) 60
III.5. Application et Résultats 61
III.5.1. Reconnaissance en temps réel 62
III.5.2. Taux de reconnaissance 63
III.6. Conclusion 64
Chapitre IV :**Commande vocale d'un bras manipulateur**
IV.1. Introduction 65

IV.2. Architecture générale de la commande vocale 66
IV.3. le model de robot RRR utilisé 67
IV.3.1. Configuration des articulations 67
IV.3.2. Le choix des trajectoires 69
IV.3.2.1. Les mots Gauche et Droite 69
IV.3.2.2. Les mots Haut et Bas 70
IV.3.2.3. Les mots Avant et Arrière 71
IV.4. Généralité sur la commande en mode glissant 71
IV.5.Conclusion 75
Conclusion Générale 76
Références bibliographique 77

LISTE DES FIGURES

CHAPITRE I : généralité sur les robots

Figure I.1. Robotique industrielle (unimate, puma)……..5
Figure I.2. Spirit, NASA, 2003 sur Mars……..6
Figure I.3. Robot fanuc……7
Figure I.4. Structure fonctionnelle d'un robot……..8
Figure I.5. Structure mécanique à chaines ouvertes simples……...10
Figure I.6. Structure mécanique à chaines fermées……10
Figure I.7. Architecture d'un robot……..11
Figure I.8. Parties principales dans un robot……...11
Figure I.9. Représentation d'une articulation rotoïde……..12
Figure I.10. Représentation d'une articulation prismatique……13
Figure I.11. Manipulateur à commande manuelle……...15
Figure I.12. Manipulateur à cycle pré-réglé……15
Figure I.13. Robot programmable……16
Figure I.14. Robot intelligent……...17
Figure I.15. Robot cartésien……....18
Figure I.16. Robot cylindrique……18
Figure I.17. Robot sphérique……...19
Figure I.18. Robot SCARA……19
Figure I.19. Robot 3R……..20
Figure I.20. Robot soudeurs Par points……22
Figure I.21. Robot soudeurs a l'arc……..22
Figure I.22. Robot pompiste……22
Figure I.23. Robot de construction……..22
Figure I.24. robot Computer motion……23
Figure I.25. robot Assistance aux personnes Handicapées……..24

Chapitre II :**Le traitement et la synthèse de Parole**

Figure II.1. Représentation de l'appareil phonatoire……26
Figure II.2. Le spectre du son voisé / i /……...28
Figure II.3. Le spectre du son non voisé / f /……29
Figure II.4. Coupe de l'appareil auditif humain……...30
Figure II.5. Réponse en fréquence d'une cellule ciliée……31
Figure II.6. Le champ auditif humain……...32

Figure II.7. Les courbes d'égale impression de puissance auditive en fonction de la fréquence……...32

Figure II.8. Enregistrement numérique d'un signal acoustique…….....33

Figure II.9. Audiogramme de signaux de parole en haut le mot « gauche» et en bas le mot «bas »……35

Figure II.10. Evolution temporelle (en haut) et transformée de Fourier discrète (en bas) 36

Figure II.11. Spectrogrammes à large bande (en haut), à bande étroite (en bas), et évolution temporelle du mot 'Droite'……38

Figure II.12. Evolution de la fréquence de vibration des cordes vocales dans la phrase "les techniques de traitement numérique de la parole"……39

Figure II.13. Schéma de principe d'un analyseur de parole. En pratique, l'étape De synthèse peut être implicite……42

Figure II .14. Le modèle auto-régressif……...43

Figure II.15. Transmission de parole basée sur le codage LPC……..45

Chapitre III : **La Reconnaissance De La Parole**

Figure III.1. Les étapes d'une paramétrisation MFCC……49

Figure III.2. Détecteur appliqué à la phrase « marche avant »……...50

Figure III.3. La phrase « marche avant » après la suppression du silence……..51

Figure III.4. Fenêtre de Hemming……...52

Figure III.5. Représentation du signal de la phrase « marche avant » après La segmentation Et le fenêtrage……...52

Figure III.6. Représentation du signal du mot «droite » après la segmentation et Le fenêtrage……53

Figure III.7. Echelle de Mel……55

Figure III.8. Type de banc filtre à échelle de Mel utilisant des filtres triangulaires……...55

Figure III.9. Le principe de fonctionnement de l'approche Statistique pour la reconnaissance de la parole……57

Figure III.10. Exemple d'un HMM à cinq états……..60

Figure III.11. Architecture de reconnaissance des mots isolés……...62

Figure III.12. Procédure de la reconnaissance en temps réel……..63

Chapitre IV :**Commande vocale d'un bras manipulateur**

Figure IV.1. L'architecture générale de la commande vocale…….66

Figure IV.2. Architecture de bras manipulateur RRR…….67

Figure IV.3. Projection de bras selon xy……67

Figure IV.4. projection de bras selon yz……..68

Figure IV.5. les deux postures des robots : à gauche (coude bas) et à droite (coude haut)….69

Figure IV.6. Rotation autour de l'axe Z avec l'angle θ_1…….70

Figure IV.7. Déplacement verticale…….70

Figure IV.8. Déplacement horizontal……..71

Figure IV.9. Plan de phase et la fonction de commutation de la commande par régime Glissant……..73

Figure IV.10. la fonction de commutation de la commande par régime glissant à bande limite…….74

LISTE DES TABLEAUX

Tableau III.1. Taux de reconnaissance…….....63

Résumés :

Dans les milieux hostiles, inaccessibles par l'homme, tels les milieux sous marins à profondeur importante, les milieux toxiques etc., la commande à distance des robots constitue un outil d'une importance capitale. Nous proposons dans le cadre de ce travail d'examiner une commande vocale d'un opérateur. Cette commande doit être exécutée par le manipulateur par une action de déplacement dans les trois axes. Les étapes de ce travail peuvent être résumées comme suit :

- Développement d'un système d'acquisition de la parole. Ce système devrait permettre d'enregistrer des commandes de type : HAUT, BAS, GAUCHE, DROITE, AVANT, ARRIERE etc. Il devrait être ouvert pour permettre l'enrichissement de la base des commandes avec de nouvelles commandes au fur et à mesure.
- Développement d'un module d'extraction des caractéristiques de chaque commande, Les caractéristiques les plus discriminantes sont à retenir comme sortie de ce module pour garantir une séparation claire entre les différentes commandes.

Mots-clés :

Signal de la parole, reconnaissance vocale, bras manipulateur.

Abstract:

In hostile environments inaccessible by human, such as marine significant depth, toxic environments etc.., the remote control of the robots constitute a tool of capital importance. We propose in this work to examine a voice command of an operator. This command must be executed by the manipulator by an action of displacement in all three axes. The stages of this work can be summarized as follows.

- Developpement of speech acquisition system. This system should make it possible to record orders of the type: HAUT, BAS, GAUCHE, DROITE, AVANT, ARRIERE etc. It should be open to allow the enrichment of the base with new orders.
- Developpement of a module of extraction of the characteristics of every command. The more discriminant features are to remember as exit of this module to ensure a clear separation between the various commands.

Keywords:

Speech signal, speech recognition, the manipulator arm.

الملخص:

في البيئات الوعرة ، لا يمكن الوصول إليها من قبل البشر ، مثل بيئات تحت الماء مع أعماق كبيرة والبيئات السامة وغيرها ، والتحكم عن بعد من الروبوتات هو أداه ذات أهمية قصوى. وفي سياق هذا العمل نقترح فحص الأمر الصوتي للمشغل. يجب تنفيذ هذا الأمر من قبل المتلاعب بواسطة إجراء نقل في المحاور الثلاثة. ويمكن تلخيص الخطوات التي اتخذت في هذا العمل علي النحو التالي :

- ✓ تطوير نظام الحصول علي الكلام. وينبغي أن يسمح هذا النظام تسجيل الأوامر من النوع: اعلي، أسفل، اليسار، اليمين، أمام، وراء.. .الخ. وينبغي فتحها للسماح بإثراء قاعدة الطلبات بأوامر جديدة.
- ✓ تطوير وحدة استخراج من خصائص كل عنصر تحكم، والسمات الأكثر تميزا هي التي ينبغي الاحتفاظ بها كناتج هذه الوحدة النمطية لضمان فصل واضح بين مختلف الضوابط.

الكلمات المفتاحية : إشارة الكلام ، التعرف علي الصوت ، الذراع المناور.

Introduction Générale

INTRODUCTION GÉNÉRALE

Traditionnellement, les robots sont contrôlés avec des dispositifs de saisie manuelle comme joysticks, écrans tactiles, et trackballs…etc., bien que ces appareils sont omniprésents et ont été utilisés dans de nombreuses applications, ils viennent avec leur propre ensemble de problèmes lorsqu'ils sont utilisés pour contrôler des robots dans des contextes opérationnels. Certains robots sont grands, d'autres nécessitent un fonctionnement à partir de positions fixes; beaucoup exigent une dextérité, coordination oeil-main, une formation importante, et le temps de pratique, et tous nécessitent l'utilisation d'au moins une main. A l'opposé, la commande vocale vient fournir de nombreux avantages. La liste suivante en fournit quelques uns :

- Plus efficace que le contrôle manuel pour le menu de navigation (ne nécessite pas la navigation à travers plusieurs couches du menu pour accéder à l'élément souhaité).
- Plus rapide et plus précis que le contrôle manuel pour le choix des options.
- Améliore l'efficacité en temps partagé lorsqu'il est utilisé en conjonction avec des commandes manuelles.
- Les mains et les yeux sont libres.
- Réduit les effets nuisibles de fonctionnement de l'appareil mobile sur l'exécution des tâches voiture, à pied, etc.)
- Efficace pour la performance de tâches simultanées (par exemple, l'abaissement du bras robotique pendant la conduite vers l'avant).
- Intuitive si les commandes sont adaptées au public cible.

La commande vocale a été explorée pour de nombreuses applications. Par exemple, elle a été examiné pour le contrôle des hélicoptères , l'exécution de fonctions de téléphonie et de l'automobile, l'automatisation du traitement des appels des clients, la conversion de la langue parlée en langue des signes, la conversion et l'analyse de gros volumes de documents parlés, l'exécution de commandes et ordinateur requête, traduire ou résumer l'information d'une langue à l'autre, créant des dossiers médicaux et l'édition.

Dans les situations où les opérateurs d'équipement ont les yeux et les mains occupées, le déchargement des tâches de contrôle de la parole a été montré pour être efficace. Afin de commander un robot avec des commandes vocales, le système doit d'abord être en mesure de reconnaître la commande. Un système de Reconnaissance Automatique de la Parole (RAP) est un système de mémorisation et de codage de mots parlés. Une fois identifiée l'information orale peut contrôler les actions d'un système ou d'une machine.

Dans ce travail nous allons traiter une application très importante de la commande vocale: la commande des bras manipulateurs. L'approche générale de ce travail est d'entrainer un bras manipulateur à se déplacer dans les trois axes à l'aide des commandes vocales. Les techniques de traitement du signal de parole seront utilisées pour extraire les paramètres utilisables par les algorithmes de la reconnaissance adoptés par le model de Markov caché (MMC) que nous allons détailler par la suite. Une phase de la commande sera effectuée pour contrôler le déplacement de notre bras manipulateur.

Outre l'introduction et la conclusion générales, ce mémoire est organisé en quatre chapitres, répartis comme suit :

- Le premier chapitre est donné un aperçu sur les robots. Nous définissons le robot industriel et la robotique et nous donnons un bref historique sur la robotique industrielle et les différents types de robots. Ce premier chapitre analyse aussi les différents constituants d'un robot manipulateur qui est l'objet de notre étude et explique ce qu'une structure mécanique articulée (SMA).
- le deuxième chapitre présentera une étude générale sur le traitement automatique et la synthèse de parole.
- Dans le troisième chapitre, après avoir exposé un rappel générale sur la reconnaissance automatique de la parole, nous allons appliquer le model de Markov caché pour la reconnaissance.
- Le quatrième chapitre traite l'application des développements précédents pour la commande vocale d'un bras manipulateur.

CHAPITRE I

Généralités sur Les robots

I.1. Introduction

L'utilisation des systèmes robotiques apparait aujourd'hui dans plusieurs domaines d'activités : la médecine, la défense, la recherche scientifique etc.… Les robots sont utilisés de manière privilégiée pour des missions où les objectifs sont quantifiables et clairement définis. Ils sont destinés à faciliter les tâches pour l'homme et à amplifier le rendement.

Dans ce premier chapitre, on va donner un aperçu non exhaustive sur les robots, un bref historique sur l'évolution de la robotique industrielle et présenter les différents types de robots et les éléments constitutifs de ces derniers.

I.2. Définition générale

I.2.1. Le robot industriel

Le terme robot a été introduit pour la première fois par l'auteur Tchèque Capek en 1920 dans sa pièce de théâtre R.U.R. (Rossum's Universal Robot) et il est dérivé du mot robota qui signifie travailleur (de force). [1]

Un robot est un dispositif mécatronique (alliant mécanique, électronique et informatique) accomplissant automatiquement des tâches diverses. C'est une machine intelligente fonctionnelle qui nécessite une autonomie de mouvements.

L'Organisation Internationale de Normalisation définit le robot comme étant un manipulateur à plusieurs degrés de liberté, à commande automatique, reprogrammable, multi-applications, mobile ou non, destiné à être utilisé dans les applications d'automatisation industrielle. [2]

I.2.2. La robotique

La robotique est une science qui s'intéresse aux robots. En fait, il s'agit d'un ensemble de disciplines techniques (mécanique, électronique, automatique, informatique) articulées autour d'un objectif et d'un objet communs. Cet objectif est l'automatisation flexible de nombreux secteurs de l'activité humaine réputés jusqu'à très récemment comme ne pouvant

se passer de la présence de l'homme, et l'objet est le robot, sorte de machine universelle dont l'homme rêve depuis toujours pour le remplacer dans les tâches difficiles.

I.3. L'histoire de la robotique industrielle

La robotique industrielle a connu un essore entre 1950-1970. Elle a vu le jour en 1954 lorsque Georges DEVOL a pu réaliser son brevet sur la robotique. Dans ce brevet Devol a conçu un robot qu'il a intitulé Unimate.

En 1961, le premier Unimate fut utilisé dans les usines de GENERAL MOTORS.

En 1966, l'entreprise Unimation continue de développer des robots et élaborent notamment des robots permettant de faire d'autres tâches, comme des robots de manipulation matérielle ou encore des robots conçus pour la soudure ou pour d'autres applications de ce genre.

En 1978 un nouveau robot est conçu par Unimation Inc avec l'aide de General Motors. Ensemble ils conçurent le robot PUMA 500. Le robot PUMA (Programmable Universal Machine for Assembly) a été conçu par Vic Schienman et fut financé par General Motors et par The Massachussets Institute of Technology au milieu des années 70. Le système de ce robot est composé d'un bras manipulateur permettant d'assembler des composants industriels et de son ordinateur de commande. Ce robot est le robot d'assemblage le plus rependu dans l'industrie des années 70.

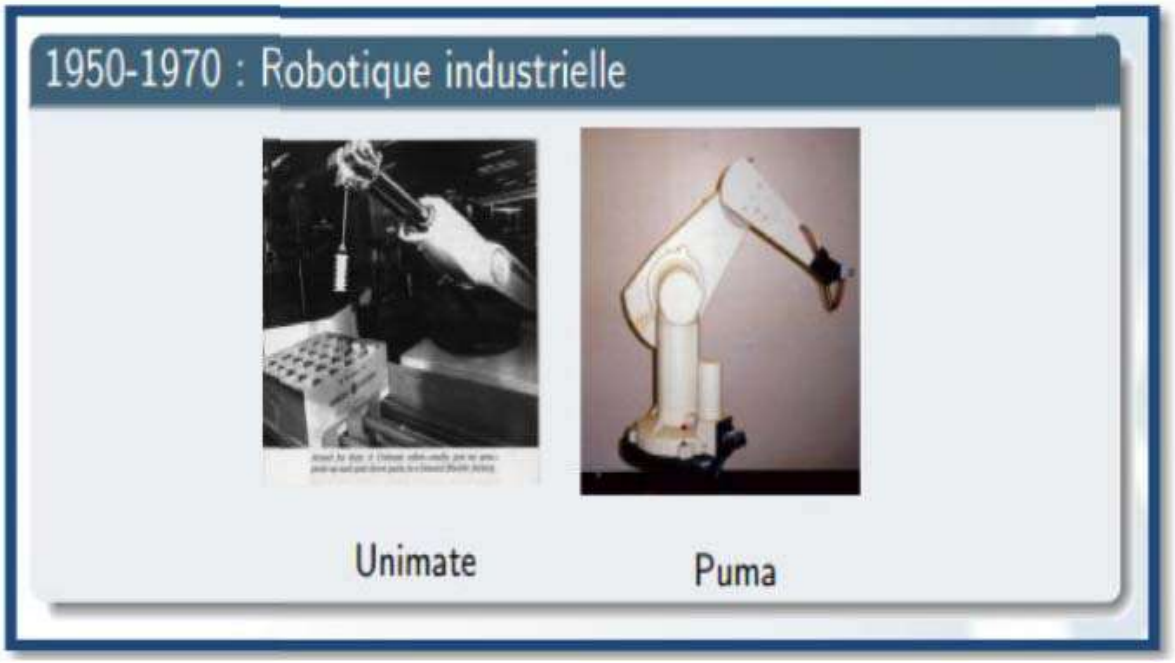


Figure I.1. *Robotique industrielle (unimate, puma).*

En 1985, Reymond Clavel a imaginé le Robot Delta qui possède un bras de manipulation formé de 3 parallélogrammes. Son brevet tombe dans le domaine public en 2007et différents constructeurs devraient alors sortir leur propre robot delta.

Le Jet Propulsion Laboratory (JPL) développe un robot industriel hexapode (à 6 pattes) du nom de Lemur. Lemur aura pour mission de monter, assembler et réparer des installations spatiales. Pesant moins de 5 kg, il offre la possibilité innovante d'adapter différents outils sur chacun de ses membres.

Selon l'étude robotique de la Fédération Internationale de Robotique (IFR) en 2012, il y a au moins 1 153 000 robots industriels opérationnels fin 2011 dans le monde.

Avec l'apparition de la robotique industrielle, les robots étaient conçus pour remplacer les ouvriers dans les tâches pénibles, répétitives ou dangereuses (peinture, soudure…). Aujourd'hui avec le développement de l'électronique, de l'informatique, de la mécanique et aussi de l'automatique, la technologie robotique a progressé. La recherche dans le domaine de la robotique est dirigée vers le développement de robots dévoués à des tâches bien différentes que celles demandées par l'industrie. Par exemple des robots travaillant en mode automatique ou semi-automatique et qui ont souvent pour objectif d'interagir avec des humains et de les

aider dans leurs tâches (surveillance, manutention d'objets lourds…). Ils sont dotés d'une intelligence qui leur donne une certaine autonomie.

Ainsi donc, le développement important de l'intelligence artificielle et de la robotique font que de nouveaux robots apparaissent constamment et l'utilisation de systèmes robotiques apparait aujourd'hui dans plusieurs domaines d'activité : la médecine, la défense, la recherche etc.…

I.4. Types de robots

Il existe deux types de robots : robots mobiles et robots manipulateurs [3].

I.4.1. Robots mobiles

Ce sont des robots capables de se déplacer dans un environnement comme le montre la figure I.2. Ils sont équipés ou non de manipulateurs suivant leur utilisation, (les robots explorateurs, les robots de services….).



Figure I.2. *Spirit, NASA, 2003 sur Mars.*

I.4.2. Robots manipulateurs

Des robots ancrés physiquement à leur place de travail et généralement mis en place pour réaliser une tâche précise répétitive, (tels que les robots industriels, médicaux….).



Figure I.3. *Robot fanuc.*

I.5 Les éléments constitutifs d'un robot

« De même qu'un trou n'est défini que par ce qui l'entoure, un robot n'a de sens que dans un environnement qu'il modifie ». [4]. C'est pourquoi on peut distinguer les ensembles interactifs dans un robot en fonctionnement [5].

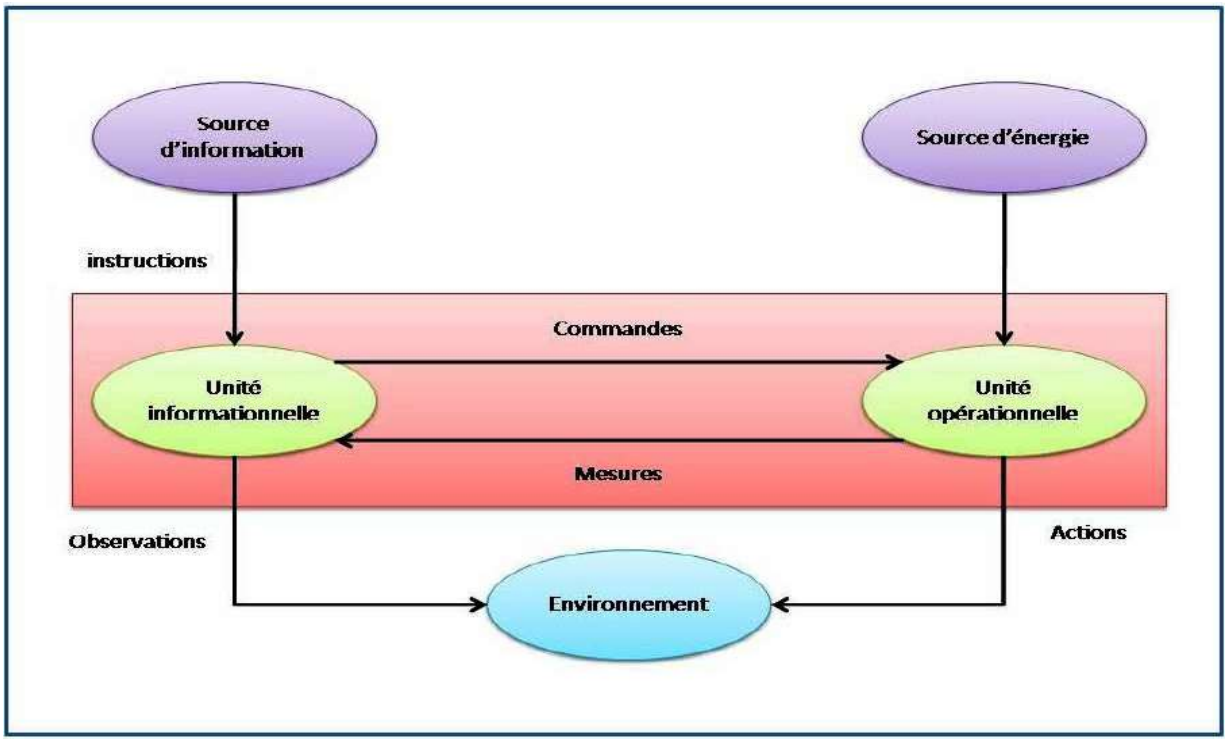


Figure I.4. *Structure fonctionnelle d'un robot. [6]*

I.5.1. Unité informationnelle

Reçoit les instructions décrivant la tache à accomplir, les mesures relatives à l'état interne de la structure mécanique qui constitue le bras manipulateur et les observations concernant son environnement. Elle élabore en conséquence les commandes de ses différentes articulations en vue de l'exécution de ses taches. Les systèmes actuels fonctionnent en interaction permanentent selon le cycle information-décision-action [6].

I.5.2. Unité opérationnelle

Exerce les actions commandées en empruntant la puissance à la source d'énergie. Cette partie, qui constitue le robot physique, intègre la structure mécanique (segments, articulations, architecture,..), les modules d'énergie (amplificateurs, variateurs, servovalves,…), les convertisseurs d'énergie (moteurs, vérins,…), les chaines cinématiques de transmission mécanique (réducteurs, vis à billes, courroies crantées,…), les capteurs de proprioceptifs placés sur chaque axe pour mesurer en permanence leur position et leur vitesse, et enfin l'effecteur, ou organe terminal, qui est en interaction avec l'environnement [6].

I.6. La structure mécanique articulée

Un robot manipulateur est constitué généralement par deux sous-ensembles distincts, un organe terminal qui est le dispositif destiné à manipulateur des objets, et une structure mécanique articulée (SMA), constituée d'un ensemble de solides reliés entre eux, généralement les uns à la suite des autres où chaque solide est mobile par rapport un précédent. Cette mobilité s'exprime en termes de degrés de liberté (d.d.l) qui est par définition le nombre de mouvements indépendants possibles d'un solide C_1 par rapport un solide qui lui est directement relié C_2**.**

Une structure mécanique articulée peut être représentée par une architecture composée de plusieurs chaînes de corps rigides assemblés par des liaisons appelées articulations. Les chaînes peuvent être dites soit ouvertes ou en série dans les quelles tous les corps ont au plus deux liaisons, ou bien arborescentes où au moins l'un des corps a plus de deux liaisons. Les chaînes peuvent aussi être fermées dans les quelles l'organe terminal est relié à la base du mécanisme par l'intermédiaire de plusieurs chaînes [7].

I.6.1. Structure mécanique articulée à chaîne cinématique simple

C'est une chaîne cinématique dont chaque membre possède un degré de connexion (nombre de liaisons mécaniques) inférieur ou égal à deux. Un robot sériel est formé d'une chaîne cinématique simple dont la base et l'organe effecteur possèdent un degré de connexion de un (c'est-à-dire qu'il n'est relié qu'à un seul corps) et les autres éléments un degré de connexion de deux.

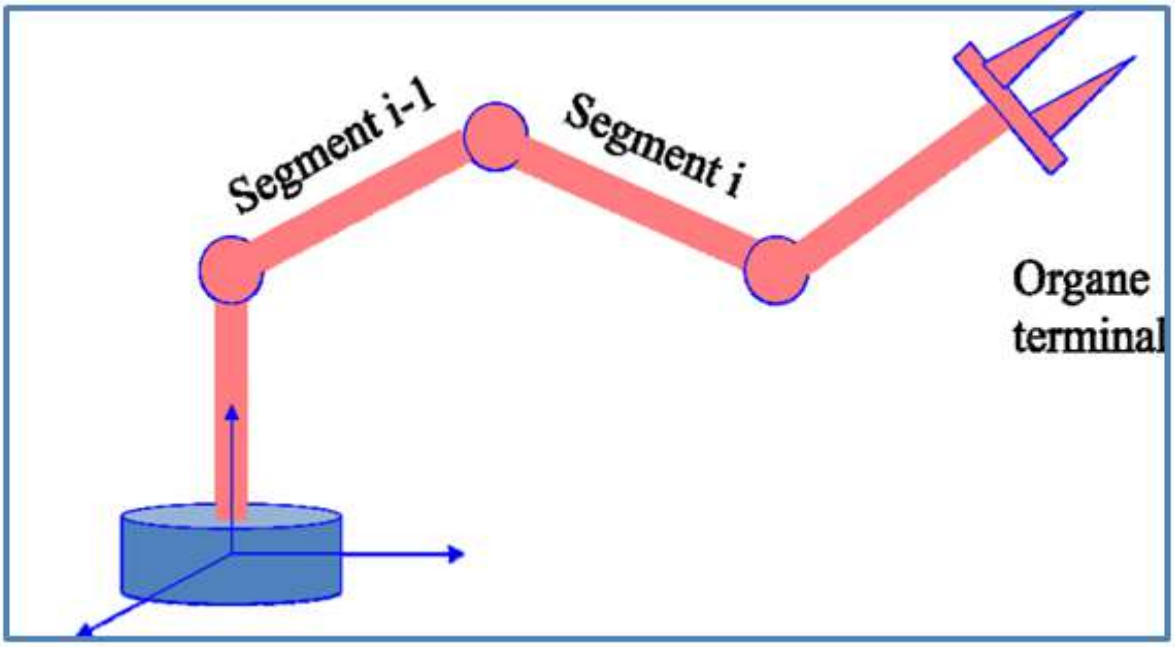


Figure I.5. *Structure mécanique à chaines ouvertes simples.*

I.6.2. Structure mécanique articulée à chaîne cinématique fermée

C'est une chaîne cinématique qu'il existe un retour mécanique d'un ou plusieurs segments à un autre dans la chaine.

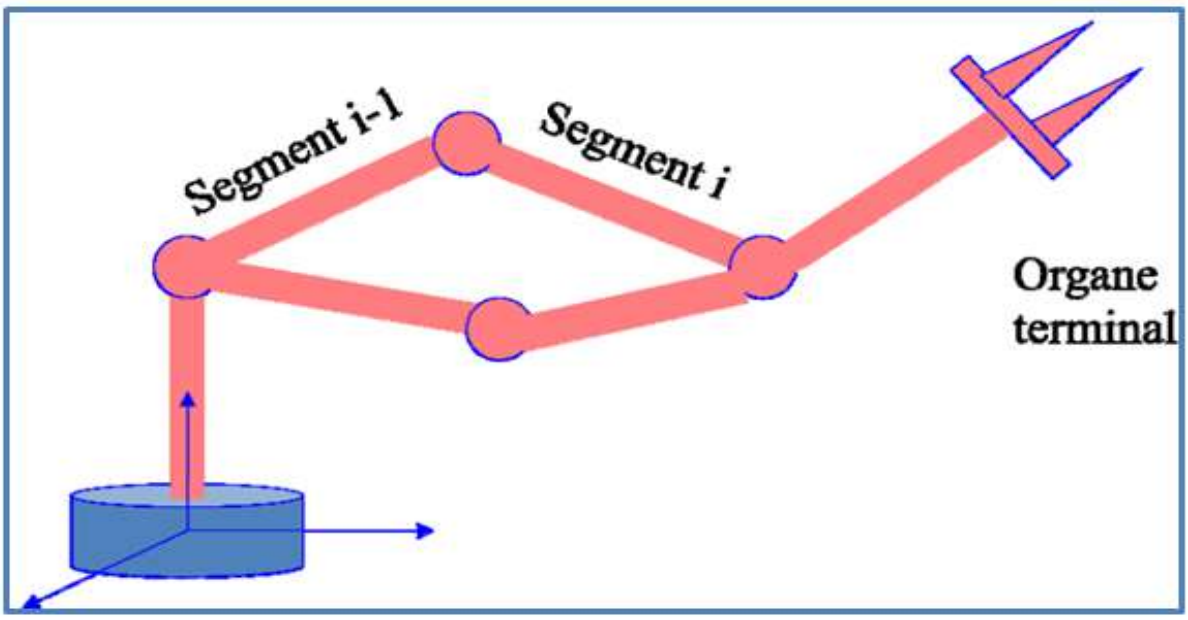


Figure I.6. *Structure mécanique à chaines fermées.*

I.7. Architecture des robots

La figure I.7 représente les vocabulaires du robot :



Figure I.7. *Architecture d'un robot.*

On distingue classiquement quatre (4) parties principales dans un robot manipulateur

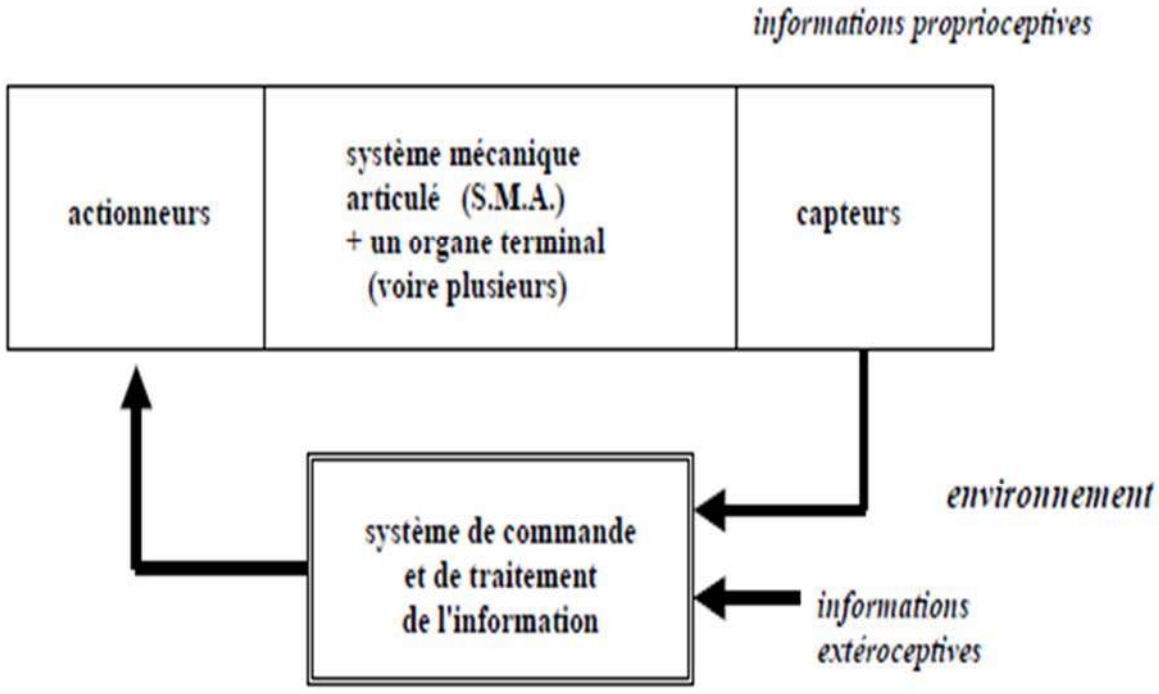


Figure I.8. *Parties principales dans un robot.*

- **La base :**

La base du manipulateur est fixée sur le lieu du travail. Ceci est le cas la quasi-totalité des robots industriels.

- **Le porteur :**

Le porteur représente l'essentiel du système mécanique articulé, il à pour rôle d'amener l'organe terminal dans une situation données imposées par la tache (la situation d'un corps peut être définie comme la position et l'orientation d'un repère attaché à ce corps par rapport à un repère de référence). Il est constitué de :

- ✓ **Segment :** corps solides rigides susceptibles d'être en mouvement par rapport à la base du porteur, et les uns par rapport aux autres.
- ✓ **Articulation :** Une articulation lie deux corps successifs en limitant le nombre de degré de liberté, de l'un par rapport à l'autre. Soit m le nombre de degré de liberté résultant, encore appelé mobilité de l'articulation. La mobilité d'une articulation est telle que : $0 \leq m \leq 6$

  Lorsque m =1, ce qui est fréquemment le cas en robotique, l'articulation est dite simple : soit rotoïde, soit prismatique
- ✓ **Articulation rotoïde :** Il s'agit d'une articulation de type pivot, notée R, réduisant le mouvement entre deux corps à une rotation autour d'un axe commun. La situation relative entre les deux corps est donnée par l'angle θ autour de cet axe (Voir la figure ci-dessus)

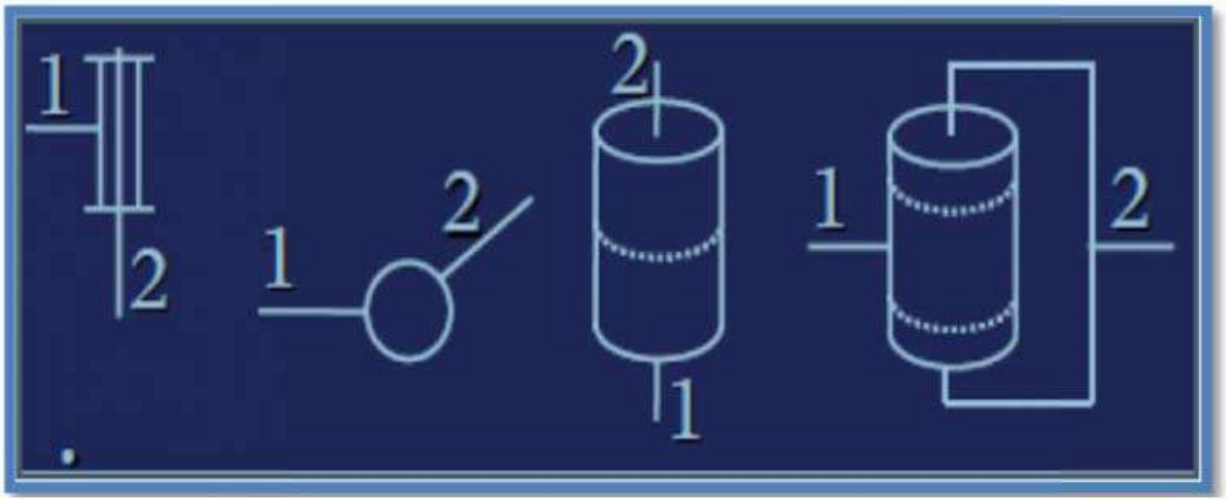


Figure I.9. *Représentation d'une articulation rotoïde.*

- ✓ **Articulation prismatique :** il s'agit d'une articulation de type glissière, notée P, réduisant le mouvement entre deux corps à une translation le long d'un axe commun. La situation relative entre les deux corps est mesurée par la distance le long de cet axe (voir la figure si dessus).

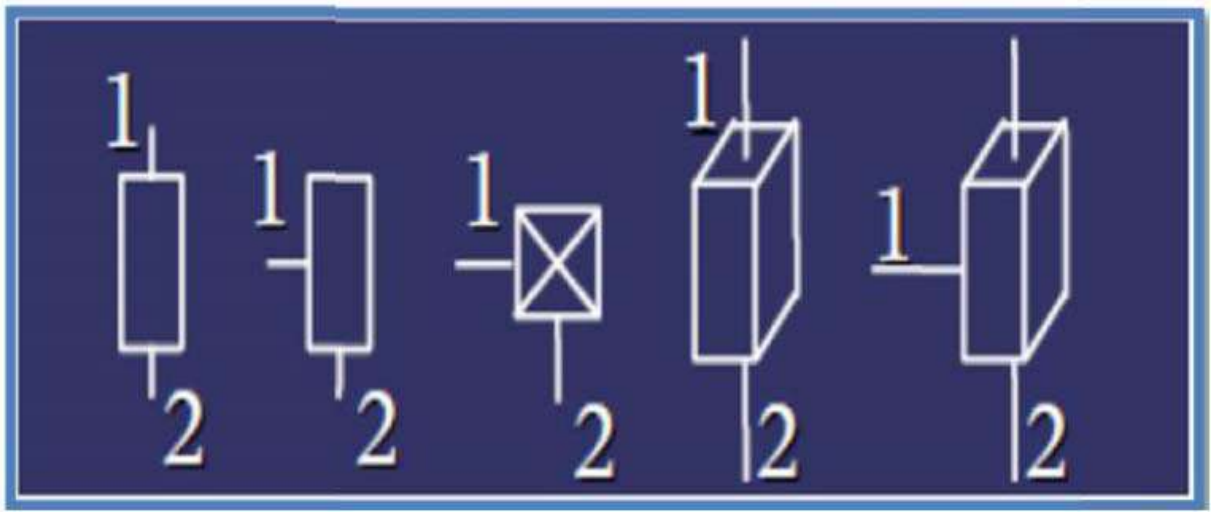


Figure I.10. *Représentation d'une articulation* prismatique.

- **L'actionneur :**

Pour être animé, la structure mécanique articulée comporte des moteurs le plus souvent associés à des transmissions (courroies crantées), l'ensemble constitue les actionneurs. Les actionneurs utilisent fréquemment des moteurs électriques à aimant permanent, à courant continu, à commande par l'induit. On trouve de plus en plus de moteurs à commutation électronique (sans balais), ou pour de petits robots, des moteurs pas à pas.

Pour les robots devant manipuler de très lourdes charges (par exemple, une pelle mécanique), les actionneurs sont le plus souvent hydrauliques, agissant en translation (vérin hydraulique) ou en rotation (moteur hydraulique).

Les actionneurs pneumatiques sont d'un usage général pour les manipulateurs à cycles (robots tout ou rien). Un manipulateur à cycles est une structure mécanique articulée avec un nombre limité de degrés de liberté permettant une succession de mouvements contrôlés uniquement par des capteurs de fin de course réglable manuellement à la course désirée (asservissement en position difficile du à la compressibilité de l'air). [8]

- **L'organe terminal :**

On regroupe tout dispositif destiné à manipuler des objets (dispositifs de serrage, dispositifs magnétiques, à dépression, …), ou à les transformer (outils, torche de soudage, pistolet de peinture, …). En d'autres termes, il s'agit d'une interface permettant au robot d'interagir avec son environnement. Un organe terminal peut être multifonctionnel, au sens où il peut être équipé de plusieurs dispositifs ayant des fonctionnalités différentes. Il peut aussi être monofonctionnel, mais interchangeable. Un robot, enfin, peut-être multi-bras, chacun des bras portant un organe terminal différent. On utilisera indifféremment le terme organe terminal, préhenseur, outil ou effecteur pour nommer le dispositif d'interaction fixé à l'extrémité mobile de la structure mécanique, exemple : pistolet pour la soudure dans les robots industriels. [8]

I.8. Classification des robots

On peut classer les robots d'un point de vue fonctionnel ou d'après leur structure géométrique.

I.8.1. Classification fonctionnelle

Le nombre de classe et les distinctions entre celles-ci varient de pays à pays (6 classes au Japon, 4 en France). L'A.F.R.I. distingue 4 classes illustrées ci-dessous [9] :

I.8.1.a. Manipulateurs à commande manuelle ou télécommande

La Figure I.11 représente les manipulateurs à commande manuelle :

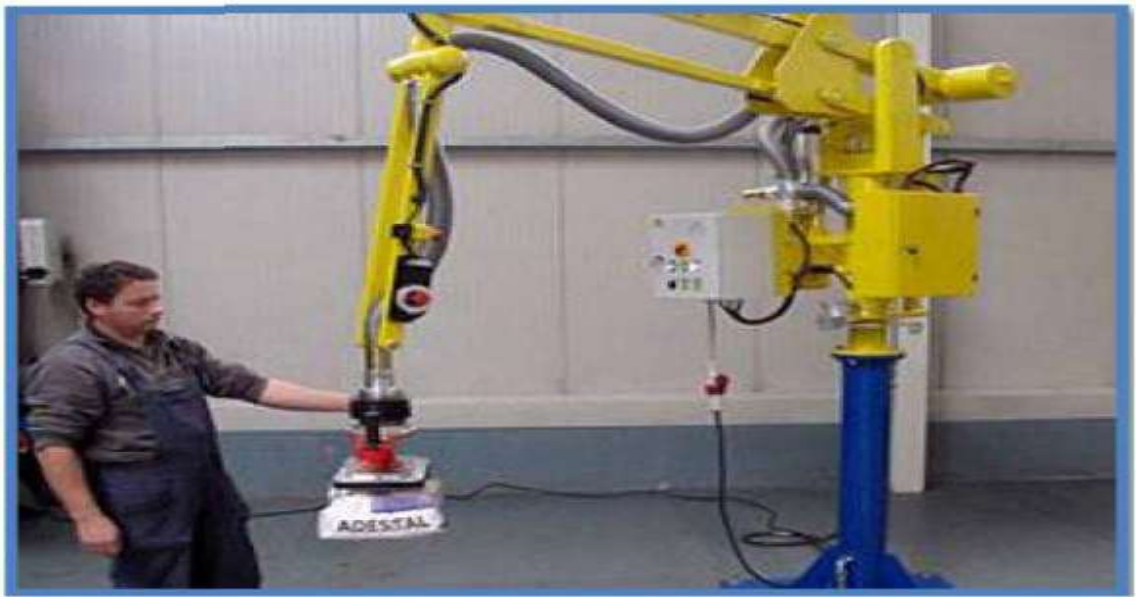


Figure I.11. *Manipulateur à commande manuelle.*

I.8.1.b. Manipulateurs automatiques à cycles préréglés

Le réglage se fait mécaniquement par cames butées comme le montre la Figure I.12, la commande peut se faire par automate programmable on peut distinguer entre manipulateurs à cycle fixe et manipulateurs à cycle programmable.



Figure I.12. *Manipulateur à cycle pré-réglé.*

I.8.1.c. Robots programmables

C'est la première génération de robots industriels ; ils répètent les mouvements qu'on leur a appris ou programmés sans informations sur l'environnement ou la tâche effectuée. On peut aussi faire la distinction entre robots « play-back » qui reproduisent la tâche apprise et robots à commande numérique qui peuvent être programmés hors-ligne.

Pour certains robots, par exemple les robots de peinture, qui doivent suivre une trajectoire complexe qu'il est difficile d'exprimer mathématiquement, un opérateur humain spécialiste de la tâche effectue la trajectoire en guidant le bras du robot à l'aide d'un « pantin », et l'entièreté de la trajectoire est mémorisée par le robot et la figure I.13 représente les robots programmable :

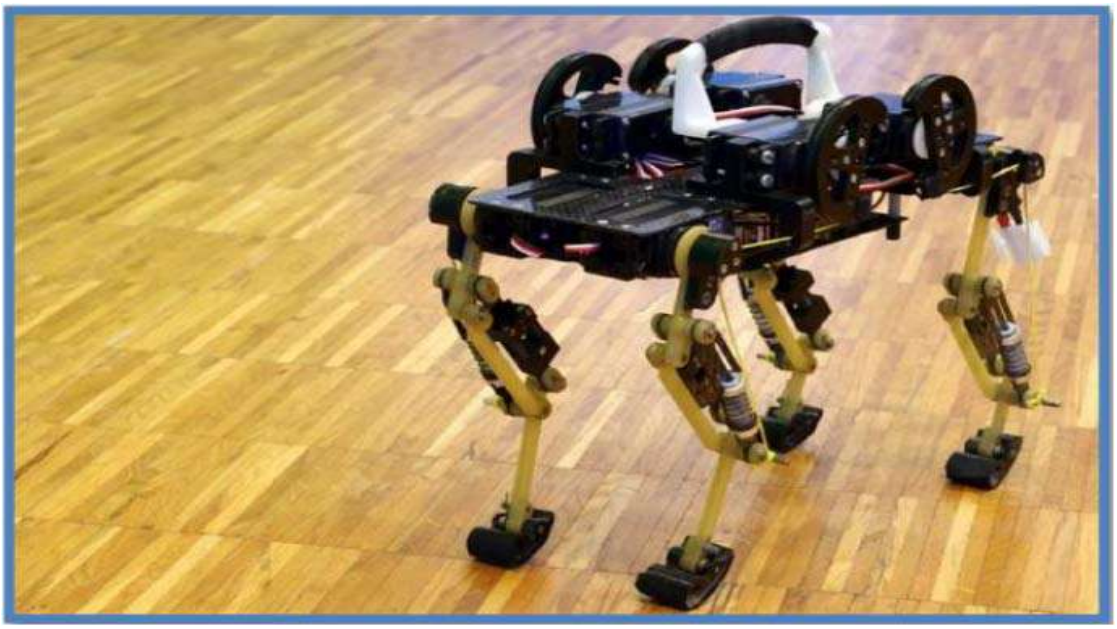


Figure I.13. *Robot programmable.*

I.8.1.d. Robots intelligents

On trouve actuellement des robots de seconde génération qui sont capables d'acquérir et d'utiliser certaines informations sur leur environnement (systèmes de vision, détecteurs de proximité, capteurs d'efforts,...) comme le montre la Figure I.14. On étudie des robots de troisième génération, capables de comprendre un langage oral proche du langage naturel et de se débrouiller de façon autonome dans un environnement complexe, grâce à l'utilisation de l'intelligence artificielle.



Figure I.14. *Robot intelligent.*

I.8.2. Classification géométrique

On peut aussi classer les robots suivant leur configuration géométrique, autrement dit l'architecture de leur porteur.

- **Structure cartésienne (PPP) :**

C'est une structure à trois liaisons prismatiques, est la plus ancienne. historiquement, elle découle logiquement de la conception traditionnelle d'une machine-outil à trois axes, type rectifieuse ou fraiseuse par exemple. Cette structure et relativement peu utilisée, sauf dans quelques applications particulière, robots pratiques, robots magasinage. [6]

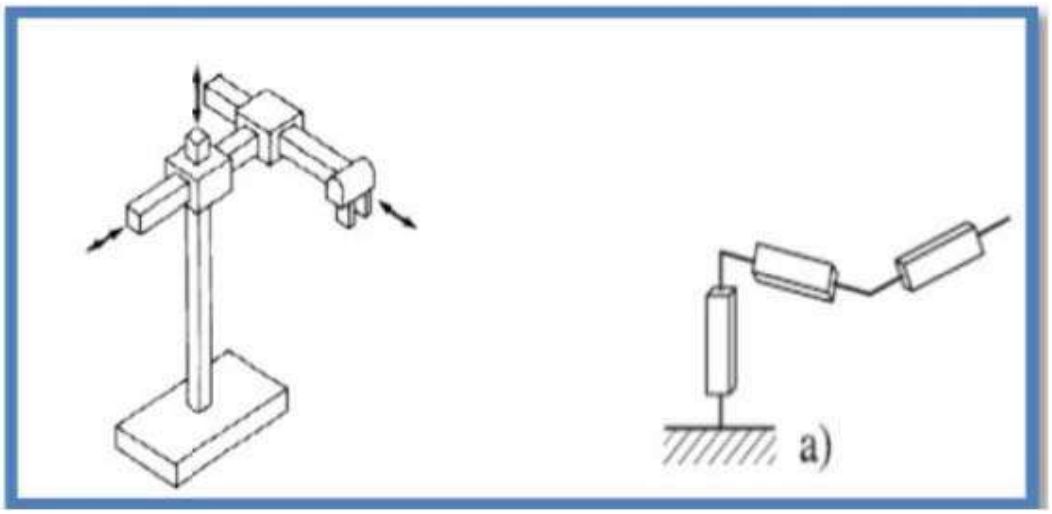


Figure I.15. *Robot cartésien.[10]*

- **La structure cylindrique (RPP) ou (PRP) :**

Associe une rotation et deux translations. Elle présente l'inconvénient d'offrir un volume de travail faible devant un encombrement total important. Elle n'est pratiquement plus utilisée [6].

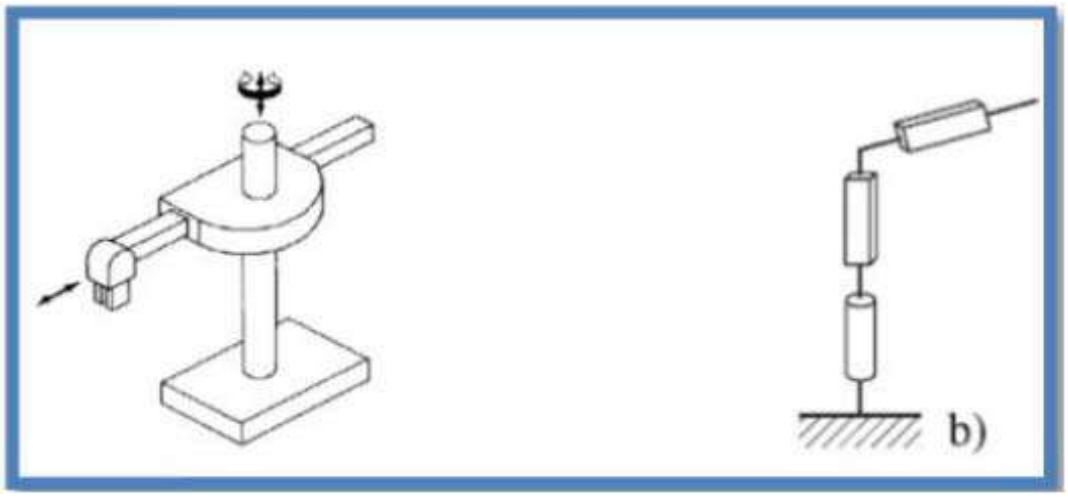


Figure I.16. *Robot cylindrique.[10]*

- **La structure sphérique ou polaire à axe de rotation orthogonale :**

Est une structure quasiment abandonnée pour des raisons similaires à l'abondon de structure cylindrique [6].

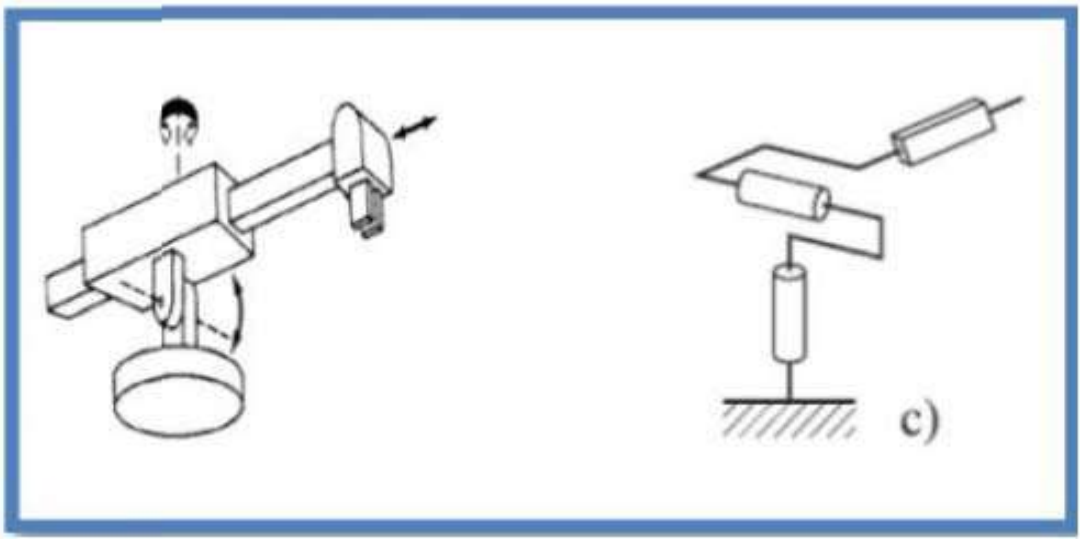


Figure I.17. *Robot sphérique[10].*

- **La structure dite SCARA :**

A axes de rotation parallèles est l'une des plus utilisées, en particulier pour des taches de manutention ou d'assemblages très fréquents dans l'industrie. Ce succès commercial est lie au fait que le ratio entre le volume de travail et l'encombrement est très favorable et aussi que la structure SCARA est très adaptée à ce type de taches [6].

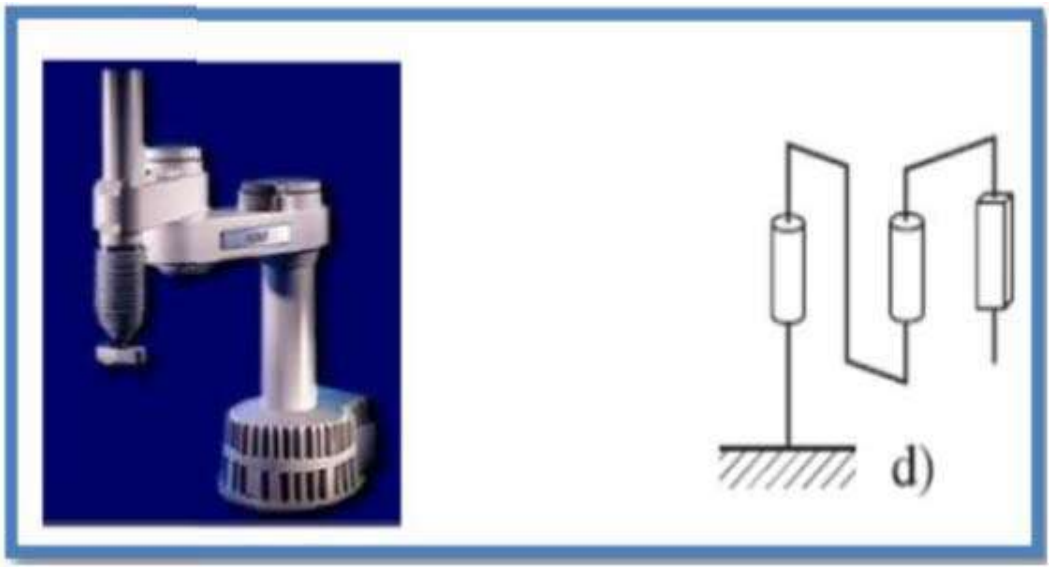


Figure I.18. *Robot SCARA[10].*

- **La structure 3R (anthropomorphe) :**

Elle Permet d'amener un solide en un point de l'espace par trois rotations, généralement une à axe vertical et deux à axes horizontaux et parallèles c'est le porteur « généraliste par excellence, pouvant se programmer facilement pour différent types de taches et disposant d'un volume de travail conséquent [6].

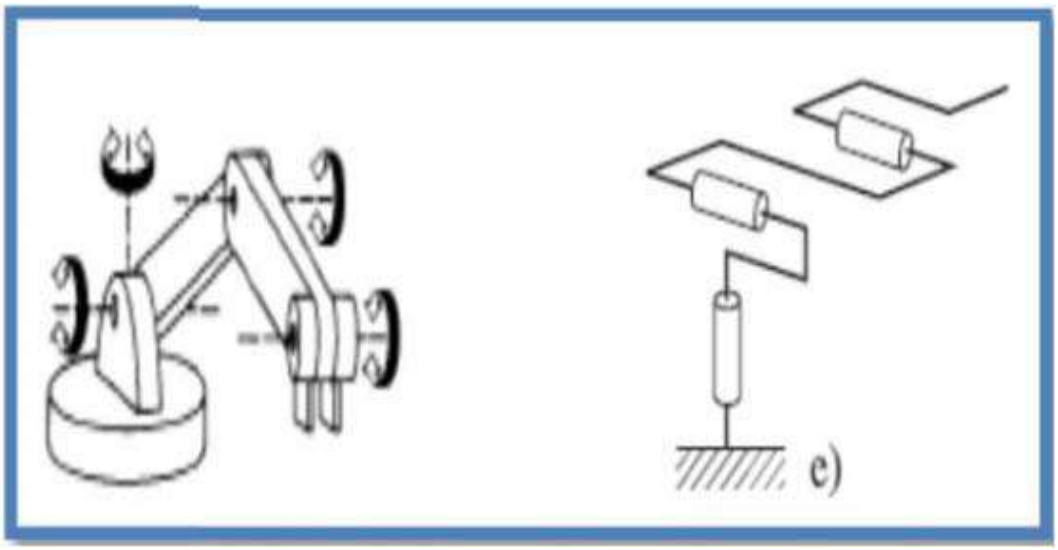


Figure I.19. *Robot 3R [10].*

I.9. Caractéristiques d'un robot

Un robot doit être choisi en fonction de l'application qu'on lui réserve. Voici quelques paramètres à prendre, éventuellement, en compte :

- La charge maximum transportable (de quelques kilos à quelques tonnes), à déterminer dans les conditions les plus défavorables (en élongation maximum).

- Le volume de travail, défini comme l'ensemble des points atteignables par l'organe terminal. Tous les mouvements ne sont pas possibles en tout point du volume de travail. L'espace de travail (reachable workspace), également appelé espace de travail maximal, est le volume de l'espace que le robot peut atteindre via au moins une orientation de l'organe terminal. L'espace de travail est le volume de l'espace que le robot peut atteindre avec toutes les orientations possibles de l'organe terminal. Cet espace de travail est un sous-ensemble de l'espace de travail maximal.

- Le positionnement absolu, correspondant à l'erreur entre un point souhaité (réel) – défini par une position et une orientation dans l'espace cartésien – et le point atteint et calculé via le modèle géométrique inverse du robot. Cette erreur est due au modèle utilisé, à la quantification de la mesure de position, à la flexibilité du système mécanique. En général, l'erreur de positionnement absolu, également appelée précision, est de l'ordre de 1 mm.

- La répétabilité, ce paramètre caractérise la capacité que le robot à retourner vers un point (position, orientation) donné. La répétabilité correspond à l'erreur maximum de positionnement sur un point prédéfini dans le cas de trajectoires répétitives. En général, la répétabilité est de l'ordre de 0,1 mm.

- La vitesse de déplacement (vitesse maximum en élongation maximum), accélération.

- La masse du robot.

- Le coût du robot.

- La maintenance.

I.10. Domaines d'applications

Les robots industriels peuvent être utilisés dans plusieurs domaines et peuvent accomplir différentes tâches.[11]

- **Tâches simples**

La grande majorité des robots est utilisée pour des tâches simples et répétitives. Ils sont programmés une fois pour toute au cours de la procédure d'apprentissage. Les critères de choix de la solution robotique sont:

- La tâche est assez simple pour être robotisée.
- Les critères de qualité sur la tâche sont importants.
- Pénibilités de la tâche (peinture, charge lourde, environnement hostile, ...)

Exemples :

- **Robot soudeurs :**



Figure I.20. *Robot soudeurs Par points*



Figure I.21. *Robot soudeurs a l'arc*

- **Tâches complexes :**
  - **Robotique de service :**



Figure I.22. *Robot pompiste*



Figure I.23. *Robot de construction.*



Figure I.24. *Robot Computer motion.*



Figure I.25. *Robot Assistance aux personnes Handicapées.*

I .11. Avenir de la robotique

- Stagnation du nombre de robots utilisés pour des tâches simples.

- Forte croissance du nombre des robots utilisés pour des tâches complexes :

  - robotique de service.
  - robotique d'assistance aux manipulations dans la recherche biologique et génétique.
  - robotique médicale. [12]

I .12. Conclusion

Dans ce chapitre, nous avons donner une idée générale sur la robotique, l'historique des robots, leurs structures, leurs utilisations et les différents types de robots ainsi que leurs classifications et leurs domaines d'applications et cela va nous servir pour la construction de notre bras manipulateur qui va respecter les classifications et les définitions donner dans ce chapitre .

Chapitre II

Le traitement et la synthèse de parole

II.1. Introduction

La parole est un système structuré qui permet aux êtres humains de communiquer entre eux. L'information d'un message parlé est transmise par les fluctuations de la pression de l'air qui sont émises par l'appareil phonatoire, c'est le signal vocal. Ce signal est analysé par l'oreille et les informations résultantes sont transmises au cerveau qui les interprète. Au sens strict, le contenu d'un signal vocal est représenté uniquement par son intelligibilité. Dans un sens plus large, il faut aussi tenir compte de tout ce qui représente la qualité du signal vocal, c'est-à-dire les intonations, la prosodie et les perturbations du milieu ambiant. Nous allons ci-après décrire certaines propriétés de l'appareil phonatoire humain [13].

Dans ce chapitre on va voir une approche générale sur l'identité du traitement de la parole commençant par tout ce qui concerne le signal de parole, et le mécanisme de la phonation seront aussi abordé dans le contenu de ce chapitre. Nous allons aussi voir le niveau acoustique du signal de parole particulièrement son spectre sa fréquence fondamentale son énergie….etc.

II.2.Le traitement de la parole

Le traitement de la parole est aujourd'hui une composante fondamentale des sciences de l'ingénieur. Située au croisement du traitement du signal numérique et du traitement du langage (c'est-à-dire du traitement de données symboliques), cette discipline scientifique a connu depuis les années 60 une expansion fulgurante, liée au développement des moyens et des techniques de télécommunications.

L'importance particulière du traitement de la parole dans ce cadre plus général s'explique par la position privilégiée de la parole comme vecteur d'information dans notre société humaine.

L'extraordinaire singularité de cette science, qui la différencie fondamentalement des autres composantes du traitement de l'information, tient sans aucun doute au rôle fascinant que joue le cerveau humain à la fois dans la production et dans la compréhension de la parole et à l'étendue des fonctions qu'il met, inconsciemment, en œuvre pour y parvenir de façon pratiquement instantanée.

La parole est en effet un des seuls à être à la fois produit et perçu instantanément par le cerveau. La parole est en effet produite par le conduit vocal, contrôlé en permanence par le cortex moteur. L'étude des mécanismes de phonation permettra donc de déterminer, dans une certaine mesure, ce qui est parole et ce qui n'en est pas. De même, l'étude des mécanismes d'audition et des propriétés perceptuelles qui s'y rattachent permettra de dire ce qui, dans le signal de parole, est réellement perçu. Mais l'essence même du signal de parole ne peut être cernée de façon réaliste que dans la mesure où l'on imagine, bien au-delà de la simple mise en commun des propriétés de production et de perception de la parole, les propriétés du signal dues à la mise en boucle de ces deux fonctions. Mieux encore, c'est non seulement la perception de la parole qui vient influer sur sa production par le biais de ce bouclage, mais aussi et surtout sa compréhension. On ne parle que dans la mesure ou l'on s'entend et où l'on se comprend soi-même; la complexité du signal qui en résulte s'en ressent forcément S'il n'est pas en principe de parole sans cerveau humain pour la produire, l'entendre, et la comprendre, les techniques modernes de traitement de la parole tendent cependant à produire des systèmes automatiques qui se substituent à l'une ou l'autre de ces fonctions :

- Les ***reconnaisseurs*** ont pour mission de décoder l'information portée par le signal vocal à partir des données fournies par l'analyse.
- Les ***synthétiseurs*** ont quant à eux la fonction inverse de celle des reconnaisseurs de parole : ils produisent de la parole artificielle.
- Enfin, le rôle des ***codeurs*** est de permettre la transmission ou le stockage de parole avec un débit réduit, ce qui passe tout naturellement par une prise en compte judicieuse des propriétés de production et de perception de la parole.

On comprend aisément que, pour obtenir de bons résultats dans chacune de ces tâches, il faut tenir compte des caractéristiques du signal étudié. Et, vu la complexité de ce signal, due en grande partie au couplage étroit entre production, perception, et compréhension, il n'est pas étonnant que les recherches menées par les spécialistes soient directement liées aux progrès obtenus dans de nombreuses autres disciplines scientifiques, progrès dont elles sont par ailleurs souvent à la fois les bénéficiaires et les instigatrices. Comme le remarque très justement Allen: "Le traitement de la parole fournit d'excellents exemples pour l'étude de systèmes complexes, dans la mesure où il soulève des questions fondamentales dans les domaines du partitionnement des systèmes, du choix d'unités descriptives, des techniques de représentation, des niveaux d'abstraction, des formalismes de représentation de la

connaissance, de l'expression d'interactions entre contraintes, des techniques de modularité et de hiérarchisation, des techniques d 'estimation de vraisemblance, des techniques de mesure de la qualité et du naturel d'un stimulus, de la détermination de classes d'équivalence, de la paramétrisation de modèles adaptatifs, de l'étude des compromis entre représentations procédurales et déclaratives, de l'architecture des systèmes, et de l'exploitation des technologies modernes pour produire des systèmes qui fonctionnent en temps réel pour un coût acceptable" [14].

II.2.1.Mécanisme de la phonation

Le signal vocal est le résultat de plusieurs actions conjuguées. En effet, la parole résulte de l'action coordonnée des appareils respiratoires et masticatoires contrôlés par le système nerveux central. Il existe deux types élémentaires d'excitation qui produisent les sons voisés et non voisés.

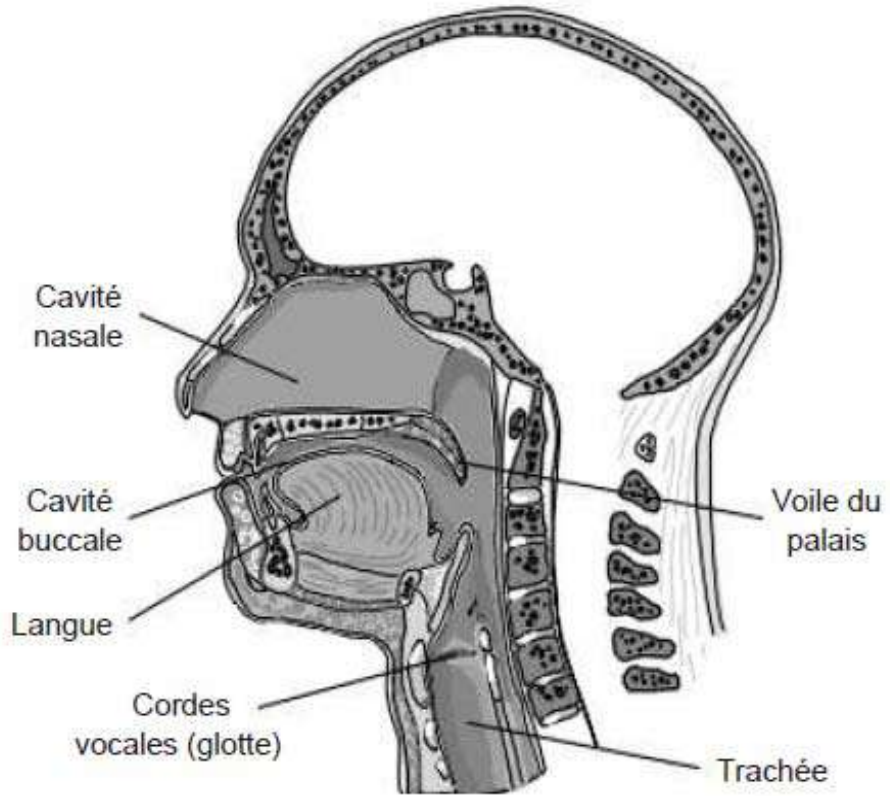


Figure II.1. *Représentation de l'appareil phonatoire*

II.2.1.1. Phonation de sons voisés

Les sons voisés sont produits à partir d'une excitation qui actionne sur le conduit vocal et qui consiste en une suite d'impulsions périodiques d'air fournies par le larynx. Les cordes vocales au début sont fermées. Sous la pression continue de 1' air qui vient des poumons elle s'ouvrent graduellement délivrant cette énergie potentielle. Pendant cette ouverture la vitesse de l'air et l'énergie cinétique augmentent jusqu'à ce que la tension élastique des cordes vocales égale la force de séparation du courant d'air. A ce point l'ouverture de la glotte est maximale. L'énergie cinétique qui a été accumulée comme tension élastique dans les cordes vocales commence à rétrécir cette ouverture et de plus la force de Bernoulli accélère encore la fermeture abrupte de la glotte [15]. Ce processus périodique est caractérisé par une fréquence propre à chaque personne, connue sous le nom de fréquence du fondamental (F0) où pitch et il donne la hauteur normale de la voix. La fréquence fondamentale (*F*0) peut varier :

- de 80 à 200 Hz pour une voix masculine,
- de 150 à 450 Hz pour une voix féminine,
- de 200 à 600 Hz pour une voix d'enfant. [16]

Cette fréquence fondamentale peut varier suite à des facteurs liés au stress, intonation et émotions. Le timbre de la voix est déterminé par les amplitudes relatives des harmoniques du fondamental.

L'intensité du son émis est liée à la pression de l'air en amont de larynx. Tous ces aspects pour un son voisé peuvent être observés dans la figure II.2.

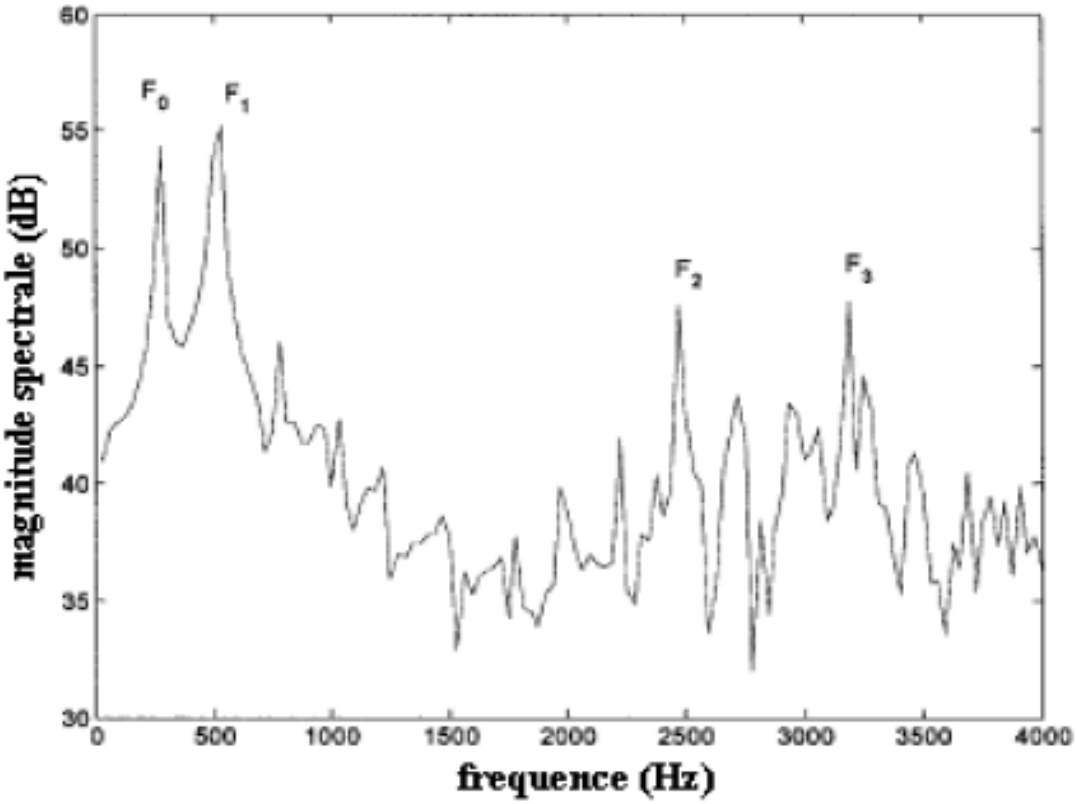


Figure II.2. *Le spectre du son voisé / i /*

II.2.1.2. Phonation de sons non voisés

Les sons non voisés sont générés par le passage de l'air dans une constriction étroite situé en un point du conduit vocal. Ils sont générés sans l'apport du larynx et ne présentent pas de structure périodique [15]. Ces caractéristiques d'un son non voisé peuvent être observées dans la figure II.3

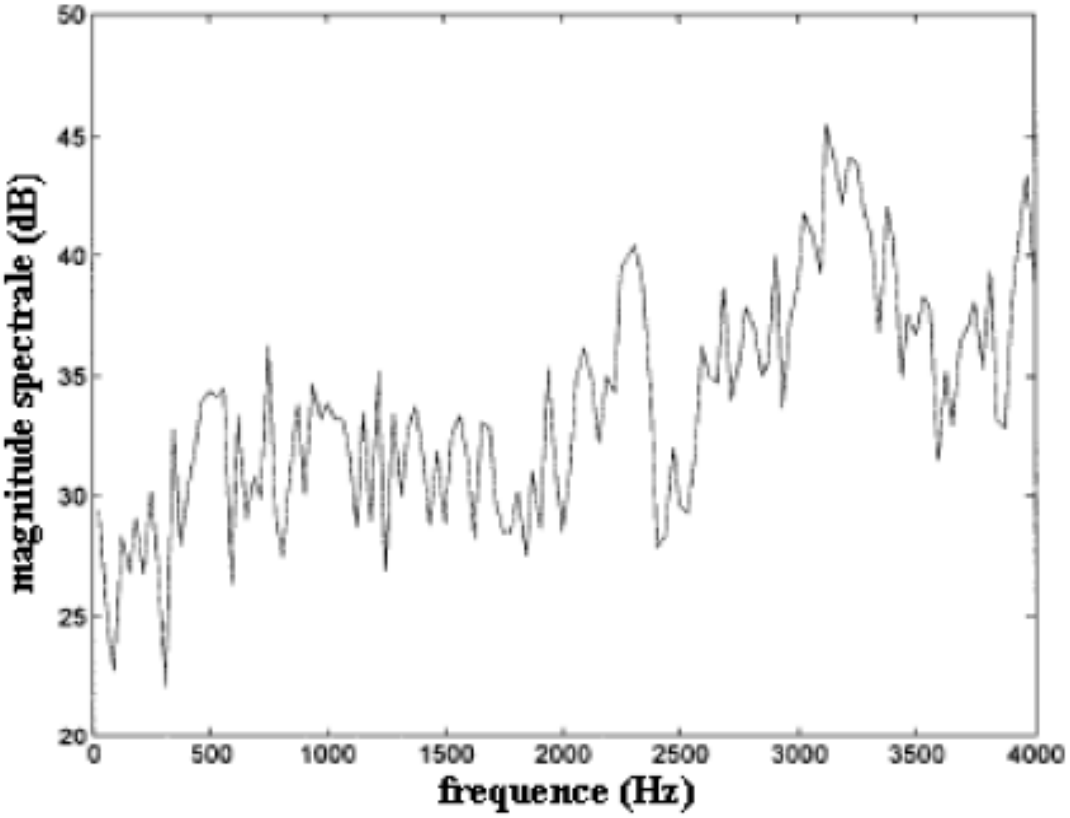


Figure II.3. *Le spectre du son non voisé / f /*

II.2.2. Mécanisme de l'audition

Dans le cadre du traitement de la parole, une bonne connaissance des mécanismes de l'audition et des propriétés perceptuelles de l'oreille est aussi importante qu'une maîtrise des mécanismes de production.

Les ondes sonores sont recueillies par l'appareil auditif, ce qui provoque les sensations auditives. Ces ondes de pression sont analysées dans l'oreille interne qui envoie au cerveau l'influx nerveux qui en résulte; le phénomène physique induit ainsi un phénomène psychique grâce à un mécanisme physiologique complexe.

L'appareil auditif comprend l'oreille externe, l'oreille moyenne, et l'oreille interne (Figure. II.4). Le conduit auditif relie le pavillon au tympan : c'est un tube acoustique de section uniforme fermé à une extrémité, son premier mode de résonance est situé vers 3000 Hz, ce qui accroît la sensibilité du système auditif dans cette gamme de fréquences. Le mécanisme de

l'oreille interne (marteau, étrier, enclume) permet une adaptation d'impédance entre l'air et le milieu liquide de l'oreille interne. Les vibrations de l'étrier sont transmises au liquide de la cochlée. Celle-ci contient la membrane basilaire qui transforme les vibrations mécaniques en impulsions nerveuses. La membrane s'élargit et s'épaissit au fur et à mesure que l'on se rapproche de l'apex de la cochlée; elle est le support de l'organe de Corti qui est constitué par environ 25000 cellules ciliées raccordées au nerf auditif. La réponse en fréquence du conduit au droit de chaque cellule est esquissée à la figure II.5. La fréquence de résonance dépend de la position occupée par la cellule sur la membrane; au-delà là de cette fréquence, la fonction de réponse s'atténue très vite. Les fibres nerveuses aboutissent à une région de l'écorce cérébrale appelée aire de projection auditive et située dans le lobe temporal. En cas de lésion de cette aire, on peut observer des troubles auditifs. Les fibres nerveuses auditives afférentes (de l'oreille au cerveau) et efférentes (du cerveau vers l'oreille) sont partiellement croisées : chaque moitié du cerveau est mise en relation avec les deux oreilles internes.

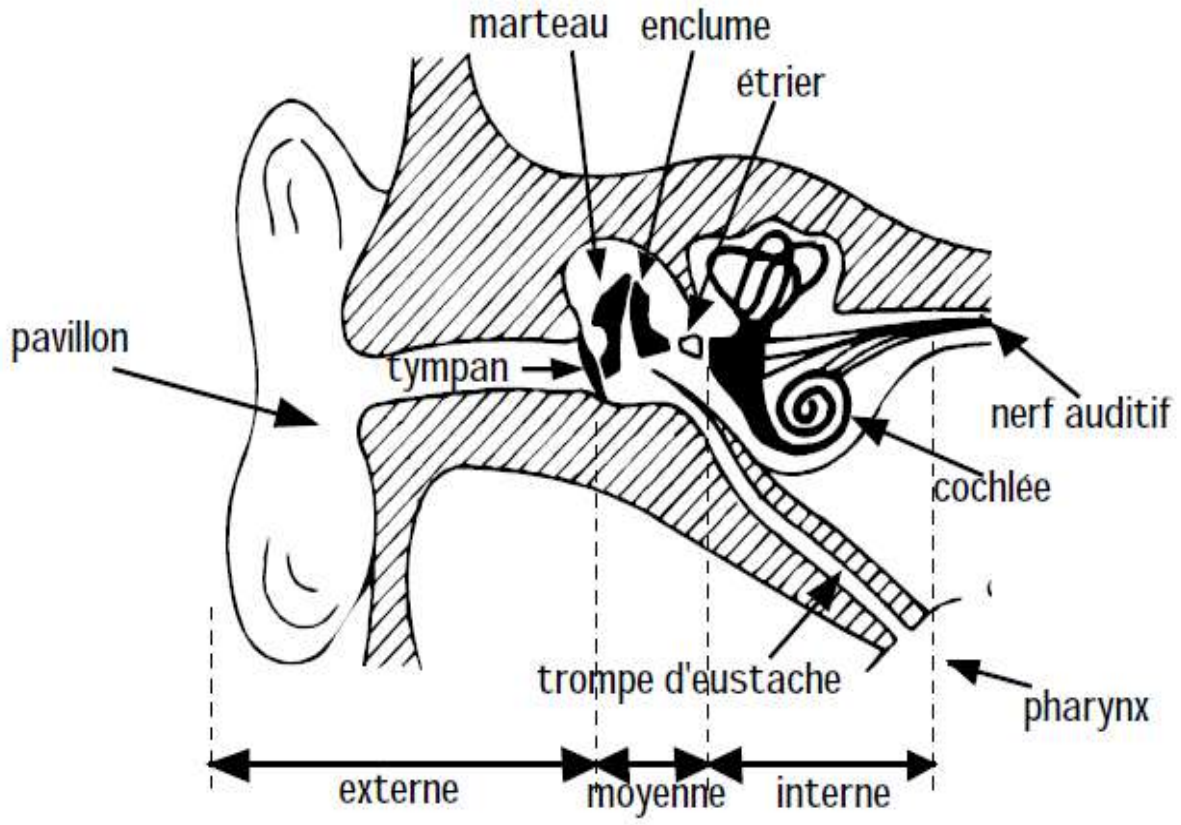


Figure II.4. *Coupe de l'appareil auditif humain [14]*

Il reste très difficile de nos jours de dire comment l'information auditive est traitée par le cerveau. On a pu par contre étudier comment elle était finalement perçue, dans le cadre d'une science spécifique appelée psycho acoustique. Sans vouloir entrer dans trop de détails sur la contribution majeure des psychos acousticiens dans l'étude de la parole, il est intéressant d'en connaître les résultats les plus marquants.

Ainsi, l'oreille ne répond pas également à toutes les fréquences. La figure II.6 présente le champ auditif humain, délimité par la courbe de seuil de l'audition et celle du seuil de la douleur. Sa limite supérieure en fréquence (≈ 16000 Hz, variable selon les individus) fixe la fréquence d'échantillonnage maximale utile pour un signal auditif (≈ 32000 Hz). [14]

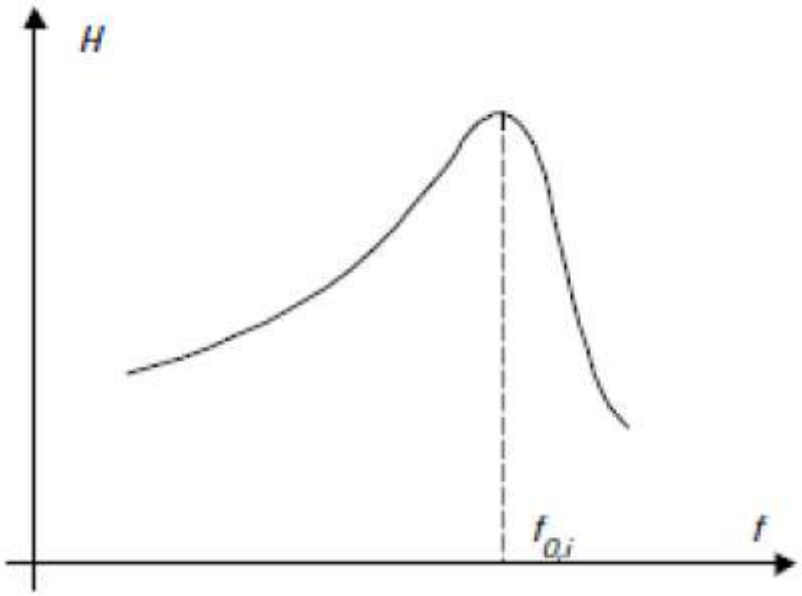


Figure II.5. *Réponse en fréquence d'une cellule ciliée. [14]*

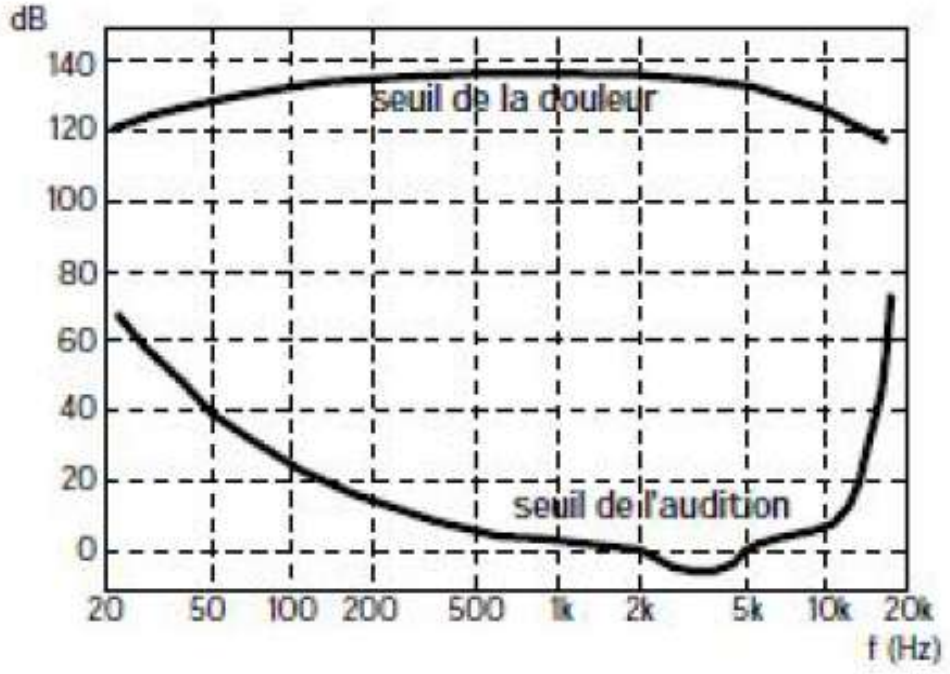


Figure II.6. *Le champ auditif humain. [14]*

A l'intérieur de son domaine d'audition, l'oreille ne présente pas une sensibilité identique à toutes les fréquences. La figure II.7 a fait apparaître les courbes d'égale impression de puissance auditive (aussi appelée sonie, exprimée en sones) en fonction de la fréquence. Elles révèlent un maximum de sensibilité dans la plage [500 Hz, 10 kHz], en dehors de laquelle les sons doivent être plus intenses pour être perçus [14].

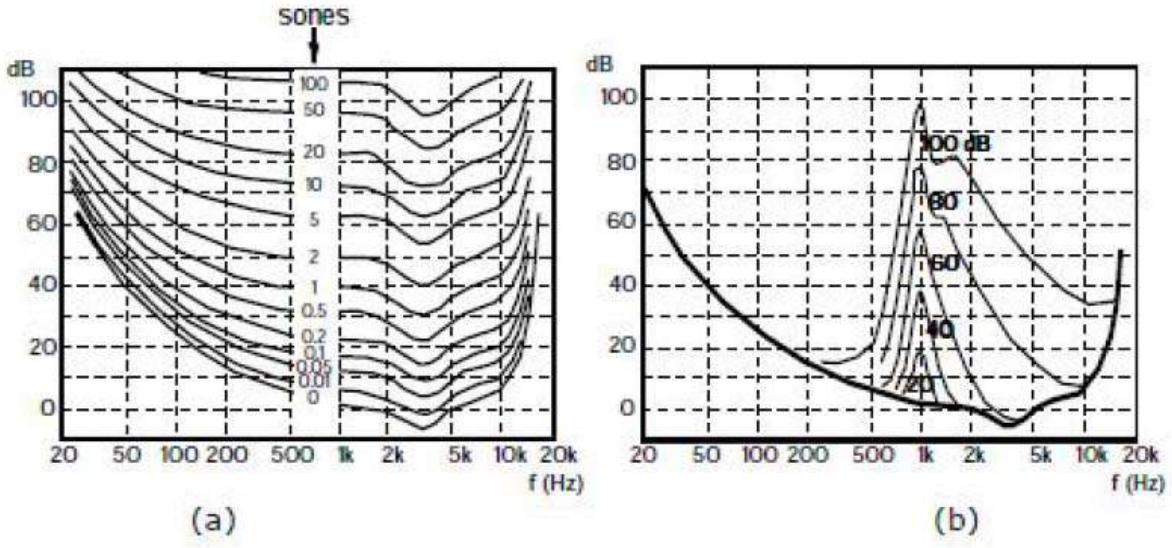


Figure II.7. *(a) : Courbes isosoniques en champ ouvert (b) : Masquage auditif par un bruit à bande étroite : limite d'audibilité en fonction de la puissance du bruit masquant [14]*

II.3. Le niveau acoustique de la parole

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

De nos jours, le signal électrique résultant est le plus souvent numérisé. Il peut alors être soumis à un ensemble de traitements statistiques qui visent à en mettre en évidence les traits acoustiques : sa fréquence fondamentale, son énergie, et son spectre. Chaque trait acoustique est lui-même intimement lié à une grandeur perceptuelle : pitch, intensité, et timbre. L'opération de numérisation, schématisée à la figure II.8, requiert successivement : un filtrage de garde, un échantillonnage, et une quantification [14].

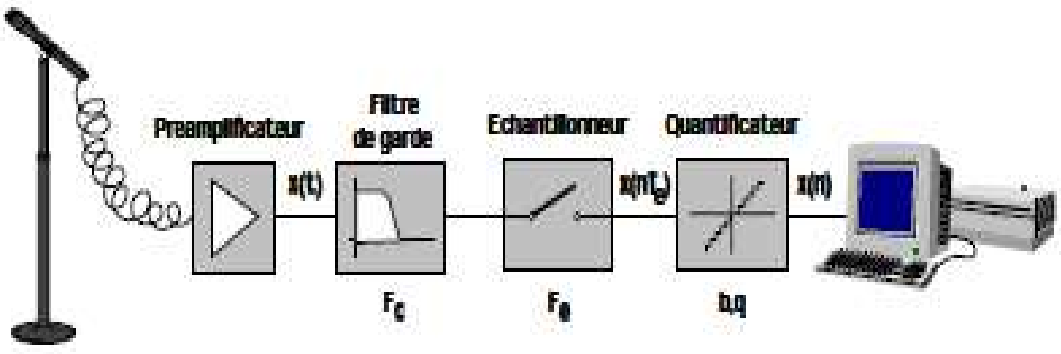


Figure II.8. *Enregistrement numérique d'un signal acoustique. [14]*

II.3.1. Audiogramme

L'échantillonnage transforme le signal à temps continu *x(t)* en signal à temps discret $x(nT_e)$ défini aux instants d'échantillonnage, multiples entiers de la période d'échantillonnageT_e; celle-ci est elle-même l'inverse de la fréquence d'échantillonnage *fe*. Pour ce qui concerne le signal vocal, le choix de f_e résulte d'un compromis. Son spectre peut s'étendre jusque 12 kHz. Il faut donc en principe choisir une fréquence $f_e$égale à 24 kHz au moins pour satisfaire raisonnablement au théorème de Shannon. Cependant, le coût d'un traitement numérique, filtrage, transmission, ou simplement enregistrement peut être réduit

d'une façon notable si l'on accepte une limitation du spectre par un filtrage préalable. C'est le rôle du filtre de garde, dont la fréquence de coupure f_c est choisie en fonction de la fréquence d'échantillonnage retenue. Pour la téléphonie, on estime que le signal garde une qualité suffisante lorsque son spectre est limité à 3400 Hz et l'on choisit f_e= 8000 Hz. Pour les techniques d'analyse, de synthèse ou de reconnaissance de la parole, la fréquence peut varier de 6000 à 16000 Hz. Par contre pour le signal audio (parole et musique), on exige une bonne représentation du signal jusque 20 kHz et l'on utilise des fréquences d'échantillonnage de 44.1 ou 48 kHz. Pour les applications multimédia, les fréquences sous-multiples de 44.1 kHz sont de plus en plus utilisées : 22.5 kHz, 11.25 kHz. Parmi le continuum des valeurs possibles pour les échantillons $x(nT_e)$, la quantification ne retient qu'un nombre fini 2b de valeurs (b étant le nombre de bits de la quantification), espacées du pas de quantification q. Le signal numérique résultant est noté *x(n).* La quantification produit une erreur de quantification qui normalement se comporte comme un bruit blanc; le pas de quantification est donc imposé par le rapport signal à bruit à garantir. Si le pas de quantification est constant, ce rapport est fonction de l'amplitude du signal; les signaux de faible amplitude sont dès lors mal représentés. Aussi adopte-t-on pour la transmission téléphonique une loi de quantification logarithmique et chaque échantillon est représenté sur 8 bits (256 valeurs). Par contre, la quantification du signal musical exige en principe une quantification linéaire sur 16 bits (65536 valeurs) [14]. Une caractéristique essentielle qui résulte du mode de représentation est le débit binaire, exprimé en bits par seconde (b/s), nécessaire pour une transmission ou un enregistrement du signal vocal. La transmission téléphonique classique exige un débit de 8 kHz x 8 bits = 64 kb/s; la transmission ou l'enregistrement d'un signal audio exige en principe un débit de l'ordre de 48 kHz x 16 bits = 768 kb/s (à multiplier par deux pour un signal stéréophonique) [14]. La figure II.9 représente l'évolution temporelle, ou audiogramme, du signal vocal pour les mots 'gauche', et 'bas'.

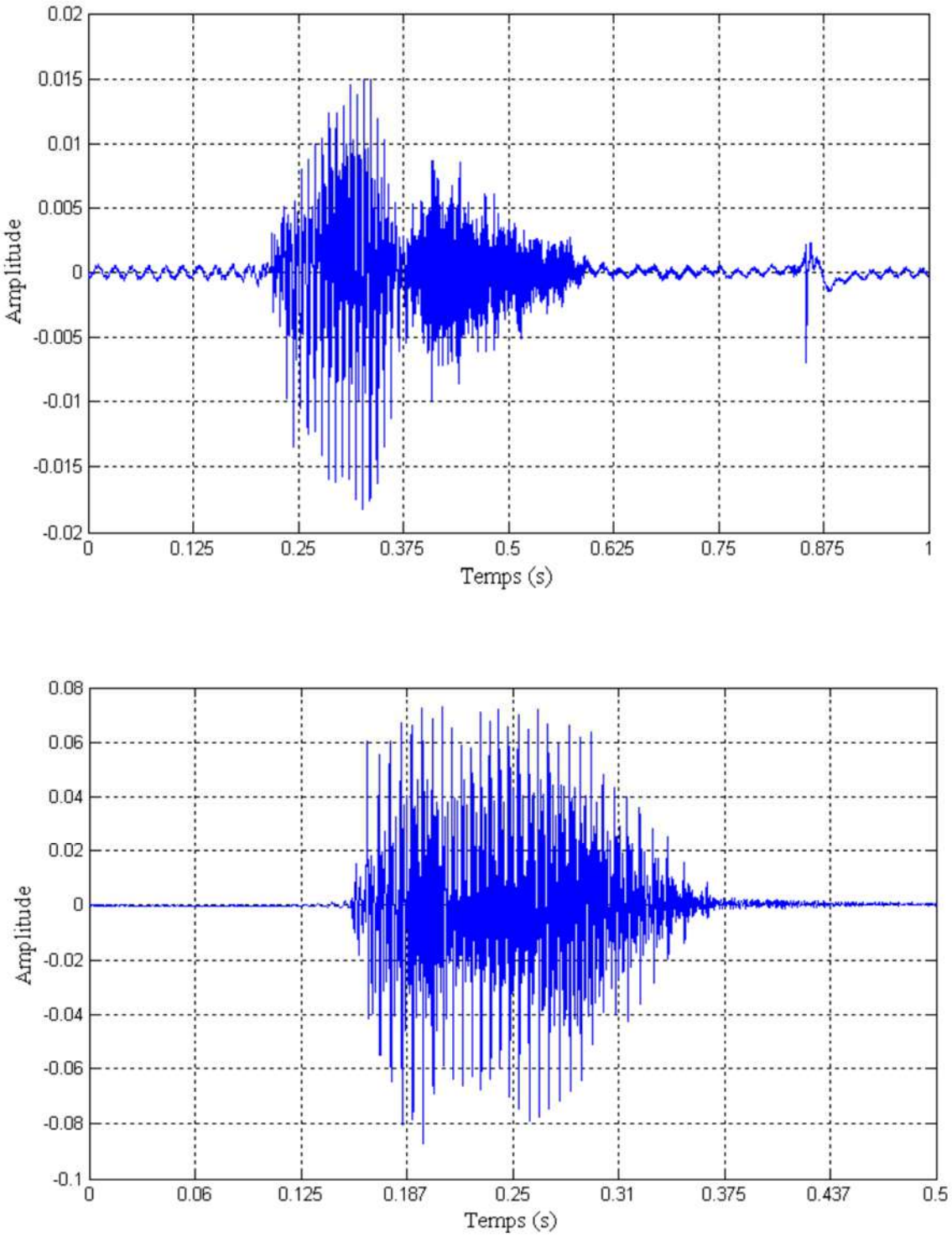


Figure II.9. *Audiogramme de signaux de parole en haut le mot « gauche» et en bas le mot « bas »*

II.3.2.Transformée de Fourier à court terme

La transformée de Fourier à court terme est obtenue en extrayant de l'audiogramme une 30aine de ms de signal vocal, en pondérant ces échantillons par une fenêtre de pondération (souvent une fenêtre de Hemming) et en effectuant un transformée de Fourier sur ces échantillons [14].

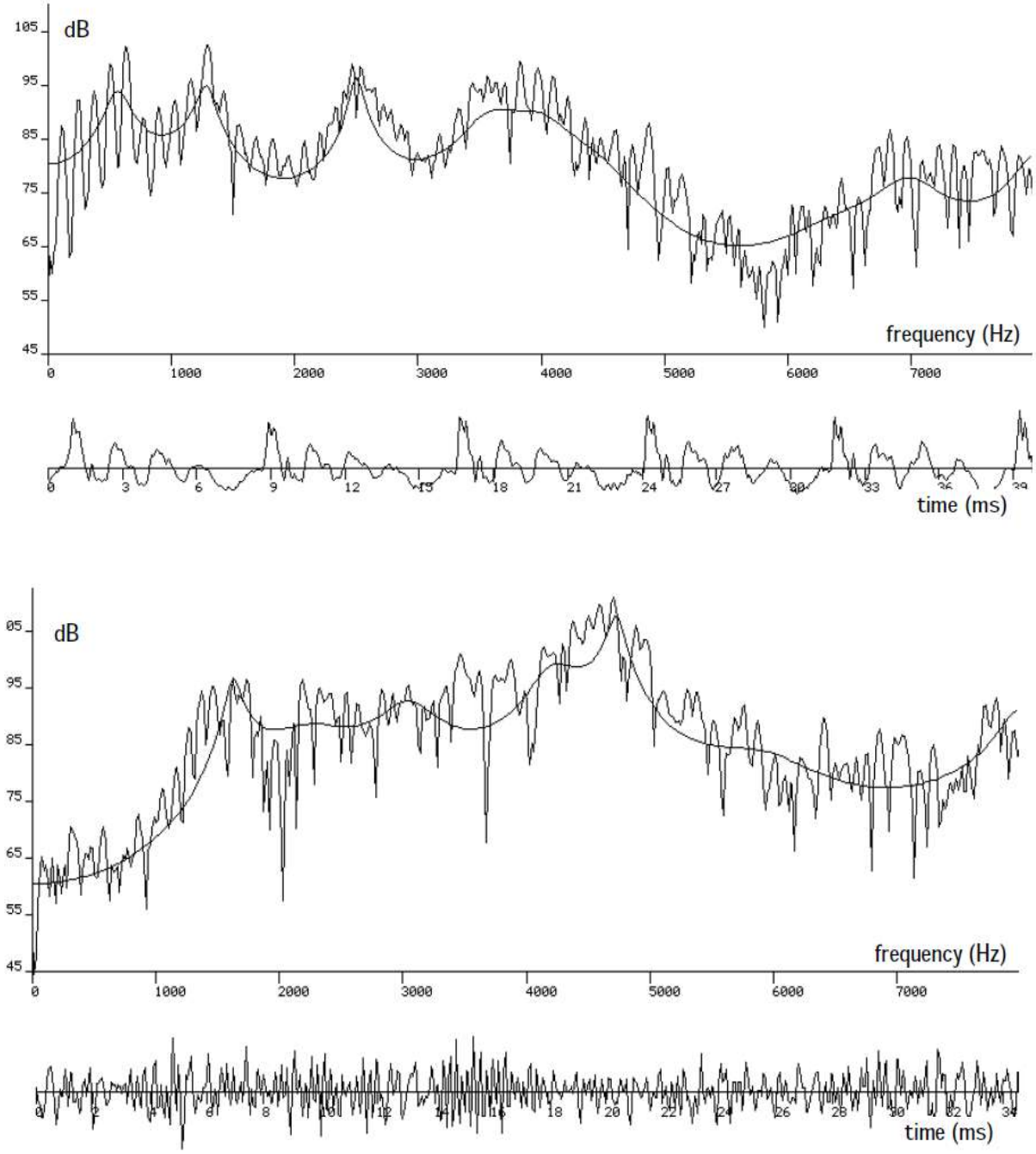


Figure II.10. *Evolution temporelle (en haut) et transformée de Fourier discrète (en bas) du [a] et du [ʃ] de 'baluchon' (signaux pondérés par une fenêtre de Hamming de 30 ms) [14].*

La figure. II.10 illustre-la transformée de Fourier d'une tranche voisée et celle d'une tranche non-voisée. Les parties voisées du signal apparaissant sous la forme de successions de pics spectraux marqués, dont les fréquences centrales sont multiples de la fréquence fondamentale. Par contre, le spectre d'un signal non voisé ne présente aucune structure particulière. La forme générale de ces spectres, appelée enveloppe spectrale, présente elle-même des pics et des creux qui correspondent aux résonances et aux anti-résonances du conduit vocal et sont appelés formants et anti-formants. L'évolution temporelle de leur fréquence centrale et de leur largeur de bande détermine le timbre du son. Il apparaît en pratique que l'enveloppe spectrale des sons voisés est de type passe- bas, avec environ un formant par kHz de bande passante, et dont seuls les trois ou quatre premiers contribuent de façon importante au timbre. Par contre, les sons non-voisés présentent souvent une accentuation vers les hautes fréquences.

II.3.3. Spectrogramme

Il est souvent intéressant de représenter l'évolution temporelle du spectre à court terme d'un signal, sous la forme d'un spectrogramme. L'amplitude spectre y apparaît sous la forme de niveaux de gris dans un diagramme en deux dimensions temps-fréquence. On parle de spectrogramme à large bande ou à bande étroite selon la durée de la fenêtre de pondération (Figure II.11). Les spectrogrammes à bande large sont obtenus avec des fenêtres de pondération de faible durée (typiquement 10 ms); ils mettent en évidence l'enveloppe spectrale du signal, et permettent par conséquent de visualiser l'évolution temporelle des formants. Les périodes voisées y apparaissent sous la forme de bandes verticales plus sombres. Les spectrogrammes à bande étroite sont moins utilisés. Ils mettent plutôt la structure fine du spectre en évidence : les harmoniques du signal dans les zones voisées y apparaissent sous la forme de bandes horizontales [14].

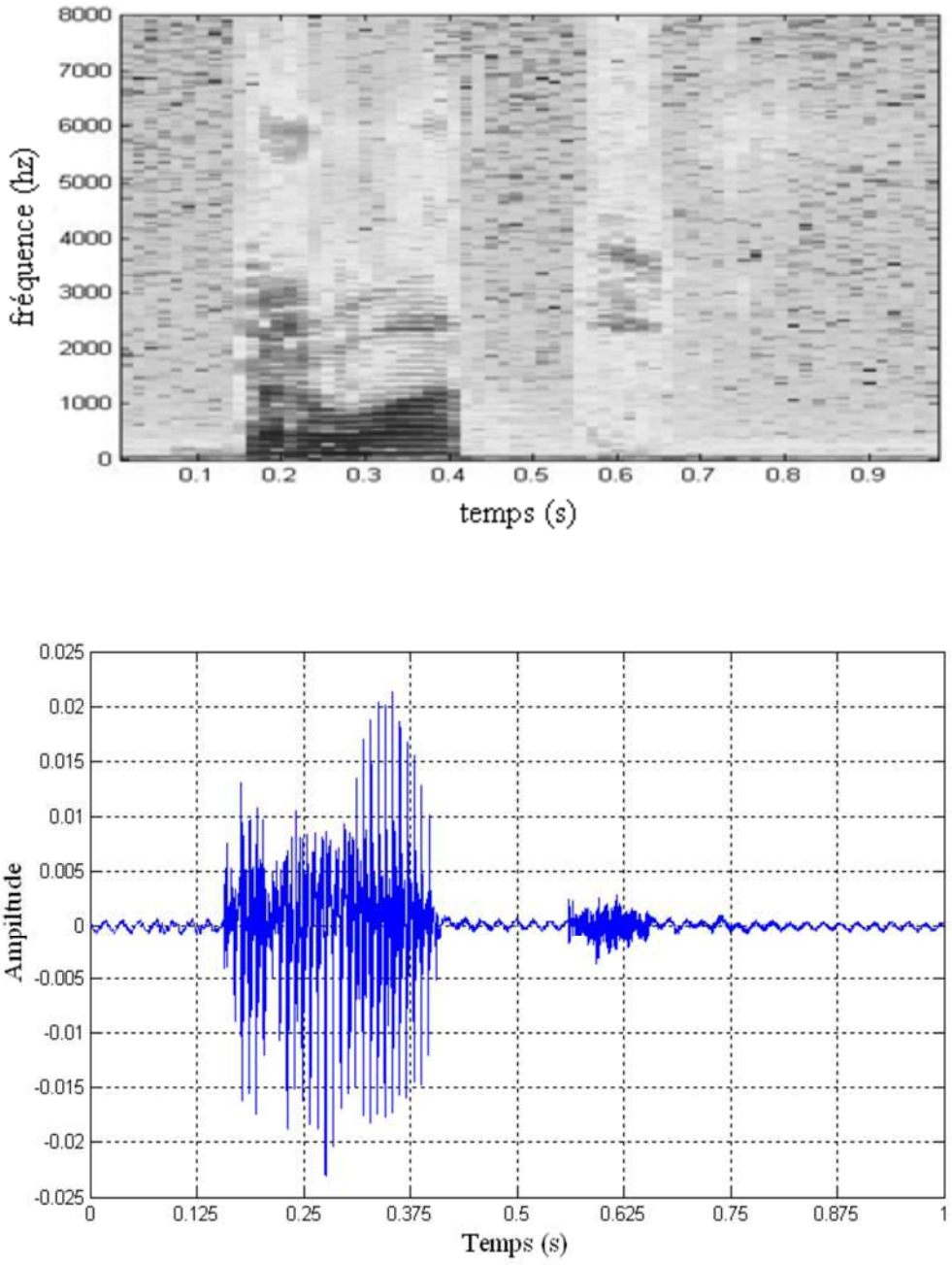


Figure II.11. *Spectrogrammes à large bande (en haut), à bande étroite (en bas), et évolution temporelle du mot 'Droite', échantillonnée à 16 kHz (calcul avec fenêtres de Hamming de 10 et 30 ms respectivement.*

II.3.4. Fréquence fondamentale

Une analyse d'un signal de parole n'est pas complète tant qu'on n'a pas mesuré l'évolution temporelle de la fréquence fondamentale ou pitch. En acoustique, la fréquence fondamentale ou son fondamental est l'harmonique de premier rang d'un son.

La figure II.12 donne l'évolution temporelle de la fréquence fondamentale de la phrase "les techniques de traitement de la parole". On constate qu'à l'intérieur des zones voisées la fréquence fondamentale évoluent lentement dans le temps. Elle s'étend approximativement de 70 à 250 Hz chez les hommes, de 150 à 400 Hz chez les femmes, et de 200 à 600 Hz chez les enfants [14].

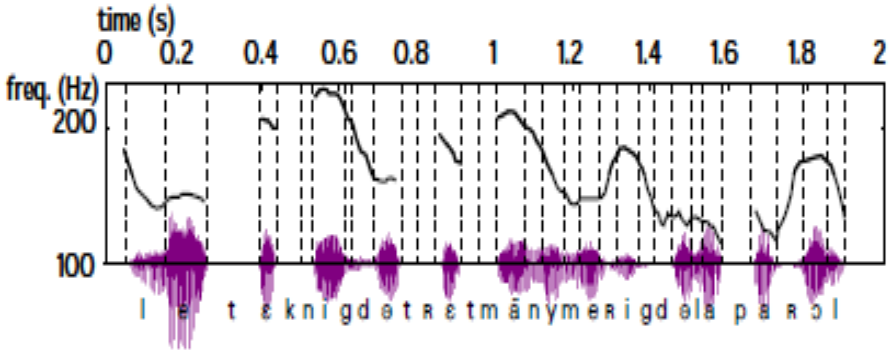


Figure II.12. *Evolution de la fréquence de vibration des cordes vocales dans la phrase "les techniques de traitement numérique de la parole". La fréquence est donnée sur une échelle logarithmique; les* sons non-voisés sont associés à une *fréquence nulle*

II.4. Synthèse de parole

Par définition Un système de synthèse à partir du texte (TTS : Text-To-Speech) est une machine capable de lire a priori n'importe quel texte à voix haute, que ce texte ait été directement introduit par un opérateur sur un clavier alpha-numérique, qu'il ait été scanné et reconnu par un système de reconnaissance optique des caractères (OCR : Optical Character Recognition), ou qu'il ait été produit automatiquement par un système de dialogue homme-machine. Un tel système diffère fondamentalement d'autres machines parlantes en ceci qu'il est destiné à donner lecture de phrases qui n'ont en principe jamais été lues auparavant. Il est

en effet possible de produire automatiquement de la parole en concaténant simplement des mots ou des parties de phrases préalablement enregistrées, mais il est clair dans ce cas que le vocabulaire utilisé doit rester très limité et que les phrases à produire doivent respecter une structure fixe, afin de maintenir dans des limites raisonnables la quantité de mémoire nécessaire à stocker les éléments vocaux de base. C'est le cas par exemple des annonceurs vocaux automatiques dans les gares. On définira donc plutôt la synthèse TTS comme la production automatique de phrases par calcul de leur transcription phonétique.

II.4.1 Applications

Les applications des systèmes de synthèse à partir du texte ne manquent pas. En voici quelques exemples :

- **Services de télécommunications :**

La libéralisation du marché des télécommunications en Europe a récemment rendu les opérateurs de télécommunications plus sensibles au confort de leurs clients. En particulier, on cherche désormais à fournir un maximum de services, à moindre coût. Les synthétiseurs permettent précisément de rendre tout type d'information écrite disponible via le téléphone. On peut ainsi créer des serveurs vocaux diffusant les horaires des cinémas, des informations routières, l'état d'un compte en banque, ou encore des explications automatisées concernant la dernière facture de téléphone. Les requêtes se font soit par la voix (en combinant le synthétiseur avec un reconnaisseur), soit par le clavier du téléphone. AT&T a récemment testé certains services de ce type auprès de ses clients, et constaté un réel engouement, à condition que l'intelligibilité des voix de synthèse soit suffisante; il s'est avéré que le naturel n'est pas un facteur déterminant pour la plupart de ces services.

- **Apprentissage (ou perfectionnement) de langues étrangères :**

Une synthèse de très bonne qualité couplée à un logiciel d'apprentissage constitue un outil très utile à l'apprentissage d'une nouvelle langue, en complément d'un cours avec un professeur. Si ce type de produit n'a pas encore percé sur le marché, c'est à cause de la mauvaise qualité des voix disponibles jusqu'à il y a peu. On voit par contre se multiplier les petits dictionnaires électroniques de poche, qui devraient rapidement être dotés de voix de synthèse. Il en va de même des traducteurs électroniques mot-à-mot qui sont apparus récemment. On pourra par exemple bientôt lire un ouvrage dans un langue étrangère et utiliser

un stylo à lecture optique (intégrant un mini-scanner) pour obtenir instantanément la traduction d'un mot inconnu et sa prononciation.

- **Aide aux personnes handicapées :**

Les handicaps liés à la parole sont soit d'origine mentale, soit d'origine motrice ou sensorielle. La machine peut être d'un grand secours dans le second cas. Avec l'aide d'un clavier spécialement adapté et/ou d'un logiciel d'assemblage rapide de phrases, un handicapé peut s'exprimer par la voix de son synthétiseur. Le célèbre astrophysicien Stephen Hawking donne tous ses cours à l'université de Cambridge de cette façon. La synthèse offre également des services aux personnes malvoyantes, en leur donnant accès à l'information écrite "en noir"16, à condition de coupler le synthétiseur à un logiciel de reconnaissance des caractères.

- **Livre et jouets parlants :**

Le marché du jouet a déjà été touché par la synthèse vocale. De nombreux ordinateurs pour enfants possèdent une sortie vocale qui en augmente l'attrait, particulièrement chez les jeunes enfants (pour qui la voix est le seul moyen de communication avec la machine).

- **Monitoring vocal :**

Dans certains cas, l'information orale est plus efficace qu'un message écrit. L'utilisation d'une voix de synthèse dans un centre de contrôle de site industriel, par exemple, permet d'attirer l'attention du personnel de surveillance sur un problème urgent. De la même manière, l'intégration d'un synthétiseur dans la cabine de pilotage d'un avion permet d'éviter au pilote d'être dépassé par la quantité d'informations visuelles qu'il a à analyser. Et quand on voit à quoi ressemblera bientôt le tableau de bord de nos voitures, on comprend qu'elles ne tarderont pas à nous parler. La maison de demain, quant à elle, fera bien de se doter d'une voix de synthèse si elle veut avertir ses occupants d'une anomalie constatée sur un de ses circuits de surveillance.

- **Communication homme-machine, multimédia :**

A plus long terme, le développement de synthétiseurs de haute qualité (ainsi que la mise au point de reconnaisseurs fiables et robustes) permettra à l'homme de communiquer avec la machine de manière plus naturelle. L'explosion récente du marché du multimédia prouve bien l'intérêt du grand public en la matière.

- **Recherche fondamentale et appliquée :**

Enfin, les synthétiseurs possèdent aux yeux des phonéticiens une qualité qui nous fait défaut : ils peuvent répéter deux fois exactement la même chose. Ils sont par conséquent utiles pour la validation de théories relatives à la production, à la perception, ou à la compréhension de la parole.

II.4.2. Le modèle prédictif Linéaire du signal vocal

L'analyse de la parole est une étape indispensable à toute application de synthèse, de codage, ou de reconnaissance. Elle repose en général sur un modèle. Celui-ci possède un ensemble de paramètres numériques, dont les plages de variation définissent l'ensemble des signaux couverts par le modèle. Pour un signal et un modèle donné, l'analyse consiste en l'estimation des paramètres du modèle dans le but de lui faire correspondre le signal analysé. Pour ce faire, on met en œuvre un algorithme d'analyse, qui cherche généralement à minimiser la différence, appelée erreur de modélisation, entre le signal original et celui qui serait produit par le modèle s'il était utilisé en tant que synthétiseur (Figure II.13) [14].

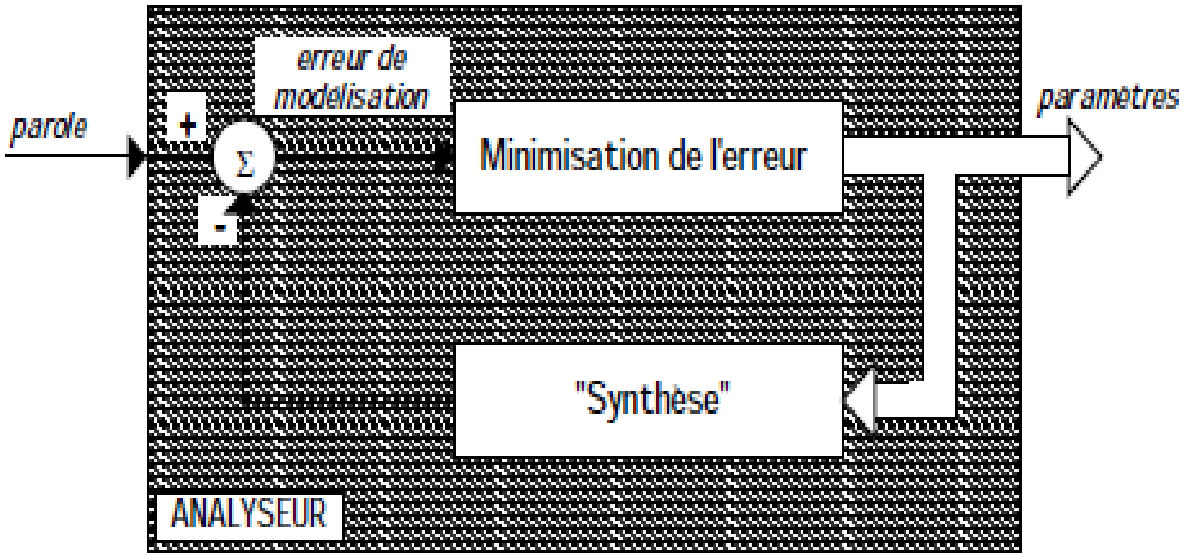


Figure II.13. *Schéma de principe d'un analyseur de parole. En pratique, l'étape De synthèse peut être implicite.*

S'il existe de nombreux modèles de parole, il en est un que l'on retrouve partout, et dans un nombre croissant d'appareils « grand-public » : le modèle prédictif linéaire (LPC : Linear Predictive Coding).

De la même façon qu'un signal de parole réel est produit par le passage, à travers le filtre que constitue notre conduit vocal, d'un signal d'excitation créé par les poumons et les cordes vocales, ce même signal de parole peut être modélisé par le passage d'un signal d'excitation numérique à travers un filtre numérique récursif. Le signal d'excitation sera tantôt une suite d'impulsions numériques (qui serviront à simuler les impulsions de débit créées par les cordes vocales) tantôt du bruit numérique (qui reproduira le souffle poussé par les poumons).

Ce modèle est appelé « prédictif linéaire » en raison du fait qu'il correspond à une régression linéaire très simple entre le signal d'excitation et le signal vocale produit. Les coefficients de cette régression linéaire sont les coefficients du filtre numérique récursif.

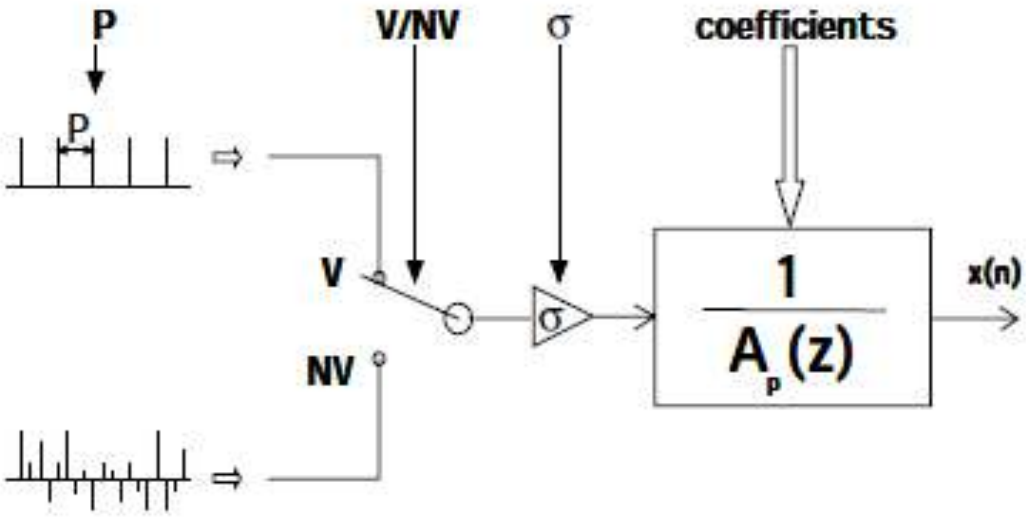


Figure II .14. *Le modèle auto-régressif.*

Les paramètres du modèle LPC sont tout simplement : la période du train d'impulsions (sons voisés uniquement), la position de l'interrupteur Voisé/Non Voisé (V/NV), le gain de l'amplificateur σ, et les coefficients du filtre numérique de synthèse.

Le problème de l'estimation d'un modèle LPC, souvent appelée analyse LPC revient à déterminer les coefficients du filtre connaissant le signal de sortie, mais pas l'entrée. Il est par

conséquent nécessaire d'adopter un critère, afin de faire un choix parmi l'infinité de solutions possibles. Le critère classiquement utilisé est celui de la minimisation de l'énergie de l'erreur de prédiction. La mise en équation de ce problème conduit aux équations dites de Yule-Walker :

$\phi_a = -\varphi$

Avec

$\varphi = [\varphi_{xx}(1), \varphi_{xx}(2), \dots\dots. \varphi_{xx}(p)]^T$

$a = [a_1, a_2, \dots\dots., a_p]^T$

$$\phi = \begin{bmatrix} \phi_{xx}(0) & \phi_{xx}(1) & \phi_{xx}(2) & \dots & \phi_{xx}(p-1) \\ \phi_{xx}(1) & \phi_{xx}(0) & \phi_{xx}(1) & \dots & \phi_{xx}(p-2) \\ \phi_{xx}(2) & \phi_{xx}(1) & \phi_{xx}(0) & \dots & \phi_{xx}(p-3) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \phi_{xx}(1) & \phi_{xx}(p-2) & \phi_{xx}(p-3) & \cdots & \phi_{xx}(0) \end{bmatrix}$$

Les coefficients de la matrice et du terme indépendant du système sont les coefficients d'auto corrélation du signal à modéliser. Les inconnues sont les coefficients du filtre de synthèse, dont le nombre, *p*, est typiquement de l'ordre de 10. Le signal de parole étant fortement non-stationnaire, ce type de modélisation ne reste guère valable plus d'une dizaine de millisecondes. On retiendra donc que l'analyse LPC d'un signal de parole implique la résolution d'un système de 10 d'équations à 10 inconnues toutes les 10 ms**.**

II.5. Codage LPC

La figure ci-dessous donne le schéma de principe d'un codeur LPC, tel qu'il peut être utilisé pour les transmissions de voix par satellite (ex : voix d'un journaliste en mission dans un pays lointain) ou plus commun ément dans un GSM. Le signal vocal mesuré par le micro est découpé en trames, analysé par l'algorithme de Schur et par un algorithme d'analyse de la fréquence des cordes vocales. Les paramètres qui en résultent sont quantifiés, c.-à-d. qu'ils sont codés sur un ensemble fini de nombres entiers (ce qui permet d'associer à chaque paramètre un nombre fini de bits par trame).

En d'autres termes, lors d'un appel par GSM, le GSM émetteur (qui n'est rien d'autre qu'un ordinateur de poche spécialisé dans l'analyse, le codage, le décodage, et la synthèse LPC) enregistre la parole à transmettre, en réalise toutes les 10 ms une analyse LPC (par laquelle il trouve les coefficients de prédiction qui « collent » le mieux au conduit vocal de l'appelant, pour la tranche de parole considérée), et transmet ces coefficients (et non la voix originale de l'appelant). Le GSM récepteur reçoit quant à lui les paramètres du conduit vocal de l'appelant, produit un signal de synthèse simulant ce conduit vocal, et le fait entendre au correspondant, qui croit entendre l'appelant. Il s'agit pourtant bien de parole de synthèse, au même titre qu'on pourrait imaginer une caméra inspectant l'appelant et ne transmettant d'un modèle 3D de don visage, lequel serait reproduit en image de synthèse coté récepteur.[14]

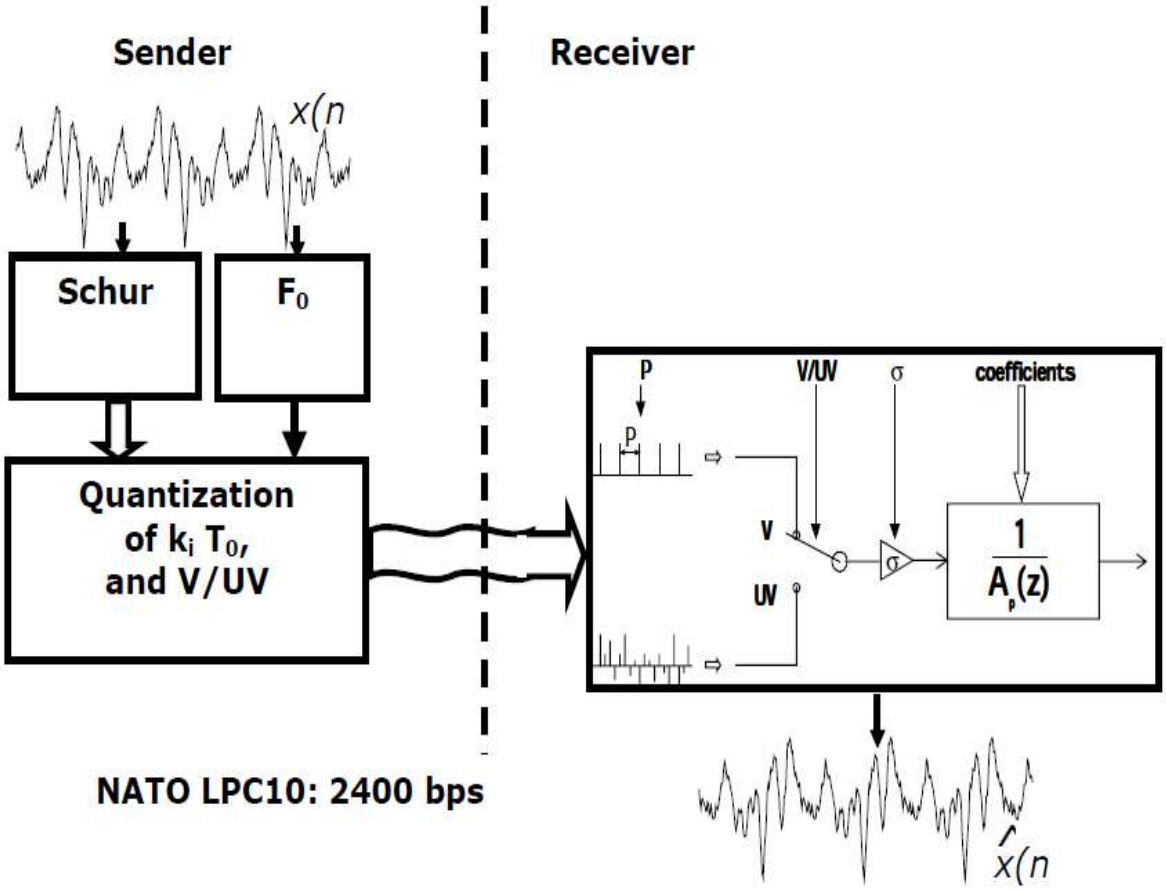


Figure II.15. *Transmission de parole basée sur le codage LPC*

II.6. Conclusion

Le traitement de la parole est aujourd'hui une composante fondamentale des sciences de l'ingénieur, mais avant le traitement il faut savoir les caractéristiques et le comportement du signal de parole. Dans ce chapitre, nous avons passé en revue les principales caractéristiques acoustique-phonétiques du signal de parole. Au niveau phonétique nous avons vu l'architecture de l'appareil vocal, le mécanisme de phonation et l'audition…etc. Au niveau acoustique on s'est intéressé essentiellement par l'évolution temporel du signal de parole, sa transformer de Fourier à court terme et sa fréquence fondamentale et la synthèse de parole.

Chapitre III

La reconnaissance de la parole

III.1. Introduction

La reconnaissance automatique de la parole est un domaine d'étude actif depuis le début des années 50. Il est clair qu'un outil de reconnaissance de la parole efficace facilitera l'interaction entre les hommes et les machines .les application possible associées à un tel outil sont nombreuses et sont amenées à connaître un grand essor.la plupart des applications en reconnaissance de la parole peuvent être regroupées en quatre catégories : commande et contrôle, accès à des bases de données ou recherche d'informations, dictée vocale et transcription automatique de la parole.

La technologie la plus utilisée depuis plus de 20 ans est basée sur des modèles statistique les modèles de Markov cachés (en anglais Hidden Markov Modèles : HMM) capable de modéliser simultanément les caractéristiques fréquentielles et temporelles du signal de parole depuis l'introduction de ces modèles, de nombreux progrès ont été réalisés dans le domaine de la reconnaissance de la parole.

III.2. Le traitement automatique de la parole

Le traitement automatique de la parole est un champ de recherche riche mais difficile en effet, le signal de la parole est complexe car sa structure résulte de l'interaction entre la production des sons et leur perception par l'oreille. Les chercheurs du domaine ont réussi malgré tout à réaliser quelque système de reconnaissance. Ou de synthèse exploitable dans des conditions réelles.

Dans les sections suivantes, nous nous intéressons au traitement automatique de la parole en vue de sa reconnaissance .nous commençons tout d'abord par une vue sur les niveaux de complexité liées a sa reconnaissance. Ensuite nous détaillons les différentes étapes de son traitement qui visent à extraire des coefficients MFCC (Mel Freconcy Castrum Coefficient) utilisées dans la majorité des systèmes de reconnaissance fondée sur l'utilisation des modèles de Markov caches [17].

III.2.1. Niveaux de complexité

Pour bien appréhender le problème de la reconnaissance automatique de la parole, il est bon d'en comprendre les différents niveaux de complexités et les différents facteurs qui en font un problème difficile [14].

Il y a d'abord le problème de la variabilité intra et interlocuteurs. Le système est-il dépendant des locuteurs (optimise pour un locuteur bien particulier) ou indépendant du locuteur (pouvant reconnaitre n'importe quel utilisateur) ?

Evidement, les systèmes dépendant des locuteurs sont plus faciles a développer et sont caractérise par de meilleurs taux de reconnaissance que le système indépendants du locuteur étant donne que la variabilité du signal de parole est plus limitée. Cette dépendance au locuteur est cependant acquise au prix d'un entrainement spécifique à chaque utilisateur. Ceci n'est cependant pas toujours possible. Par exemple, dans le cas d'application téléphonique, il est évident que les système doivent pouvoir être utilise par n'importe qui et doivent donc être indépendantes du locuteur est cependant obtenue par l'acquisition de nombreux locuteurs (couvant si possible les différent dialectes) qui sont utilisées simultanément pour l'entrainement de modèles susceptible d'en extraire toutes les caractéristiques majeures.une solution intermédiaire parfois utilisée et de développer des systèmes capable de s'adapter (de façon supervisée ou non supervisée) rapidement au nouveau locuteur.

Le système reconnaît-il des mots ou de isolés bien sépares par des périodes de silence que de reconnaître la séquence de mots constituant une phrase. En effet, dans ce dernier cas, non seulement la frontière entre mots n'est plus connue mais, de plus, les mots deviennent fortement articulés (c'est-à-dire que la prononciation de chaque mot est affectée par le mot qui précède ainsi que par celui qui suit –un exemple simple et bien connu étant les liaisons du français).

Dans le cas de la parole continue, le niveau de complexité varie également selon qu'il s'agisse de texte lu, de texte parlé ou, beaucoup plus difficile, de langage naturel avec ses hésitation. Phrases grammaticalement incorrectes, faux départs, etc. Un autre problème, qui commence à être bien maîtrisé, concerne la reconnaissance de mots clés en parole libre. Dans ce dernier cas, le vocabulaire à reconnaître est relativement petit et bien défini mais le

locuteur n'est pas contraint de parler en mots isolés. Par exemple, si un utilisateur est invité à répondre par « oui » ou « non », il peut répondre « oui, s'il vous plaît ». Dans ce contexte, un problème qui reste particulièrement difficile est le rejet de phrases ne contenant aucuns mots clés.

La taille du vocabulaire et son degré de confusion sont également des facteurs importants. Les petits vocabulaires sont évidemment plus faciles à reconnaître que les grands vocabulaires, étant donné que dans ce dernier cas, les possibilités de confusion augmentent. Certains petits vocabulaires peuvent cependant s'avérer particulièrement difficiles à traiter; ceci est le cas, par exemple, pour l'ensemble des lettres de l'alphabet, contenant surtout des mots très courts et acoustiquement proches [14].

III.2.2. Extraction des paramètres

Pour traiter un signal vocal pour la reconnaissance on doit d'abord extraire des paramètres qui caractérisent l'information caché derrière ce signal. Pour cela les coefficients mél centraux (MFCC) sont utilisés. Pour l'extraction de ces coefficients les étapes illustrées par la figure III-1 sont exécutées.

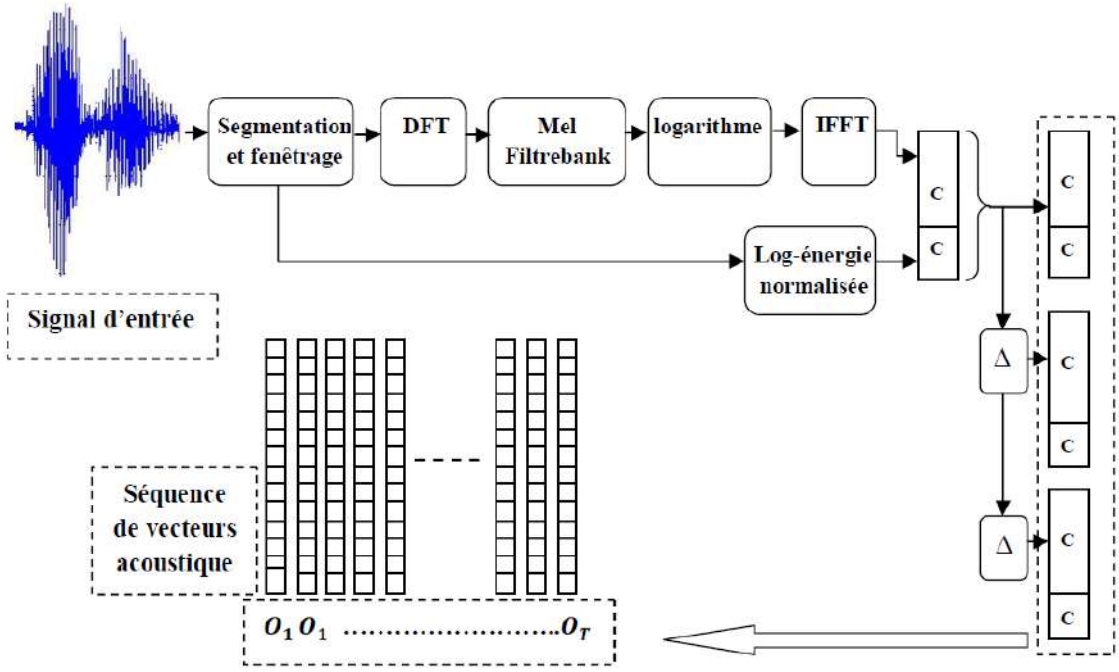


Figure III.1. *Les étapes d'une paramétrisation* MFCC.

III.2.2.1. Acquisition et préaccentuation du signal

L'acquisition du signal de parole constitue la première étape à franchir. Il s'agit de numériser un signal analogique (la parole) pour qu'il soit prêt à des traitements numériques ultérieurs. Cette étape est généralement réalisée à l'aide d'une carte d'acquisition spécialisée.

Une fois capté par un microphone, le signal est tout d'abord filtré, ensuite échantillonné et enfin quantifié. Ces opérations successives permettent de transformer un signal continu x(t) (où t désigne le temps) en un signal numérique x(n) où n correspond à des instants discrets [17].

Le format du signal ainsi obtenu doit être encore traité pour qu'on puisse l'exploiter à des fins de reconnaissance. Pour ce faire, une préaccentuation est tout d'abord effectuée pour relever les hautes fréquences [17].

Une analyse supplémentaire est par la suite appliquée au signal pour séparer le silence du signal de parole. Cette analyse et la détection automatique de la parole (VAD), dans le but de réduire la parti à traiter du signal et annuler la parti qui représente le silence. Le détecteur de parole sert à la discrimination des zones de parole et de pauses utilisant deux mesures: la puissance et le taux de passages par zéro.

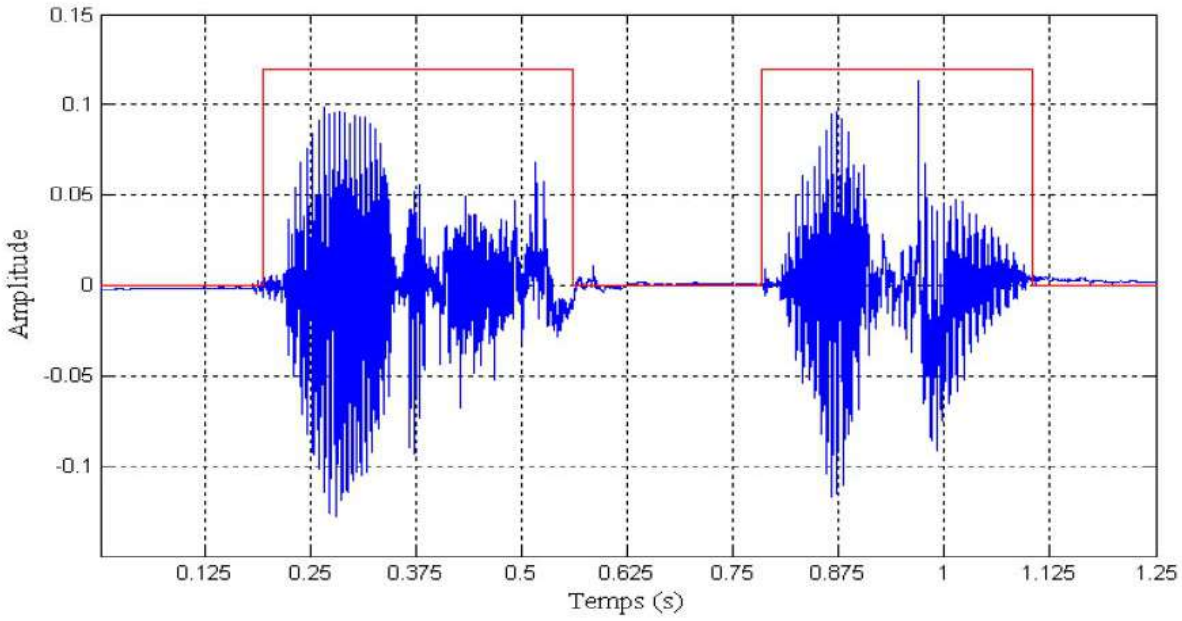


Figure III.2. *Détecteur appliqué à la phrase « marche avant »*

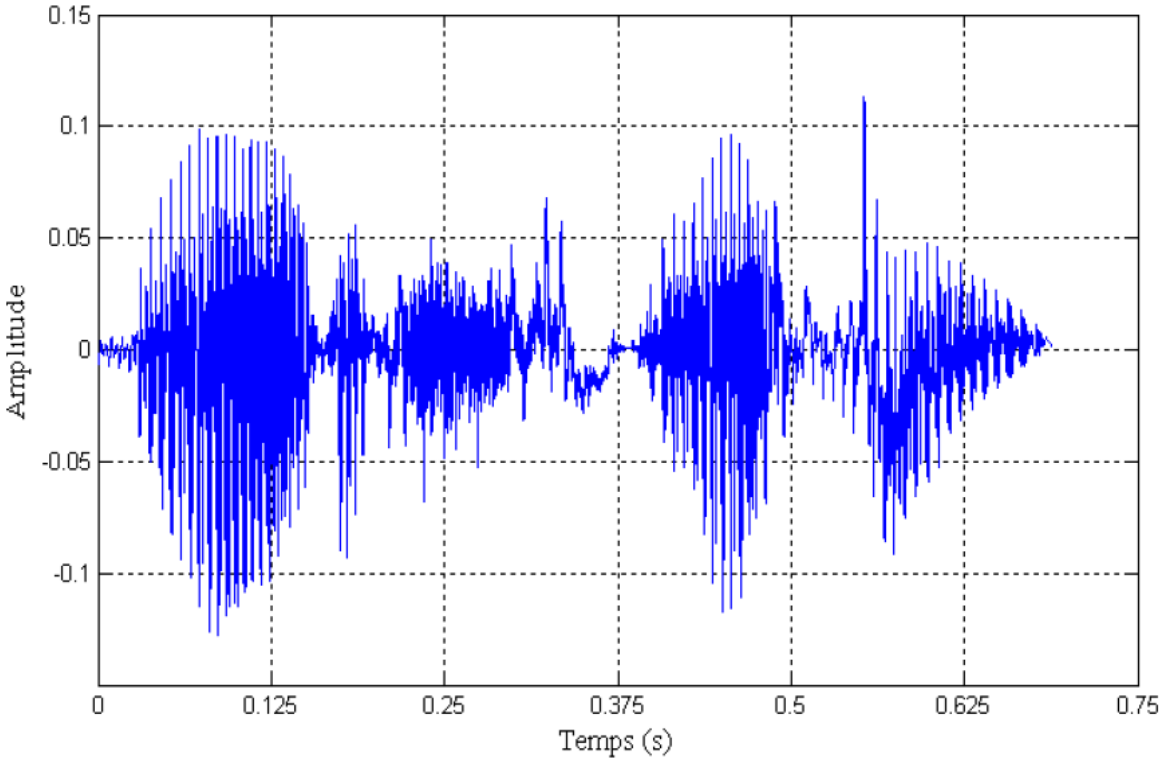


Figure III.3. *La phrase « marche avant » après la suppression du silence*

III.2.2.2. Segmentation et Fenêtrage

Le signal est segmenté en trames où chaque trame est constituée d'un nombre fixe K d'échantillons de parole. En général, K est fixé de telle manière que chaque trame corresponde à environ 30 ms de parole. Cette segmentation est réalisée à l'aide de fenêtres temporelles glissantes, de taille 256 ou 512 points. En général, les fenêtres successives se recouvrent de moitié de leur taille soit de 128 ou 256 points respectivement. Le découpage du signal en trames produit des discontinuités aux frontières des trames. Pour réduire ces problèmes, des fenêtres de pondération sont appliquées. Ce sont des fonctions que l'on applique à l'ensemble des échantillons prélevés dans la fenêtre du signal original de façon à diminuer les effets de bord. Parmi les fenêtres les plus courantes, nous pouvons citer la fenêtre de Hemming [17] :

$$w(n) = 0.54 + 0.46 \times \cos\left(2\pi \frac{n}{n-1}\right) \quad \textbf{(III. 1)}$$

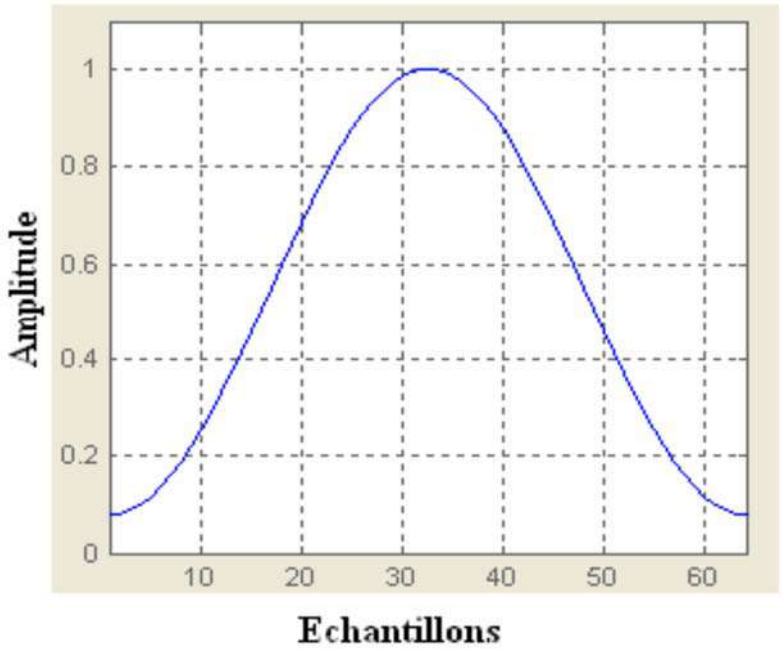


Figure III.4. *Fenêtre de Hemming*

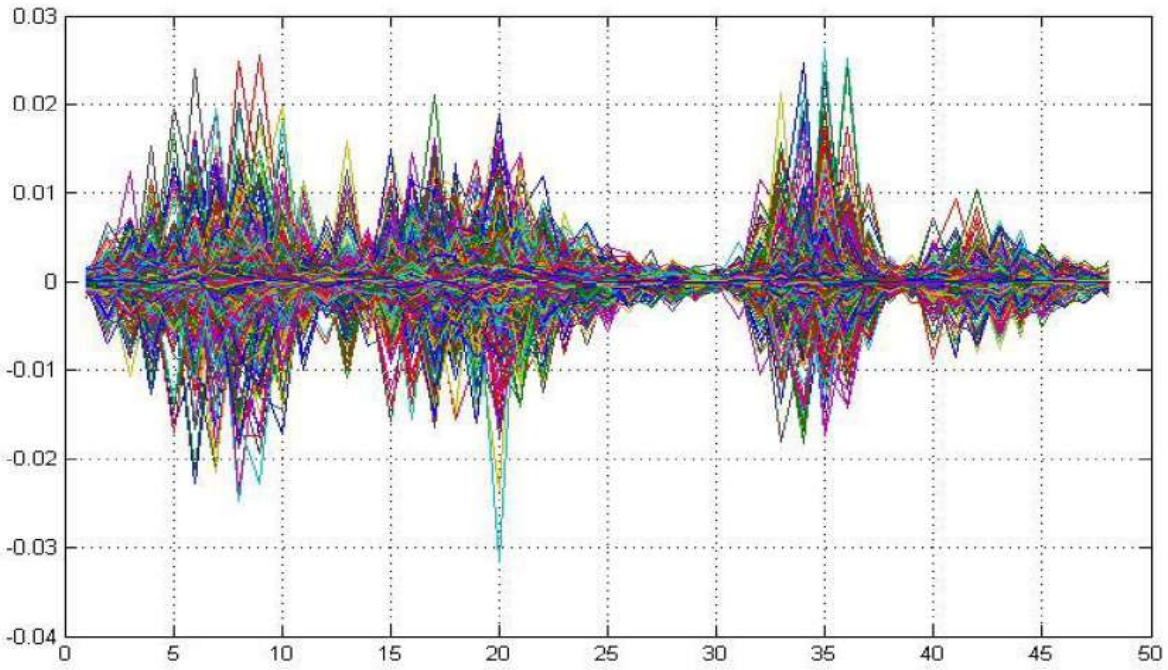


Figure III.5. *Représentation du signal de la phrase « marche avant » après La segmentation Et le fenêtrage.*

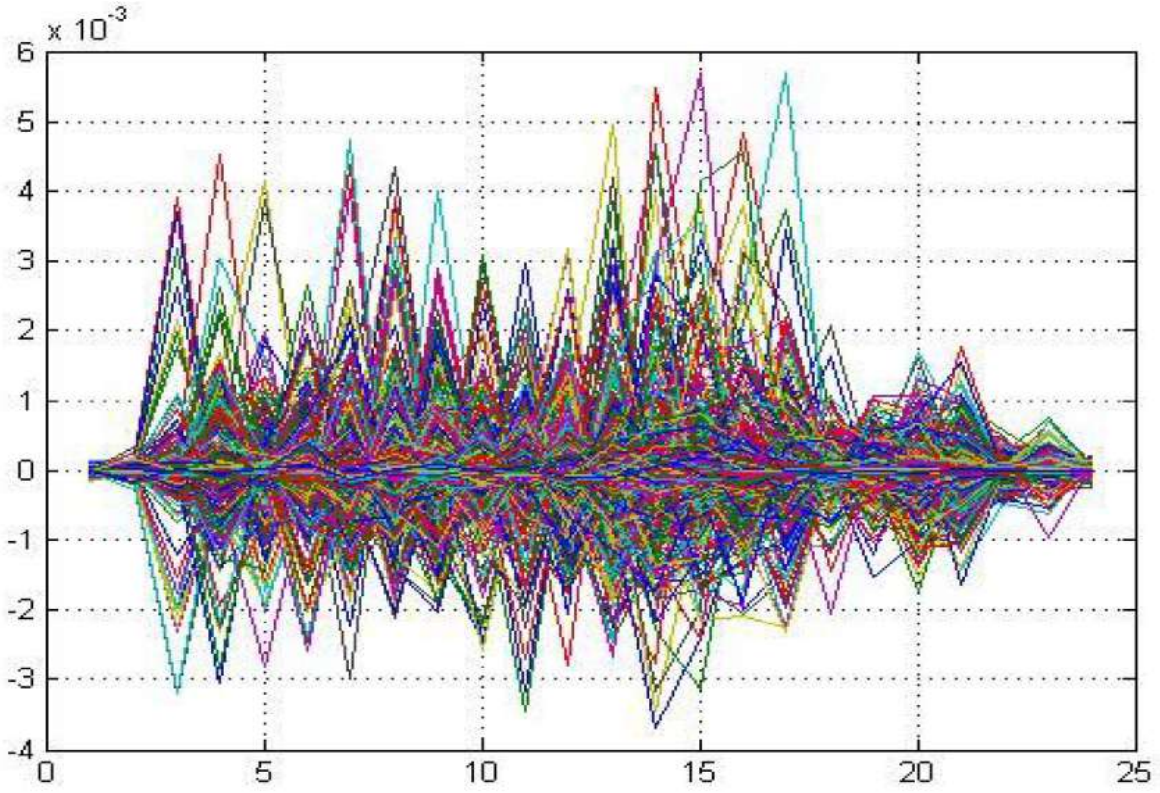


Figure III.6. *Représentation du signal du mot «droite » après la segmentation et Le fenêtrage.*

III.2.2.3. Transformée de Fourier discrète et l'énergie du signal

Après cette mise en forme du signal (commune à la plupart des méthodes d'analyse de la parole), des analyses temporelles ou spectrales peuvent être appliquées au signal. L'énergie du signal est le paramètre le plus intuitif utilisé pour caractériser le signal de parole. Elle correspond à la puissance du signal et est calculée directement dans le domaine temporel sur une trame de parole [17].

L'analyse spectrale reste cependant le moyen le plus utilisé pour caractériser le signal de parole. Elle permet de mettre en évidence certains phénomènes caractéristiques de la production de ce dernier. Les spectrogrammes ont été utilisés pour représenter la parole dès les années 40 en utilisant des filtres analogiques. Actuellement, les spectres sont obtenus numériquement par une transformée de Fourier discrète (DFT : Discret Fourier Transform), en particulier grâce à l'algorithme de transformée de Fourier rapide (FFT : Faste Fourier

transform). Une fois appliquée, cette transformée nous permet de passer du domaine temporel au domaine fréquentiel. La formule de la DFT est donnée par l'équation suivante :

$$x(n) = \frac{1}{n}\sum_{K=0}^{N-1} x(k)e^{-jk2\pi\frac{n}{N}} \quad \textbf{(III. 2)}$$

III.2.2.4. l'échelle Mel

Le nombre de paramètres spectraux calculés sur une trame par la FFT reste trop élevé pour un traitement automatique ultérieur. L'énergie du spectre est donc calculée à travers un banc de filtres numériques couvrant la bande passante, ce qui permet de ne conserver qu'un sous ensemble de ces paramètres. Les filtres triangulaires sont les plus utilisés. Ils sont préférés pour leur simplicité et leur effet de lissage sur le spectre. Ces filtres sont le plus souvent répartis sur l'échelle Mel qui est non linéaire. La relation entre la fréquence en échelle Hertz et sa correspondance en méls est la suivante :

$$M_{mels} = x.\log(1 + \frac{f_{HZ}}{y}) \quad \textbf{(III. 3)}$$

Où fHz est la fréquence, x = 2595 et y = 700. L'intérêt de l'échelle Mel est qu'elle est assez proche d'échelles issues d'études sur la perception sonore et sur les bandes passantes critiques de l'oreille, comme l'échelle Bark [18].

Les sorties du banc de filtres peuvent être utilisées directement comme paramètres d'un système de reconnaissance, comme dans le système d'IBM [19]. Cependant, dans les systèmes les plus récents, d'autres coefficients dérivés de ces sorties sont le plus souvent utilisés. Il s'agit des coefficients cepstraux MFCC. Ces coefficients sont linéairement répartis sur l'échelle Mel d'où leur nom. Ils sont considérés plus discriminants que les sorties des filtres et plus robustes au bruit ambiant; en plus ils sont moins corrélés entre eux [17].

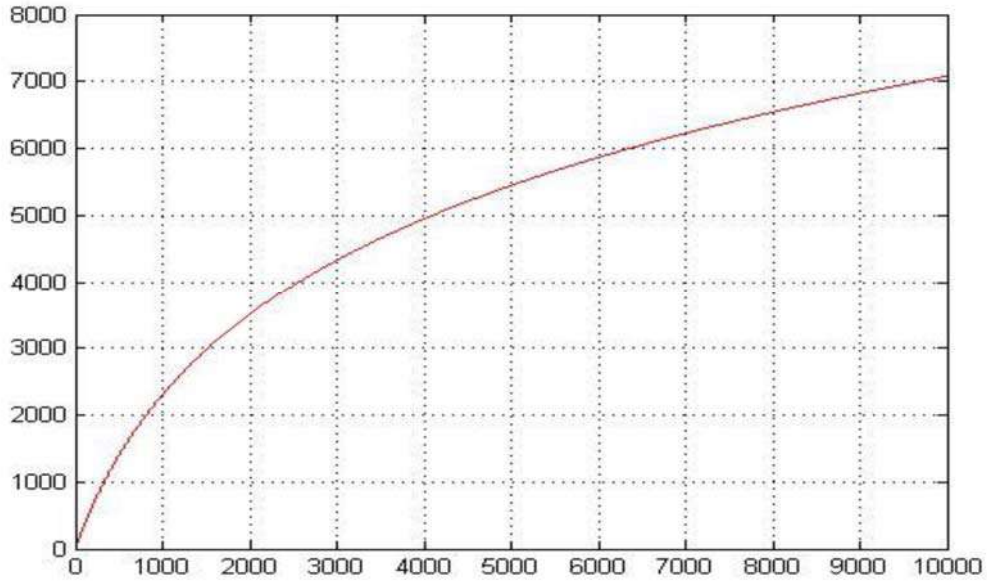


Figure III.7. *Echelle de Mel*

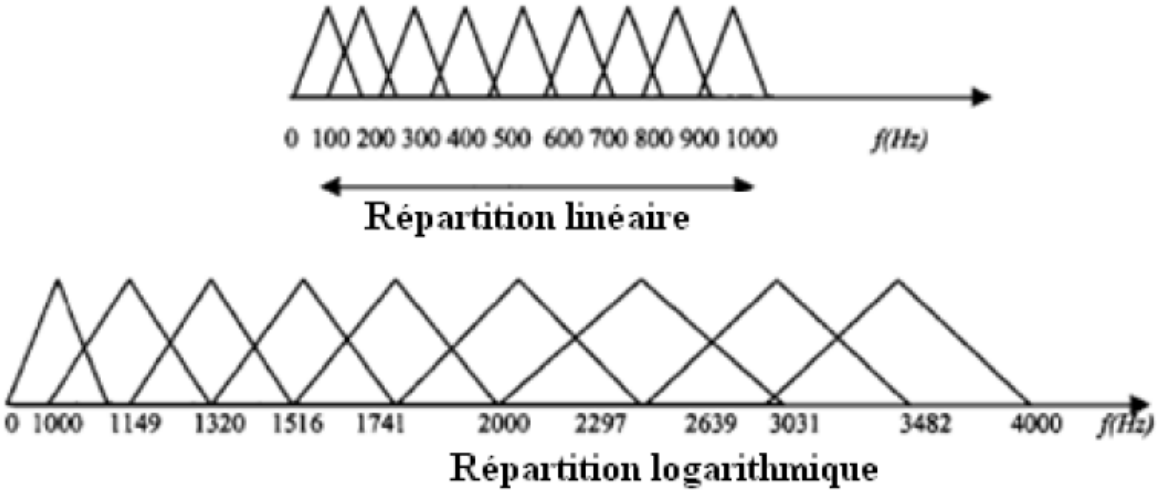


Figure III.8. *Type de banc filtre à échelle de Mel utilisant des filtres triangulaires.*

III.2.3. La reconnaissance

La reconnaissance automatique de la parole, vue comme un problème de la théorie de communication, a pour but de reconstruire un message W inconnu à partir d'une séquence

d'observations O [20]. Ceci revient à retrouver, parmi tous les messages possibles, celui qui selon toute vraisemblance, correspond à la suite d'observations acoustiques O. Cette dernière correspond à une suite de vecteurs permettant de caractériser le signal de parole. Le message W est une suite de mots prononcés. Il s'agit donc de trouver W qui correspond au message le plus probable connaissant la suite des observations acoustiques O :

$$\widehat{W} = arg_w \max \mathrm{P}(W|O) \quad \textbf{(III. 4)}$$

La probabilité P$(W|O)$ est très difficile à déterminer, d'où la nécessité de la décomposer. En utilisant la règle de Bayes, il est possible de reformuler la probabilité $P(W|O)$comme suit :

$$p(w|o) = \frac{p(w)p(O|W)}{p(o)} \quad \textbf{(III. 5)}$$

Puisque P (O) ne dépend pas de W, l'équation (III-6) sera équivalente à :

$$\hat{w} = arg_w max\ p(w)p(o|w) \quad \textbf{(III. 6)}$$

Ainsi, l'étape de reconnaissance consiste à déterminer la suite de mots W qui maximise le produit des deux termes P (W) et. Le premier terme représente la probabilité a priori d'observer la suite de mots W indépendamment du signal. Cette probabilité est déterminée par le modèle de langage. Le deuxième terme indique la probabilité d'observer la séquence de vecteurs acoustiques O sachant une séquence de mots spécifiques W. Cette probabilité est estimée par le modèle acoustique. La qualité d'un tel système de reconnaissance de la parole peut être caractérisée par la précision et la robustesse des deux modèles qui permettent de calculer ces deux termes P (W) et$P(O|W)$ Il s'agit donc d'intégrer les niveaux acoustiques et linguistiques dans un seul processus de décision permettant de retrouver le message prononcé.

La figure III.9 montre les différentes étapes nécessaires à la reconnaissance d'un message m prononcé en entrée. Tout d'abord, le signal de parole est subdivisé en vecteurs acoustiques. En utilisant ces vecteurs, le modèle acoustique se charge, à partir des HMM de phonèmes appris sur un corpus d'apprentissage, de construire la suite des phonèmes hypothèses du signal prononcé. Un seul modèle HMM représentant l'hypothèse, sera construit par la concaténation d'un ensemble de HMM de phonèmes. La suite de mots obtenue sera aussi évaluée par le modèle de langage qui permet d'estimer la probabilité p(w). En principe,

ce processus est répété pour toutes les hypothèses possibles. Le système donne enfin les N meilleures hypothèses comme résultat de la reconnaissance.

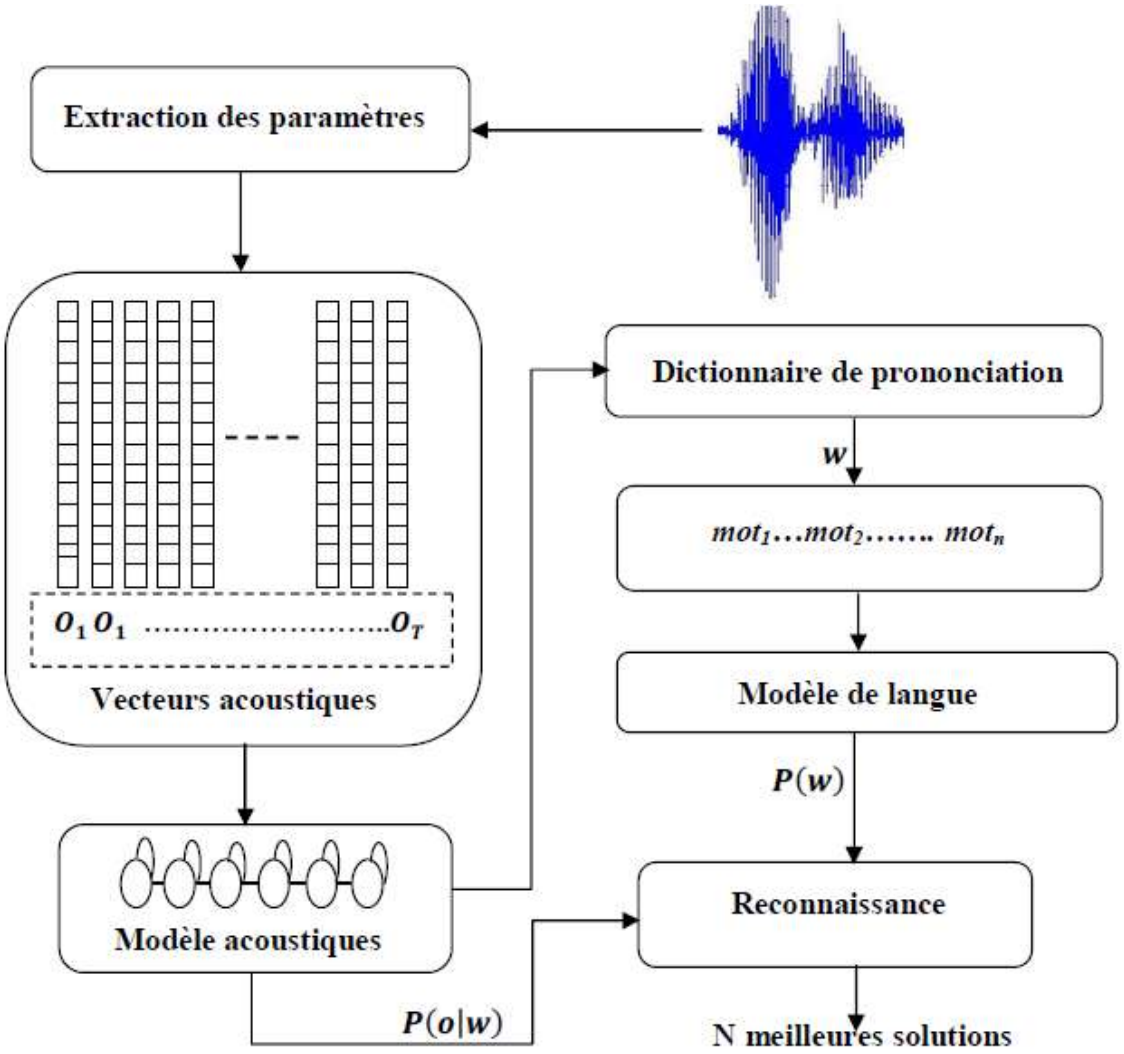


Figure III.9. *Le principe de fonctionnement de l'approche Statistique pour la reconnaissance de la parole.*

En conclusion, la reconnaissance de la parole passe par trois grandes étapes. La première concerne la modélisation acoustique du signal à l'aide des HMM. Ces derniers sont estimés à partir d'un corpus d'apprentissage, ils ont pour rôle de modéliser les phonèmes dans tous leurs contextes. Comme résultat, ils doivent pouvoir évaluer la probabilité P (O|W). La

deuxième étape s'intéresse à la modélisation du langage. Elle a pour but l'estimation de la probabilité P (W). Pour ce faire, les modèles de langage de type statistique sont le plus souvent utilisés. Ces modèles cherchent à estimer les probabilités d'apparition des mots en se basant sur leurs historiques. Enfin, la troisième étape utilise des algorithmes de recherche pour délivrer les *N* meilleures solutions. L'utilisation de tels algorithmes est nécessaire car il n'est pas possible avec les machines actuelles d'énumérer toutes les solutions pour trouver les *N* meilleures. Il est indispensable d'écarter des hypothèses et d'alléger ainsi le processus de recherche d'où la nécessité d'algorithmes précis et rapides pour accomplir cette tâche de sélection.

III.3. Modélisation acoustique à base de HMM

Les modèles de Markov cachés sont les plus communément utilisés en reconnaissance automatique de la parole pour estimer la probabilité P(O|W). Un modèle de Markov caché est un automate stochastique capable, après une phase d'apprentissage, d'estimer la probabilité qu'une séquence d'observations ait été générée par ce modèle. L'objectif des HMM en reconnaissance de la parole est de modéliser au mieux les unités représentatives du signal de reconnaissance de la parole est de modéliser au mieux les unités représentatives du signal de types d'unités comme les allophones, les syllabes, les tréphones, les mots, etc. Dans ce chapitre, nous nous intéressons seulement aux mots, plus d'informations sur les autres unités phonétiques peuvent être trouvées dans [21].

Dans la suite, nous commençons par définir mathématiquement les modèles de Markov cachés. Ensuite, nous mettons l'accent sur les problèmes fondamentaux des HMM sans nous attarder sur les détails algorithmiques des solutions apportées. Enfin, nous montrons comment mettre en pratique la théorie des HMM dans le cadre d'une application de reconnaissance de la parole.

III.3.1. Définitions des modèles de Markov cachés

Les modèles de Markov cachés sont utilisés notamment en reconnaissance de formes, en intelligence artificielle ou encore en traitement automatique du langage naturel. Chaque modèle est constitué de nœuds cachés et de nœuds d'observation.

Un HMM peut être vu comme un ensemble discret d'états et de transitions entre ces états. Formellement, il peut être défini par l'ensemble des paramètres λ [22] :

$$\lambda = (N, A, B, \pi) \quad \textbf{(III.7)}$$

Où :

- N est le nombre de nœuds ou d'états du modèle.
- $A = \{aij\} = \{\mathrm{P}(qj|qi)\}$ est une matrice de taille N × N. Elle contient les probabilités de transition sur l'ensemble des états du modèle. La probabilité de transition est la probabilité de choisir la transition aij pour accéder à l'état qj en partant de l'état q_i. Pour un HMM d'ordre 1, cette probabilité ne dépend que de l'état précédent :

  $$\mathrm{P}(q_t = j|q_{t-1} = i, q_{t-2} = k, \ldots) = \mathrm{P}(q_t = j|q_{t-1} = i) \quad \textbf{(III. 8)}$$

  Elle dépend des deux états précédents dans le cas d'un HMM d'ordre deux :

  $$P(q_t = j|q_{t-1} = i, q_{t-2} = k, \ldots\ldots) = \mathrm{P}(q_t = j|q_{t-1} = i, q_{t-2} = k) \quad \textbf{(III. 9)}$$

  En d'autres termes, l'évolution du système entre deux instants t − 1 et t ne dépend que de l'état de ce système au temps t − 1 (ordre un) ou des deux instants précédents t – 1 et t – 2 (d'ordre 2).

- $\mathrm{B} = \{bj(ot)\} = \{P(ot|qj)\}$, où $j\epsilon[1, \mathrm{N}]$, est l'ensemble des probabilités d'émission de l'observation $\mathrm{O_t}$ sachant qu'on est dans l'état qj La forme que prend cette distribution détermine le type du HMM. C'est ainsi qu'on parle de HMM discrets, semi-continus, continus, etc.
- π est la distribution initiale des états, $\pi j = \mathrm{P}(q0 = j), et\ j\epsilon[1, N]$.q0 représente l'état initial du modèle HMM. Il ne peut acoustiques émettre de vecteur

III.4. Application des HMM à la reconnaissance des mots isolés

En reconnaissance de la parole, des modèles de Markov gauche-droite d'ordre n sont le plus souvent utilisés du fait de l'aspect séquentiel du signal de la parole.

La figure III.10 illustre un exemple d'un HMM à 5 états typique utilisé pour la modélisation d'un mot.

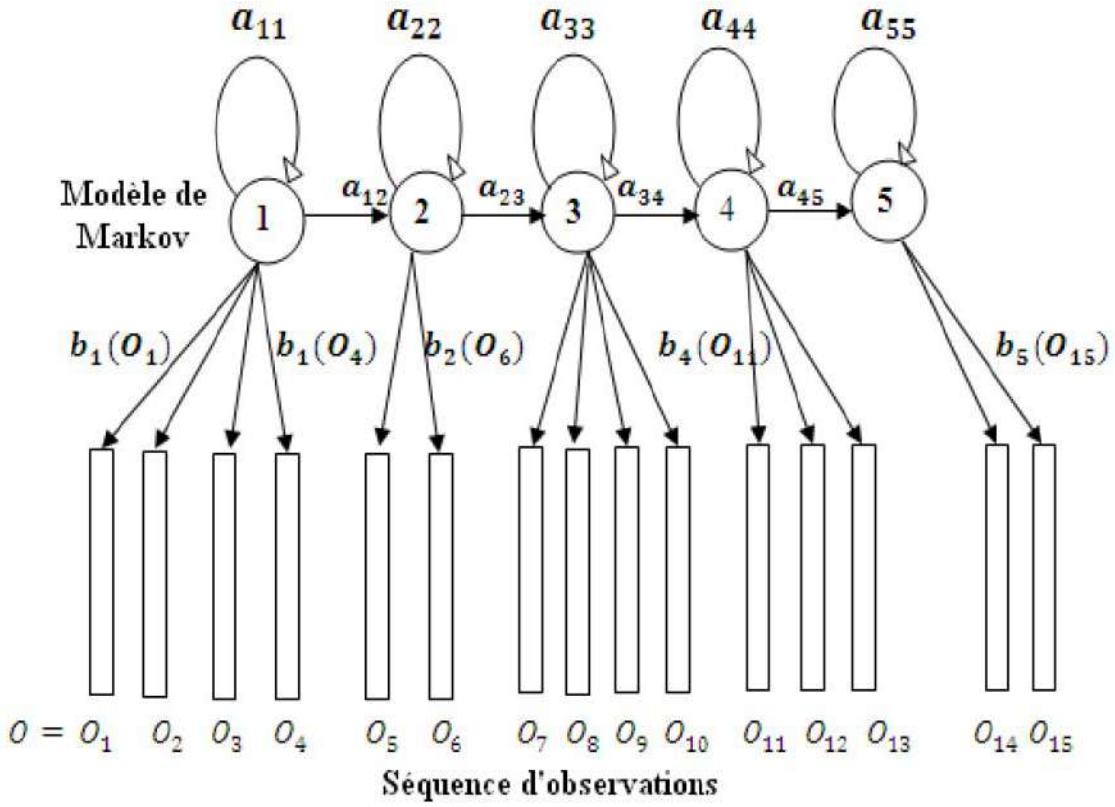


Figure III.10. *Exemple d'un HMM à cinq états. Caractérisé par une distribution de Probabilités pour chaque état associé à une observation et par des probabilités de transition entre les états*.

III.4.1. Reconnaissance de mots isolés en nombre limité (<100 mots)

Chaque mot m du vocabulaire *M* est modélisé par un HMM. La phase de reconnaissance consiste, pour une suite d'observations acoustiques données, à calculer la vraisemblance de ces observations par rapport à chacun des modèles et à considérer comme mot reconnu le mot correspondant à la vraisemblance maximum. Cette approche n'est qu'une version améliorée de l'approche de type programmation dynamique.

Cette approche n'est réalisable qu'avec un vocabulaire limité dans la mesure où l'algorithme de reconnaissance considère tous les chemins modélisant chaque mot *M*, ce qui entraîne un important coût de calcul.

III.4.2. Reconnaissance de mots isolés en nombre inférieur à 1000

Chaque mot m du vocabulaire *M* est modélisé par un HMM .Tous les HMM sont liés entre eux par une entrée et une fin commune dans un réseau global. L'entrée correspond à un HMM représentant un silence précédant le mot, tandis que la fin correspond à un HMM modélisant un silence suivant le mot.

Chaque chemin du réseau global correspond à la prononciation d'un mot. En phase de reconnaissance, le mot reconnu correspond alors au chemin le plus probable étant donné la suite d'observations.

Cette approche se différencie de la précédente par l'algorithme de reconnaissance : dans ce cas, celui-ci ne prend en compte que le chemin le plus vraisemblable parmi tous ceux modélisant les mots du vocabulaire *M*, réduisant ainsi considérablement le coût de calcul.

III.5. Application et Résultats

Pour notre travail on a utilisé l' HMM pour la reconnaissance de six mots isolés. Donc on a un vocabulaire de six mots (gauche, droite, avant, arrière, haut, bas) , chaque mot est modélisé par un HMM distinct, en outre ,on a pour chaque mot du vocabulaire, un ensemble de d'occurrences *K* (prononcé par 1 ou plusieurs locuteurs), chaque occurrence du mot représente une observation.

Pour réaliser la reconnaissance des mots isolés les étapes suivantes sont exécutées:

- Pour chaque mot dans le vocabulaire, nous devions construire un HMM, nous devions aussi estimer les Paramètres du modèle $(\mathrm{N}, \mathrm{A}, \mathrm{B}, \pi)$qui optimisent la vraisemblance des vecteurs d'apprentissage d'observation de la consigne pour chaque mot.
- Pour chaque mot inconnu le traitement de la figure III.11 doit être effectué, à savoir la mesure de séquence d'observation $O = \{\mathrm{O}_1, \mathrm{O}_2, \ldots\ldots, \mathrm{O}_3\}$, par une analyse des caractéristiques de la parole correspondant à la parole, suivi par le calcul de la vraisemblance du modèle pour tous les modèles possibles, $\mathrm{P}(\mathrm{O}|\lambda)$; suivie par la sélection de la parole qui a la plus haute probabilité de modèle.

$$M^* = \text{argmax}[P(O|\lambda_m)] \text{ où } 1 > m > M \quad \textbf{(III.10)}$$

- l'étape de calcul de probabilité est réalisée par l'algorithme de Viterbe (l'algorithme cité à l'III-3-2-2) (le chemin de vraisemblance maximale est utilisé).

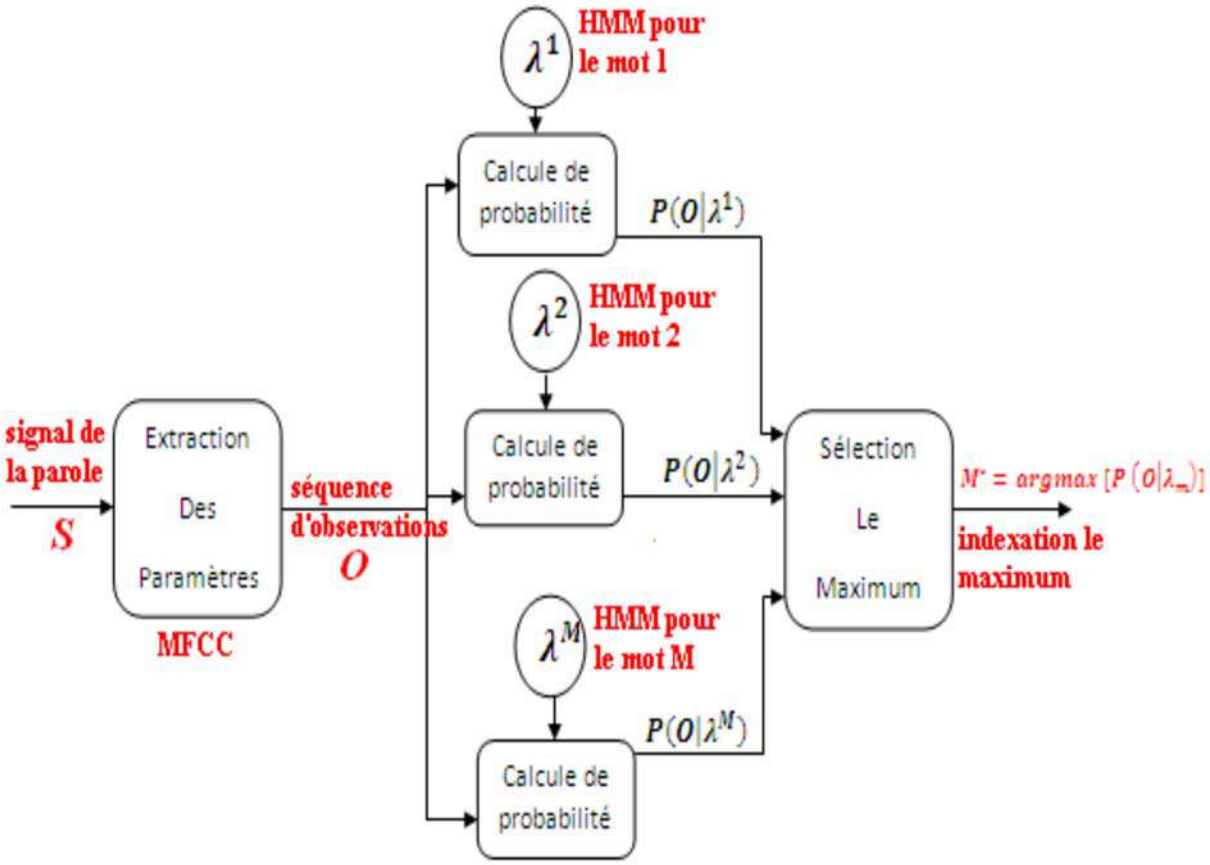


Figure III.11. *Architecture de reconnaissance des mots isolés*

III.5.1. Reconnaissance en temps réel

Le robot mentor ou bras manipulateur à 5 axes est un robot industriel. Lorsqu'un opérateur veut utiliser ce type de robot, il doit prononcer un mot isolé, ce mot est d'abord transformé en série de vecteurs acoustiques (MFCC), cet observation d'entrée est reconnue par un algorithme de Viterbi, qui aligne les observations avec les HMM, si cette commande est reconnue, l'angle correspondante à cette commande est envoyée via un réseau local au

serveur qui envoie à son tour les ordres au robot (Faire tourner les servomoteurs de robot). Cette procédure est schématisée dans la Figure III.12.

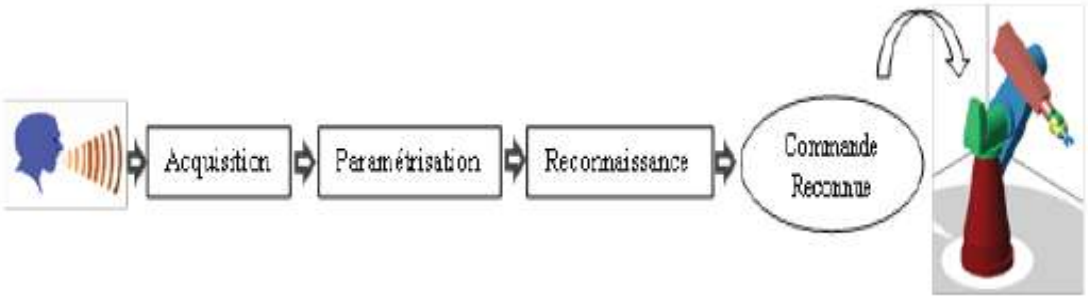


Figure III.12. *Procédure de la reconnaissance en temps réel*

III.5.2. Taux de reconnaissance

Dans ce que suit, nous présentons les résultats expérimentaux concernant le taux de reconnaissance de chaque mot de notre vocabulaire. Ces résultats sont obtenus à partir de plusieurs essais de reconnaissance pour chaque mot :

Tableau III.1. *Taux de reconnaissance*

Les mots	**Taux de reconnaissance**
Avant	92 %
Arrière	96 %
Droite	96 %
Gauche	94 %
Haut	82 %
Bas	86 %

III.6. Conclusion

A partir des expériences précédentes et les résultats obtenus on remarque que le model de Markov caché assure un bon résultat dans le domaine de la reconnaissance de la parole et en particulier dans la reconnaissance des mots isolés. Les coefficients MFCC sont derrières ces bon résultats car ils ont une bonne représentation du comportement du signal de la parole. Il est à signaler que des mauvais résultats sont obtenus pendant la phase d'apprentissage et de reconnaissance probablement dûs à la qualité de l'enregistrement des sons ou aux mauvaises prononciations ce qui nous a mené à réenregistrer les mots qui donnes des résultats illogique.

Chapitre IV

Commande vocale d'un bras manipulateur

IV.1. Introduction

A la sortie de la phase de la reconnaissance vocale et après l'identification du mot prononcé qui correspond à un ordre de mouvement, notre robot manipulateur est tenu d'appliquer l'instruction demandé parmi les six existants et poursuivre la trajectoire lié à cet ordre. La fiabilité de notre système est jugée à partir de testes globales sur la commande vocale en temps réel d'un bras manipulateur à 3 degrés de libertés de model RRR.

L'une des techniques de la commande et de la régulation a été adopté pour contrôlé l'efficacité de poursuivre la trajectoire à moindres erreurs, c'est la méthode de commande en mode glissant.

Les tests de simulations considèrent deux cas principales pour la position initiale du bras manipulateur notamment la parti haut qui comprend les deux articulations *q2* et *q3*. Les différentes postures sont posture à coude bas et posture à coude haut qui correspondent respectivement à *q3* positive et à *q3* négative.

IV.2. Architecture générale de la commande vocale

L'architecture générale adoptée pour la commande est représenté selon la figure IV.1

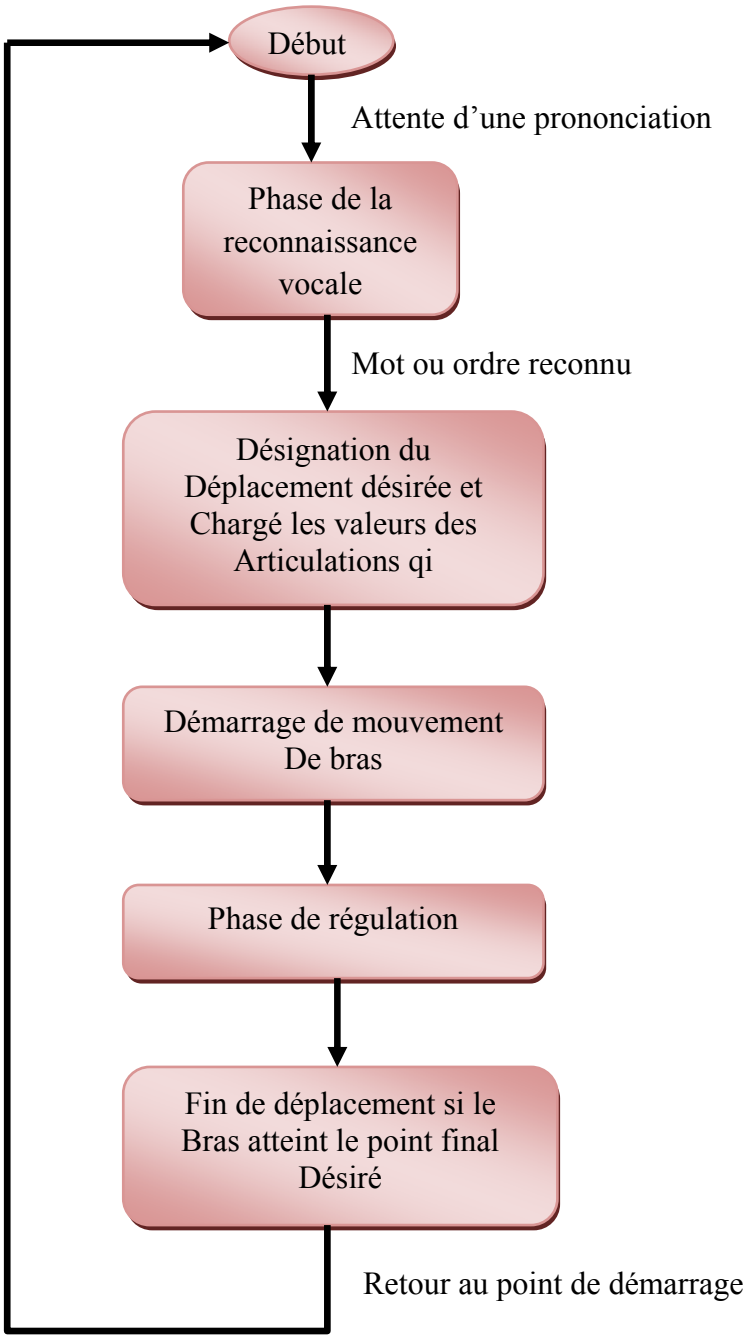


Figure IV.1. *L'architecture générale de la commande vocale*

IV.3. le model de robot RRR utilisé

Le bras manipulateur utilisé dans ce travail est illustré par la figure IV.2.

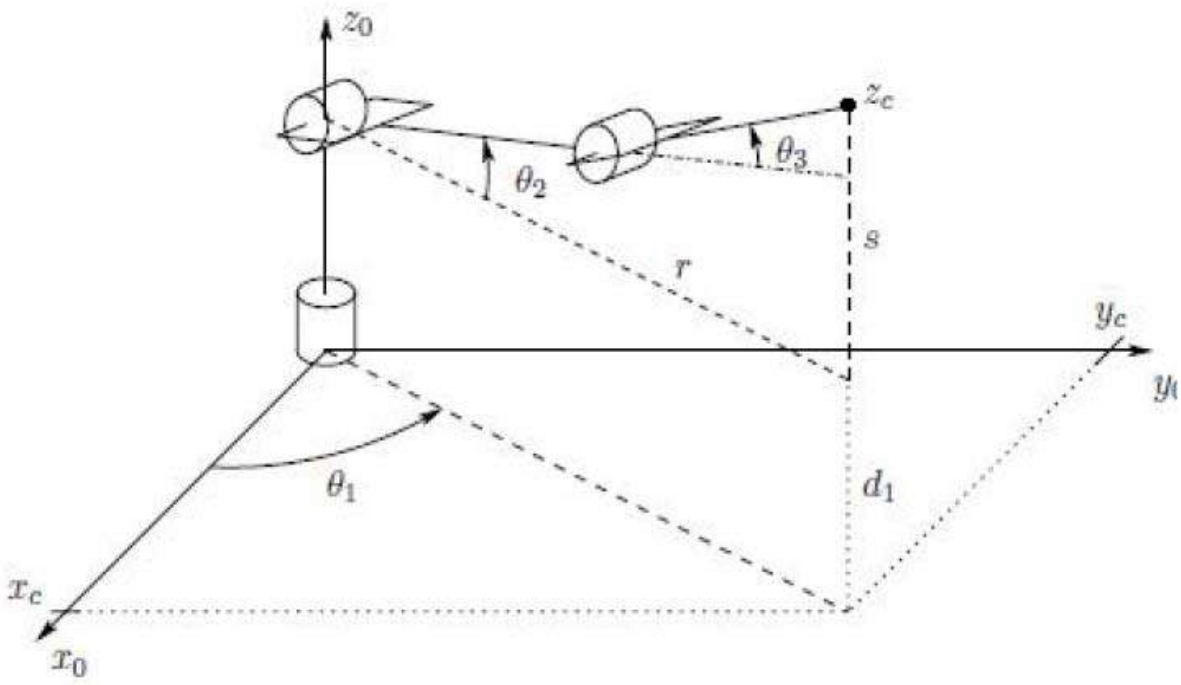


Figure IV.2. *Architecture de bras manipulateur RRR*

IV.3.1. Configuration des articulations

Considérons le manipulateur présenté à la figure IV.2, avec les composantes de O_c noté x_c , y_c et z_c ; la projection d'O_c sur le plan xy et présenté sur la figure IV.3 :

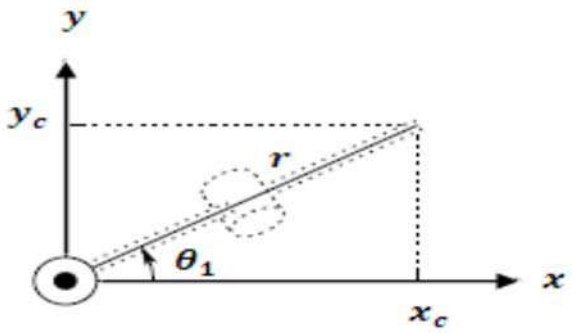


Figure IV.3. *Projection de bras selon xy*

On voit dans cette projection que:

$$\theta_1 = \arctan(y_c/x_c) \quad \textbf{(IV. 1)}$$

Pour trouver les angles θ_2 et θ_3 du manipulateur, on considère le plan formé de la deuxième et troisième liaison comme le montre la figure IV.4.

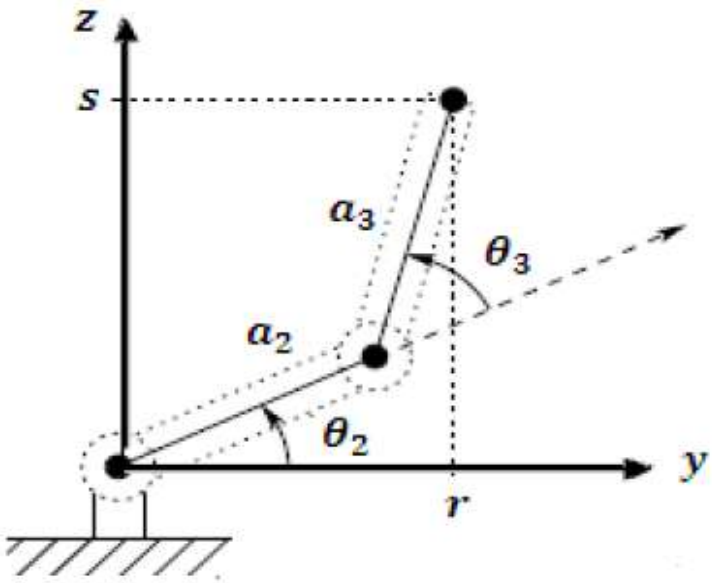


Figure IV.4. *projection de bras selon yz*

$$\cos\theta_3 = \frac{x_c^2+y_c^2+(z_c-d1)^2-a_2^2-a_3^2}{2a_2a_3} \quad \textbf{(IV. 2)}$$

$$\theta_3 = \pm Acos\frac{x_c^2+y_c^2+(z_c-d1)^2-a_2^2-a_3^2}{2a_2a_3} \quad \textbf{(IV. 3)}$$

- Lorsque θ_3 est positif, le robot a une posture ***coude bas,*** et si θ_3 est négatif le robot a une posture ***coude haut,*** voir la figure IV.5.

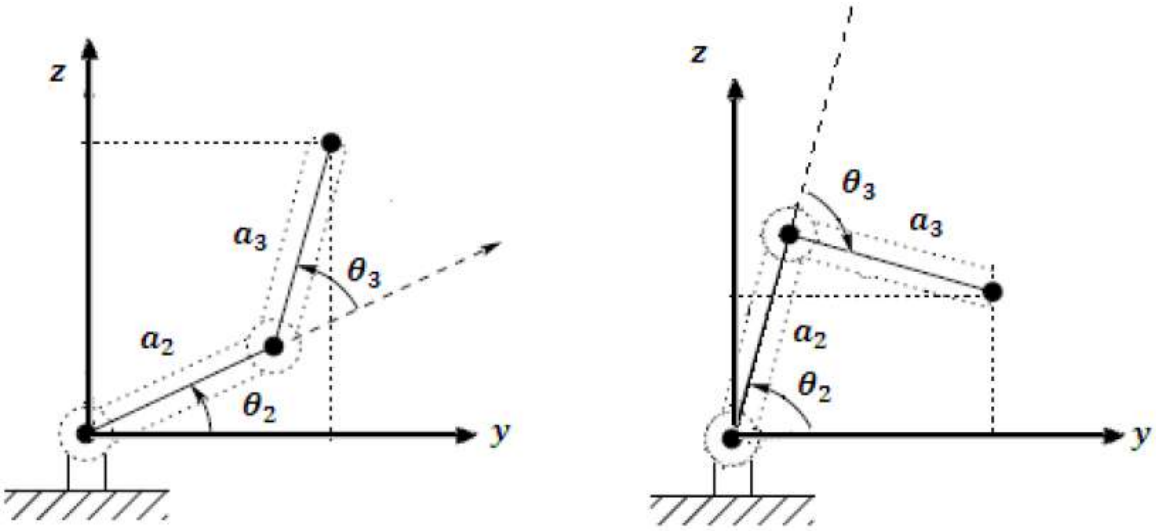


Figure IV.5. *les deux postures des robots : à gauche (coude bas) et à droite (coude haut)*

$$\theta_2 = \text{Atan}\left(\frac{z_c - d1}{\sqrt{x_c^2 + y_c^2}}\right) - Atan\left(\frac{a_3 \sin\theta_3}{a_2 + a_3 \cos\theta_3}\right) \qquad \textbf{(IV. 4)}$$

IV.3.2. Le choix des trajectoires

Pour le déplacement du bras on a choisi trois trajectoires simples. Chaque trajectoire a un sens opposé ce qui implique six déplacements c'est-à-dire un sens pour chaque mot ou ordre de notre dictionnaire.

IV.3.2.1. Les mots Gauche et Droite

Ces deux mots ordonnent le manipulateur de tourner autour de l'axe Z dans les deux sens avec l'angle θ_1.

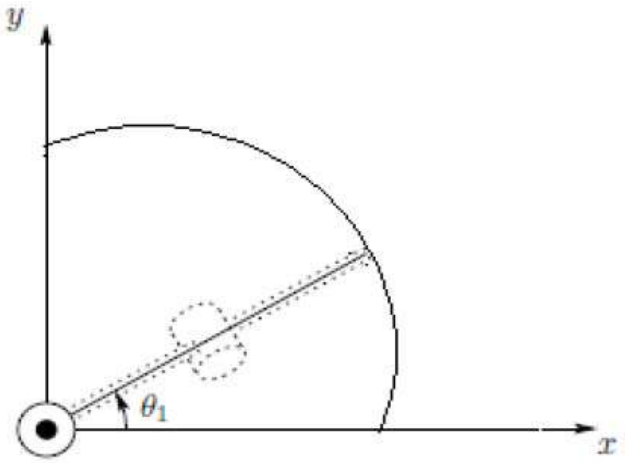


Figure IV.6. *Rotation autour de l'axe Z avec l'angle* θ_1

IV.3.2.2. Les mots Haut et Bas

Ces deux mots provoquent un déplacement vertical selon l'axe Z. Les valeurs de x_c et y_c dans ce cas sont fixes et les angles concernés dans ce déplacement sont $\theta_2$et θ_3.

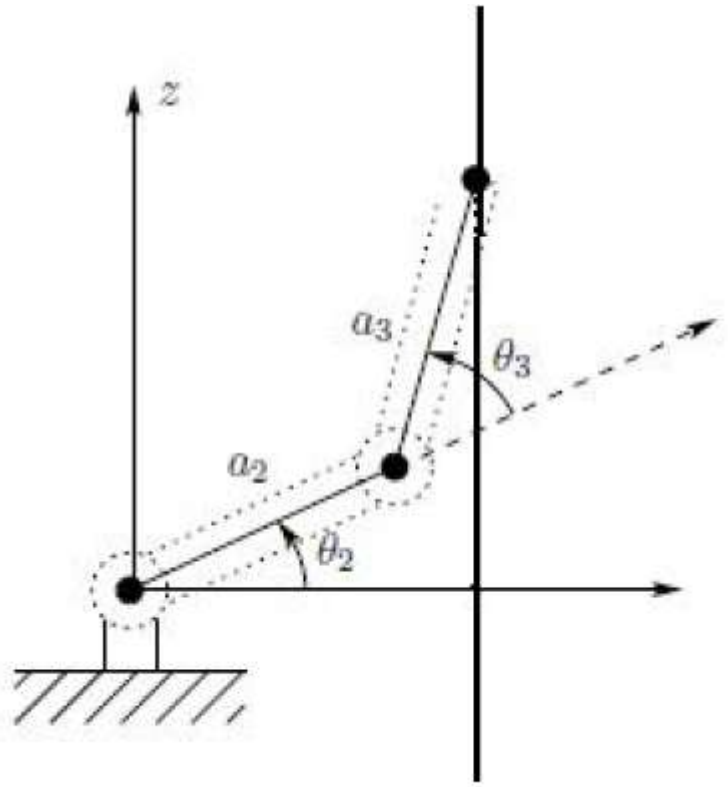


Figure IV.7. *Déplacement verticale*

IV.3.2.3. Les mots Avant et Arrière

Pour les derniers mots ***avant*** et ***arrière*** le bras doit se déplacer horizontalement en parallèle au plan XY avec la z_c fixe.

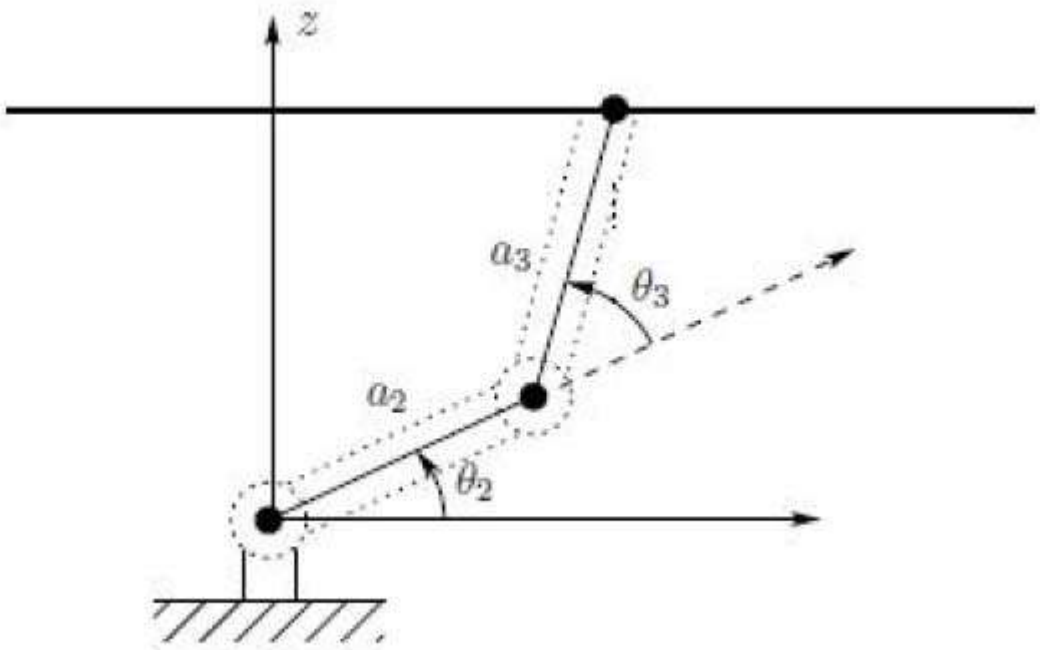


Figure IV.8. *Déplacement horizontal*

IV.4. Généralité sur la commande en mode glissant

La phase de régulation est une phase très importante pour ajuster les paramètres des articulations afin de suivre la trajectoire désirée à moindres erreurs. C'est la commande en mode glissant qui est utilisée pour assurer cet ajustement.

Considérant un système non linéaire décrit par :

$$x^{(n)} = f(x,t) + g(x,t).u \qquad \textbf{(IV. 5)}$$

Avec :

$x(t) = \left(x, \dot{x}, \dots, x^{(n-1)}\right)^T$ Le vecteur d'état.

$f(x,t)$ Et $g(x,t)$ des fonctions non linéaires du vecteur d'état décrivant le système.

Le but du contrôle en poursuite est de trouver une loi de commande tel que, étant donné une trajectoire désirée $x_d(t)$, l'erreur de poursuite $x(t) - x_d(t)$ tend vers zéro malgré la présence des perturbations [23].

L'erreur de poursuite est définie par :

$$e(t) = x(t) - x_d(t) = \left(e, \dot{e}, \dots, e^{(n-1)}\right)^T \quad \textbf{(IV. 6)}$$

La mise en œuvre d'une commande par mode glissant passe par trois étapes :

- **Le choix de la surface de glissement**

Slotine [24] a proposé une forme générale qui consiste à définir une fonction scalaire des surfaces de glissement dans le plan de phase dans le but d'assurer la convergence d'une variable d'état x vers sa valeur de consigne x_d. Cette fonction est donnée par l'équation:

$$s(x,t) = \left(\frac{\partial}{\partial t} + \lambda\right)^{n-1} . e \quad \textbf{(IV. 7)}$$

Où :

$\boldsymbol{\lambda}$: est un scalaire qui représente la pente de la surface de glissement, cette dernière est obtenue pour un système du deuxième ordre lorsque : $s(x,t) = 0$

Une fois la fonction de commutation est établie le problème de la poursuite nécessite la conception d'une loi de commande tel que le vecteur d'état $e(t)$reste sur la surface de glissement $s(x,t) = 0$ pour tout $t \geq 0$.

- **La condition d'existence d'une surface de glissement**

Elle peut être déduite de la fonction énergétique de *Liapunov* donnée par la relation suivante :

$$V = \frac{1}{2} s^2 \quad \textbf{(IV. 8)}$$

Une condition suffisante pour que le système (IV-5) soit stable [23] :

$$\dot{V} = \frac{1}{2}\frac{\partial}{\partial x}(s^2) \leq -\eta.|s| \quad \textbf{(IV. 9)}$$

Où : $\eta > 0$

- **L'établissement de la loi de commande :**

La loi de commande par mode glissant est donnée par la formule suivante :

$$u = u_e + k.sign(s) = u_e + u_n \quad \textbf{(IV. 10)}$$

Sachant que :

u_e: est la commande équivalente.

$sign()$: est la fonction signe.

k: est une constante positive qui représente le gain de la commande discontinue.

La figure IV.6 représente le plan de phase d'un système du 2ème ordre ainsi que la commande discontinue de l'équation (IV.9).

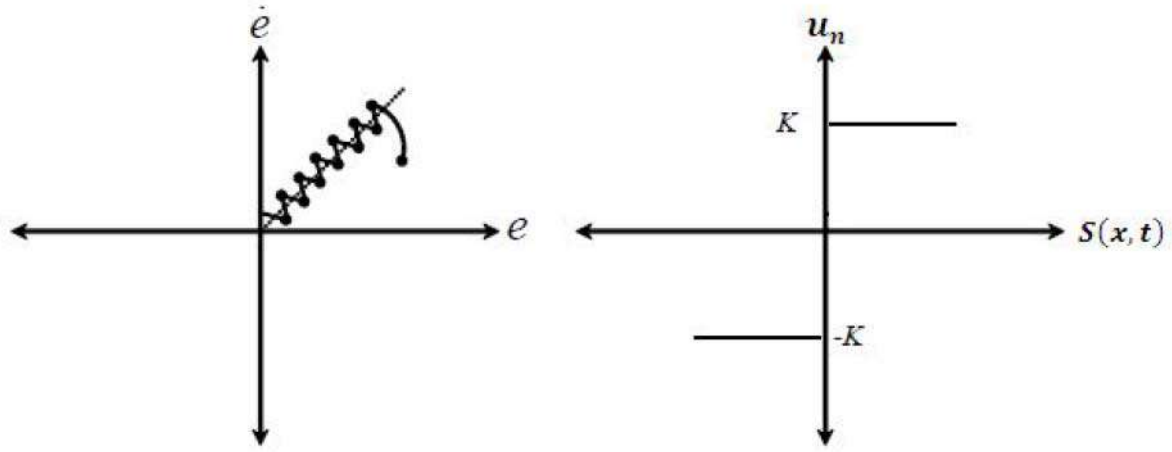


Figure IV.9. *Plan de phase et la fonction de commutation de la commande par régime Glissant.*

Une méthode qui permet de réduire l'effet du broutement est de remplacer la fonction discontinue par une fonction de saturation, qui consiste à déterminer une bande limite autour de la surface de glissement assurant ainsi le lissage de la commande et le maintient de l'état du système dans cette bande. La loi de commande devient alors [23]:

$$u = \begin{cases} u_e + k.sign(s) & si|s| > \varepsilon \\ u_e + k.\frac{1}{\varepsilon}.s & si|s| > \varepsilon \end{cases} \quad \textbf{(IV. 11)}$$

Cette fonction est connue aussi par, la fonction saturation « *SAT* ».

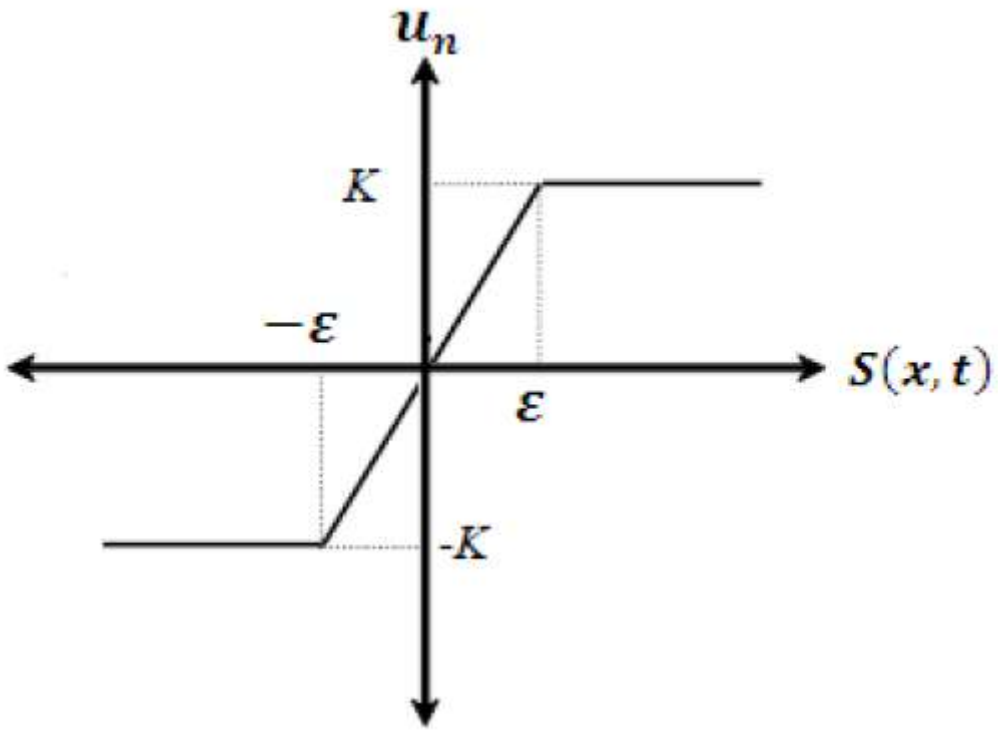


Figure IV.10. *la fonction de commutation de la commande par régime glissant à bande limite*

IV.5. Conclusion

Dans ce chapitre nous avons vu les différentes réactions de bras manipulateur devant les ordres vocaux prononcés par l'opérateur. On a testé tous les déplacements liés aux ordres prononcés dans les deux situations possibles de l'état initial (coude bas et coude haut).

Conclusion générale

Conclusion Générale

L'objectif de ce travail était de trouver un moyen simple et efficace pour la commande vocale d'un bras manipulateur.

Toute la chaine du traitement du signal de parole est alors considérée, depuis la production de ce signal jusqu'à son exploitation. L'acquisition du signal de parole est faite avec un simple microphone.

Un traitement particulier est appliqué, à ce signal, pour extraire les paramètres exploitables pour la reconnaissance. Ce sont les coefficients mel-cepstraux (MFCC), basé sur l'analyse à court terme du signal, qui ont étés retenus pour cet objectif.

La reconnaissance proprement dite est opérée par utilisation d'une approche statistique basé sur le model de Markov caché. Dans une première phase, on opère un entrainement des paramètres du model de Markov, l'algorithme de Viterbi est en suite utilisé dans la phase de reconnaissance.

Pour la validation des modèles, un bras manipulateur à trois degrés de libertés commande en mode glissant est utilisé.

Les travaux future peuvent se focaliser sur:

- L'augmentation de la banque des ordres pour répondre à d'autres situations.
- Pousser la reconnaissance à la personnalisation, de sorte à ce que le robot ne répond qu'aux ordres du maitre dont il a fait la reconnaissance.

RÉFÉRENCES BIBLIOGRAPHIQUE

[1] BOUZIANE Fatima Zohra « Rétro-conception du bras horizontal de robot manipulateur de la cellule flexible (Tlemcen)», UNIVERSITE ABOU BEKR BELKAID-TLEMCEN Option Ingénierie des systèmes mécaniques productiques, (2013).

[2] PRIEL MARC «les robots industriels: caractéristiques, performances et choix »: Edition AFNOR (1990).

[3] Laetitia Matignon, «Introduction a la robotique», Université de Caen France 2011/2012.

[4] P. Coiffet, Hermes, La Robotique, principes et applications.

[5] H. Hamdi, «Introduction à la Robotique»,Les éditons de l'université Mentouré Constantine 2002-2003.

[6] E. Dombre, «Analyse et Modélisation des Robots Manipulateurs», 22 janvier 2002.

[7] A. Chrifa, « Analyse Cinématique des Robots Parallèles», Thèse Docteur D'état. Université Mentouri Constantine 2007/2008.

[8] Jean-Louis Boimond, «Robotique», ISTIA Université Angers.

>> www.istia.univ-angers.fr/~boimond/Cours_robotique.pdf

[9] P. Fissette, H.Buyse, J.C.Samin « Introduction à la robotique », 29 juin 2009

>>http://meca2732.pdf

[10] R.Gourdeau, «ELE4203 — Robotique : Modélisation des Robots Manipulateurs», Département de génie ´électrique, Ecole Polytechnique de Montréal, 4 novembre 2010.

[11] O.Ahcene, NEMMICHE Belkacem, BELHAMITI hadj Afif, BERRAHO Ismail «Commande en 3D d'un bras robot à 5 DL par positionnement en temps réels par deux caméras », Université ABDELHAMID IBN BADIS De Mostaganem.

[12] Jacques Gangloff « Cours de Robotique ».

>> http://eavr.u-strasbg.fr/wiki/upload/a/a4/Cours_rob_intro.pdf

[13] R. Boite, et M. Kunt, «Traitement de la Parole», Presses Polytechniques Romandes, Complément au Traité d'électricité, Première édition, 1987.

[14] René Boite, Hervé Bourlard, T. Dutoit, «Traitement de la parole», presses polytechniques et Universitaires de Romandes, Lausanne, 2000.

[15] Deller, J.R., Hansen, J.H. L., Proakis J. G, «Discrete Time Processing of Speech Signals» IEEE Press 1999.

[16] Calliope, «La parole et son traitement automatique», Paris, Masson 1989.

[17] Salma Jamoussi,«Méthodes statistiques pour la compréhension automatique de la parolethèse »Doctorat de l'université Henri Poincare-Nancy décembre 2004.

[18] E. Zwicker et R. Feldtkeller. Psycho acoustique, «L'oreille, récepteur d'information» Masson, 1981.

[19] L. R. Bahl, P. F. Brown, P. V. de Souza, R. L. Mercer, et M. A. Picheny, ,«Acoustic Markov models used in the Tangora speech recognition system » , In Proceedings of the International Conference on Acoustics, Speech and Signal Processing, 1988.

[20] F., Jelinek, «Continuous speech recognition by statistical methods», Proceeding of IEEE, 1976.

[21] J. P. Haton, J. M. Pierrel, G. Perennou, J. Caelen, et J. L. Gauvain,«Reconnaissance automatique de la Parole »dunod, 1991.

[22] L. Rabiner et B. H. Juang, «A tutorial on hidden Markov models and selected application in speech recognition», Proceeding of IEEE, 1989.

[23] R.Ouiguini, R.Bouzid et Y.Sellami, «Une commande robuste par mode glissant flou Appliquée à la poursuite de trajectoire d'un robot mobile non holonome» CISTEMA 2003.

[24] J.J.Slotine, W.Li, «Applied nonlinear control», Englewood Cliffs, NJ: prentice Hall, 1991.