

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE
MINISTERE DE L'ENSEIGNEMENT SUPERIEUR
ET DE LA RECHERCHE SCIENTIFIQUE
UNIVERSITE DE BATNA
FACULTE DE TECHNOLOGIE
DEPARTEMENT D'ELECTRONIQUE

Mémoire

En Vue de l'Obtention du Diplôme de

Magister En Electronique

Option

Contrôle Industriel

Thème

Commande vocale d'un robot manipulateur

Par

BOUAGAL Fawzi

(Ingénieur d'état en électronique)

Jury

S. AOUGHELLANET	(M.C) Université de BATNA	Président
D. MALAAB	(M.C) Université de Batna	Rapporteur
Y. ABDESSEMED	(M.C) Université de Batna	Examineur
A. MESSAI	(M.C) Université de Constantine	Examineur

Remerciements

Mes premiers remerciements s'adressent à mon rapporteur monsieur MELAAB Djamel Maître de Conférences au Département d'Electronique de l'Université de Batna pour la confiance qu'il m'a fait, son aide son encouragement et ses conseils durant la réalisation de ce travail.

Je voudrais exprimer mes remerciements à monsieur AOUGHELLANET Saïd , Maître de Conférences au Département d'Electronique de l'Université de Batna, d'avoir accepté la présidence du jury.

Je tiens à remercier également monsieur ABDESSEMED Yassine, Maître de Conférences au Département d'Electronique de l'Université de Batna, d'avoir accepté d'examiner ce travail.

J'adresse mes vifs remerciements à monsieur MESSAI Abderraouf, Maître de Conférences au Département d'Electronique de l'Université de Constantine, de m'avoir honoré de faire partie de ce jury.

Je ne pourrai oublier d'adresser mes remerciements les plus distingués à toutes les personnes qui ont contribué de près ou de loin à la réalisation de ce travail.

F. BOUAGAL

Résumé

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, extraction des paramètres, 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, Feature extraction, speech recognition, the manipulator arm.

Table des matières

<i>Introduction Générale</i>	01
------------------------------------	----

CHAPITRE I

MODELISATION DES BRAS MANIPULATEURS

I-1 Introduction	03
I-2 Description des bras manipulateurs.....	04
I-2-1 morphologie et Chaîne cinématique d'un bras manipulateur	04
I-2-1-1 Configuration Articulée (RRR).....	04
I-2-1-2 Configuration Sphérique (RRP).....	05
I-2-1-3 Configuration SCARA (RRP).....	06
I-2-1-4 Configuration Cylindrique (RPP).....	06
I-2-1-5 Configuration Cartésienne (PPP).....	07
I-2-2 Structure mécanique	07
I-2-3 Espace articulaire, espace opérationnel et modèle de transformation entre les espaces	08
I-2-4 degré de liberté et de mobilité	08
I-3 structure géométrique	09
I-3-1 Paramètres de Denavit-Hartenberg modifiés	09
I-3-2 Relations géométriques.....	12
I-4 La Modélisation	12
I-4-1 Modèle géométrique	12
I-4-1-1 Modèle géométrique Directe.....	12
I-4-1-2 Modèle géométrique inverse.....	13
I-4-2 Modèle cinématique.....	14
I-4-2-1 Modèle cinématique directe.....	14
I-4-2-1 Modèle cinématique inverse	15

I-4-3 Modèle dynamique	15
I-4-3-1 Modèle dynamique inverse.....	15
I-4-3-2 Modèle dynamique direct.....	21
I-4-3-3 Problèmes du modèle dynamique.....	21
I-5 Conclusion	21

CHAPITRE II

LE SIGNAL DE PAROLE

II-1 Introduction	22
II-2 Production de la parole	23
II-2-1 Architecture de l'appareil vocal.....	23
II-2-1-1 La soufflerie et Le vibrateur.....	24
II-2-1-2 Le larynx.....	24
II-2-1-3 Le corps sonore.....	24
II-2-1-4 Le système articulatoire.....	25
II-2-2 Mécanisme de la phonation.....	25
II-2-2-1 Phonation de sons voisés.....	25
II-2-2-1 Phonation de sons non voisés.....	26
II-2-3 Classification des phonèmes	27
II-2-3-1 Les voyelles	27
II-2-3-2 Les diphtongues.....	28
II-2-3-3 Les semi-consonnes.....	29
II-2-3-4 Les consonnes.....	29
II-3 Le système auditif.....	30
II-4 niveau acoustique de la parole	32
II-4-1 Audiogramme.....	33
II-4-2 Transformée de Fourier à court terme.....	35
II-4-3 Spectrogramme.....	35
II-4-4 Fréquence fondamentale.....	35
II-5 Conclusion	37

CHAPITRE III

LA RECONNAISSANCE DE LA PAROLE

III-1 Introduction	38
--------------------------	----

III-2 le traitement automatique de la parole	39
III-2-1 niveaux de complexité	39
III-2-2 Extraction des paramètres	40
III-2-2-1 acquisition et préaccentuation du signal	41
III-2-2-2 segmentation et fenêtrage	43
III-2-2-3 transformer de Fourier discrète et l'énergie du signal	45
III-2-2-4 l'échelle Mel.....	45
III-2-3 La reconnaissance.....	49
III-3 Modélisation acoustique à base de HMM.....	51
III-3-1 Définitions des modèles de Markov cachés.....	51
III-3-2 Les problèmes fondamentaux des HMM	52
III-3-2-1 Problème d'évaluation.....	53
III-3-2-2 Problème de décodage.....	55
III-3-2-3 Problème d'apprentissage.....	57
III-3-3 Densités d'observation continues dans les modèles de Markov cachés.....	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-4-2 Reconnaissance de mots isolés en nombre inférieur à 1000.....	60
III-5 Application et Résultats	61
III-5-1 Taux de reconnaissance	62
III-6 Conclusion	63

CHAPITRE IV

APPLICATION SUR LES BRAS MANIPULATEURS

IV-1 Introduction	64
IV-2 architecture générale de la commande vocale.....	65
IV-3 le model de robot RRR utilisé	66
IV-3-1 Configuration des articulations.....	66
IV-3-2 le choix des trajectoires	68
IV-3-2-1 les mots <i>Gauche</i> et <i>Droite</i>	68
IV-3-2-2 les mots <i>Haut</i> et <i>Bas</i>	68

IV-3-2-3 les mots <i>Avant</i> et <i>Arrière</i>	69
IV-4 généralité sur la commande en mode glissant	69
IV-5 résultats et simulation	72
IV-5-1 première situation « posture coude haut ».....	72
IV-5-2 deuxième situation « posture coude bas ».....	78
IV-5-3 Rotation de bras autour de l'axe Z	84
IV-6 Conclusion.....	87
Conclusion Générale	88
Bibliographie	89
Annexe.....	92

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 œil-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 examinée 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'entraîner 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 modèle 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 dédié à une présentation des différents modèles utilisés pour décrire le mouvement des articulations d'un bras manipulateur.
- le deuxième chapitre présentera une étude générale sur le signal de parole (son production, ses caractéristiques et son traitement)
- Dans le troisième chapitre, après avoir exposé un rappel générale sur la reconnaissance automatique de la parole, nous allons appliquer le modèle 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

Modélisation Des Bras Manipulateurs

I-1-Introduction

Quel que soit l'objectif qu'on se fixe, concevoir, simuler ou commander un système mécanique, il est nécessaire, entre autres, de disposer de modèles du mécanisme à traiter. Plusieurs niveaux de modélisation sont possibles, qui dépendent des spécifications du cahier des charges de l'application envisagée : modèles *géométriques*, *cinématiques* ou *dynamiques* à partir desquels peuvent être engendrés les mouvements du robot, ou bien des modèles *statiques* qui décrivent les interactions du mécanisme avec son environnement.

I-2 Description des bras manipulateurs

I-2-1 morphologie et Chaîne cinématique d'un bras manipulateur

On suppose les bras manipulateurs constitués de n corps mobiles, supposés parfaitement rigides, reliés entre eux par n liaisons rotoïdes et/ou prismatiques formant une structure de chaîne simple (voir figure I.1).

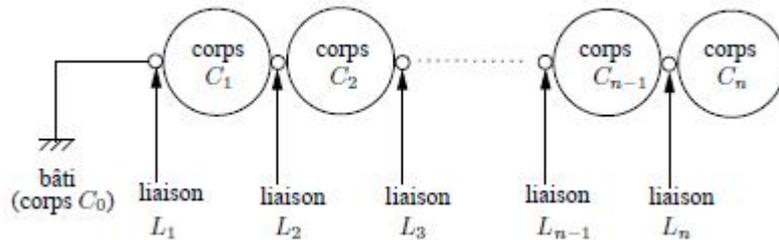


Figure I-1 Chaîne cinématique d'un bras manipulateur série

Pour identifier la nature de la i -ème liaison du bras manipulateur, on définit le paramètre :

$$\sigma_i = \begin{cases} 0, & \text{pour une liaison rotoïdes,} \\ 1, & \text{pour une liaison prismatique.} \end{cases}$$

On désigne fréquemment les bras manipulateurs en accolant les lettres R (pour rotoïdes) et P (pour prismatique) pour décrire la succession des liaisons. Par exemple un bras manipulateur de type anthropomorphe sera désigné par le sigle $6R$ alors qu'un bras manipulateur de type SCARA sera de type RRP ou $RRPR$ (il existe plusieurs variantes) [5].

I-2-1-1 Configuration Articulée (RRR)

Le manipulateur articulé appelé aussi le manipulateur rotoïde, ou anthropomorphe est un design des plus utilisés, la structure terminologie associée à ce type de manipulateur est montrée par la figure I-2 [19] :

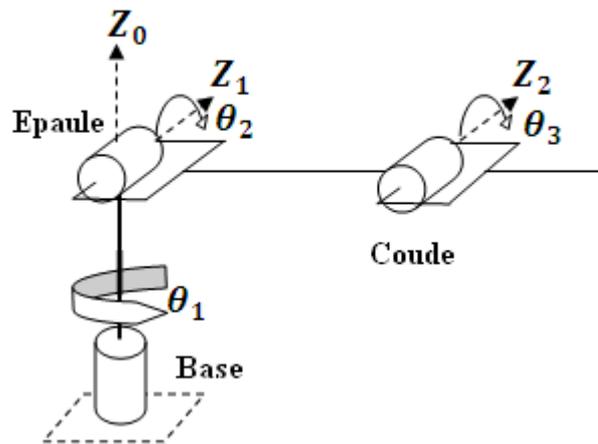


Figure I-2 Configuration RRR

I-2-1-2 Configuration Sphérique (RRP)

En remplaçant la troisième articulation (le coude) dans la configuration rotoïde par une prismatique on obtient la configuration sphérique montrée dans la Figure 1-3.

Le terme configuration sphérique dérive du fait que les coordonnées sphériques définissant la position de l'organe terminal, en respectant que l'origine de la base coïncide avec l'intersection des deux axes Z_1 et Z_2 , sont les mêmes que ceux des trois premières articulations [19].

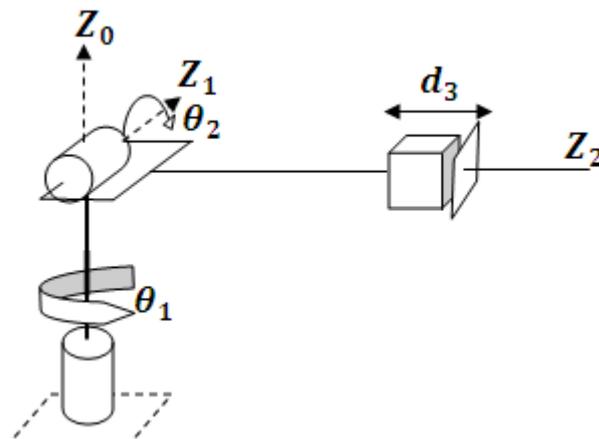


Figure I-3 Configuration RRP

I-2-1-3 Configuration SCARA (RRP)

SCARA (pour Selective Compliant Articulated Robot for Assembly) montré dans la figure I-4 est une configuration très connue. Comme son nom sous-entend il est spécialement conçu pour les opérations d'assemblage. Quoique le robot SCARA a une structure RRP, il est tout à fait différent de la configuration sphérique ni dans l'apparence ni dans le domaine d'application. A la différence du design sphérique qui a z_0 , z_1 , z_2 mutuellement perpendiculaires, le robot SCARA a z_0 , z_1 , z_2 parallèles [19].

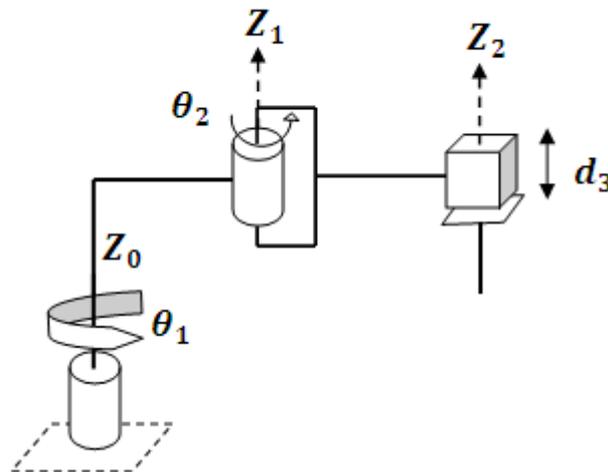


Figure I-4 Configuration SCARA (RRP)

I-2-1-4 Configuration Cylindrique (RPP)

La configuration cylindrique est montrée dans la figure 1-5. La première articulation est rotoïde et produit une rotation autour de la base, alors que la seconde et la troisième sont prismatiques. Comme le nom le suggère, les variables articulaires sont les coordonnées cylindriques de l'organe terminal en prenant compte de la base [19].

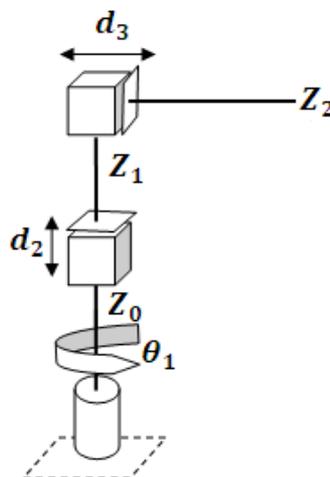


Figure I-5 Configuration cylindrique (RPP)

I-2-1-5 Configuration Cartésienne (PPP)

Un manipulateur dont les trois premières articulations sont prismatiques est connu sous le nom de manipulateur cartésien. Il est montré dans la figure I-6.

Pour le manipulateur cartésien les variables articulaires sont les cartésiennes de l'organe terminal en respectant la base. Comme on peut le deviner la description cinématique de ce manipulateur est la plus simple de toutes les configurations [19].

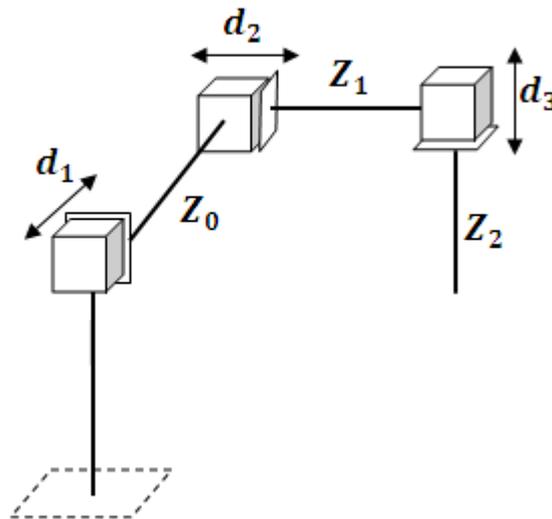


Figure I-6 Configuration cartésienne (PPP)

I-2-2 Structure mécanique

La structure mécanique du robot manipulateur peut être divisée en deux parties distinctes comme le montre la figure I-7 [2] [3]:

- **Organe terminal** : les tâches qui sont dévolues aux robots sont très variées. Pour chaque opération ou travail spécifique, l'organe terminal prend un aspect particulier.
- **Élément porteur** : il est composé d'un ensemble de corps souples ou rigides liés par des articulations, servant à déplacer l'organe terminal d'une configuration à une autre.

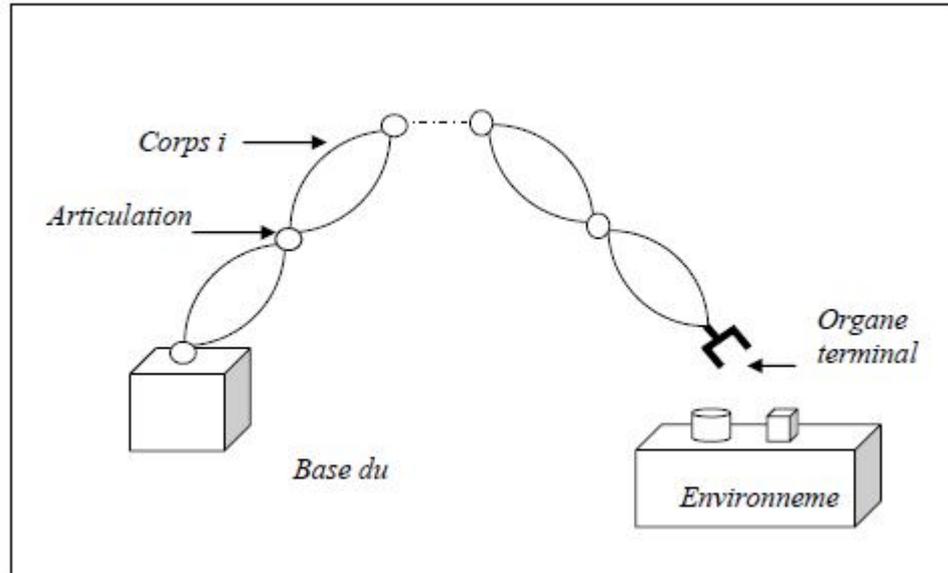


Figure I-7 Robot à chaîne ouverte simple

I-2-3 Espace articulaire, espace opérationnel et modèle de transformation entre les espaces

Définition1 : on appelle espace de configuration articulaire d'un robot manipulateur (ou simplement configuration) l'état du robot représentant la situation de ses différents corps. Les variables ou coordonnées articulaires sont aussi appelées coordonnées généralisées. L'espace de ces variables noté NB, est appelé espace articulaire ou espace des configurations.

Définition2 : l'espace opérationnel est celui dans lequel est représentée la situation de l'organe terminal. Les variables ou coordonnées au nombre minimum permettant de définir la situation de l'organe terminale sont appelées les coordonnées opérationnelles. L'espace opérationnel de dimension m est noté MB [1].

I-2-4 degré de liberté et de mobilité

- **Degré de liberté** : nombre de paramètres utilisés pour spécifier la configuration d'un élément de la chaîne cinématique par rapport à un autre [4].

- **Degré de mobilité** : toute articulation est caractérisée par son degré de mobilité m , c'est-à-dire le nombre de degrés de liberté entre deux corps successifs de la chaîne cinématique ($0 \leq m \leq 6$) [4].

I-3 structure géométrique

I-3-1 Paramètres de Denavit-Hartenberg modifiés [5]

La technique la plus répandue pour décrire la géométrie d'un bras manipulateur consiste à utiliser les paramètres de Denavit-Hartenberg modifiés. On associe au i -ème corps le repère $R_i = (O_i, x_i, y_i, z_i)$, avec $i = 0, 1, \dots, n$. Pour placer les repères R_1 à R_{n-1} selon la règle des paramètres de Denavit-Hartenberg modifiés, il convient de suivre les règles suivantes (voir figure I-8), pour $i = 2, 3, \dots, n$:

- O_{i-1} est le pied de la perpendiculaire commune aux axes des liaisons L_{i-1} et L_i , situé sur L_{i-1} . Si les axes des liaisons sont parallèles, il faut choisir arbitrairement une perpendiculaire commune ;
- x_{i-1} est le vecteur unitaire de cette perpendiculaire commune, orienté de L_{i-1} vers L_i . Si les axes des deux liaisons sont concourants ou confondus, l'orientation est arbitraire (règle avant, droite, haut) ;
- z_{i-1} est un vecteur unitaire de l'axe de la liaison L_{i-1} , librement orienté. On privilégie généralement les débattements positifs pour les liaisons prismatiques et on symétrise généralement par rapport à 0 le débattement des liaisons rotoïdes ;
- y_{i-1} est tel que le repère R_{i-1} soit orthonormé direct.

Ensuite, le repère R_0 est défini librement en suivant des considérations de simplicité.

De même, le point O_n est choisi de façon à annuler un maximum de paramètres. Un point O_{n+1} est associé à l'OT du bras manipulateur, selon des critères opérationnels : typiquement, si le bras manipulateur est équipé d'une pince, le point O_{n+1} représentera le centre de ses mors. On choisit alors enfin R_n de sorte que le point O_{n+1} se situe dans le plan

$R_i = (O_n, x_n, z_n)$, conformément à la figure I-9.

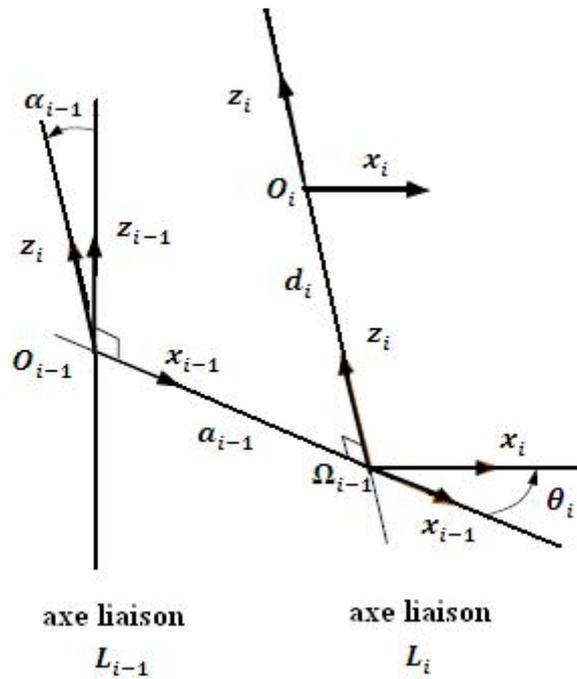


Figure I-8 Paramètres de Denavit-Hartenberg modifiés

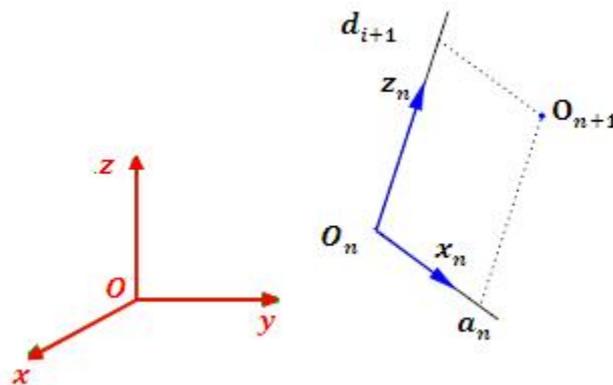


Figure I-9 Placements respectifs des repères R_n et du point O_{n+1}

Les repères étant choisis, la situation du repère d'un corps respectivement à celle du repère du corps suivant peut être définie conformément à la figure I-8. Les paramètres α_{i-1} , a_{i-1} , θ_{i-1} et d_i , avec $i = 1, 2, \dots, n$, placés sur cette figure sont appelés paramètres de Denavit-Hartenberg modifiés. Textuellement, on peut les définir comme suit :

- α_{i-1} : angle algébrique entre z_{i-1} et z_i , mesuré autour de x_{i-1} ;
- a_{i-1} : distance arithmétique de la perpendiculaire commune aux axes des liaisons L_{i-1} et L_i mesurée le long de x_{i-1} ;
- θ_i : angle algébrique entre x_{i-1} et x_i , mesuré autour de z_i ;
- d_i : distance algébrique du point O_i à la perpendiculaire, mesurée le long de z_i .

Les paramètres de Denavit-Hartenberg modifiés permettent de déterminer de manière unique la configuration des corps et donc la géométrie du bras manipulateur. Lorsque l'on modélise un bras manipulateur, il est souhaitable de représenter celui-ci dans une configuration particulière, pour visualiser plus facilement les paramètres et annuler certains paramètres constants. On obtient alors les paramètres constants α_{i-1} , a_{i-1} , θ_i et d_i et la configuration des liaisons décrite par les paramètres θ_{i-1} et d_i variables. A la figure I-10, on a explicité un exemple de paramétrage dans le cas d'un bras manipulateur de type 6R.

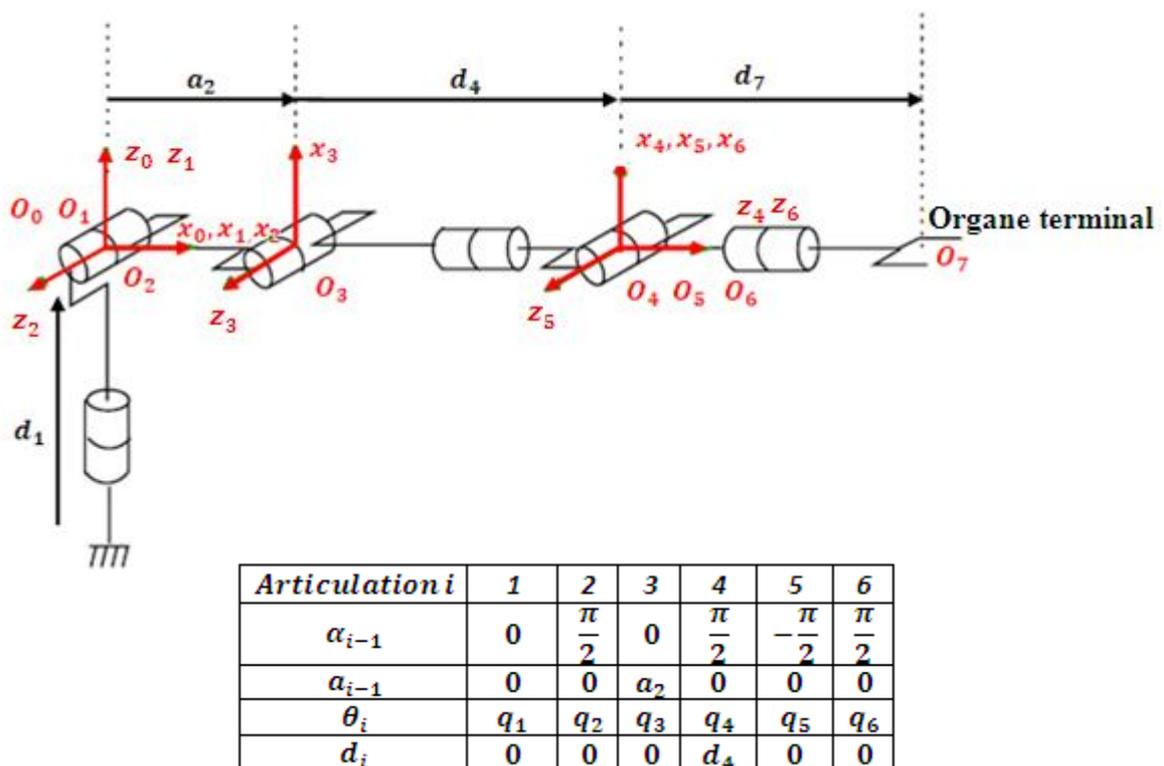


Figure I-10 Repérage et paramétrage d'un bras manipulateur de type 6R selon la méthode des paramètres de Denavit-Hartenberg modifiés

I-3-2 Relations géométriques

Conformément au paramétrage de Denavit Hartenberg modifié, la matrice de transformation homogène est donnée comme suit [5] :

$$T_i^{i-1} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos \alpha_{i-1} & -\sin \alpha_{i-1} & 0 \\ 0 & \sin \alpha_{i-1} & \cos \alpha_{i-1} & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & a_{i-1} \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \cos \theta_i & -\sin \theta_i & 0 & 0 \\ \sin \theta_i & \cos \theta_i & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & d_i \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Soit :

$$T_i^{i-1} = \begin{bmatrix} \cos \theta_i & -\sin \theta_i & 0 & a_{i-1} \\ \cos \alpha_{i-1} \sin \theta_i & \cos \alpha_{i-1} \cos \theta_i & -\sin \alpha_{i-1} & -d_i \sin \alpha_{i-1} \\ \sin \alpha_{i-1} \sin \theta_i & \sin \alpha_{i-1} \cos \theta_i & \cos \alpha_{i-1} & d_i \cos \alpha_{i-1} \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Qui prend la forme :
$$T_i^{i-1} = \begin{bmatrix} A_i^{i-1} & P_i^{i-1} \\ 0 & 1 \end{bmatrix} \quad (I.1)$$

I-4 La Modélisation

I-4-1 Modèle géométrique

I-4-1-1 Modèle géométrique Directe

Le modèle géométrique direct (MGD) est l'ensemble des relations qui permettent d'exprimer la situation de l'organe terminal, c.à.d. les coordonnées opérationnelles du robot en fonction des coordonnées articulaires, le modèle s'écrit :

$$X = f(q) \quad (I.2)$$

q étant le vecteur des variables articulaires tel que

$$q = [q_1 \ q_2 \ \dots \ q_n]^T \quad (I.3)$$

les coordonnées opérationnelles sont définies par

$$X = [x_1 \ x_2 \ \dots \ x_m]^T = [P_x \ P_y \ P_z \ n_x \ n_y \ n_z \ a_x \ a_y \ a_z]^T \quad (I.4)$$

Le calcul du MGD consiste donc à exprimer la position du point O_{n+1} et l'orientation du repère R_n lié à l'OT, en fonction de sa configuration. Il faut pour cela multiplier les matrices de passage homogènes successives reliant le repère R_0 lié au bâti au repère R_n lié à l'OT :

$$T_n^0(q) = T_1^0(q_1) T_2^1(q_2) \dots T_n^{n-1}(q_n) \quad (I.5)$$

I-4-1-2 Modèle géométrique inverse

Le MGD du robot permet de calculer les coordonnées opérationnelles en fonction des coordonnées articulaires. Le problème inverse consiste à calculer les coordonnées articulaires qui amènent l'organe terminal dans une situation désirée, spécifiée par ses coordonnées opérationnelles. Cette opération est très souvent appelée transformation de coordonnée (ou changement de coordonnées). Lorsqu'elle existe, la forme explicite qui donne toutes les solutions possibles au problème inverse constitue ce que nous appelons le modèle géométrique inverse (MGI), le modèle s'écrit

$$q = f^{-1}(X) \quad (I.6)$$

Problèmes du modèle géométrique inverse

➤ Existence de solution

L'existence de solution de l'équation $X = f(q)$ est conditionnée par le fait que l'effecteur évolue dans le domaine atteignable. Ce domaine est défini d'une part par des limitations dimensionnelles des éléments mécaniques formant le vecteur X et d'autre part d'éventuelles limitations structurelles.

a) limitation dimensionnelles

Ces limitations sont fonction :

- des amplitudes finies des rotations et translations que peuvent réaliser les actionneurs.
- des longueurs finies des segments constituant l'architecture mécanique du vecteur.

b) limitations structurelles

Ces limitations sont caractérisées par des architectures particulières du vecteur X , et aussi que si le nombre n d'actionneurs est inférieur ou égal à six le domaine atteignable est alors une variété de \mathbb{R}^6 de dimension inférieure ou égale à n [28].

➤ **problèmes liés à l'inversion de f**

Dans la plupart des cas d'architecture complexe du vecteur X , l'application f est fortement non linéaire et son inversion n'est possible que dans certaines régions du domaine atteignable. Dans le cas où f est déduite de l'une des méthodes de description citées précédemment. Alors les problèmes suivants se posent :

- la solution analytique n'existe pas toujours et lorsqu'elle existe son expression est souvent assez difficile à déterminer.
- lorsque deux axes de rotation sont confondus ou lorsque deux axes de translation deviennent parallèles, l'organe terminal perd un degré de liberté.
- la prise en considération des butées articulaires et des obstacles est, en général, très difficile et nécessite l'élaboration d'algorithmes spéciaux.

I-4-2 Modèle cinématique

I-4-2-1 Modèle cinématique direct

Le modèle cinématique direct permet de déterminer la vitesse de l'organe terminal dans l'espace opérationnel en fonction de la vitesse des variables articulaires. Le modèle est décrit par l'équation :

$$\dot{X} = J(q)\dot{q} \tag{I.7}$$

où $J(q) \in R^{m \times n}$ est la matrice jacobienne.

L'une des méthodes utilisées pour le calcul de la matrice jacobéenne est la dérivation du modèle géométrique direct [3]:

$$J(q) = \begin{bmatrix} \frac{\partial f_1}{\partial q_1} & \dots & \frac{\partial f_1}{\partial q_n} \\ \vdots & & \vdots \\ \frac{\partial f_m}{\partial q_1} & \dots & \frac{\partial f_m}{\partial q_n} \end{bmatrix} \tag{I.8}$$

I-4-2-1 Modèle cinématique inverse

Le modèle cinématique inverse permet de déterminer la vitesse des variables articulaires en fonction de la vitesse des variables opérationnelles. Pour les manipulateurs non redondants ($n=m$), le modèle s'écrit

$$\dot{q} = J^{-1}(q)\dot{X} \quad (I.9)$$

La solution de l'équation (I.9) existe si J est de rang plein, cela est valable tant que le manipulateur ne passe pas par une configuration singulière. Pour les manipulateurs redondants, le modèle cinématique inverse admet plusieurs solutions possibles. Le choix d'une parmi plusieurs est guidé par l'optimisation d'une fonction objective. Les méthodes employées pour résoudre un tel problème peuvent être classées en deux groupes [3].

- Méthodes d'optimisation cinématique.
- Méthodes d'optimisation dynamique.

I-4-3 Modèle dynamique

I-4-3-1 Modèle dynamique inverse

L'opération de la dynamique inverse consiste à calculer le couple résultant à l'articulation à partir des variables articulaires (positions, vitesses et accélérations). Le résultat est utile pour calculer la dynamique directe (modèle de robot) et certains contrôleurs (commande en avance, linéarisation par retour d'état etc.).

Les formalismes les plus utilisés pour le calcul du modèle dynamique inverse sont :

- Formalisme de Lagrange.
- Formalisme de Newton-Euler.
- Formalisme de Lagrange

Les équations de Lagrange opèrent à partir de l'énergie cinétique et l'énergie potentielle d'un système. Le Lagrangien L s'écrit :

$$L = K - V \quad (I.10)$$

Où :

K et V représentent respectivement les énergies cinétique et potentielle du système.

Les équations de Lagrange sont définies par :

$$\frac{d}{dt} \left(\frac{\partial L}{\partial \dot{q}} \right) - \frac{\partial L}{\partial q} = \tau_i \quad (\text{I.11})$$

Pour le calcul du modèle dynamique d'une chaîne cinématique (figure I-6), on définit les paramètres suivants pour chaque corps i de la chaîne cinématique :

m_i : masse du corps i de la chaîne cinématique

$r^i = [r_x^i, r_y^i, r_z^i]^T$ Vecteur du centre d'inertie du corps i exprimé dans le repère R_i .

$m^i r = [m^i r_x, m^i r_z, m^i r_z]^T$ Premier moment du corps i autour de O_i , exprimé dans R_i .

$I_i^i = \begin{bmatrix} I_{ixx}^i & -I_{ixy}^i & -I_{ixz}^i \\ -I_{iyx}^i & I_{iyy}^i & -I_{iyz}^i \\ -I_{izx}^i & -I_{izy}^i & I_{izz}^i \end{bmatrix}$ Tenseur d'inertie du corps i en O_i , exprimé dans R_i .

Où :

$I_{ixx}^i = \int [(r_y^i)^2 + (r_z^i)^2] dm_i$: Moment d'inertie.

$I_{ixy}^i = \int r_x^i r_y^i dm_i$: Produit d'inertie.

$\hat{S}_i^i = [\hat{S}_{ix}^i, \hat{S}_{iy}^i, \hat{S}_{iz}^i]^T$: Position du centre de masse du corps i exprimée dans R_i .

$H_i = \int r^i (r^i)^T dm_i$: Matrice des pseudo-inerties du corps i .

$$H_i = \begin{bmatrix} \frac{-I_{ixx} + I_{iyy} + I_{izz}}{2} & I_{xy} & I_{ixz} & m_i \hat{S}_{ix} \\ I_{iyx} & \frac{I_{ixx} - I_{iyy} + I_{izz}}{2} & I_{iyz} & m_i \hat{S}_{iy} \\ I_{izx} & I_{izy} & \frac{I_{ixx} + I_{iyy} - I_{izz}}{2} & m_i \hat{S}_{iz} \\ \hat{S}_{ix} & m_i \hat{S}_{iy} & m_i \hat{S}_{iz} & m_i \end{bmatrix} \quad (\text{I.12})$$

La définition reste la même pour les autres éléments du tenseur

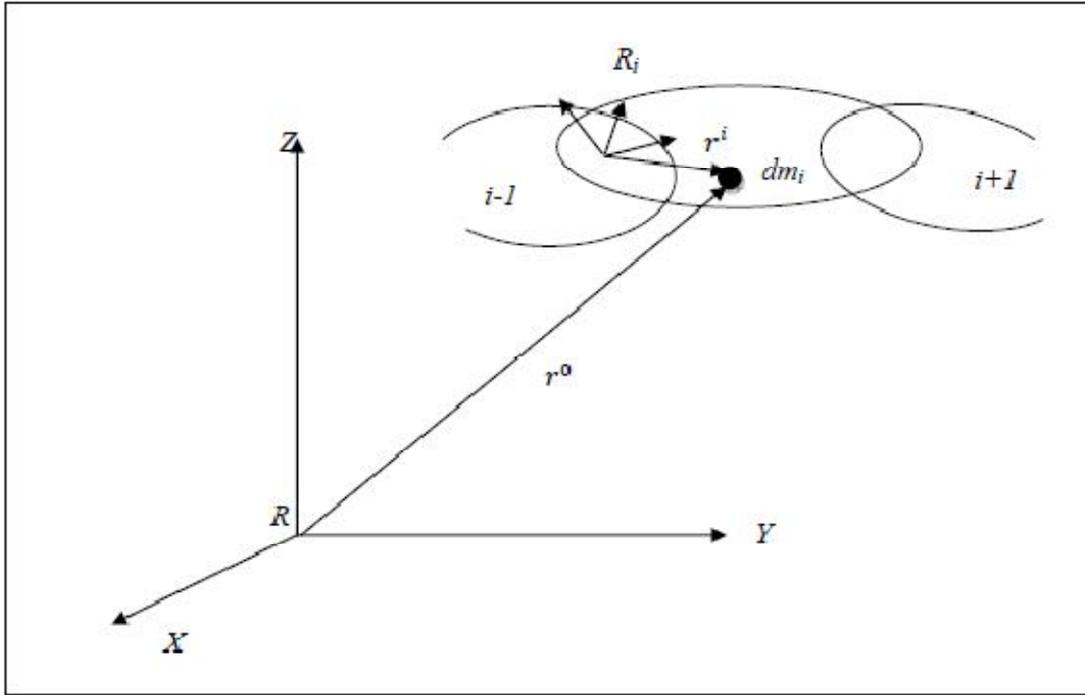


Figure I-11 Représentation d'un point r_i de la chaîne cinématique

L'énergie cinétique du corps i est donnée par :

$$K_i = \int_{\text{corps } i} \frac{1}{2} (\dot{r}^0)^T \dot{r} dm \quad (\text{I.13})$$

$$K_i = \frac{1}{2} \sum_{j=1}^i \sum_{k=1}^i \text{tr} \left(\frac{\partial T_i^0}{\partial q_k} H_i \frac{\partial (T_i^0)^T}{\partial q_k} \right) \dot{q}_j \dot{q}_k \quad (\text{I.14})$$

tr : signifie trace de la matrice ; elle est égale à la somme des éléments de la diagonale.

L'énergie de potentiel du corps i est donnée par :

$$V_i = -m_i \tilde{g}^T T_i^{0,i} \hat{S}_i \quad (\text{I.15})$$

$\tilde{g} = [\tilde{g}_x, \tilde{g}_y, \tilde{g}_z, 0]^T$: Vecteur des accélérations de la gravité exprimé dans R_0 .

Le Lagrangien pour l'ensemble des corps du manipulateur est :

$$L = \sum_{i=1}^n (K_i - V_i) \quad (\text{I.16})$$

Par substitution de (I.16) dans (I.11) on obtient :

$$\begin{aligned} \tau = & \sum_{k=i}^n \sum_{j=1}^k \text{tr} \left(\frac{\partial T_k^0}{\partial q_j} H_k \frac{\partial (T_k^0)^T}{\partial q_i} \right) \ddot{q}_j + \sum_{k=i}^n \sum_{j=1}^k \sum_{m=1}^k \text{tr} \left(\frac{\partial^2 T_k^0}{\partial q_j \partial q_m} H_k \frac{\partial (T_i^0)^T}{\partial q_i} \right) \dot{q}_j \dot{q}_m \\ & - \sum_{j=1}^n m_j \tilde{g}^T \frac{\partial T_j^0}{\partial q_i} \hat{S}_j \end{aligned} \quad (\text{I.17})$$

Le modèle obtenu peut être mis sous la forme matricielle suivante :

$$\tau = M(q)\ddot{q} + N(q, \dot{q})\dot{q} + G(q) \quad (\text{I.18})$$

Où :

$$\begin{aligned} M_{ij} = & \sum_{k=\max(i,j)}^n \text{tr} \left(\frac{\partial T_k^0}{\partial q_j} H_k \frac{\partial (T_k^0)^T}{\partial q_i} \right) \\ N(q, \dot{q}) = & \sum_{j=1}^n \sum_{m=1}^n \sum_{k=\max(i,j,m)}^k \text{tr} \left(\frac{\partial^2 T_k^0}{\partial q_j \partial q_m} H_k \frac{\partial (T_i^0)^T}{\partial q_i} \right) \dot{q}_j \dot{q}_m \end{aligned} \quad (\text{I.19})$$

$$G(q) = \sum_{j=i}^n m_j \tilde{g}^T \frac{\partial T_j^0}{\partial q_i} \hat{S}_j$$

$M(q) \in R^{n \times n}$: Matrice d'inertie du manipulateur, symétrique et régulière

$N(q, \dot{q}) \in R^n$: Vecteur des termes de Coriolis et centrifuges.

$G(q) \in R^n$: Vecteur des actions de la gravité.

➤ Formalisme de Newton-Euler

Le formalisme de Newton-Euler est basé sur les théorèmes généraux de la mécanique. Il ne nécessite pas le calcul extrinsèque des coefficients dynamiques du robot. Le modèle est obtenu par double récurrence [3] :

➤ Récurrence de la base du robot vers l'organe terminal pour le calcul des vitesses et accélérations des corps puis du tenseur dynamique :

$$\omega_i^i = \begin{cases} (A_i^{i-1})^T \omega_{i-1}^{i-1} + e_z \dot{q}_i & \text{si } R_0 \\ (A_i^{i-1})^T \omega_{i-1}^{i-1} & \text{si } P_r \end{cases} \quad (\text{I.19})$$

$$\dot{\omega}_i^i = \begin{cases} (A_i^{i-1})^T \omega_{i-1}^{i-1} + e_z \ddot{q}_i + \left((A_i^{i-1})^T \omega_{i-1}^{i-1} \right) \otimes e_z \dot{q}_i & \text{si } R_0 \\ (A_i^{i-1})^T \dot{\omega}_{i-1}^{i-1} & \text{si } P_r \end{cases} \quad (\text{I.21})$$

$$\ddot{P}_i = \begin{cases} (A_i^{i-1})^T [\ddot{P}_{i-1}^{i-1} + \dot{\omega}_{i-1}^{i-1} \otimes \hat{P}_{i-1}^{i-1} + \omega_{i-1}^{i-1} \otimes (\omega_{i-1}^{i-1} \otimes \hat{P}_{i-1}^{i-1})] & \text{si } R_0 \\ (A_i^{i-1})^T [\ddot{P}_{i-1}^{i-1} + \dot{\omega}_{i-1}^{i-1} \otimes \hat{P}_{i-1}^{i-1} + \omega_{i-1}^{i-1} \otimes (\omega_{i-1}^{i-1} \otimes \hat{P}_{i-1}^{i-1})] + \\ 2 \left((A_i^{i-1})^T \omega_{i-1}^{i-1} \right) \otimes (e_z \dot{q}_i) + e_z \ddot{q}_i & \text{si } P_r \end{cases} \quad (\text{I.22})$$

$$\ddot{S}_i^i = \ddot{P}_i + \dot{\omega}_i^i \otimes \hat{S}_i^i + \omega_i^i \otimes (\omega_i^i \otimes \hat{S}_i^i) \quad (\text{I.23})$$

$$\hat{f}_i^i = m_i \ddot{S}_i^i \quad (\text{I.24})$$

$$\hat{n}_i^i = I_i^i \omega_i^i + \omega_i^i \otimes (I_i^i \omega_i^i) \quad (\text{I.25})$$

➤ Récurrence de l'organe terminal vers la base pour le calcul des couples actionneurs :

$$f_i^i = A_{i+1}^i f_{i+1}^{i+1} + \hat{f}_i^i \quad (\text{I.26})$$

$$n_i^i = A_{i+1}^i n_{i+1}^{i+1} + \hat{n}_i^i + \hat{S}_i^i \otimes \hat{f}_i^i + \hat{P}_{i+1}^i \otimes (A_{i+1}^i f_{i+1}^{i+1}) \quad (\text{I.27})$$

$$\tau_i = \begin{cases} e_z^T n_i^i & \text{si } R_0 \\ e_z^T f_i^i & \text{si } P_r \end{cases} \quad (\text{I.28})$$

Avec

$$e_z = [0 \ 0 \ 1]^T \quad (\text{I.29})$$

R_0 : rotation rotoïde.

P_r : rotation prismatique.

ω_i^i et \ddot{P}_i^i : vitesse et accélération du repère R_i exprimées dans R_i .

f_i^i et n_i^i : force et moment appliqués à l'articulation i .

\hat{f}_i^i et \hat{n}_i^i : force et moment appliqués au centre de masse du corps i .

$$\text{et } \omega \otimes = \begin{bmatrix} 0 & -\omega_z & \omega_y \\ \omega_z & 0 & -\omega_x \\ -\omega_y & \omega_x & 0 \end{bmatrix} \quad (\text{I.30})$$

Si aucune force et aucun moment n'est exercé sur le manipulateur on prend comme condition terminale

$$\begin{cases} f_{i+1}^{i+1} = 0 \\ n_{i+1}^{i+1} = 0 \end{cases} \quad (\text{I.31})$$

et les conditions initiales :

$$\ddot{P}_0^0 = -[\tilde{g}_x \ \tilde{g}_y \ \tilde{g}_z]^T \quad (\text{I.32})$$

$$\begin{cases} \omega_0^0 = 0 \\ \omega_0^0 = 0 \end{cases} \quad (\text{I.33})$$

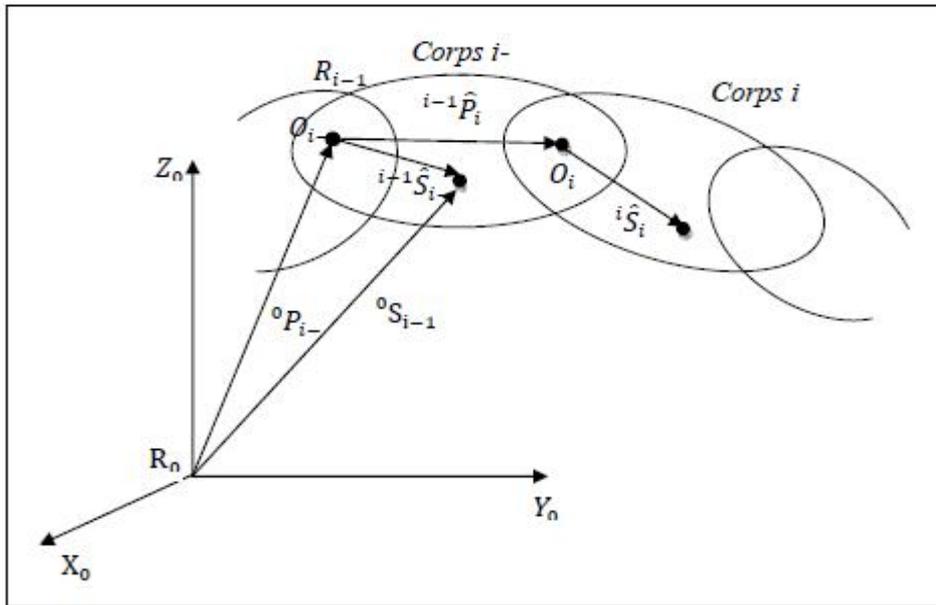


Figure I-12 Formalisme de Newton-Euler

S'il est nécessaire de considérer les frottements aux niveaux des articulations on ajoute un terme τ_f aux équations (I.17) ou (I.28). On obtient alors :

$$\tau = M(q)\ddot{q} + N(q, \dot{q})\dot{q} + G(q) + \tau_f \quad (\text{I.34})$$

$$\tau_i = \begin{cases} e_z^T n_i^i + \tau_{fi} & \text{si } R_0 \\ e_z^T f_i^i + \tau_{fi} & \text{si } P_r \end{cases} \quad (\text{I.35})$$

Différents modèles mathématiques peuvent être utilisés pour exprimer τ_f . Pour une modélisation simple, on choisit :

$$\tau_{fi} = k_{fvi}\dot{q}_i + k_{fci}\text{sgn}(\dot{q}_i) \quad (\text{I.36})$$

où k_{fvi} et k_{fci} sont respectivement le coefficient des frottements visqueux et de Coulomb.

$$\text{sgn}(\dot{q}_i) = \begin{cases} -1 & \text{si } \dot{q}_i < 0 \\ 0 & \text{si } \dot{q}_i = 0 \\ 1 & \text{si } \dot{q}_i > 0 \end{cases} \quad (\text{I.37})$$

I-4-3-2 Modèle dynamique direct

Le modèle dynamique direct exprime la position, la vitesse et l'accélération des articulations en fonction du couple appliqué [3]. Il est obtenu par inversion du modèle précédant, il s'écrit :

$$\ddot{\mathbf{q}} = \mathbf{M}^{-1}(\mathbf{q})(\boldsymbol{\tau} - \mathbf{N}(\mathbf{q}, \dot{\mathbf{q}})\dot{\mathbf{q}} - \mathbf{G}(\mathbf{q}) - \boldsymbol{\tau}_f) \quad (\text{I-38})$$

I-4-3-3 Problèmes du modèle dynamique

Le développement de la technologie des calculateurs et des logiciels du calcul symbolique peut remédier à certains problèmes de modélisation tels que le risque d'erreur lors des calculs mais, il existe toujours d'autres problèmes. Le modèle d'un robot est conçu non seulement dans un but de simulation, mais surtout de la commande. Dans ce dernier cas on se trouve face à trois types de problèmes :

- le modèle doit être valide, c'est-à-dire suffisamment précis. Or, plus le comportement devient dynamique moins on va pouvoir négliger des effets considérés comme parasites.
- le modèle dynamique étant essentiellement non linéaire, de nombreux coefficients fonction de la configuration vont devoir être évalués en ligne ce qui pose un problème d'exécution en temps réel.

I-5 Conclusion

Nous avons abordé dans ce chapitre les différentes structures et les modèles des bras manipulateurs, on a constaté d'après cette approche que l'extraction du modèle d'un bras consiste à savoir toutes ses caractéristiques géométriques et mécaniques et nous avons vu aussi les deux méthodes essentielles pour obtenir le modèle dynamique des bras manipulateurs, il s'agit de la méthode de Newton-Euler et de Lagrange.

On a constaté durant ce chapitre que l'obtention de différents modèles n'est pas aisée, la difficulté variant selon la complexité de la cinématique de la chaîne articulée. Entrent en ligne de compte le nombre de degrés de liberté, le type des articulations mais aussi le fait que la chaîne peut être ouverte simple, arborescente ou fermée.

Chapitre II

Le signal De Parole

II-1 Introduction :

La parole est la capacité de l'être humain de communiquer la pensée par l'intermédiaire de sons articulés. Dû à son importance, la parole a préoccupé depuis toujours les scientifiques. Ainsi quelques-unes des sciences qui se préoccupent de l'étude de la parole ont déjà des centaines d'années.

Dans ce chapitre on va voir une approche générale sur l'identité du signal de parole commençant par tout ce qui concerne sa production y compris l'architecture de l'appareil vocal, le mécanisme de la phonation et la classification des phonèmes 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 Production de la parole

II-2-1 Architecture de l'appareil vocal

L'appareil vocal, ou système phonatoire, comprend quatre éléments fondamentaux fonctionnant en étroite synergie pour produire des signaux acoustiques. Ce sont, dans l'ordre où ils s'élaborent :

- La soufflerie
- Le vibreur
- Le corps sonore
- Le système articulateur

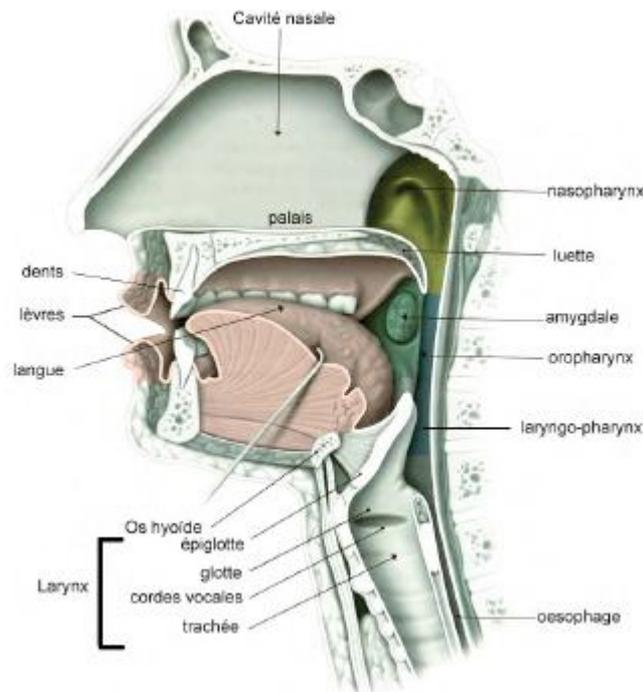


Figure II-1 L'appareil vocal

La soufflerie est constituée d'un réservoir d'air, les poumons, actionnés par les muscles du thorax et de l'abdomen, et d'un tube, la trachée artère, qui conduit l'air aux cordes vocales ; le vibreur est le larynx, qui engendre les ondes aériennes; le corps sonore est constitué d'un ensemble complexe de résonateurs, dont le pharynx et la bouche sont les principaux; le système articulateur, enfin, se compose d'éléments fixes et mobiles permettant de modifier largement la forme de l'onde laryngée. Tous ces éléments sont placés sous la dépendance étroite du système nerveux central, qui en assure le synchronisme et la coordination [6] .

II-2-1-1 La soufflerie et Le vibreur

L'air est la matière première de la voix. Si le fonctionnement de notre appareil vocal est souvent comparé à celui d'un instrument de musique, il doit être décrit comme celui d'un instrument à vent. En effet, en expulsant l'air pulmonaire à travers la trachée, le système respiratoire joue le rôle d'une soufflerie. Il s'agit du « souffle phonatoire » produit, soit par l'abaissement de la cage thoracique, soit dans le cadre de la projection vocale par l'action des muscles abdominaux [7].

II-2-1-2 Le larynx

Le larynx constitue l'extrémité supérieure de la trachée artère, situé à la hauteur de la sixième vertèbre cervicale (chez l'adulte). C'est un assemblage de cartilages articulés, reliés entre eux par des ligaments et des muscles (dont les cordes vocales), l'ensemble étant tapissé d'une muqueuse [6].

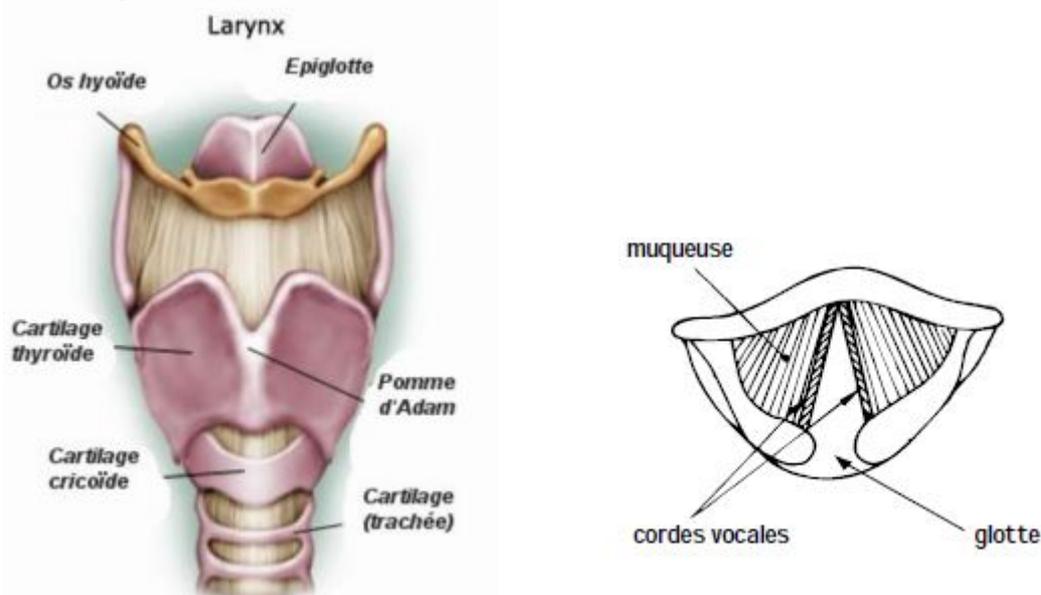


Figure II-2 Vue schématique antérieure du larynx (Gauche) et sa section, vue de haut (droite)

II-2-1-3 Le corps sonore

Les résonateurs du système phonatoire sont, pour l'essentiel, responsables du timbre de la voix. Leur originalité par rapport aux caisses de résonance des instruments de musique traditionnels est leur faculté de changer, grâce à un réseau musculaire dense et élaboré, — dans de larges proportions, et très rapidement — de forme et de volume, assurant ainsi au son vocal une variété acoustique sans équivalent.

Les résonateurs sont au nombre de cinq : le pharynx, la cavité buccale divisée en deux, la cavité labiale, et la cavité nasale. Tous communiquent entre eux par des ouvertures de tailles réglables. Tapissés de muqueuse, ils sont peu amortis.

Les sinus de la face, contrairement à une croyance largement répandue, sont trop petits pour se comporter en résonateurs, et n'ont d'ailleurs aucune fonction reconnue dans la phonation [6].

II-2-1-4 Le système articulateur

Il intègre une ensemble d'organes mobiles, le voile du palais, la mâchoire inférieure (ou mandibule), la langue et les lèvres. Les mouvements de la mâchoire inférieure contribuent largement aux variations de volume de la bouche. La langue, reliée par sa base à l'os hyoïde, est extrêmement mobile, car commandée par dix-sept muscles (huit pairs et un impair) ; elle prend appui sur différents points du conduit pharyngo-buccal pour articuler les phonèmes. Ces points d'articulation sont [6] :

- les lèvres (articulations labiales ou bilabiales)
- les dents (articulations dentales)
- les alvéoles¹⁶ (articulations alvéolaires)
- le palais dur, ou partie osseuse de la voûte (articulations palatales)
- le voile du palais ou « palais mou » (articulations vélares)
- la luette (articulations uvulaires)
- le pharynx (articulations pharyngales)
- la glotte (articulations glottales)

II-2-2 Mécanisme de la phonation

L'une de plus importantes caractéristiques du signal vocal est la nature de l'excitation. Il existe deux types élémentaires d'excitation qui produisent les sons voisés et non voisés.

II-2-2-1 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 l'air qui vient des poumons elles 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 [9]. Ce processus périodique est caractérisé par une fréquence propre à chaque personne, connue sous le nom de fréquence du fondamental (F_0) où pitch et il donne la hauteur normale de la voix. La fréquence fondamentale peut varier de 80 à 200 Hz pour une voix masculine, de 150 à 450 Hz pour une voix féminine et de 200 à 600 Hz pour une voix d'enfant [10].

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-3.

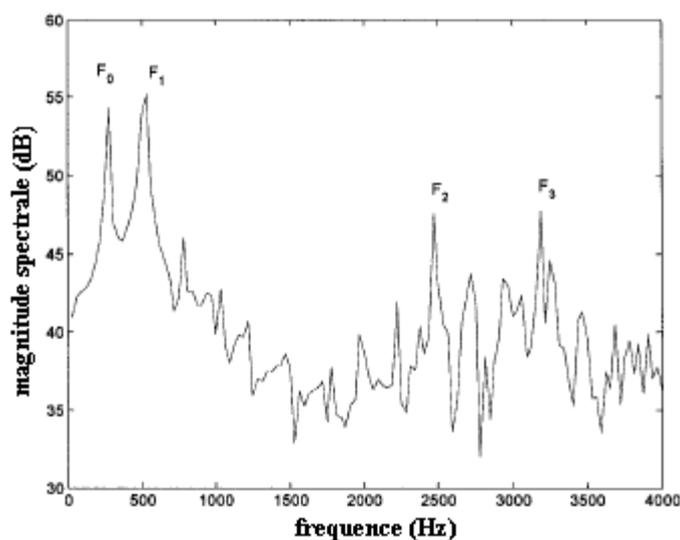


Figure II-3 Le spectre du son voisé /i/

II-2-2-1 Phonation de sons non voisés

Les sons non voisés sont générés par le passage de l'air dans une constriction étroite située en un point du conduit vocal. Ils sont générés sans l'apport du larynx et ne présentent pas de structure périodique [9]. Ces caractéristiques d'un son non voisé peuvent être observées dans la figure II-4.

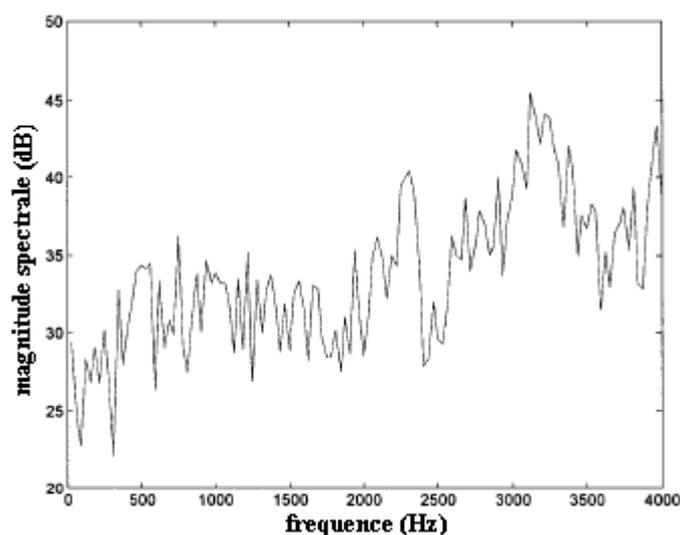


Figure II-4 Le spectre du son non voisé /f/

II-2-3 Classification des phonèmes

Il y a plusieurs façons de classer les phonèmes. Un phonème est stationnaire ou continuant si la configuration du conduit vocal ne change pas pendant la production du son. Un phonème est non continuant si pendant sa production il y a des changements dans la configuration du conduit vocal [9].

On peut grouper les 36 phonèmes de la langue française en classes et sous classes d'après le mode d'articulation de l'appareil de phonation (tableau II-1).

II-2-3-1 Les voyelles

Les voyelles sont des sons voisés, continus, normalement avec la plus grande amplitude parmi tous les phonèmes et elles peuvent varier beaucoup en durée, entre 40 et 400 ms.

Les voyelles orales sont produites sans l'intervention de la cavité nasale pendant que pour les voyelles nasales, le conduit nasal est couplé à la cavité buccale et la production de son se fait par la bouche et par les narines en même temps. Les voyelles sont différenciées en trois groupes d'après la position de la courbure de la langue et le degré de la constriction induit dans le conduit vocal.

Différenciées en trois groupes d'après la position de la courbure de la langue et le degré de constriction induit dans le conduit vocal.

L'analyse dans le domaine temporel et fréquentiel révèle plusieurs caractéristiques acoustiques qui aident à la classification de chaque son. L'analyse dans le domaine temporel montre que les voyelles sont de sons quasi périodiques dus à l'excitation.

Phonèmes								
Voyelles		Semi- Consonnes	Consonnes					
Orales	Nasales		Liquides	Nasales	Fricatives		Occlusives	
					Voisées	Non- Voisées	Voisées	Non- Voisées
i(I)	ẽ (IN)	j (Y)	l(L)	m(M)	v(V)	f(F)	b(B)	p(P)
e(E)	œ(UN)	w(W)	R(R)	n(N)	z(Z)	s(S)	d(D)	t(T)
ɛ(AI)	ã(AN)	ɥ (UI)		ɲ(GN)	ʒ(I)	ʃ (CH)	g(G)	k(K)
a(A)	õ(ON)							
ɔ(O)								
u(OU)								
y(U)								
ø(EU)								
œ(OE)								
ə(E)								
o(AU)								

Tableau II-1 Phonèmes de la langue française

Les voyelles peuvent être identifiées par les locations de leurs fermants dans le domaine fréquentiel. La position des deux premiers fermants est suffisante pour caractériser la majorité des voyelles, le troisième formant est nécessaire juste pour quelques-uns. La position de fermants de fréquence plus élevée reste presque inchangée et n'apporte pas d'information utile pour l'identification.

II-2-3-2 Les diphtongues

Les diphtongues impliquent un mouvement d'une voyelle initiale vers une autre voyelle finale. Donc les diphtongues sont essentiellement des sons non continus. La différence entre une diphtongue et les deux voyelles individuelles composantes est que la durée de la transition est plus grande que la durée de chaque voyelle. De plus la voyelle initiale est plus longue que la voyelle finale. Dans la parole les deux voyelles composant une diphtongue peuvent ne pas être réalisées entièrement ce qui accentue l'idée de non-stationnarité qui caractérise les diphtongues.

II-2-3-3 Les semi-consonnes

Les semi -consonnes sont des sons non continus et voisés qui possèdent des caractéristiques spectrales semblable aux voyelles. On peut voir les semi-consonnes comme des sons transitoires qui s'approchent, atteint et après s'éloignent d'une positionne cible. La durée des transitions est comparable à la durée passée en position cible.

II-2-3-4 Les consonnes

Les consonnes sont de sons pour lesquels le conduit vocal est plus étroit pendant la production, par rapport aux voyelles. Les consonnes impliquent les deux formes d'excitation pour le conduit vocal et elles peuvent être continuantes ou non.

➤ Les consonnes fricatives

Les fricatives non voisées résultent d'une turbulence créée par le passage de l'air dans une constriction du conduit vocal qui peut se trouver près des lèvres pour les labiales, au milieu du conduit vocal pour les dentales et au fond du conduit vocal pour le palatales.

Dans ce cas la constriction cause une source de bruit et aussi divise le conduit vocal en deux cavités. La première cavité actionne comme une enceinte anti-résonante qui atténue les basses fréquences d'où la concentration de l'énergie vers les hautes fréquences dans le domaine spectral.

Pour les fricatives voisées l'excitation est mixte et à la source de bruit s'ajoutent les impulses périodiques créés par la vibration de cordes vocales.

➤ Les consonnes occlusives

Les consonnes occlusives sont des sons non continuants qui sont des combinaisons de sons voisés, non voisés et de courtes périodes de silence. Une forte pression d'air s'accumule avant une occlusion totale dans un point du conduit vocal qui après est relâché brusquement. Cette période d'occlusion s'appelle la phase de tenue.

Pour les occlusives non voisés la phase de tenue est un silence et la période de friction qui suit est plus longue que pour les occlusives voisés. Pour les occlusives voisés, pendant la phase de tenue, un son de basse fréquence est émis par vibration des cordes vocales.

➤ Les consonnes nasales

Les consonnes nasales sont des sons continus et voisés. Les vibrations produites par les cordes vocales excitent le conduit vocal que cette fois est formé de la cavité nasale ouverte et la cavité buccale fermée. Même fermée, la cavité buccale est couplée à la cavité nasale et influence la production de sons comme une enceinte anti-résonante qui atténue certaines fréquences, en fonction du point où elle est fermée. Les formes d'onde des consonnes nasales

ressemblent aux celles des voyelles mais sont normalement plus faibles en énergie due à la capacité réduite de la cavité nasale de radier des sons par rapport à la cavité buccale.

➤ **Les consonnes liquides**

Les consonnes liquides sont des sons non continus et voisés qui possèdent des caractéristiques spectrales similaires aux voyelles. Elles sont plus faibles en énergie due au fait que le conduit vocal est plus étroit pendant leur production.

II-3 Le système auditif

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-5). 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-6. La fréquence de résonance dépend de la position occupée par la cellule sur la membrane; au-delà 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.

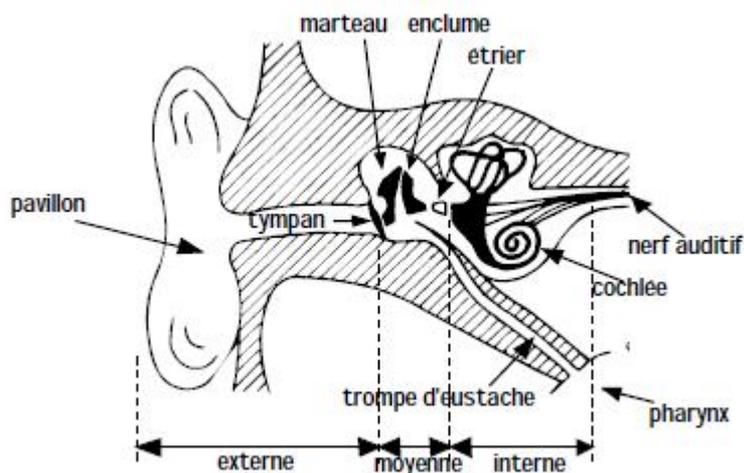


Figure II-5 Coupe de l'appareil auditif humain [11]

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 psychoacoustique. Sans vouloir entrer dans trop de détails sur la contribution majeure des psychoacousticiens 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-7 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).[11]

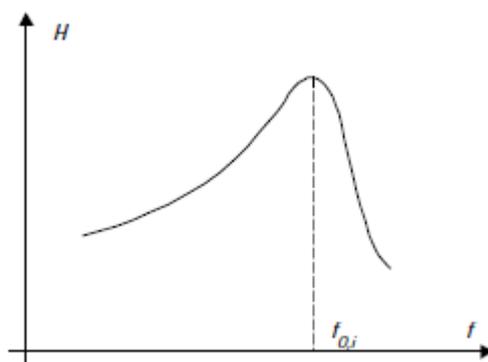


Figure II-6 Réponse en fréquence d'une cellule ciliée.[11]

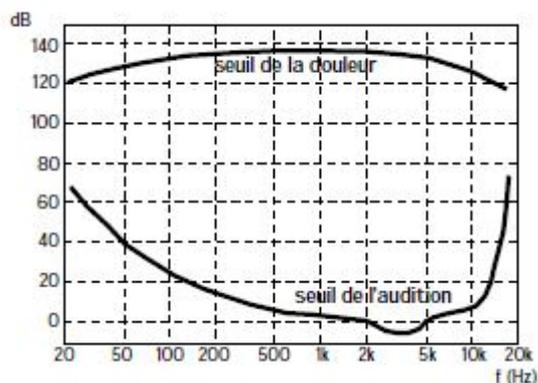


Figure II-7 Le champ auditif humain.[11]

A l'intérieur de son domaine d'audition, l'oreille ne présente pas une sensibilité identique à toutes les fréquences. La figure II-8 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[11].

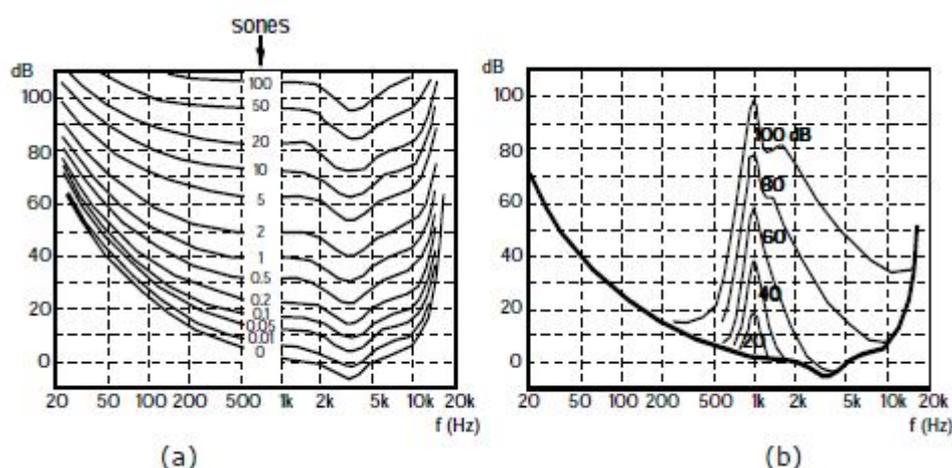


Figure II-8(a) : Courbes isononiques en champ ouvert. (b) : Masquage auditif par un bruit à bande étroite : limite d'audibilité en fonction de la puissance du bruit masquant.[11]

II-4 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-9, requiert successivement : un filtrage de garde, un échantillonnage, et une quantification.[11]

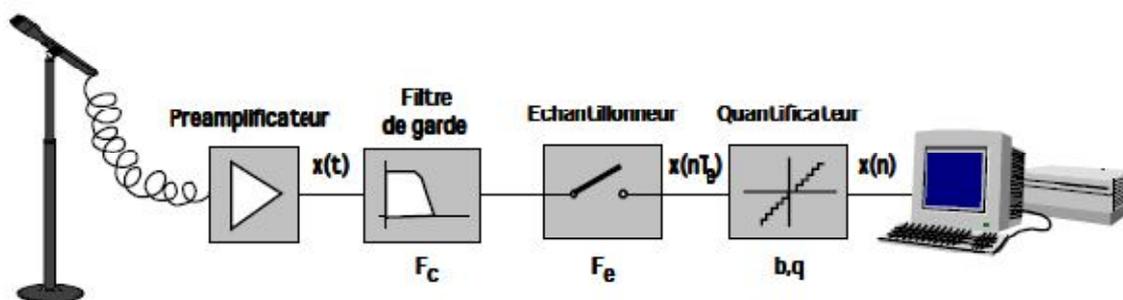


Figure II-9 Enregistrement numérique d'un signal acoustique.[11]

II-4-1 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'échantillonnage T_e ; celle-ci est elle-même l'inverse de la fréquence d'échantillonnage f_e . 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 2^b 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) [11].

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) [11]. La figure II-10 représente l'évolution temporelle, ou audiogramme, du signal vocal pour les mots 'gauche', et 'bas'.

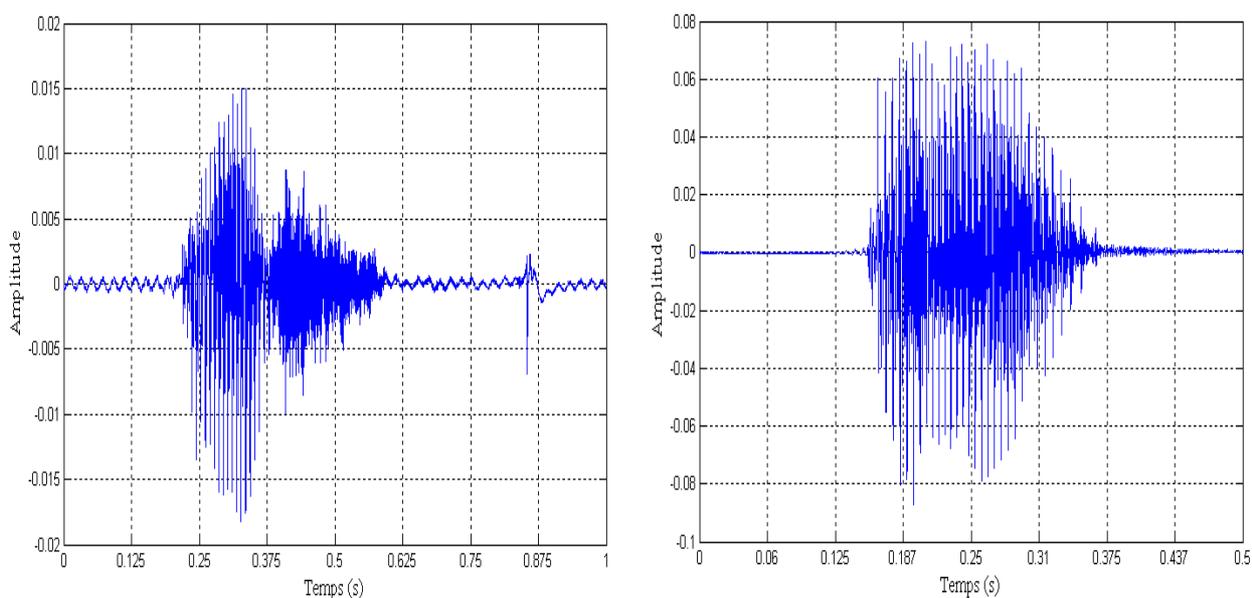


Figure II-10 Audiogramme de signaux de parole à gauche le mot « gauche » et à droite le mot « bas »

II-4-2 Transformée de Fourier à court terme

La transformée de Fourier à court terme est obtenue en extrayant de l'audiogramme une trentaine de ms de signal vocal, en pondérant ces échantillons par une fenêtre de pondération (souvent une fenêtre de Hamming) et en effectuant une transformée de Fourier sur ces échantillons[11].

II-4-3 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 du spectre y apparaît sous la forme de niveaux de gris dans un diagramme en deux dimensions temps-fréquence. On parle de spectrogramme à large bande ou à bande étroite selon la durée de la fenêtre de pondération (Figure. II-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[11].

II-4-4 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[11]. En acoustique, la **fréquence fondamentale** ou **son fondamental** est l'harmonique de premier rang d'un son.

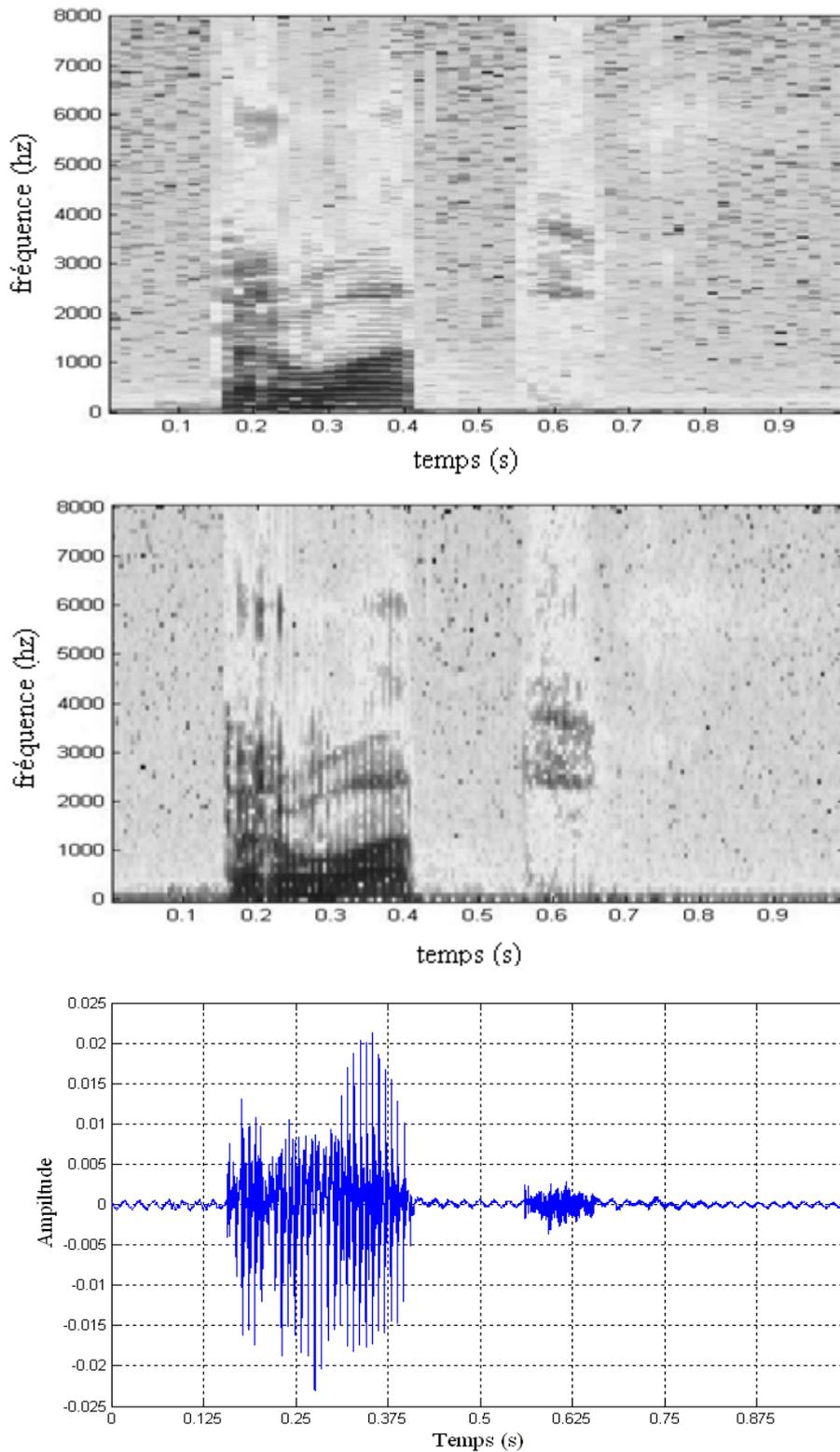


Figure II-11 Spectrogrammes à large bande (en haut), à bande étroite (en bas), et évolution temporelle du mot 'Droite', échantillonnée à 16 kHz (calcul avec fenêtres de Hamming de 10 et 30 ms respectivement).

II-5 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 acoustico-phonétiques du signal de parole. Au niveau phonétique nous avons vu le principe général de la production de la parole, l'architecture de l'appareil vocal, le mécanisme de phonation ...etc. Au niveau acoustique on s'est intéressé essentiellement par l'évolution temporel du signal de parole, sa transformée de Fourier à court terme et sa fréquence fondamentale.

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 applications possibles 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 statistiques : les modèles de Markov cachés (en anglais Hidden Markov Models : HMM) capables 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 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 quelques systèmes de reconnaissance ou de synthèse exploitables 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 liés à sa reconnaissance. Ensuite, nous détaillons les différentes étapes de son traitement qui visent à extraire des coefficients caractéristiques du signal. Nous étudions particulièrement les coefficients MFCC (Mel Frequency Cepstrum Coefficients) utilisés dans la majorité des systèmes de reconnaissance actuels. Enfin nous décrivons l'approche statistique pour la reconnaissance fondée sur l'utilisation des modèles de Markov cachés.[20]

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 [11].

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

Evidemment, les systèmes dépendants du locuteur sont plus faciles à développer et sont caractérisés par de meilleurs taux de reconnaissance que les systèmes indépendants du locuteur étant donné que la variabilité du signal de parole est plus limitée. Cette dépendance au locuteur est cependant acquise au prix d'un entraînement spécifique à chaque utilisateur. Ceci n'est cependant pas toujours possible. Par exemple, dans le cas d'applications téléphoniques, il est évident que les systèmes doivent pouvoir être utilisés par n'importe qui et doivent donc être indépendants du locuteur. Bien que la méthodologie de base reste la même, cette indépendance au locuteur est cependant obtenue par l'acquisition de nombreux locuteurs (couvrant si possible les différents dialectes) qui sont utilisés simultanément pour l'entraînement de modèles susceptibles d'en extraire toutes les caractéristiques majeures. Une solution intermédiaire parfois utilisée est de développer des systèmes capable de s'adapter (de façon supervisée ou non supervisée) rapidement au nouveau locuteur.

Le système reconnaît-il des mots isolés ou de la parole continue? Evidemment, il est plus simple de reconnaître des mots isolés bien séparés 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 hésitations, 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 plait ». 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 [11] .

III-2-2 Extraction des paramètres

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

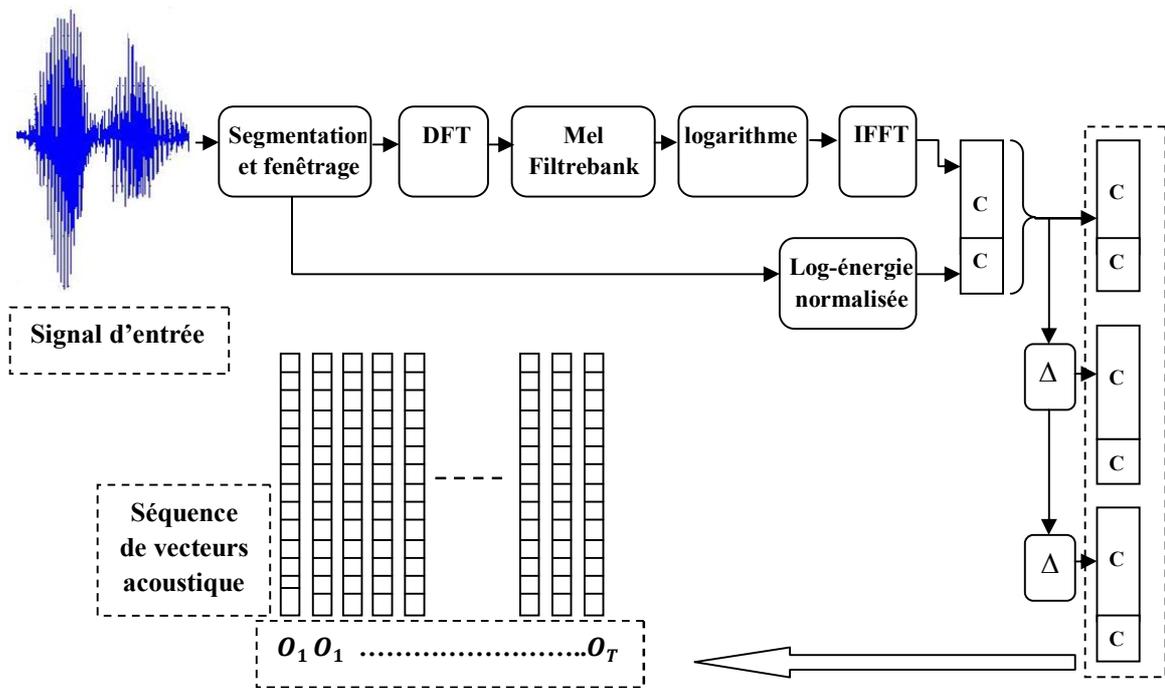


Figure III-1 Les étapes d'une paramétrisation MFCC.

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 continue $x(t)$ (où t désigne le temps) en un signal numérique $x(n)$ où n correspond à des instants discrets [20].

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 [20].

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 partie à traiter du signal et annuler la partie 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.

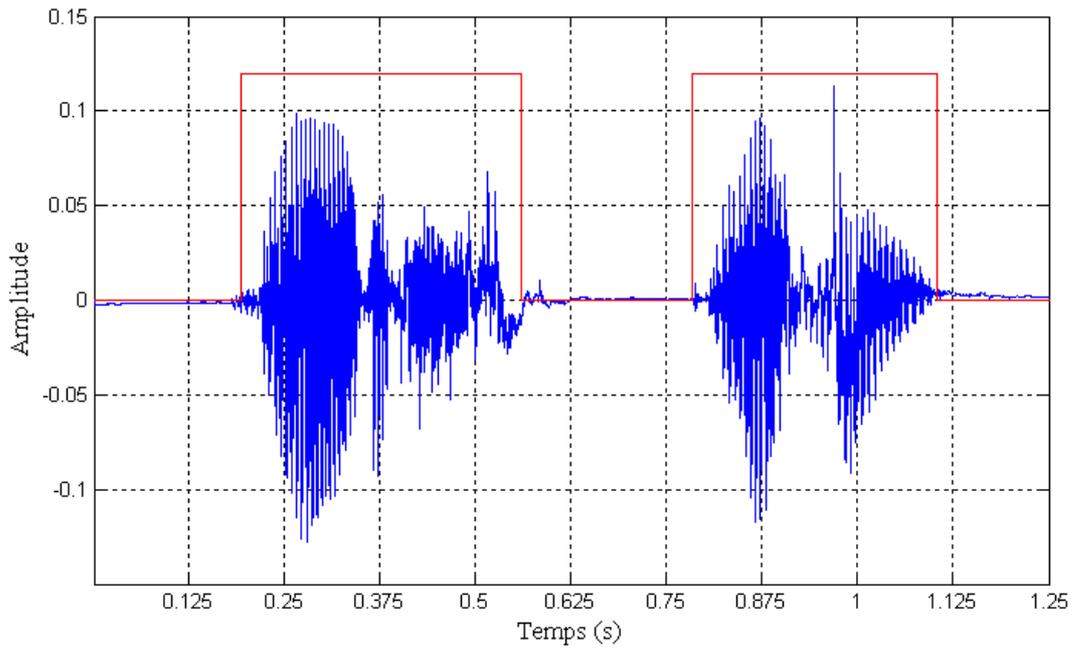


Figure III-2 détecteur appliqué à la phrase « marche avant »

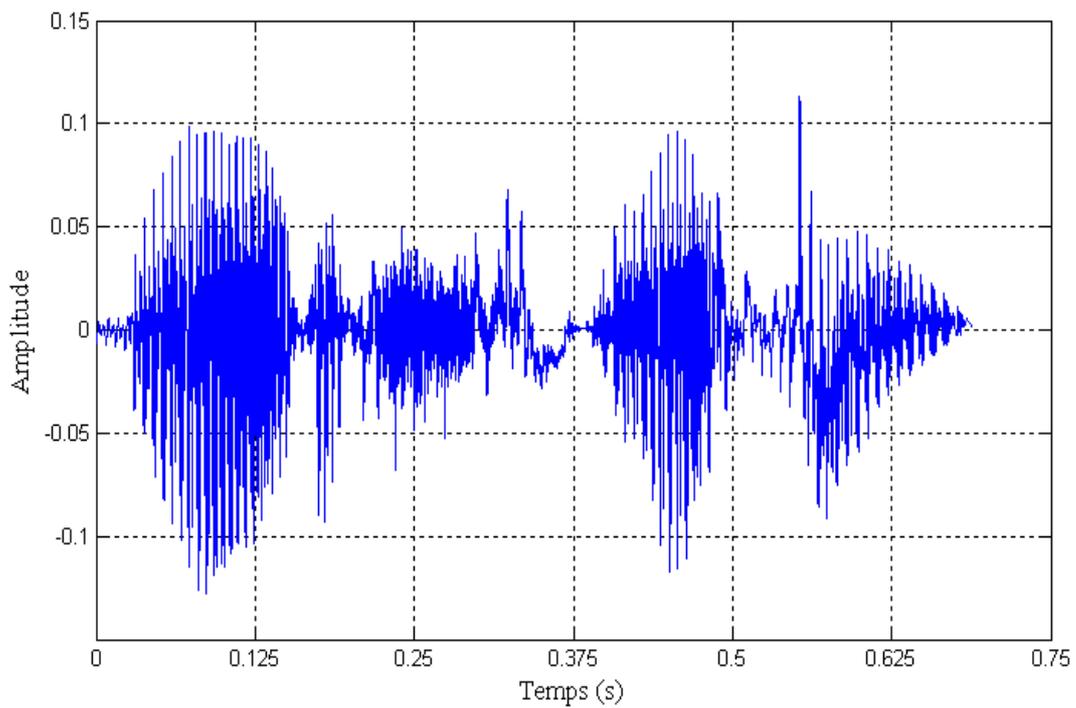


Figure III-3 la phrase « marche avant » après la suppression du silence

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 Hamming [20] :

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

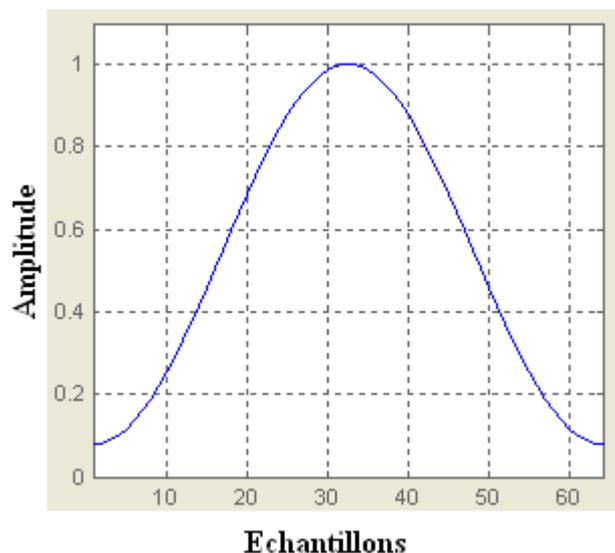


Figure III-4 Fenêtre de Hamming

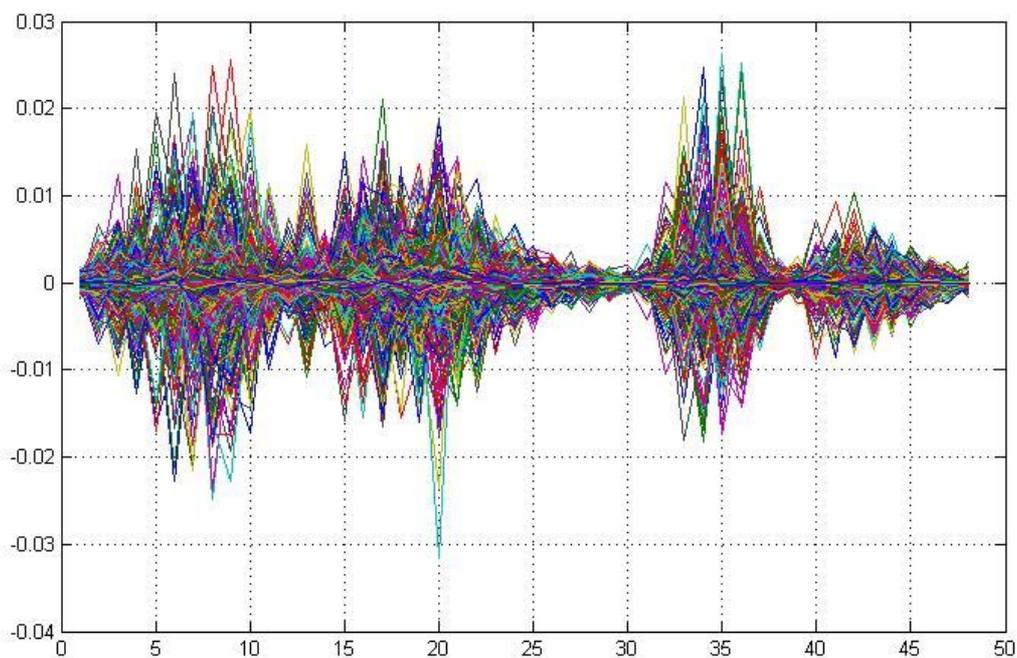


Figure III-5 représentation du signal de la phrase « marche avant » après la segmentation et le fenêtrage.

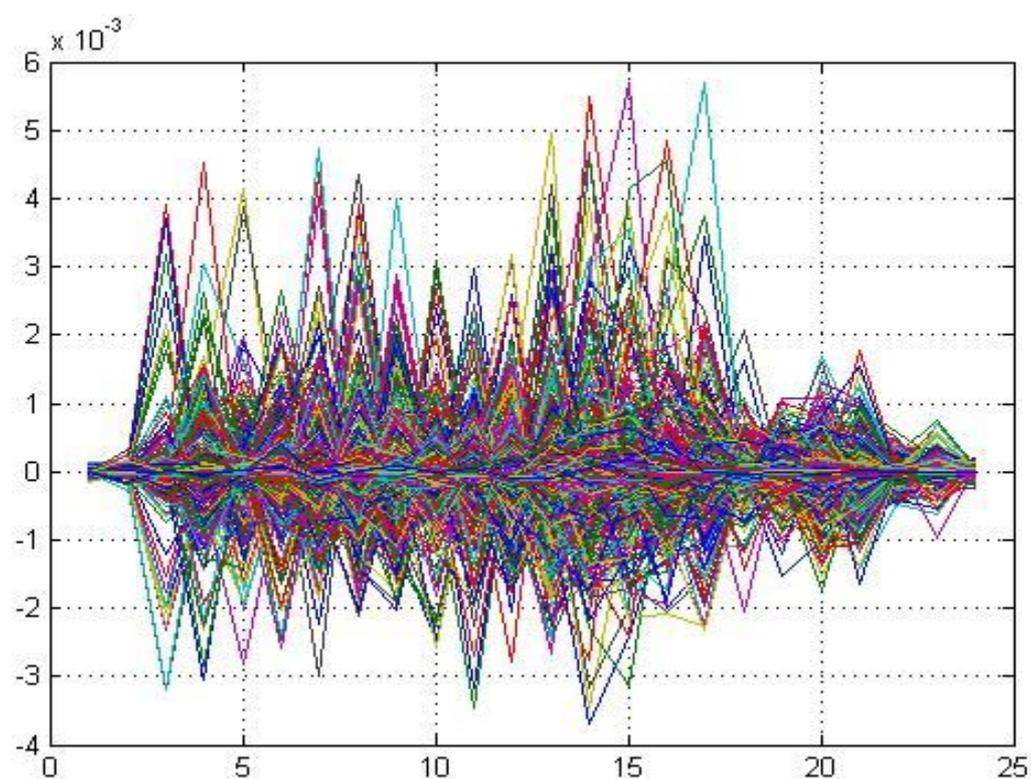


Figure III-6 représentation du signal du mot « droite » après la segmentation et le fenêtrage.

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 [20].

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 : Fast 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 :

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

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 mels est la suivante :

$$M_{mels} = x \cdot \log \left(1 + \frac{f_{Hz}}{y} \right) \quad (\text{III. 3})$$

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 [15].

Les sorties du banc de filtres peuvent être utilisées directement comme paramètres d'un système de reconnaissance, comme dans le système d'IBM [16]. 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 [20].

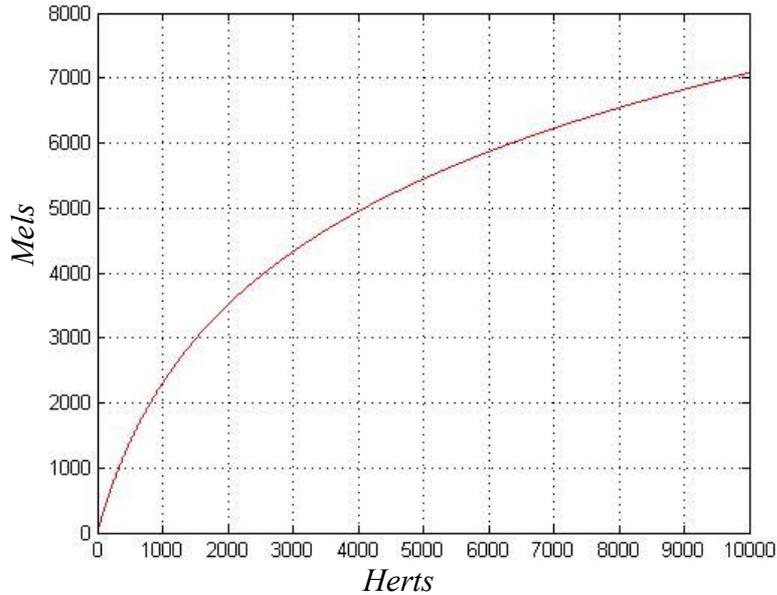


Figure III-7 Echelle de Mel

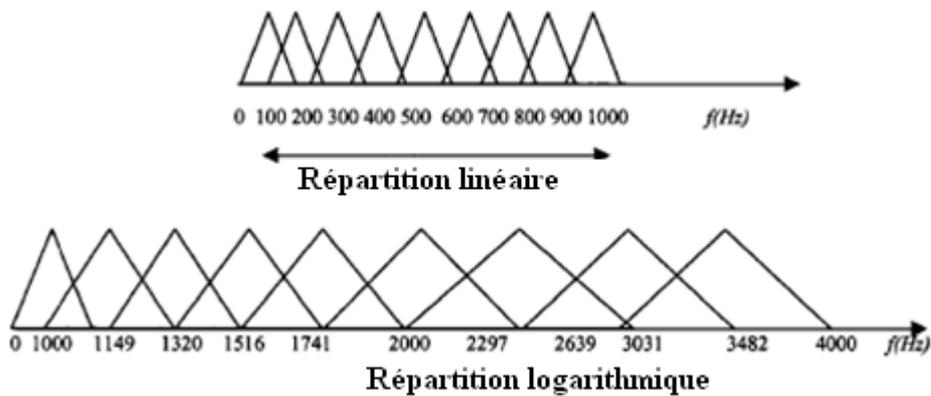


Figure III-8 Type de banc filtre à échelle de Mel utilisant des filtres triangulaires.

Le spectre du signal est donné par la transformée de Fourier inverse du logarithme de la densité spectrale. Une représentation spectrale présente l'intérêt de dissocier, dans le cadre de

la théorie d'un modèle source-filtre, l'excitation glottique et les résonances du conduit vocal. Cette séparation est réalisée par un homomorphisme qui transforme la convolution des signaux dans le domaine temporel en une addition dans le domaine cepstral. En effet, le signal $x(n)$ est considéré comme une convolution dans le domaine temporel, du signal d'excitation $g(n)$ et de la réponse impulsionnelle du conduit vocal $h(n)$:

$$x(n) = g(n) * h(n) \quad (\text{III. 4})$$

L'application du logarithme du module de la transformée de Fourier à cette équation donne :

$$\log|X(k)| = \log|G(k)| + \log|H(k)| \quad (\text{III. 5})$$

Où $X(k)$, $G(k)$ et $H(k)$ sont les spectres respectifs de $x(n)$, $g(n)$ et $h(n)$. La transformée de Fourier inverse de cette quantité donne le spectre, qui est la somme des contributions de la source et du conduit. La formule d'une FFT inverse est la suivante :

$$FFT^{-1}\{X(k)\} = x_p(n) = \frac{1}{N} \sum_{k=-\frac{N}{2}}^{\frac{N}{2}-1} X(k) e^{jk2\pi(\frac{n}{N})} \quad (\text{III. 6})$$

En appliquant cette formule à la quantité $\log|H(k)|$ nous obtenons les coefficients MFCC du signal $x(t)$.

Afin de mieux caractériser le signal de parole qui est un signal très variable, des informations sur son évolution dans le temps sont aussi utilisées. Il s'agit des dérivées premières et secondes des paramètres cepstraux du signal. Les dérivées premières permettent de représenter la variation de chacun de ces paramètres dans le temps et les dérivées secondes représentent leurs accélérations.

Plusieurs autres types de paramétrisation sont présents dans la littérature [21]. Nous pouvons citer entre autres le codage LPC (Linear Predictive Coding) et la paramétrisation PLP (Perceptual Linear Predictive). En plus, nous pouvons trouver d'autres techniques visant à améliorer la qualité de ces paramètres comme l'analyse discriminante linéaire (LDA : Linear Discriminant Analysis) et l'analyse discriminante non linéaire NLDA. Des modèles plus complexes pour l'audition et la perception existent aussi et ils ont été étudiés dans le domaine de la parole.

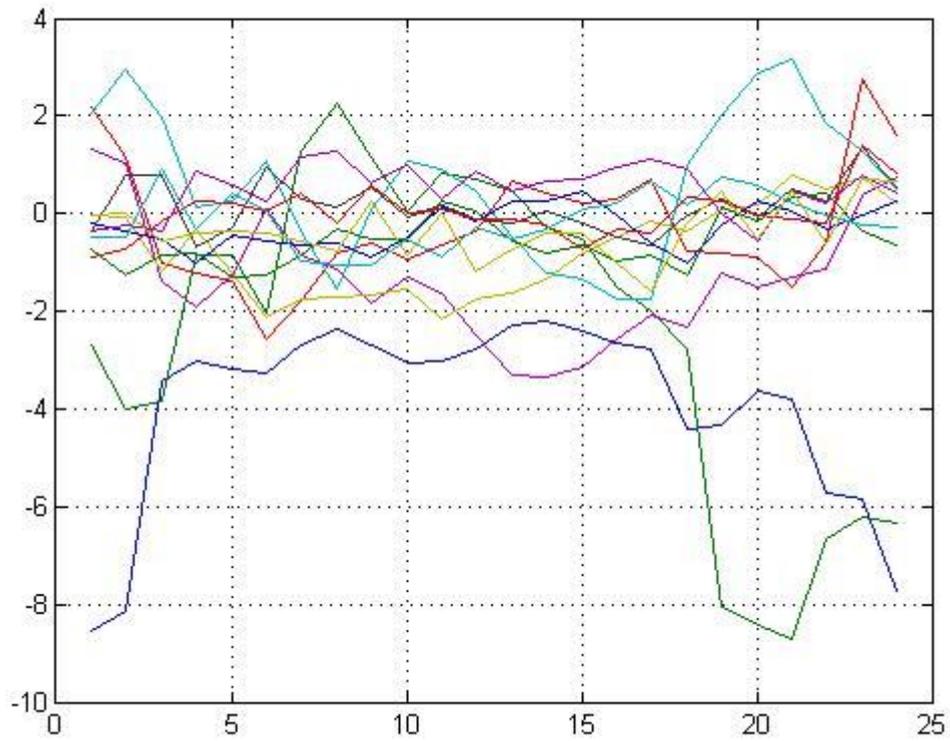


Figure III-9 représentation de MFCC du mot «droite» 'Treize coefficients'

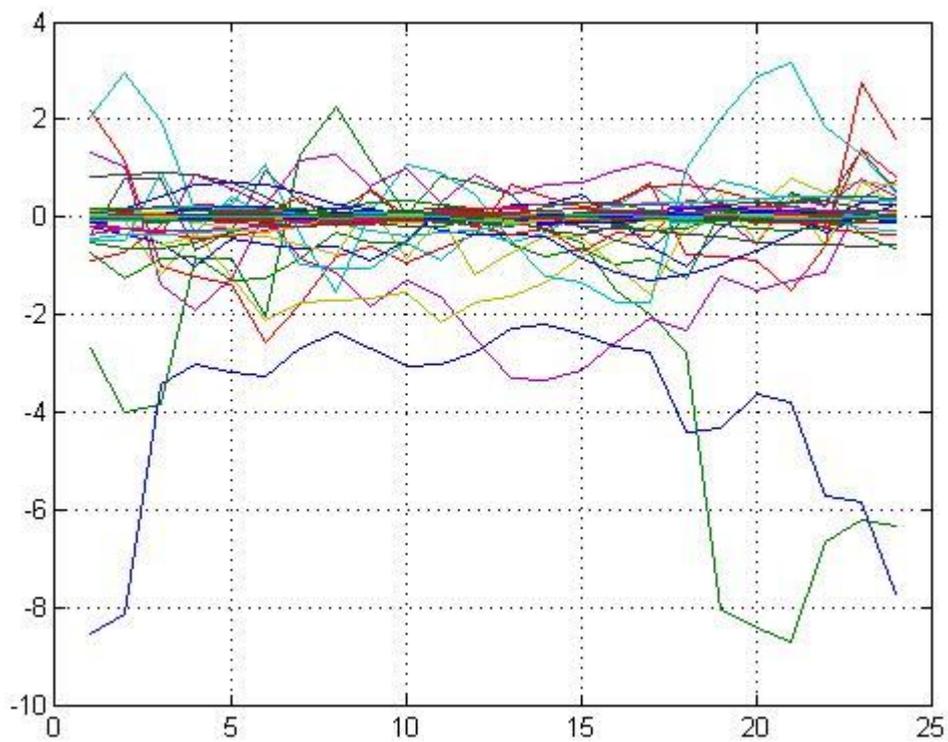


Figure III-10 représentation de MFCC du mot «droite» 'Trente neuf coefficients'

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 [17]. 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 :

$$\hat{W} = \arg_W \max P(W|O) \quad (\text{III. 7})$$

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 :

$$P(W|O) = \frac{P(W)P(O|W)}{P(O)} \quad (\text{III. 8})$$

Puisque $P(O)$ ne dépend pas de W , l'équation (III-7) sera équivalente à :

$$\hat{W} = \arg_W \max P(W)P(O|W) \quad (\text{III. 9})$$

Ainsi, l'étape de reconnaissance consiste à déterminer la suite de mots W qui maximise le produit des deux termes $P(W)$ et $P(O|W)$. 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-11 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.

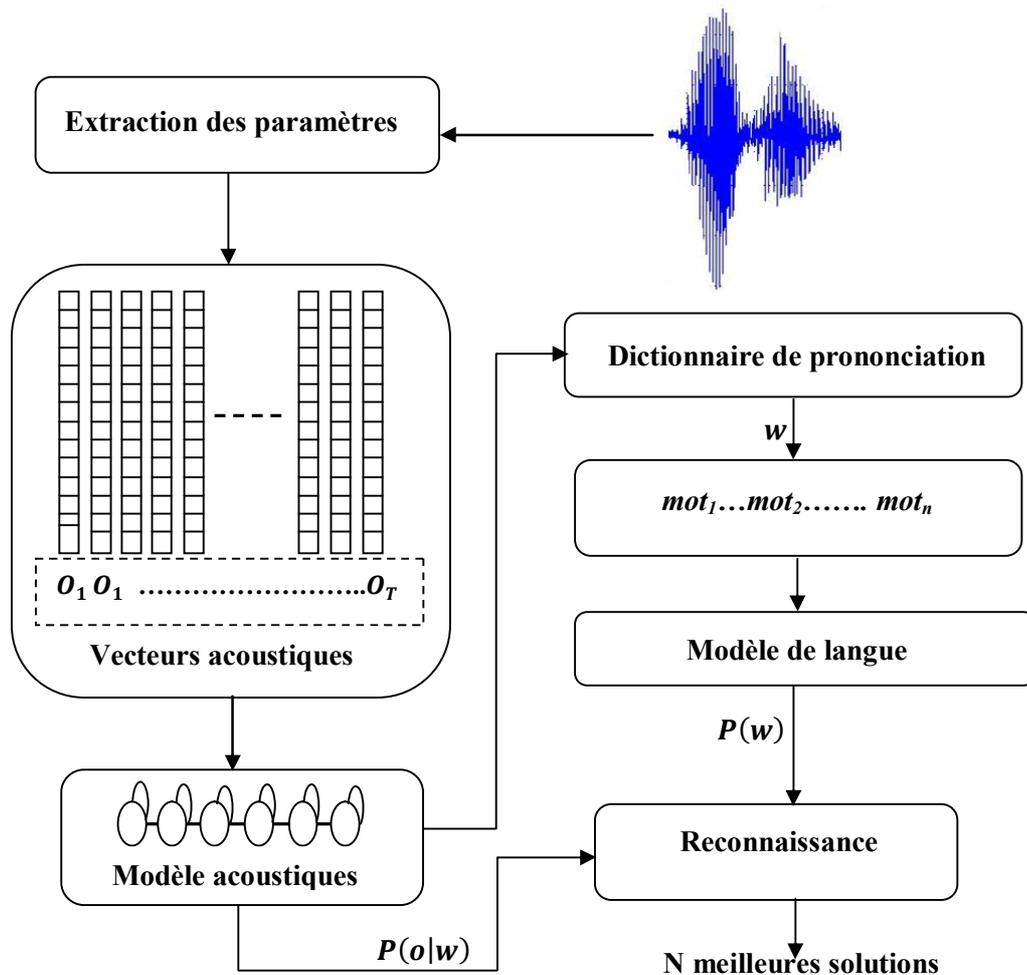


Figure III-11 Le principe de fonctionnement de l'approche statistique pour la reconnaissance automatique de la parole.

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 la parole. Les unités les plus utilisées sont les phonèmes. Cependant, il existe aussi d'autres types d'unités comme les allophones, les syllabes, les triphones, les mots, etc. Dans ce chapitre, nous nous intéressons seulement aux mots, plus d'informations sur les autres unités phonétiques peuvent être trouvées dans [14].

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

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 [21] :

$$\lambda = (N, A, B, \pi) \tag{III.10}$$

Où :

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

$$P(q_t = j | q_{t-1} = i, q_{t-2} = k, \dots) = P(q_t = j | q_{t-1} = i) \quad (\text{III. 11})$$

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

$$P(q_t = j | q_{t-1} = i, q_{t-2} = k, \dots) = P(q_t = j | q_{t-1} = i, q_{t-2} = k) \quad (\text{III. 12})$$

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).

- $B = \{b_j(o_t)\} = \{P(o_t | q_j)\}$, où $j \in [1, N]$, est l'ensemble des probabilités d'émission de l'observation o_t sachant qu'on est dans l'état q_j . 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, $\pi_j = P(q_0 = j)$, et $j \in [1, N]$. q_0 représente l'état initial du modèle HMM. Il ne peut émettre de vecteurs acoustiques.

III-3-2 Les problèmes fondamentaux des HMM

Soient λ un modèle de Markov caché et O une séquence d'observations acoustiques. La reconnaissance de cette séquence s'effectue en trouvant le modèle λ qui maximise la probabilité $P(\lambda|O)$ (probabilité qu'un modèle λ génère une séquence d'observations O). Cette probabilité est aussi appelée probabilité a posteriori. Malheureusement, il n'est pas possible d'accéder directement à cette probabilité, mais on peut calculer la probabilité qu'un modèle donné générera une certaine séquence d'observations $P(O|\lambda)$. En utilisant la loi de Bayes, il est possible de lier ces deux probabilités par :

$$P(\lambda|O) = \frac{P(\lambda)P(O|\lambda)}{P(O)} \quad (\text{III. 13})$$

- $P(O|\lambda)$. est la vraisemblance de la séquence d'observations O étant donné le modèle λ .
- $P(\lambda)$. est la probabilité a priori du modèle.
- $P(O)$. est la probabilité a priori de la séquence d'observations.

Pour une séquence d'observations connue $O = o_1, o_2, \dots, o_T$, la probabilité $P(O)$ peut être considérée constante, puisqu'elle est indépendante du modèle λ si les paramètres de ce dernier sont fixés. Ainsi, maximiser $P(\lambda|O)$ revient à maximiser $P(O|\lambda)P(\lambda)$.

Pour cela, il faut résoudre trois problèmes fondamentaux :

- **Évaluation :**
Étant donné une séquence d'observations : $O = o_1, o_2, \dots, o_T$ et le modèle $\lambda = (N, A, B, \pi)$, comment calculer efficacement $P(O|\lambda)$ la probabilité d'observer la séquence O sachant le modèle λ ?
- **Décodage :**
Étant donné une séquence d'observations : $O = o_1, o_2, \dots, o_T$ et le modèle $\lambda = (N, A, B, \pi)$, comment choisir la séquence d'états $Q = q_1, q_2, \dots, q_T$ qui a le plus de chance d'émettre la séquence d'observations O ?
- **Apprentissage :** Comment déterminer les paramètres du modèle $\lambda = (N, A, B, \pi)$ afin de maximiser $P(O|\lambda)$?

III-3-2-1 Problème d'évaluation

Soient le modèle $\lambda = (N, A, B, \pi)$, $O = o_1, o_2, \dots, o_T$ une séquence d'observations et $Q = q_1, q_2, \dots, q_T$ une séquence d'états. La probabilité d'observer la séquence O pour une séquence d'états Q est [22]:

$$P(O|Q, \lambda) = b_{q_1}(O_1) \cdot b_{q_2}(O_2) \dots b_{q_T}(O_T) \quad (\text{III. 14})$$

Or, la probabilité de la séquence Q peut s'écrire sous la forme suivante :

$$P(Q|\lambda) = \pi_{q_1} a_{q_1 q_2} a_{q_2 q_3} \dots a_{q_{T-1} q_T} \quad (\text{III. 15})$$

La probabilité conjointe du chemin Q et des observations O est :

$$P(O, Q|\lambda) = P(Q|\lambda).P(O|Q, \lambda) \quad (\text{III. 16})$$

La probabilité de la séquence d'observations O sachant le modèle λ est obtenue par la sommation de $P(O, Q|\lambda)$ sur toutes les séquences d'états Q possibles. Ainsi la probabilité d'émission des observations est :

$$P(O|\lambda) = \sum_Q P(O, Q|\lambda) \quad (\text{III. 17})$$

$$P(O|\lambda) = \sum_{q_1, q_2, \dots, q_T} \pi_{q_1} b_{q_1}(o_1) a_{q_1 q_2} b_{q_2}(o_2) \dots a_{q_{T-1} q_T} b_{q_T}(O_T) \quad (\text{III. 18})$$

Pour une machine à N états, ce calcul direct nécessite $(2T-1) \times N^T$ multiplications et N^T-1 additions, ce qui le rend trop complexe et impossible à implémenter. Il existe heureusement un algorithme rapide et efficace dit **avant-arrière** (Forward-Backward) qui donne une solution pour mener efficacement ce calcul [22].

L'algorithme avant-arrière

Soit, la probabilité avant : $\alpha_t(i) = P(o_1, o_2, \dots, o_t, q_t = i|\lambda)$, la probabilité d'observer la séquence o_1, o_2, \dots, o_t et d'être à l'état i à l'instant t sachant le modèle λ . Cette probabilité est calculée d'une manière récursive.

L'algorithme avant

- **Initialisation** : $\alpha_1(i) = \pi_i b_i(O_1) \quad 1 \leq i \leq N$
- **Récurrence** : $\alpha_{t+1}(j) = [\sum_{i=1}^N \alpha_t(i) a_{ij}] b_j(O_{t+1}), \quad t \in \{1, 2, \dots, T-1\}$ et $1 \leq j \leq N$
- **Terminaison** : $P(O|\lambda) = \sum_{i=1}^N \alpha_T(i)$

Cette récursion dépend du fait que la probabilité d'être à l'état j au temps $t+1$ et d'observer o_{t+1} peut être déduite en sommant les probabilités avant pour tous les états prédécesseurs de j pondérées par les probabilités de transition a_{ij} [22].

De la même manière, soit la probabilité arrière $\beta_t(j)$ définie par :

$$\beta_t(j) = P(O_{t+1}, O_{t+2}, \dots, O_T | q_t = j, \lambda) \quad (\text{III. 19})$$

C'est la probabilité d'observer la séquence $O_{t+1}, O_{t+2}, \dots, O_T$ sachant qu'on est à l'état i au temps t et qu'on a le modèle λ .

De la même façon cette probabilité est calculée d'une manière récursive

L'algorithme arrière

- **Initialisation** : $\beta_T(i) = 1 \quad 1 \leq i \leq N$
- **Récurrance** : $\beta_t(j) = \sum_{i=1}^N a_{ij} b_j(O_{t+1}) \beta_{t+1}(i), \quad t \in \{T-1, T-2, \dots, 1\}$
et $1 \leq j \leq N$
- **Terminaison** : $P(O|\lambda) = \sum_{i=1}^N \alpha_T(i) = \sum_{i=1}^N \pi_i b_i(O_1) \beta_1(i) = \sum_{i=1}^N \alpha_t(i) \beta_t(i)$

III-3-2-2 Problème de décodage

Étant donné une séquence d'observations O , et un modèle $\lambda = (N, A, B, \pi)$, le problème de décodage revient à la recherche d'une séquence d'états "optimale". Cela peut-être fait de différentes façons. La difficulté réside dans la définition de la séquence d'états optimale. Donc, il faut choisir un critère parmi plusieurs critères d'optimalité. Par exemple, un critère envisageable pour répartir les vecteurs de la séquence d'observations sur les états de la chaîne, consiste à optimiser séparément chaque état q_t . Pour implémenter cette solution, une variable est définie par :

$$\gamma_t(i) = P(q_t = i | O, \lambda) = \frac{\alpha_t(i) \beta_t(i)}{\sum_{i=1}^N \alpha_t(i) \beta_t(i)} = \frac{\alpha_t(i) \beta_t(i)}{P(O|\lambda)} \quad (\text{III. 20})$$

$\gamma_t(i)$ est la probabilité d'être à l'état i au temps t , étant donnée l'observation O et le modèle λ .

L'état optimal à un instant t sera donc :

$$q_t = \underset{i}{\operatorname{argmax}} [\gamma_t(i)] \quad (\text{III. 21})$$

Ce critère d'optimalité maximise le nombre d'états. Cependant, cette méthode peut aboutir à des erreurs. Par exemple, lorsque le modèle de Markov possède des probabilités de transitions égales à zéro, la séquence optimale obtenue pourrait en fait ne pas être une séquence d'états possibles puisque le critère considéré ne tient pas compte des probabilités des changements d'états.

Une solution possible est de modifier le critère d'optimalité. On pourrait par exemple chercher la séquence d'états qui maximise les paires d'états (q_t, q_{t+1}) ou même les triplets d'états (q_t, q_{t+1}, q_{t+2}) [22].

Si ces critères sont tout à fait adaptés à certaines applications, le critère le plus utilisé est celui qui cherche la meilleure séquence d'états globale (le meilleur chemin), c'est-à-dire qui maximise $P(Q|O, \lambda)$. ce qui revient à maximiser $P(Q, O|\lambda)$. Une technique formelle existe pour calculer ce chemin optimal, il s'agit de **l'algorithme de Viterbi**.

L'algorithme de Viterbi

Pour trouver la meilleure séquence d'états $Q = q_1, q_2, \dots, q_T$, connaissant une séquence d'observations $O = o_1, o_2, \dots, o_T$, on a besoin de définir la quantité $\delta_t(i)$.

$$\delta_t(i) = \max_{q_1, q_2, \dots, q_{t-1}} P(q_1, q_2, \dots, q_t = i, O_1, O_2, \dots, O_t | \lambda) \quad (\text{III. 22})$$

$\delta_t(i)$ est le meilleur résultat (probabilité la plus grande) selon un simple chemin ; ce chemin se compose des t premières observations et se termine dans l'état i . On peut déterminer les $\delta_t(i)$ de façon itérative. On a en effet :

$$\delta_{t+1}(j) = \max_{1 \leq i \leq N} [\delta_t(i) a_{ij}] b_j(o_{t+1}) \quad (\text{III. 23})$$

L'algorithme

- **Initialisation :** $\delta_1(i) = \pi_i b_i(O_1) \quad 1 \leq i \leq N$
 $\psi_1(i) = 0$
- **Réurrence :** $\delta_t(j) = \max_{1 \leq i \leq N} [\delta_{t-1}(i) a_{ij}] b_j(o_t) \quad 2 \leq t \leq T \quad 1 \leq j \leq N$
 $\psi_t(j) = \arg_{1 \leq i \leq N} \max [\delta_{t-1}(i) a_{ij}]$
- **Terminaison :** $P^* = \max_{1 \leq i \leq N} [\delta_T(i)]$
 $\psi_T^*(j) = \arg_{1 \leq i \leq N} \max [\delta_T(i)]$
- **Recherche :** $q_t^* = \psi_{t+1}^*(q_{t+1}^*) \quad t = T - 1, T - 2, \dots, 1$

Pour déterminer la séquence d'états, il est donc nécessaire de garder la trace de l'indice i qui a maximisé la formule précédente, et ceci pour tout t et tout j . On réalise ceci par l'intermédiaire d'un tableau $\psi(j)$.

III-3-2-3 Problème d'apprentissage

Le troisième problème consiste à trouver une méthode pour ajuster les paramètres du modèle $\lambda = (N, A, B, \pi)$ afin de maximiser la probabilité d'une séquence d'observations donnée, sachant le modèle λ . Ce problème n'a pas de solution analytique connue et il n'existe pas de technique optimale pour estimer les paramètres du modèle. On peut cependant choisir $\lambda = (N, A, B, \pi)$ de telle façon que $P(O|\lambda)$ soit localement maximale en utilisant une procédure itérative telle que la méthode de Baum-Welch ou la technique du gradient [16].

Dans ce qui suit nous présentons une procédure itérative basée sur la technique de Baum-Welch.

Pour décrire comment ré-estimer les paramètres du HMM, on définit la probabilité $\xi_t(i, j)$ qui représente la probabilité d'être à l'état i au temps t et de faire une transition à l'état j au temps

$t + 1$ étant donnée la séquence d'observations O et le modèle λ .

$$\xi_t(i, j) = P(q_t = i, q_{t+1} = j | O, \lambda) \quad (\text{III. 24})$$

D'après les définitions des probabilités avant et arrière, $\xi(i, j)$ peut s'écrire sous la forme suivante :

$$\xi_t(i, j) = \frac{\alpha_t(i) a_{ij} b_j(O_{t+1}) \beta_{t+1}(j)}{P(O|\lambda)} \quad (\text{III. 25})$$

Nous avons défini, précédemment $\gamma_t(i)$ comme étant la probabilité d'être à l'état i au temps t , étant donnée l'observation O et le modèle λ . Ainsi nous pouvons relier $\gamma_t(i)$ à $\xi_t(i, j)$ par une sommation sur j , d'ou la relation suivante :

$$\gamma_t(i) = \sum_{j=1}^N \xi_t(i, j) \quad (\text{III. 26})$$

L'algorithme de Baum-Welch estime les nouveaux paramètres de la chaîne de Markov cachée comme suit :

$$\bar{\pi}_i = \gamma_1(i), \quad 1 \leq i \leq N \quad (\text{III. 27})$$

$$\bar{a}_{ij} = \frac{\sum_{t=1}^{T-1} \xi_t(i, j)}{\sum_{t=1}^{T-1} \gamma_t(i)} \quad 1 \leq i \leq N, 1 \leq j \leq N \quad (\text{III. 28})$$

$$\bar{b}_j(k) = \frac{\sum_{t=1, ot=k}^T \gamma_t(j)}{\sum_{t=1}^{T-1} \gamma_t(i)} \quad 1 \leq j \leq N \quad (\text{III. 29})$$

La ré-estimation de π_i est la probabilité d'être à l'état i au temps $t = 1$. La formule de ré-estimation de a_{ij} est le rapport du nombre de transitions de l'état i vers l'état j sur le nombre de transitions partant de l'état i . La ré-estimation de $b_i(k)$ est le rapport du nombre de fois d'être à l'état i en observant k sur le nombre de fois étant dans l'état i .

Nous avons défini le modèle courant $\lambda = (N, A, B, \pi)$, et nous l'avons utilisé pour recalculer ces variables, ainsi nous avons le modèle ré-estimé $\bar{\lambda} = (N, \bar{A}, \bar{B}, \bar{\pi})$. Nous pouvons ainsi affirmer l'une ou l'autre de ces propositions :

- le modèle initial λ définit un point critique de la fonction de vraisemblance, dans ce cas $\bar{\lambda} = \lambda$
- le modèle $\bar{\lambda}$ est meilleur que le modèle λ dans le sens où $P(O|\bar{\lambda}) > P(O|\lambda)$, donc la séquence d'observations O est plus probable avec le nouveau modèle $\bar{\lambda}$.

En se basant sur cette procédure, si nous utilisons itérativement le modèle $\bar{\lambda}$ à la place de λ et si nous répétons l'étape de la ré-estimation des paramètres. Nous pouvons alors améliorer la probabilité que O soit observé sachant le modèle jusqu'à atteindre un certain point limite.

Le résultat final de la procédure de ré-estimation est appelé : l'estimation au maximum de vraisemblance du HMM (Maximum Likelihood Estimation : MLE). Il existe d'autres critères d'apprentissage, comme les critères MAP (Maximum A Posteriori) [23] ou MMI (Maximum Mutual Information)[24][25], mais leur mise en œuvre est généralement plus difficile [22].

III-3-3 Densités d'observation continues dans les modèles de Markov cachés

Jusqu'à présent nous n'avons considéré que le cas où les observations prennent des valeurs dans un alphabet fini discret et nous pouvions donc utiliser une loi de probabilité discrète dans chaque état du modèle. Une telle approche n'est pas compatible avec des observations qui sont des signaux continus. Bien sûr, quantifier le signal pourrait permettre de résoudre le problème, mais cela ne pourrait entraîner que des dégradations. Il est donc préférable d'utiliser des modèles de Markov cachés avec des densités d'observation continues. La représentation la plus générale de la fonction des densités de probabilités pour laquelle une procédure de ré-estimation a été effectuée est de la forme :

$$b_j(O_t) = \sum_{m=1}^M c_{jm} \mathcal{N}(O_t, \mu_{jm}, \Sigma_{jm}) \quad (\text{III. 30})$$

$$\text{Avec } \sum_{m=1}^M c_{jm} = 1, 1 \leq j \leq N$$

$$\text{et } \mathcal{N}(O_t, \mu_{jm}, \Sigma_{jm}) = \frac{1}{(2\pi)^{D/2} |\Sigma_{jm}|^{1/2}} e^{-\frac{1}{2}(O_t - \mu_{jm})^T \Sigma_{jm}^{-1} (O_t - \mu_{jm})}$$

Où c_{jm} est le m^{ieme} coefficient du mélange dans l'état j et $\mathcal{N}()$ est une densité gaussienne de moyenne μ_{jm} et de matrice de covariance Σ_{jm} .

Dans la pratique, bien sûr, il ya plusieurs états et plusieurs mélanges, et il n'y a pas d'affectation directe des vecteurs d'observation aux états individuels, car la séquence d'états sous-jacente est inconnue. Puisque la probabilité totale de chaque séquence d'observation est basée sur la somme de toutes les séquences d'états possibles, chacun vecteur d'observation O_t contribue au calcul de la probabilité de chaque état j . Cette probabilité, pour l'état j et mélange k au temps t (il ya M Mélanges), est donnée par:

$$\gamma_t(j, k) = \left[\frac{\alpha_t(j) \beta_t(j)}{\sum_{j=1}^N \alpha_t(j) \beta_t(j)} \right] \left[\frac{c_{jk} \mathcal{N}(O_t, \mu_{jk}, \Sigma_{jk})}{\sum_{k=1}^M c_{jk} \mathcal{N}(O_t, \mu_{jk}, \Sigma_{jk})} \right] \quad (\text{III. 31})$$

$$\mu_{jk} = \frac{\sum_{t=1}^T \gamma_t(j, k) O_t}{\sum_{t=1}^T \gamma_t(j, k)} \quad (\text{III. 32})$$

$$\Sigma_{jk} = \frac{\sum_{t=1}^T \gamma_t(j, k) (O_t - \mu_{jk})(O_t - \mu_{jk})'}{\sum_{t=1}^T \gamma_t(j, k)} \quad (\text{III. 33})$$

$$c_{jk} = \frac{\sum_{t=1}^T \gamma_t(j, k)}{\sum_{t=1}^T \sum_{k=1}^M \gamma_t(j, k)} \quad (\text{III. 34})$$

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-12 illustre un exemple d'un HMM à 5 états typique utilisé pour la modélisation d'un mot.

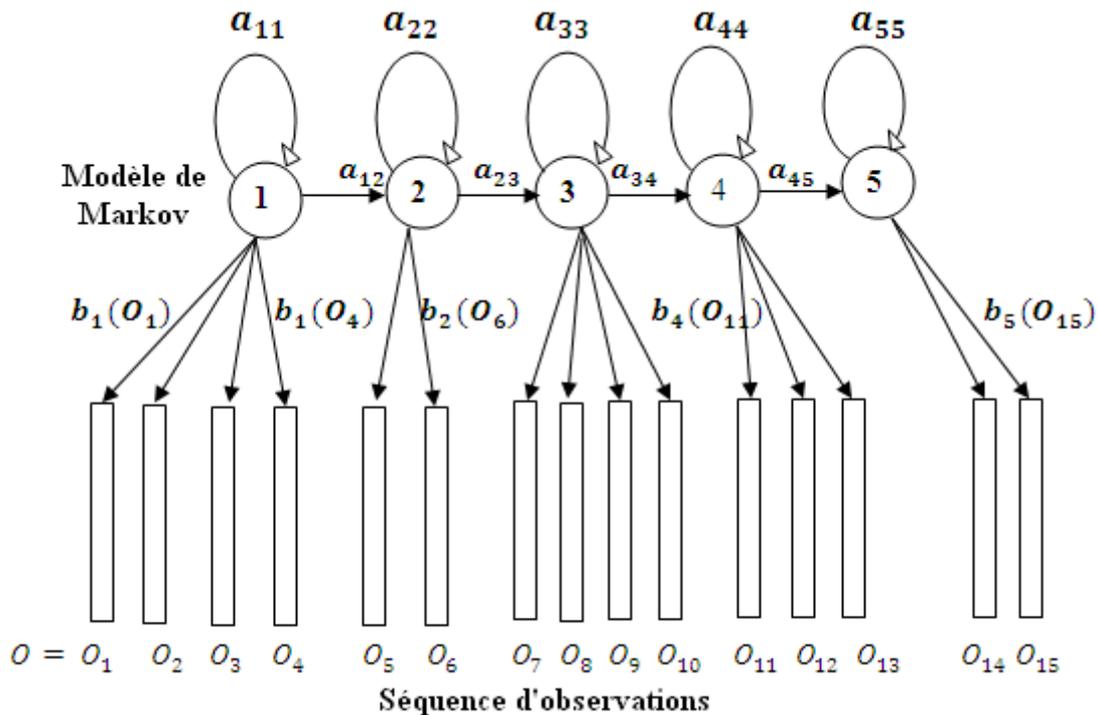


Figure III-12 Exemple d'un HMM à cinq états caractérisé par une distribution de probabilités pour chaque état associé à une observation et par des probabilités de transition entre les états.

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.

C'est la bibliothèque «Probabilistic Model Toolbox (PMT) » qui nous a servi de base pour les testes. Elle contient des fichiers de traitement statistique et probabiliste y compris le traitement du HMM (Entraînement, Initialisation, Décodage...etc), on a aussi utilisé le « VOICEBOX version v 1.17 » qui est une boîte vocale à outils MATLAB pour le divers traitement de la parole (Filtrage, Segmentation, Paramétrisation...etc).

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

- Pour chaque mot dans le vocabulaire, nous devons construire un HMM, nous devons aussi estimer les Paramètres du modèle (N, A, B, π) 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-13 doit être effectuée, à savoir la mesure de séquence d'observation $O = \{O_1, O_2, \dots, O_T\}$, 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, $P(O|\lambda)$; suivie par la sélection de la parole qui a la plus haute probabilité de modèle.

$$M^* = \underset{m}{\operatorname{argmax}} [P(O|\lambda_m)] \quad \text{où} \quad 1 > m > M \quad \text{(III-35)}$$

- l'étape de calcul de probabilité est réalisé par l'algorithme de Viterbi (l'algorithme cité au III-3-2-2) (le chemin de vraisemblance maximale est utilisée).

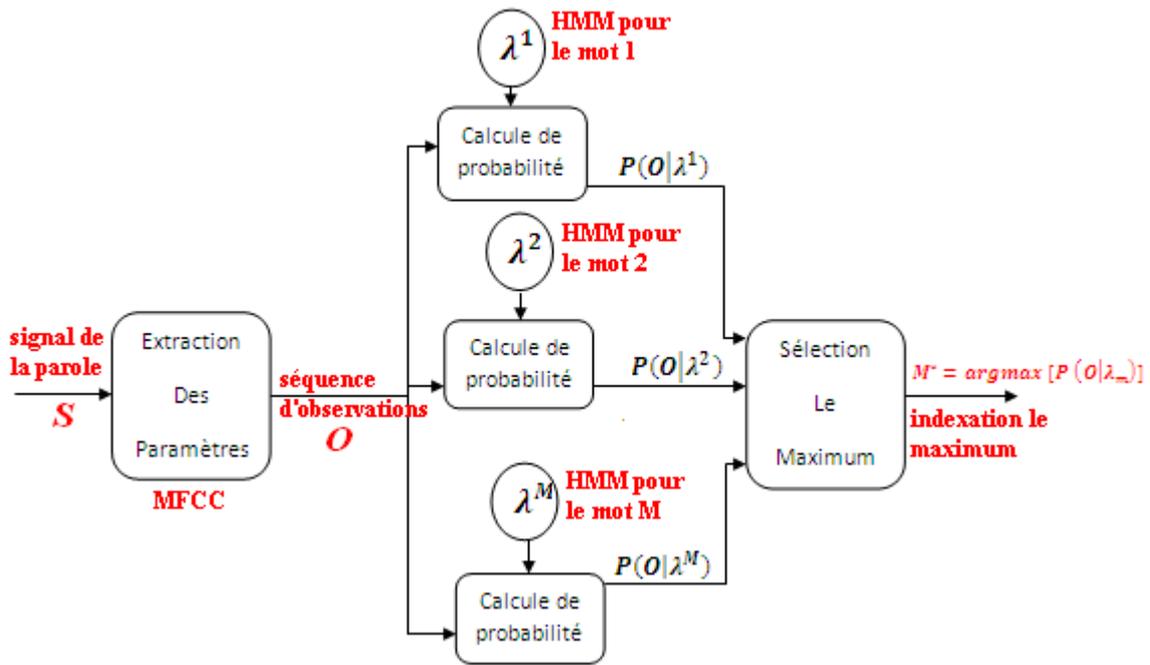


Figure III-13 architecture de reconnaissance des mots isolé

III-5-1 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 :

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

Tableau III-1 Taux de reconnaissance

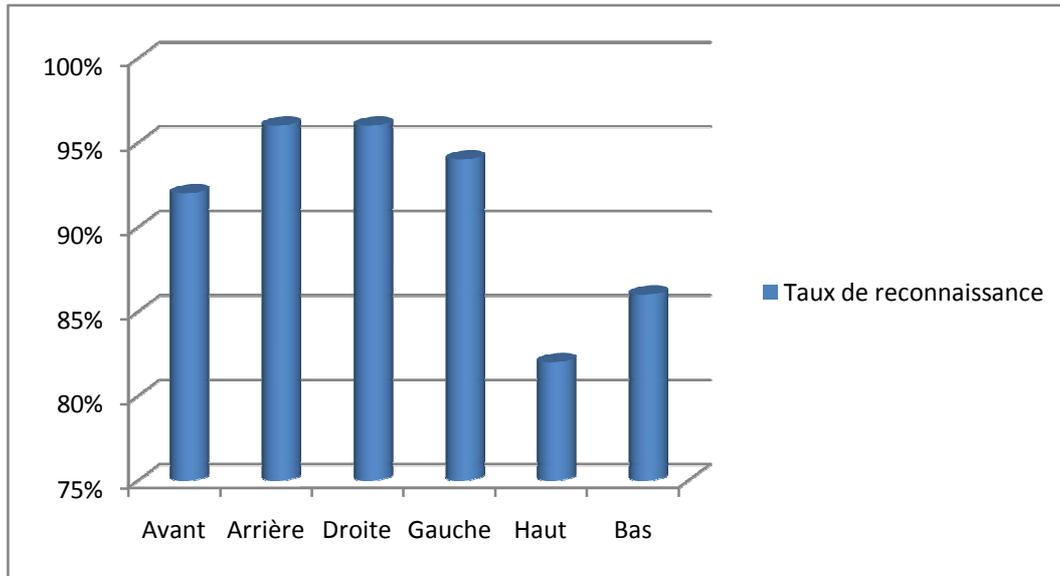


Figure III-14 Taux de reconnaissance

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 donne des résultats illogiques.

Chapitre IV

Application sur les bras manipulateurs

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 commnde 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 q_2 et q_3 . Les différentes postures sont posture à coude bas et posture à coude haut qui correspondent respectivement à q_3 positive et à q_3 négative.

IV-2 architecture générale de la commande vocale

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

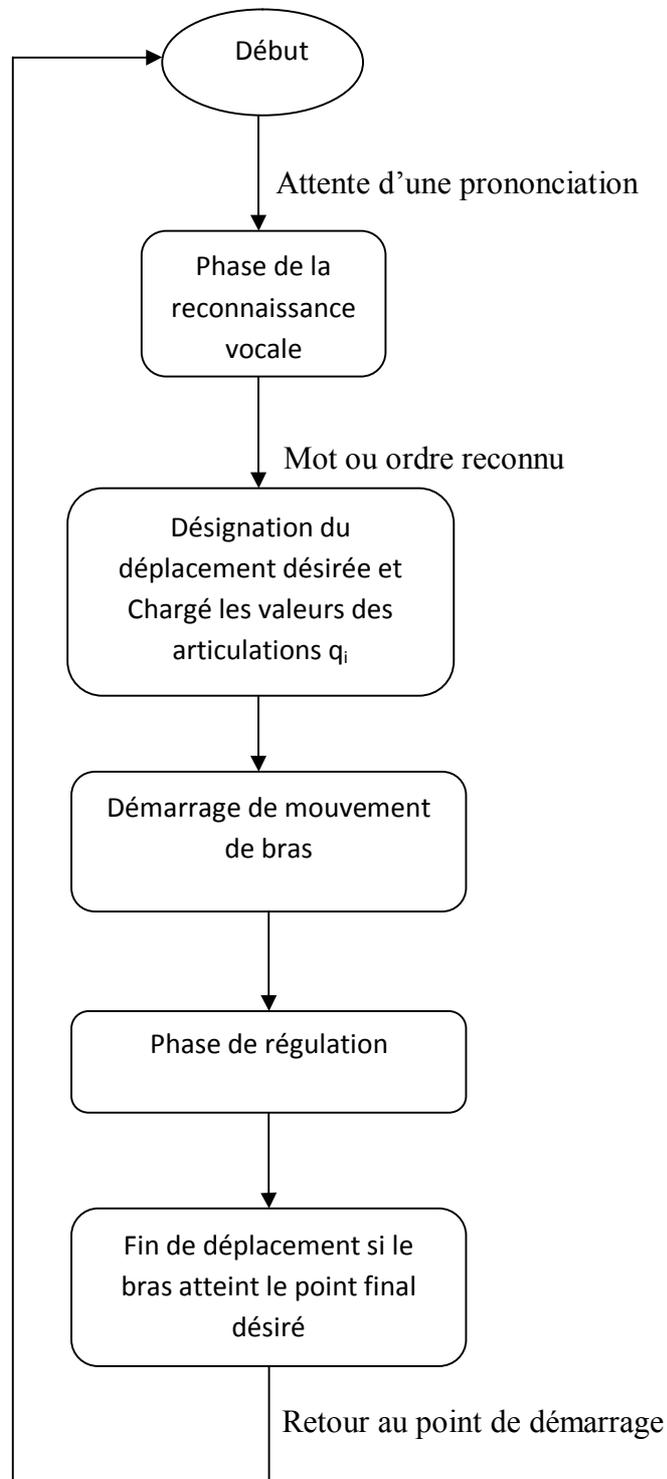


Figure IV-1 l'architecture générale de la commande vocale

IV-3 le model de robot RRR utilisé :

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

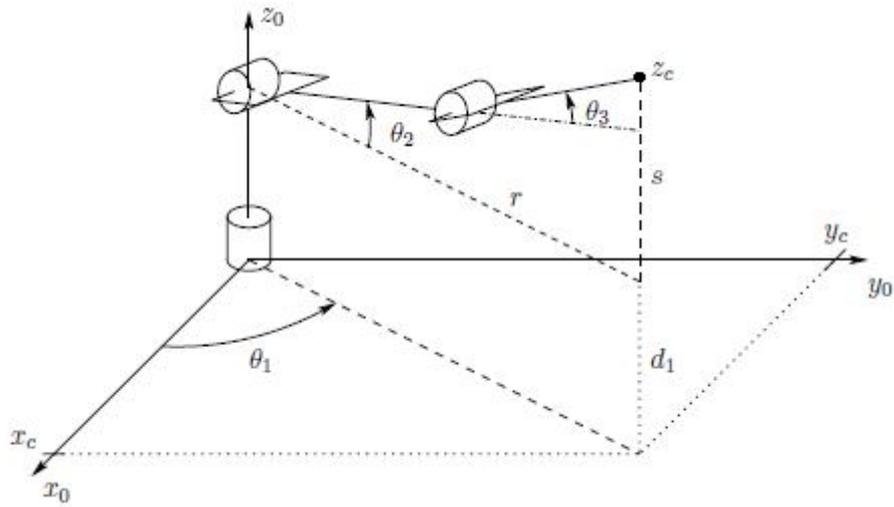


Figure IV-2 architecture de bras manipulateur RRR

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 :

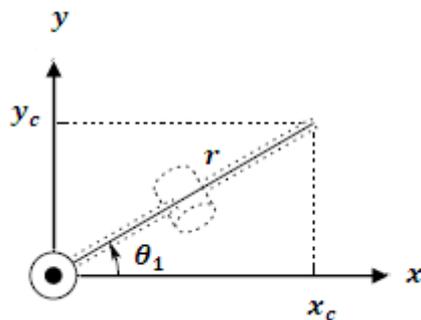


Figure IV-3 projection de bras selon xy

On voit dans cette projection que :

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

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

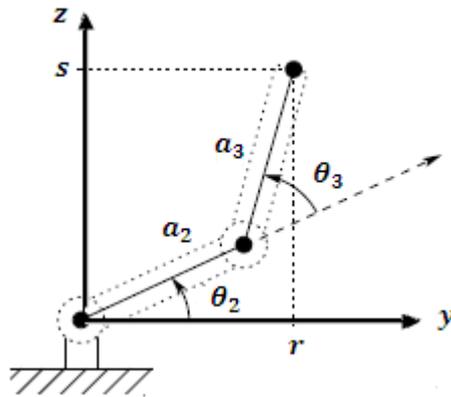


Figure IV-4 projection de bras selon yz

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

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

- 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.

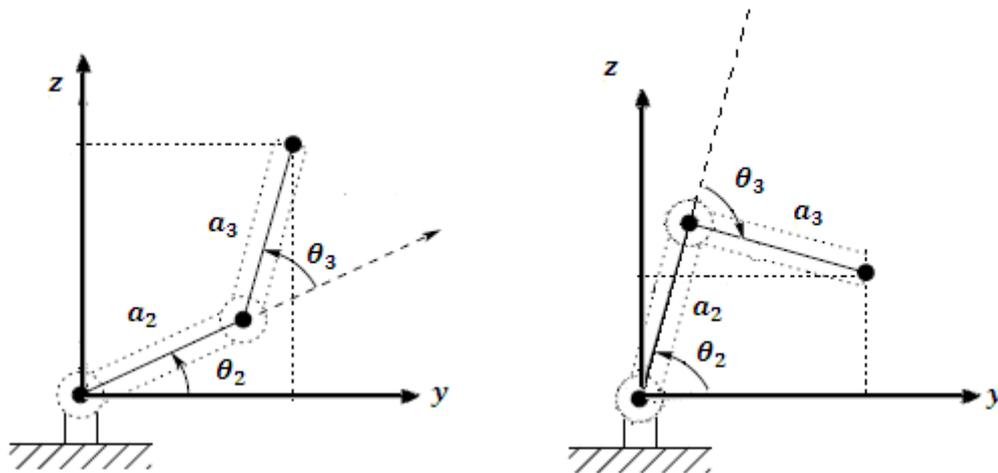


Figure IV-5 les deux postures du robots : à gauche (coude bas) et à droite (coude haut)

$$\theta_2 = \text{Atan} \left(\frac{z_c - d_1}{\sqrt{x_c^2 + y_c^2}} \right) - \text{Atan} \left(\frac{a_3 \sin \theta_3}{a_2 + a_3 \cos \theta_3} \right) \quad (\text{IV.4})$$

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

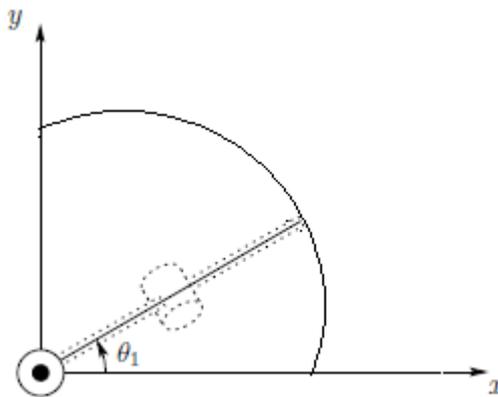


Figure IV-5 Rotation autour de l'axe Z avec l'angle θ_1

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 θ_2 et θ_3 .

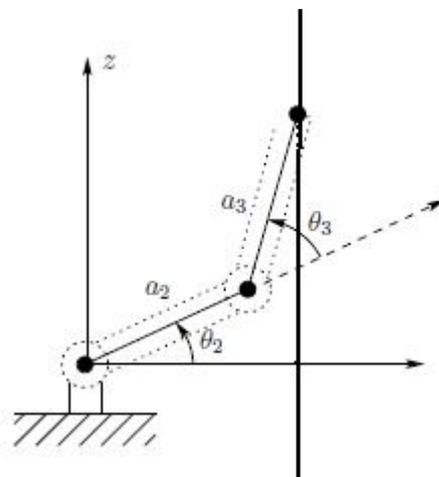


Figure IV-5 déplacement verticale

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.

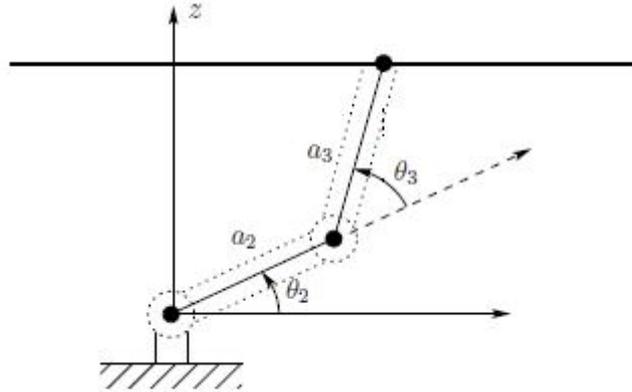


Figure IV-5 déplacement horizontal

IV-4 Généralité sur la commande en mode glissant

La phase de régulation est une phase très importante pour ajusté 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 :

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

avec :

$x(t) = (x, \dot{x}, \dots, x^{(n-1)})^T$ 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 [26].

L'erreur de poursuite est définie par :

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

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

➤ **Le choix de la surface de glissement**

Slotine [27] 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:

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

Où :

λ : 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 :

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

Une condition suffisante pour que le système (IV-5) soit stable [26] :

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

Où : $\eta > 0$.

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

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

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

Sachant que :

u_e : est la commande équivalente.

$\text{sign}()$: 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) .

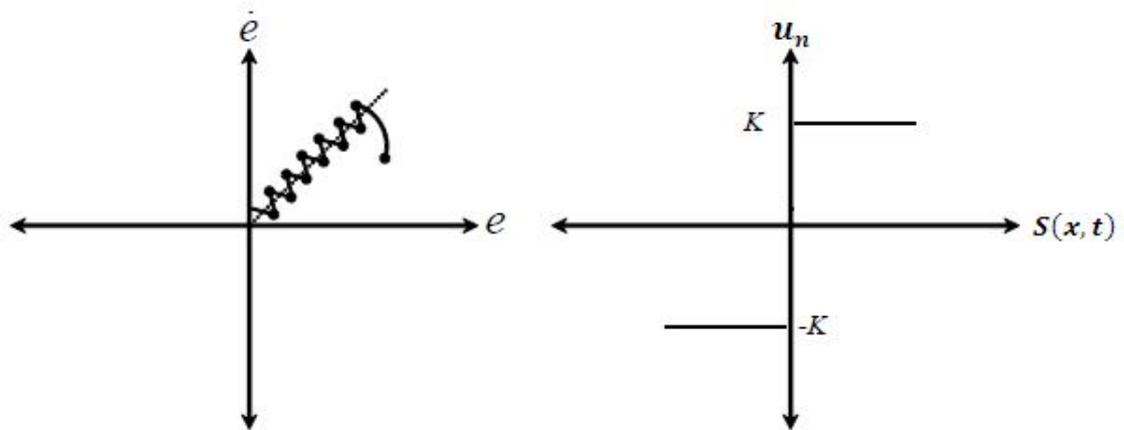


Figure IV-6 Plan de phase et la fonction de commutation de la commande par régime glissant.

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 maintien de l'état du système dans cette bande. La loi de commande devient alors [26]:

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

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

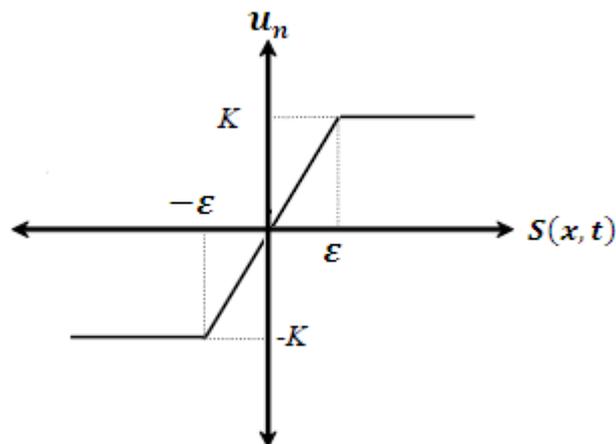


Figure IV-7 la fonction de commutation de la commande par régime glissant à bande limite

IV-5 Résultats et simulation :

IV-5-1 Première situation « *posture coude haut* »

On s'intéresse dans ce cas à la variations des articulations θ_2 et θ_3 et au déplacement du bras dans le plan YZ .

➤ *le mot Avant :*

Les états initiaux :

$$\theta_1 = \frac{\pi}{2} ; \quad \theta_2 = \frac{\pi}{4} ; \quad \theta_3 = -\frac{\pi}{2} ;$$

Le déplacement désiré

$$x_d = x_i ; \quad y_d = y_i + 0.2 \text{ m} ; \quad ; \quad z_d = z_i$$

Résultats obtenus

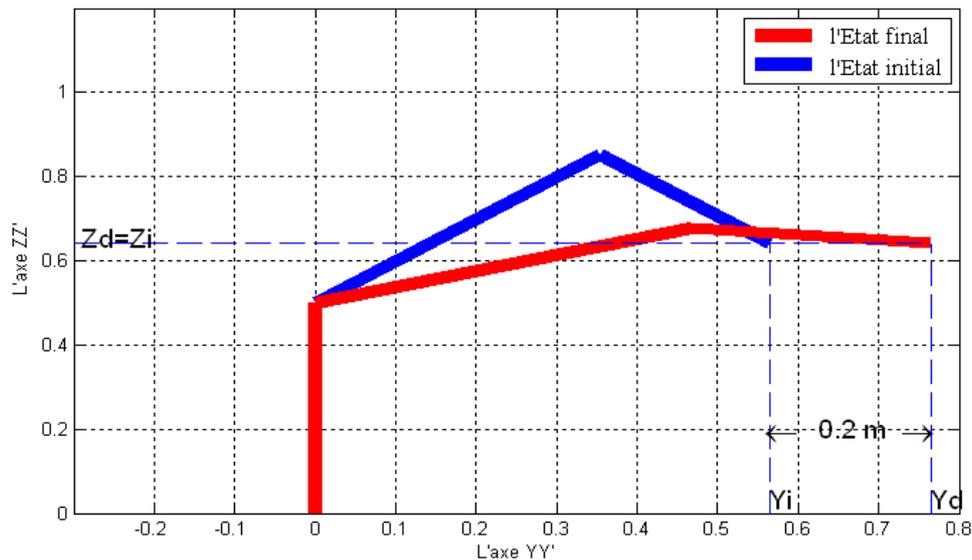


Figure IV-8 présentation du déplacement du bras en avant

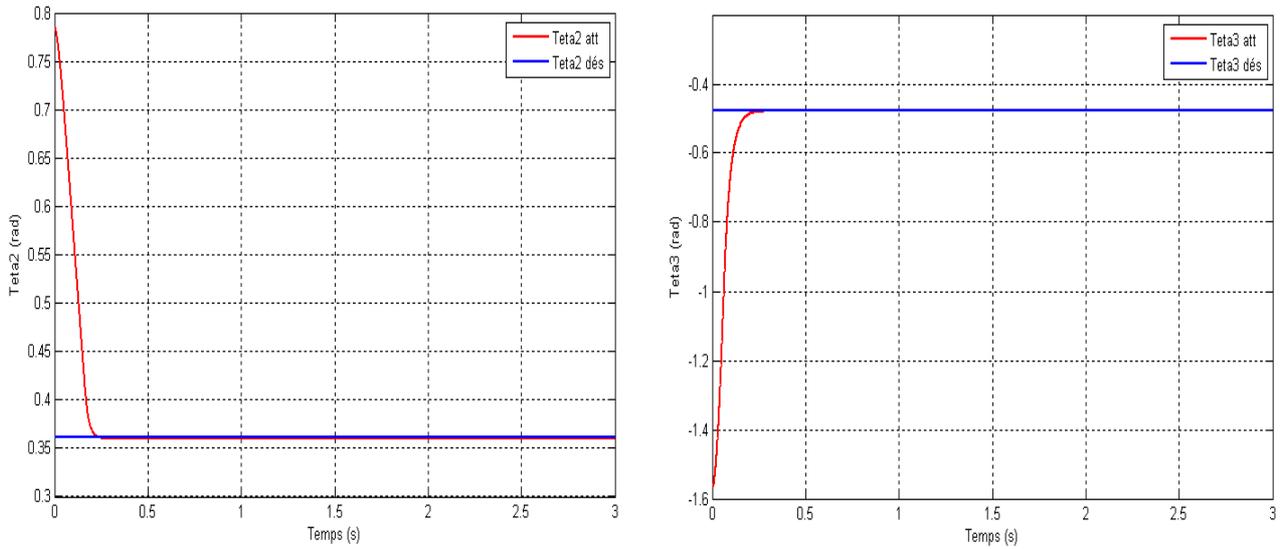


Figure IV-9 l'évolution de l'articulation θ_2 et θ_3 pour le mot « Avant » (posture coude haut)

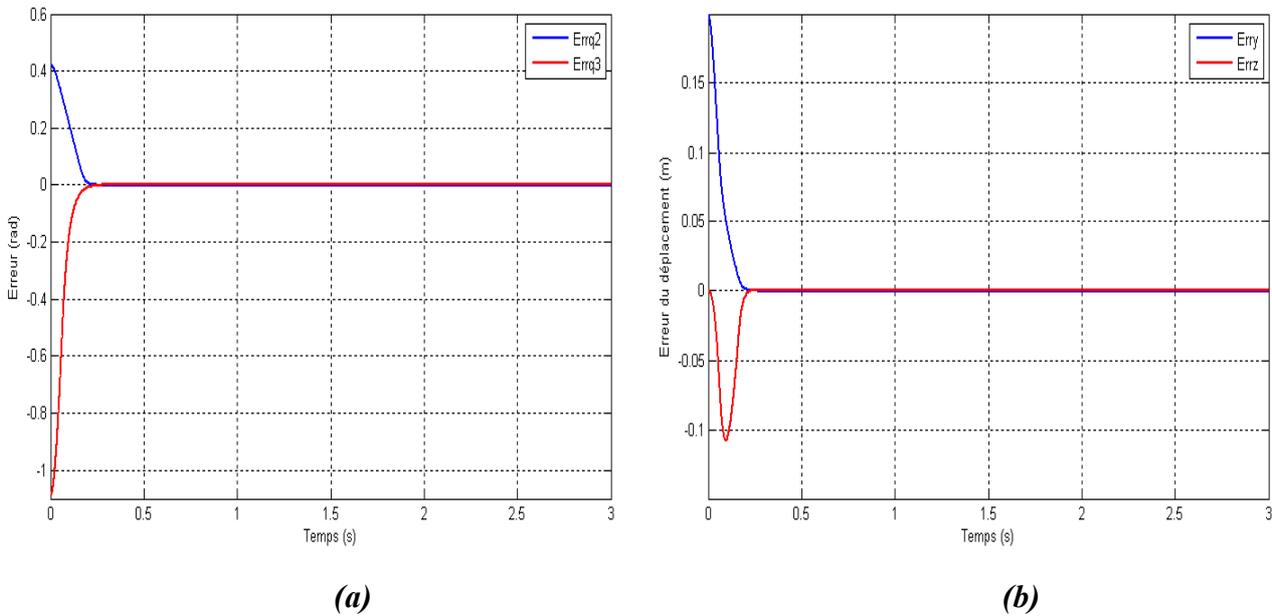


Figure IV-10 l'erreur de poursuite (a) et du déplacement (b) pour le mot « Avant » (posture coude haut)

➤ **le mot Arrière:**

Les états initiaux :

$$\theta_1 = \frac{\pi}{2}; \quad \theta_2 = \frac{\pi}{4}; \quad \theta_3 = -\frac{\pi}{2};$$

Le déplacement désiré

$$x_d = x_i; \quad y_d = y_i - 0.2 \text{ m}; \quad z_d = z_i$$

Résultats obtenus

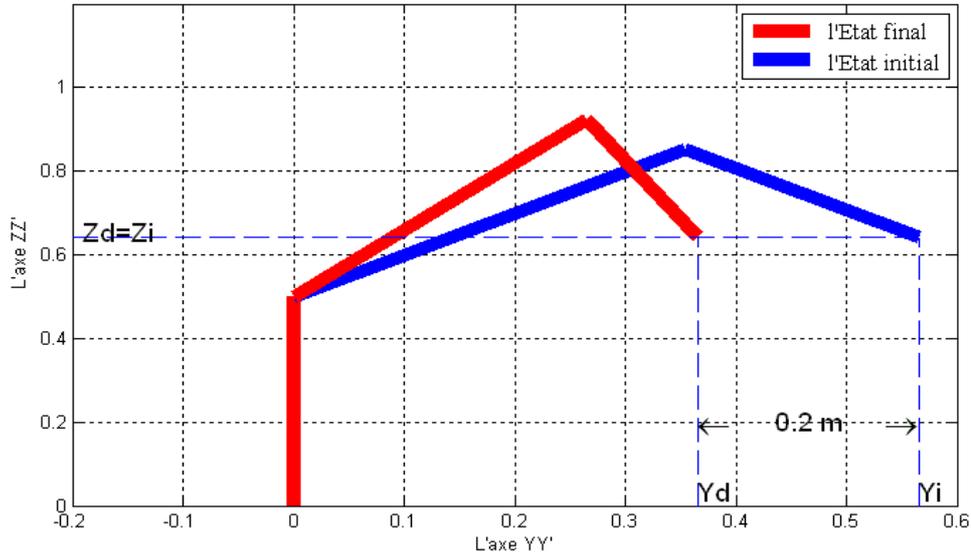


Figure IV-11 présentation du déplacement du bras en arrière

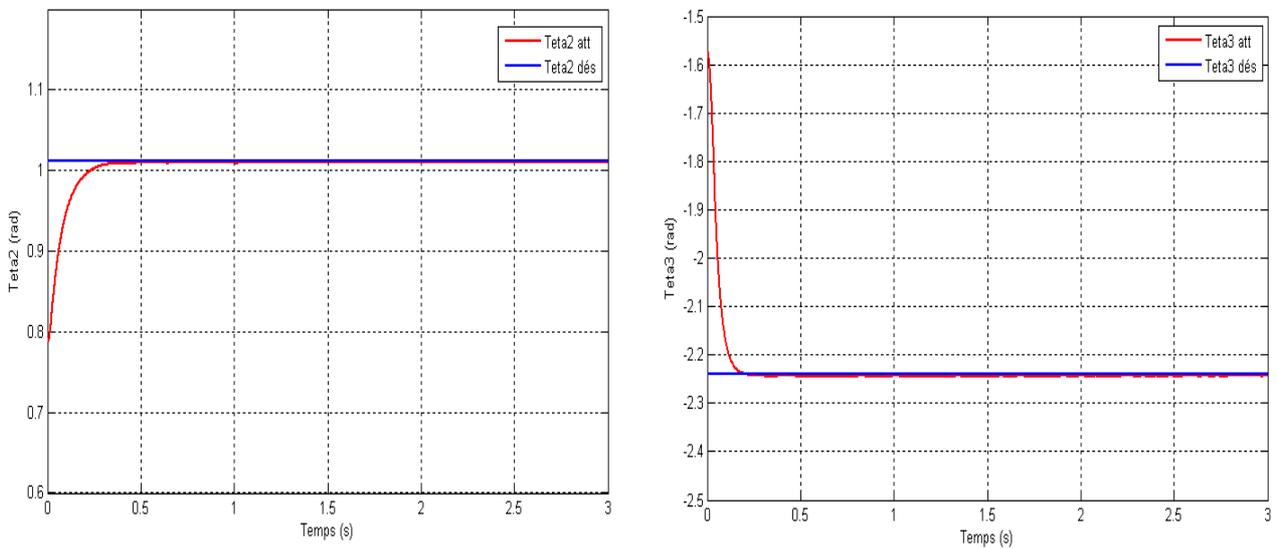


Figure IV-12 l'évolution de l'articulation θ_2 et θ_3 pour le mot « Arrière » (posture coude haut)

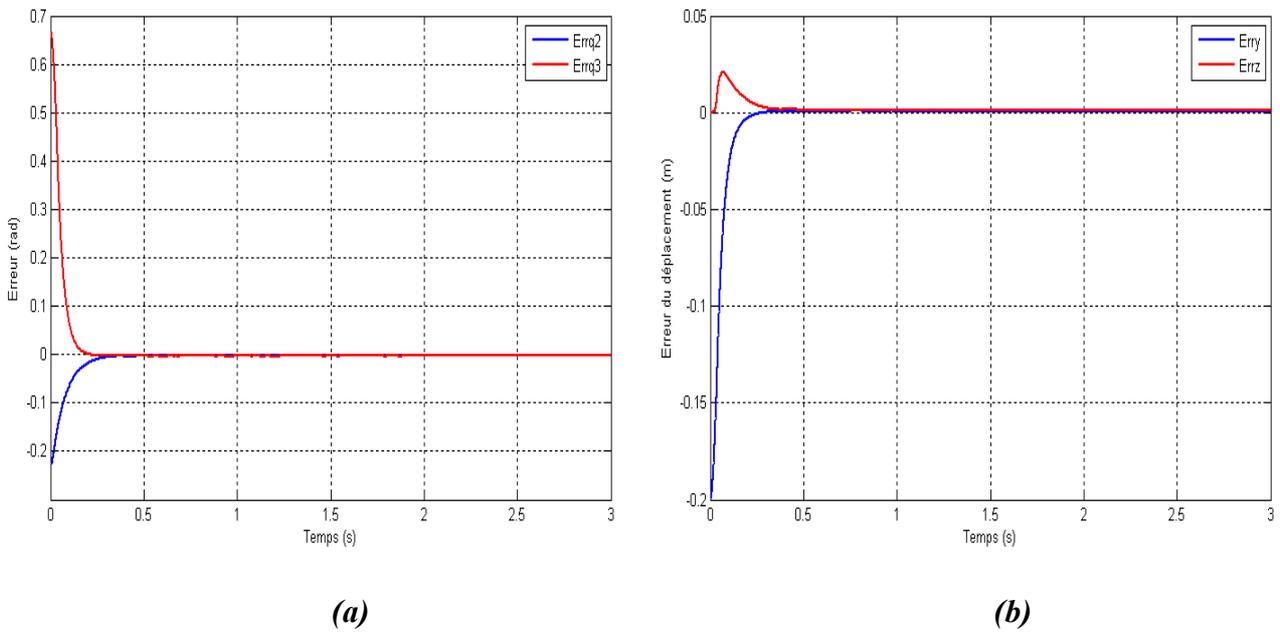


Figure IV-13 l'erreur de poursuite (a) et du déplacement (b) pour le mot « Arrière » (posture coude haut)

➤ le mot Haut:

Les états initiaux :

$$\theta_1 = \frac{\pi}{2}; \quad \theta_2 = \frac{\pi}{4}; \quad \theta_3 = -\frac{\pi}{2};$$

Le déplacement désiré

$$x_d = x_i; \quad y_d = y_i; \quad z_d = z_i + 0.2m$$

Résultats obtenus

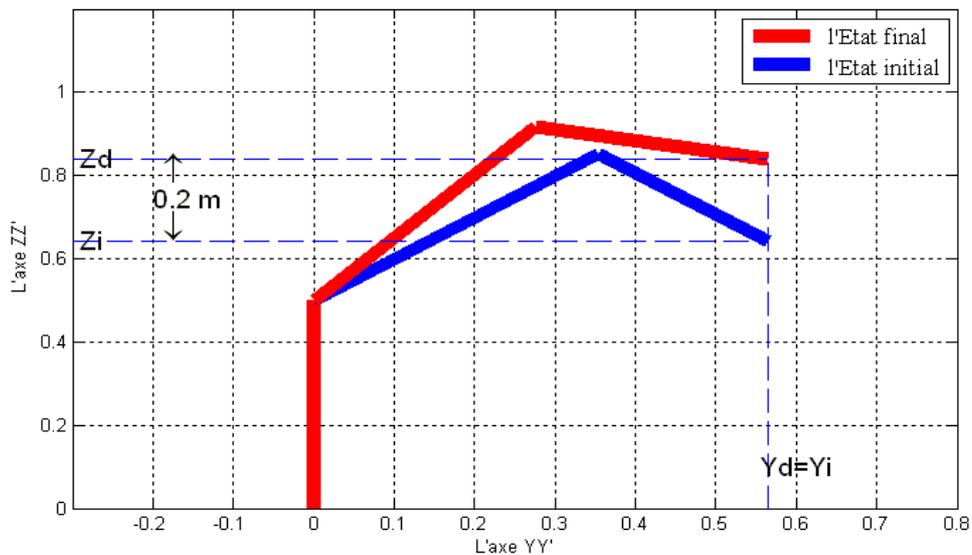


Figure IV-14 présentation du déplacement du bras en haut

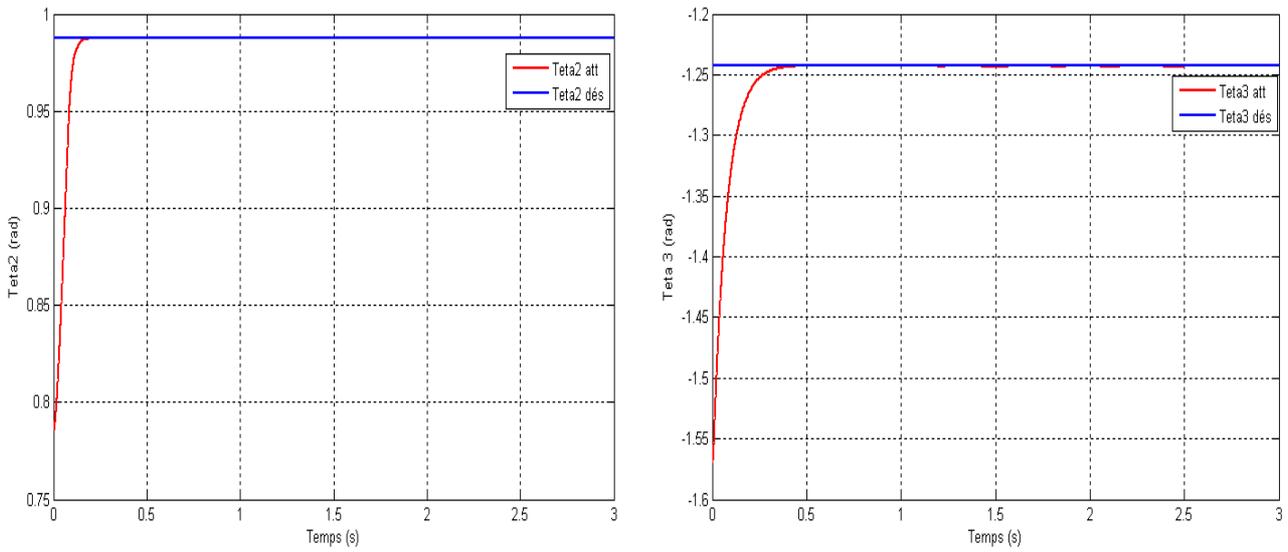


Figure IV-15 l'évolution de l'articulation θ_2 et θ_3 pour le mot « Haut » (posture coude haut)

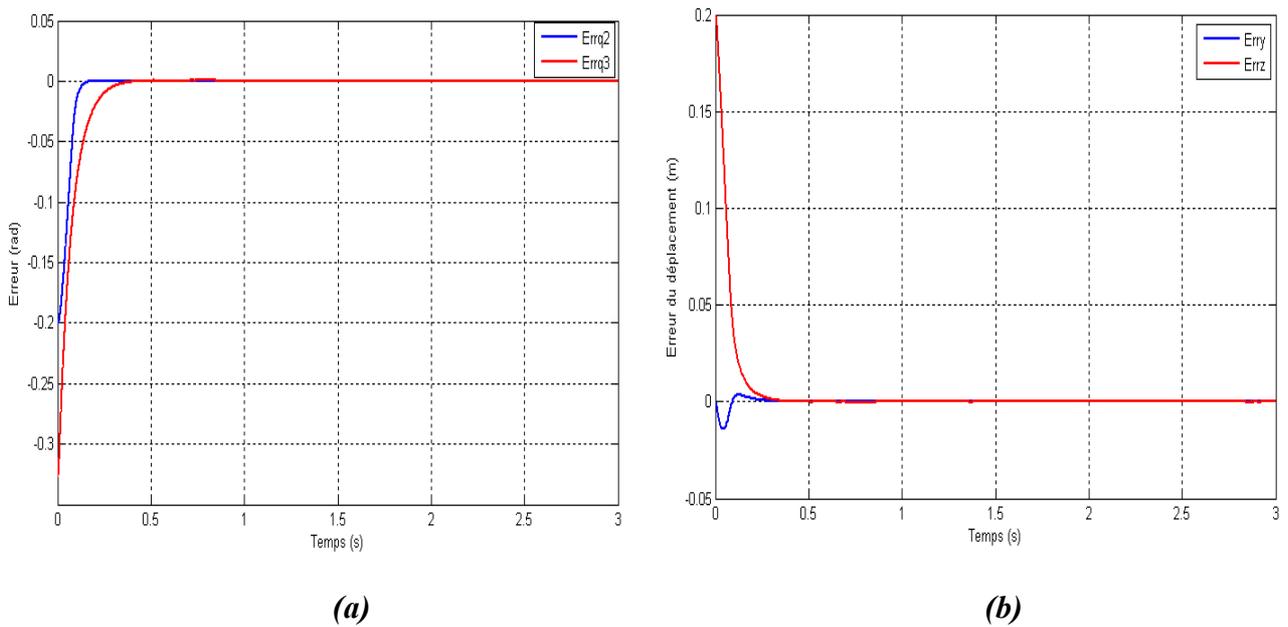


Figure IV-16 l'erreur de poursuite (a) et du déplacement (b) pour le mot « Haut » (posture coude haut)

➤ le mot Bas:

Les états initiaux :

$$\theta_1 = \frac{\pi}{2}; \quad \theta_2 = \frac{\pi}{4}; \quad \theta_3 = -\frac{\pi}{2};$$

Le déplacement désiré

$$x_d = x_i; \quad y_d = y_i; \quad z_d = z_i - 0.2m$$

Résultats obtenus

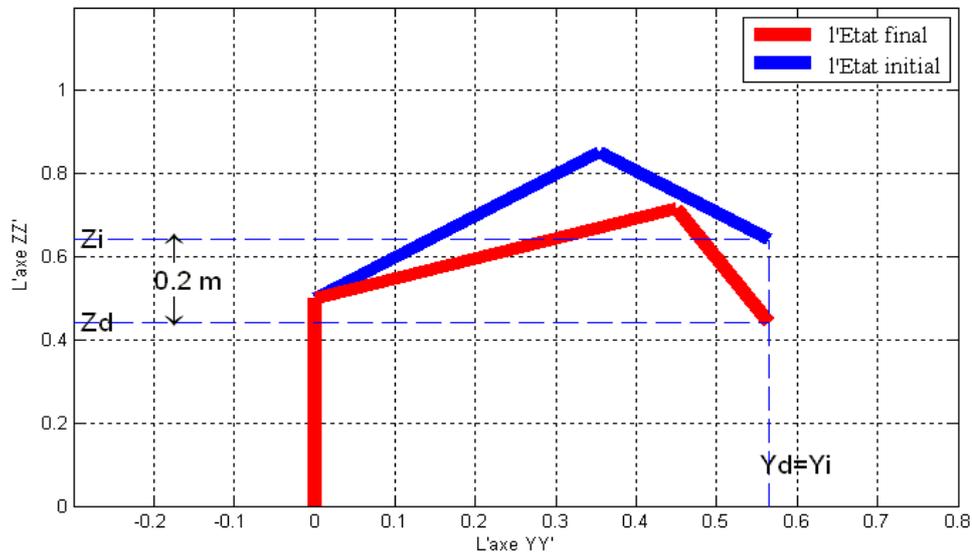


Figure IV-17 présentation du déplacement du bras en bas

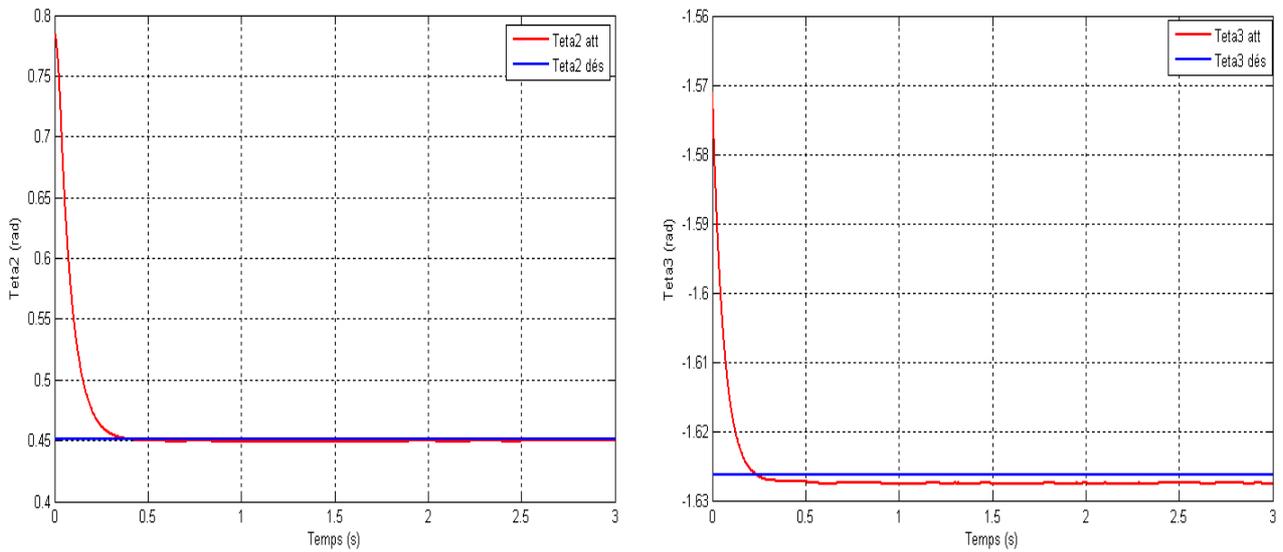


Figure IV-18 l'évolution de l'articulation θ_2 et θ_3 pour le mot « Bas » (posture coude haut)

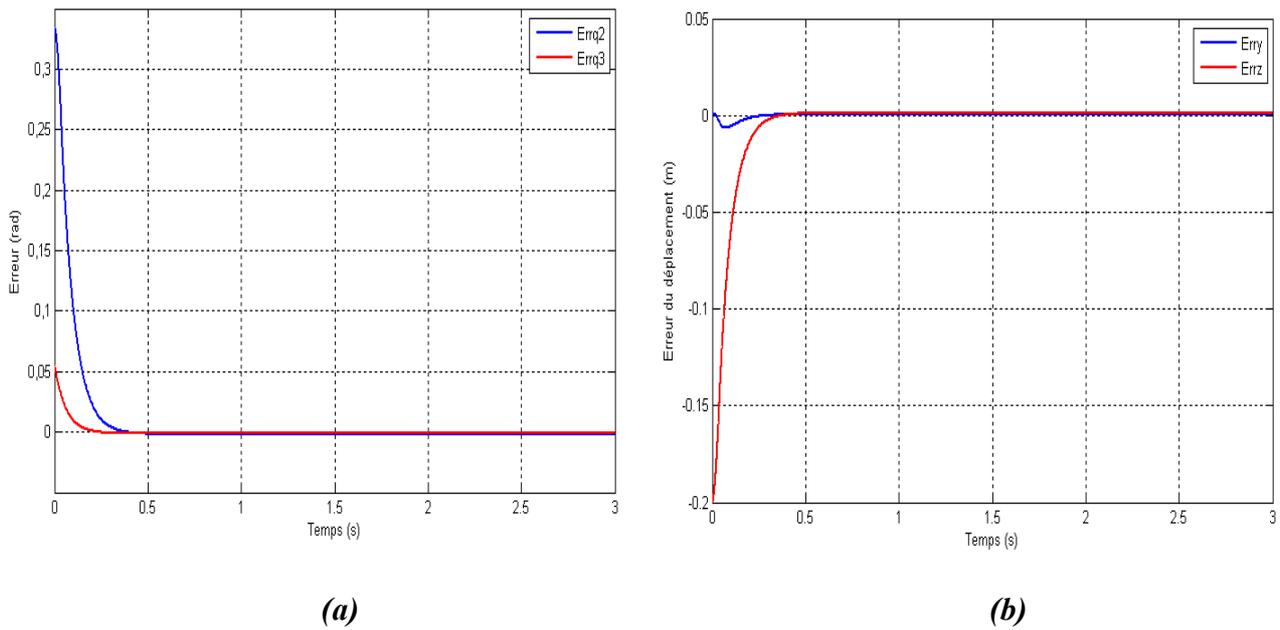


Figure IV-19 l'erreur de poursuite (a) et du déplacement (b) pour le mot « Bas » (posture coude haut)

IV-5-2 deuxième situation « posture coude bas »

➤ le mot Avant :

Les états initiaux :

$$\theta_1 = \frac{\pi}{2}; \quad \theta_2 = \frac{\pi}{4}; \quad \theta_3 = \frac{\pi}{2}$$

Le déplacement désiré

$$x_d = x_i; \quad y_d = y_i + 0.2 \text{ m}; \quad z_d = z_i$$

Résultats obtenus

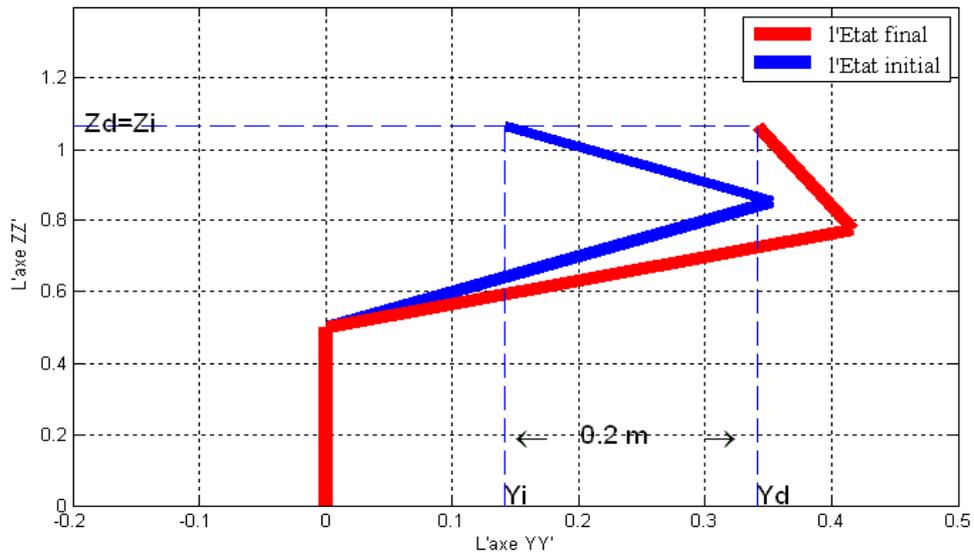


Figure IV-20 présentation du déplacement du bras en Avant

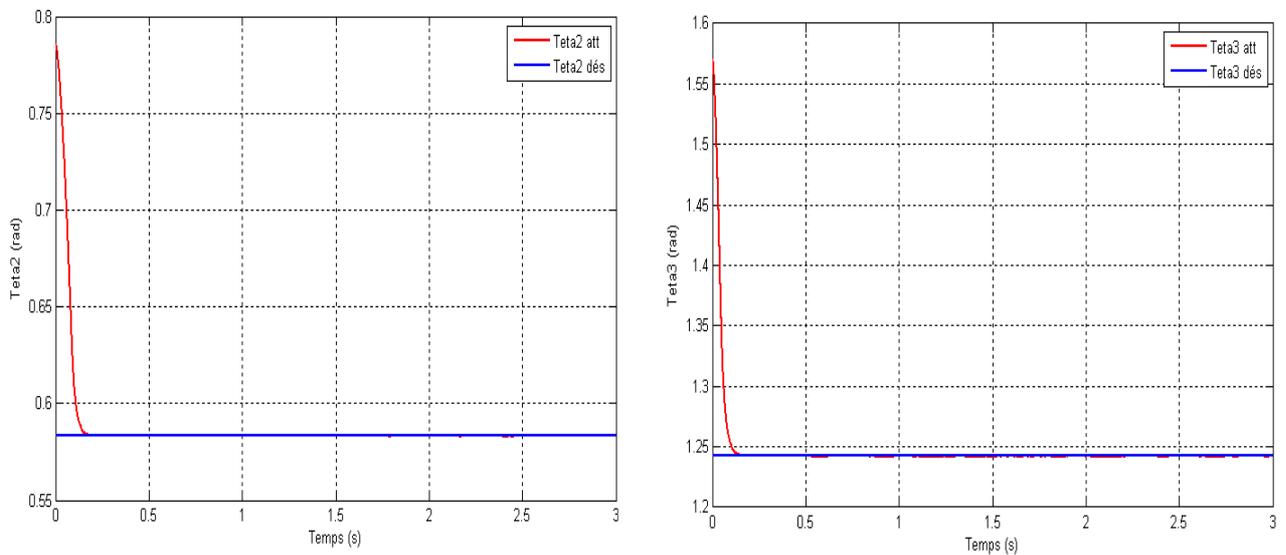
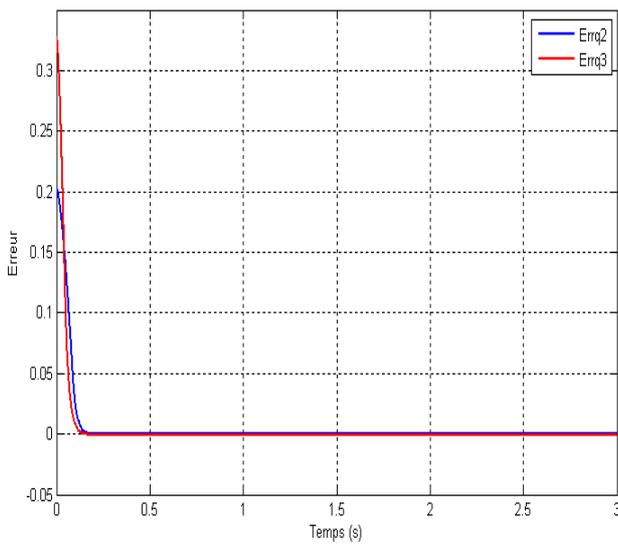
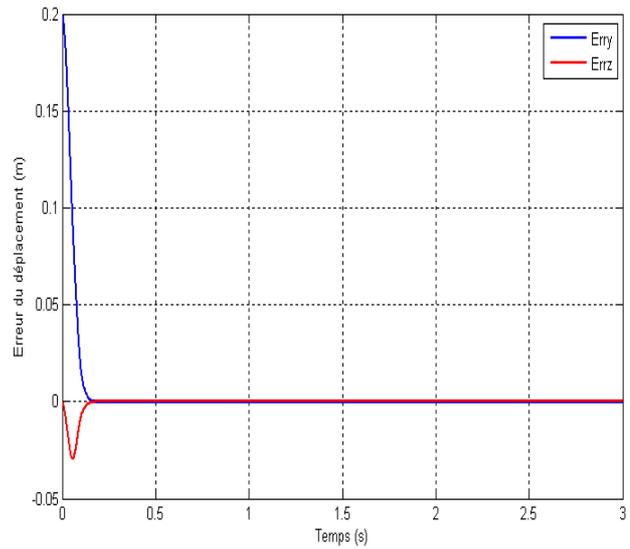


Figure IV-21 l'évolution de l'articulation θ_2 et θ_3 pour le mot « Avant » (posture coude bas)



(a)



(b)

Figure IV-22 l'erreur de poursuite (a) et du déplacement (b) pour le mot « Avant » (posture coude bas)

➤ le mot Arrière:

Les états initiaux :

$$\theta_1 = \frac{\pi}{2}; \quad \theta_2 = \frac{\pi}{4}; \quad \theta_3 = \frac{\pi}{2}$$

Le déplacement désiré

$$x_d = x_i; \quad y_d = y_i - 0.2 \text{ m}; \quad z_d = z_i$$

Résultats obtenus

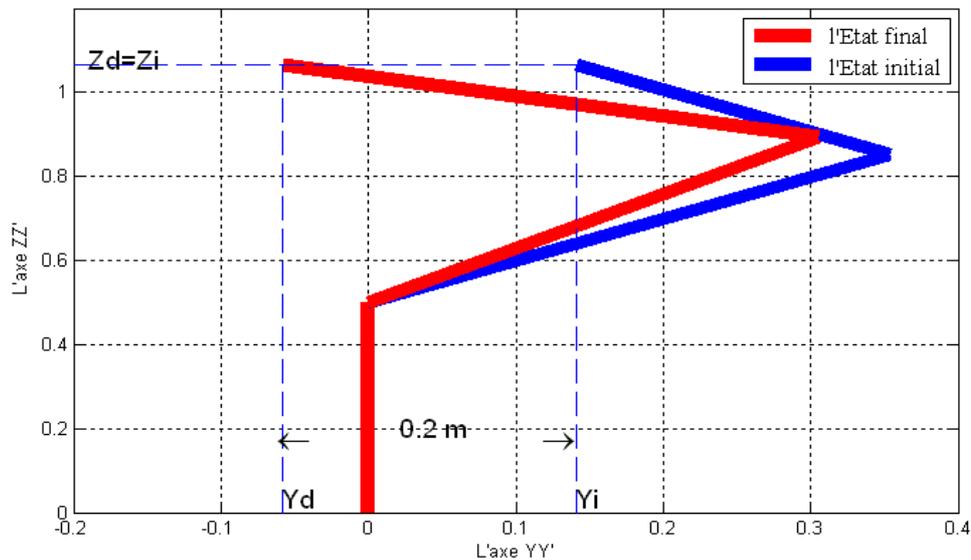


Figure IV-23 présentation du déplacement du bras en arrière

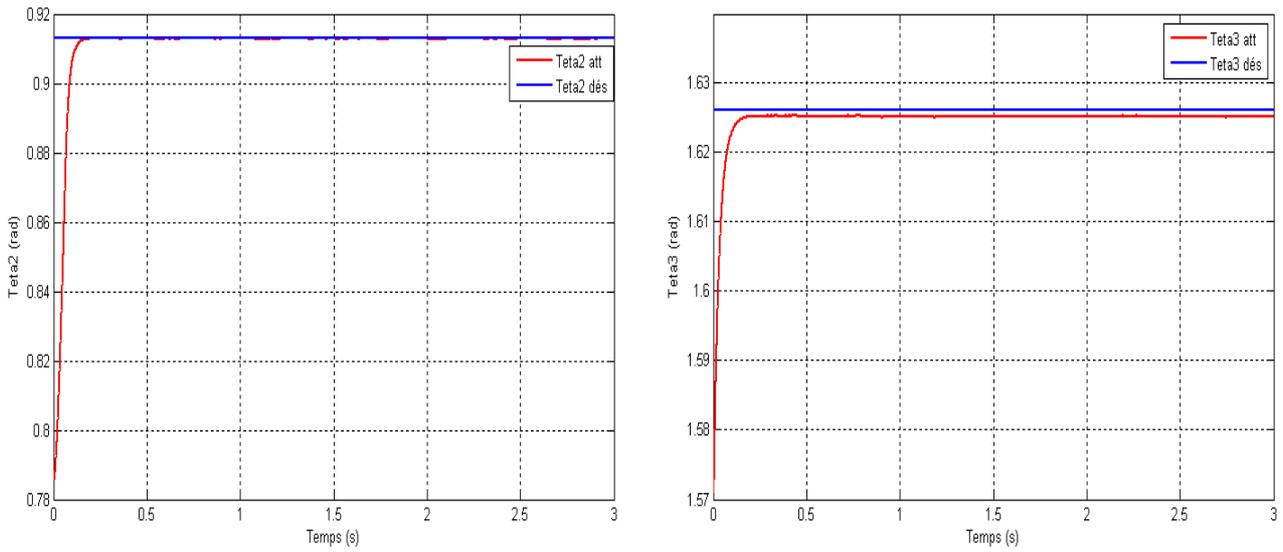
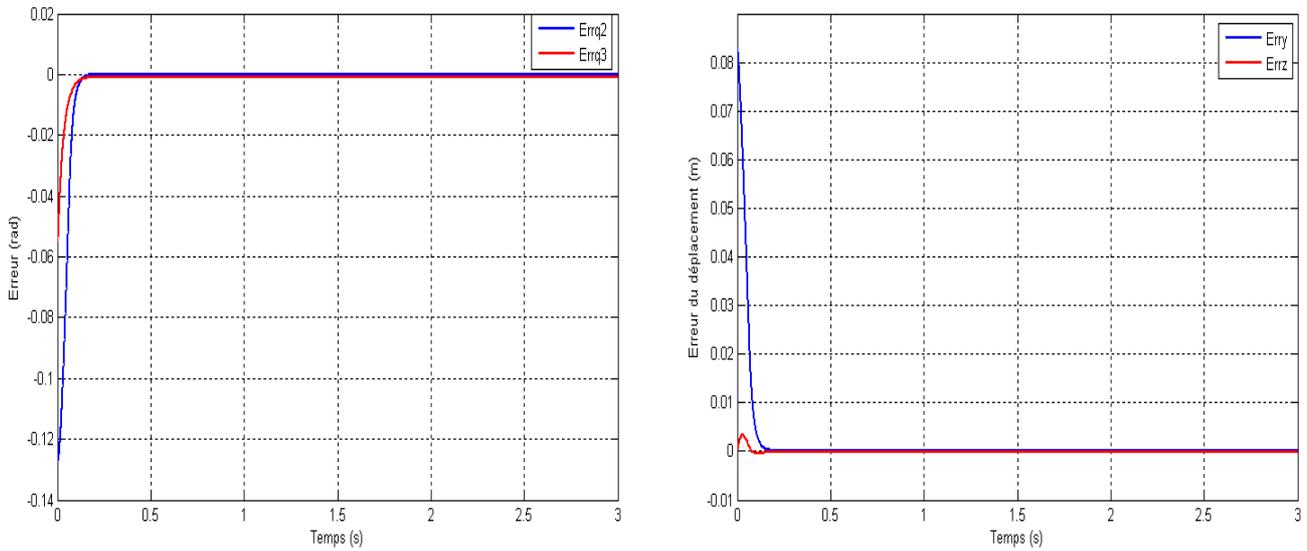


Figure IV-24 l'évolution de l'articulation θ_2 et θ_3 pour le mot « Arrière» (posture coude bas)



(a)

(b)

Figure IV-25 l'erreur de poursuite (a) et du déplacement (b) pour le mot « Arrière » (posture coude bas)

➤ *le mot Haut:*

Les états initiaux :

$$\theta_1 = \frac{\pi}{2}; \quad \theta_2 = \frac{\pi}{4}; \quad \theta_3 = \frac{\pi}{2}$$

Le déplacement désiré

$$x_d = x_i; \quad y_d = y_i; \quad z_d = z_i + 0.2$$

Résultats obtenus

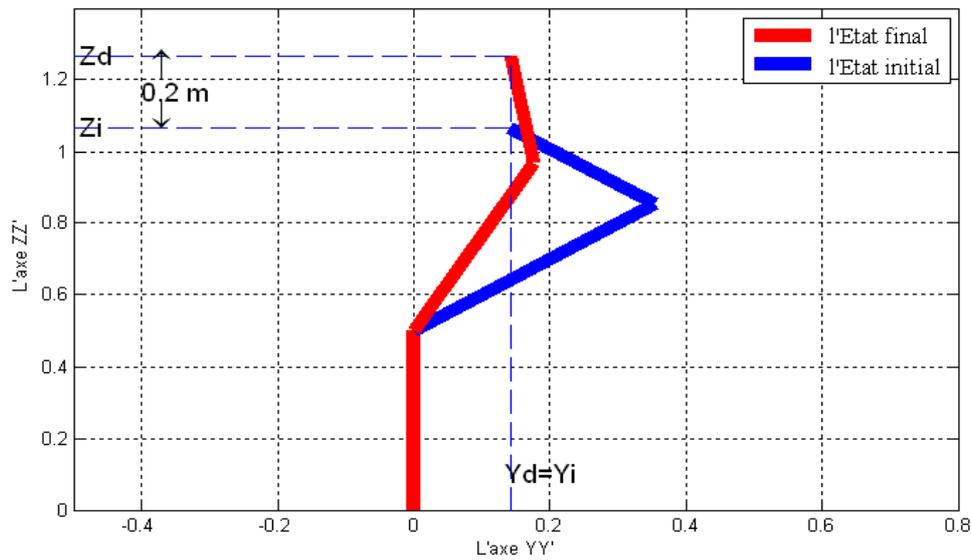


Figure IV-26 présentation du déplacement du bras en haut

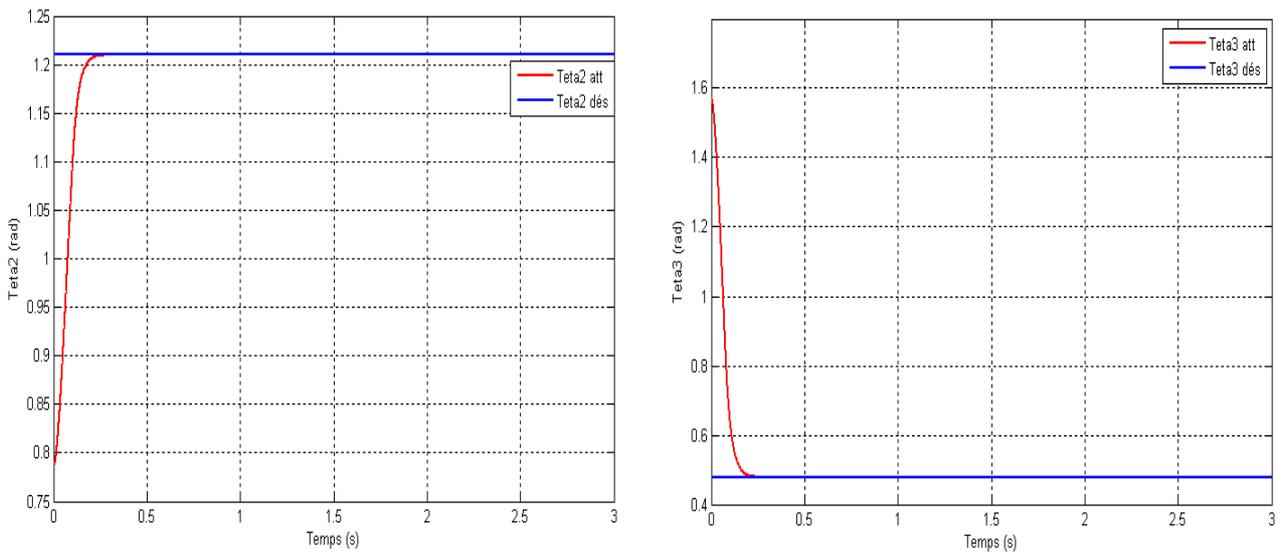


Figure IV-27 l'évolution de l'articulation θ_2 et θ_3 pour le mot « Haut» (posture coude bas)

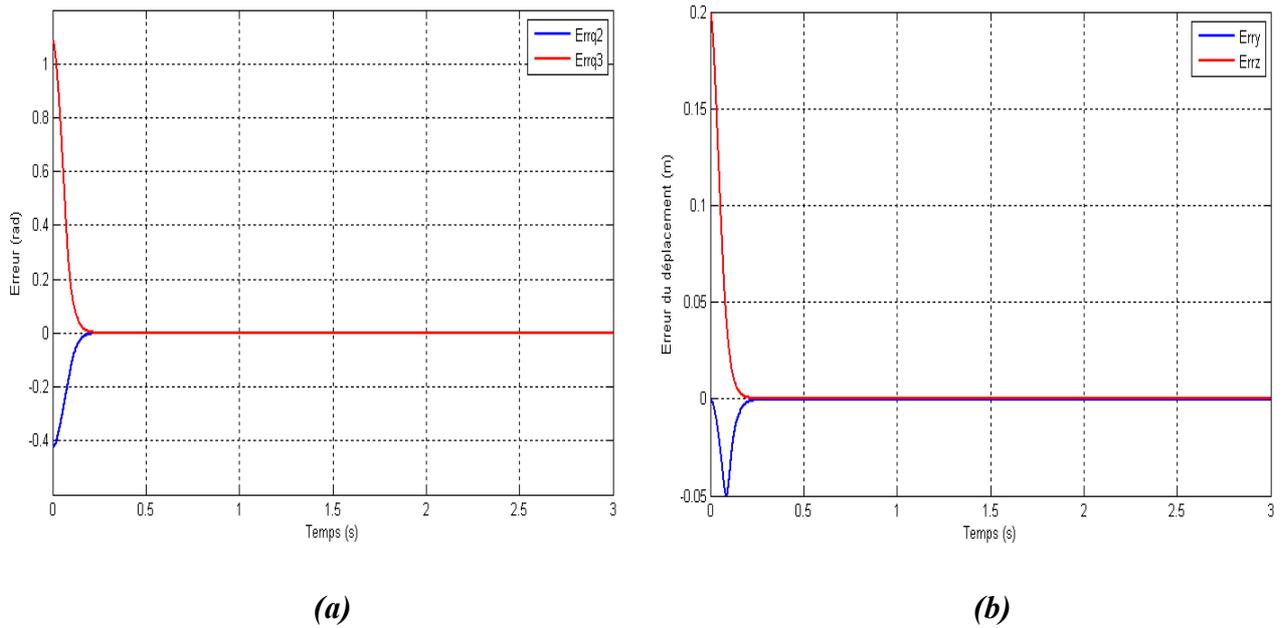


Figure IV-28 l'erreur de poursuite (a) et du déplacement (b) pour le mot « Haut » (posture coude bas)

➤ le mot Bas:

Les états initiaux :

$$\theta_2 = \frac{\pi}{4}; \quad \theta_3 = \frac{\pi}{2}; \quad \theta_3 = \frac{\pi}{2}$$

Le déplacement désiré

$$x_d = x_i; \quad y_d = y_i - 0.2; \quad z_d = z_i$$

Résultats obtenus

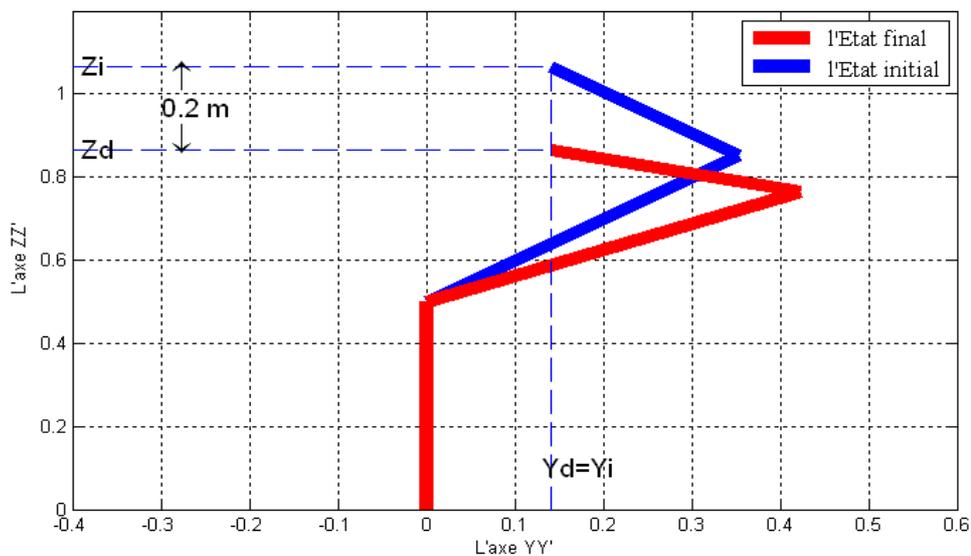


Figure IV-29 présentation du déplacement du bras en bas

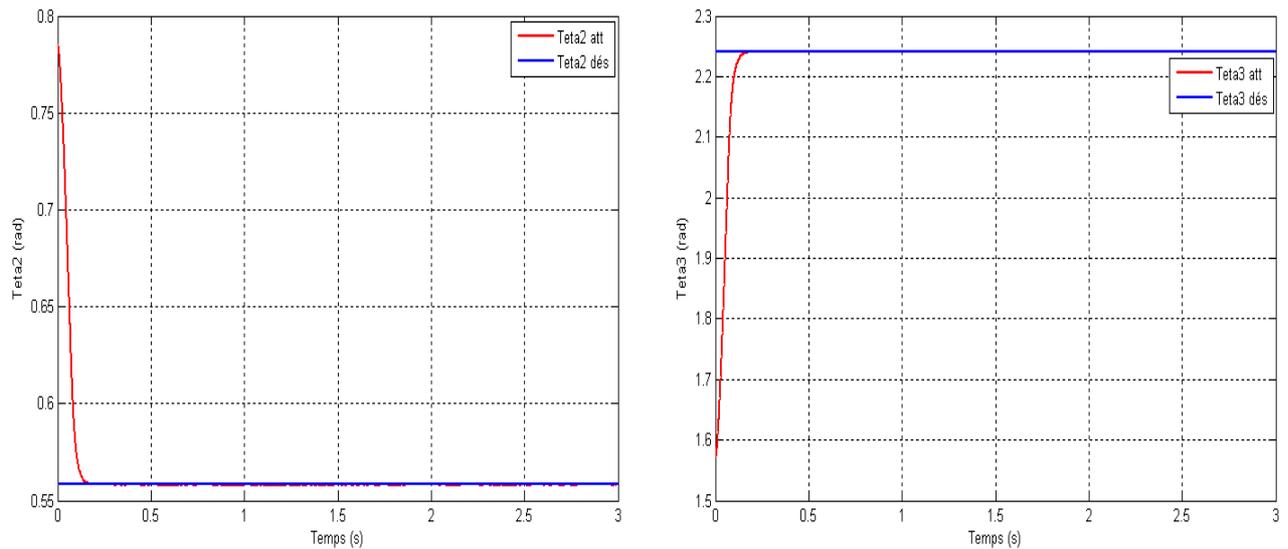


Figure IV-30 l'évolution de l'articulation θ_2 et θ_3 pour le mot « Bas » (posture coude bas)

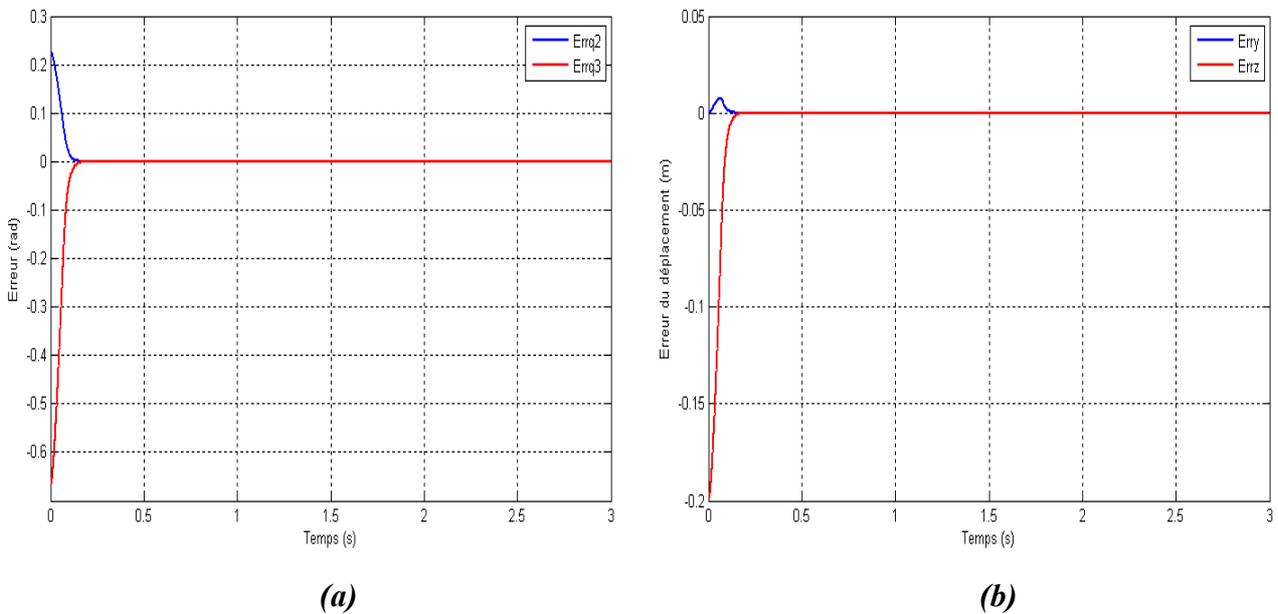


Figure IV-31 l'erreur de poursuite (a) et du déplacement (b) pour le mot « Bas » (posture coude bas)

IV-5-3 Rotation de bras autour de l'axe Z

Les mots concernés pour ce mouvement sont le mot *droite* et le mot *gauche* chaque mot provoque une rotation de $-\frac{\pi}{4}$ et $\frac{\pi}{4}$ respectivement.

➤ *le mot Gauche :*

Les états initiaux :

$$\theta_1 = \frac{\pi}{2}; \quad \theta_2 = \frac{\pi}{4}; \quad \theta_3 = -\frac{\pi}{2}$$

Le déplacement désiré

$$\theta_1 = \frac{\pi}{2} + \frac{\pi}{4}; \quad \theta_2 = \frac{\pi}{4}; \quad \theta_3 = -\frac{\pi}{2}$$

Résultats obtenus

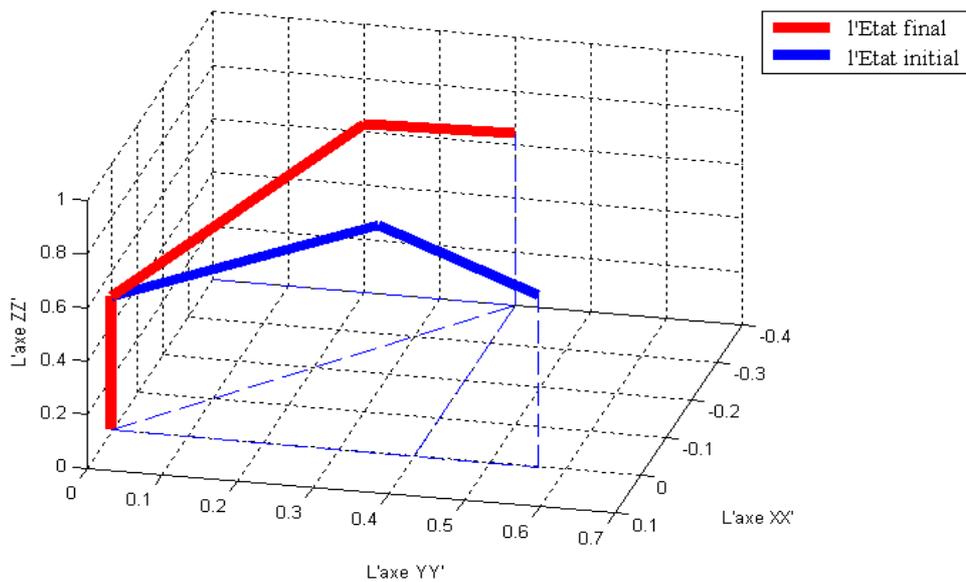


Figure IV-32 présentation de rotation du bras à gauche

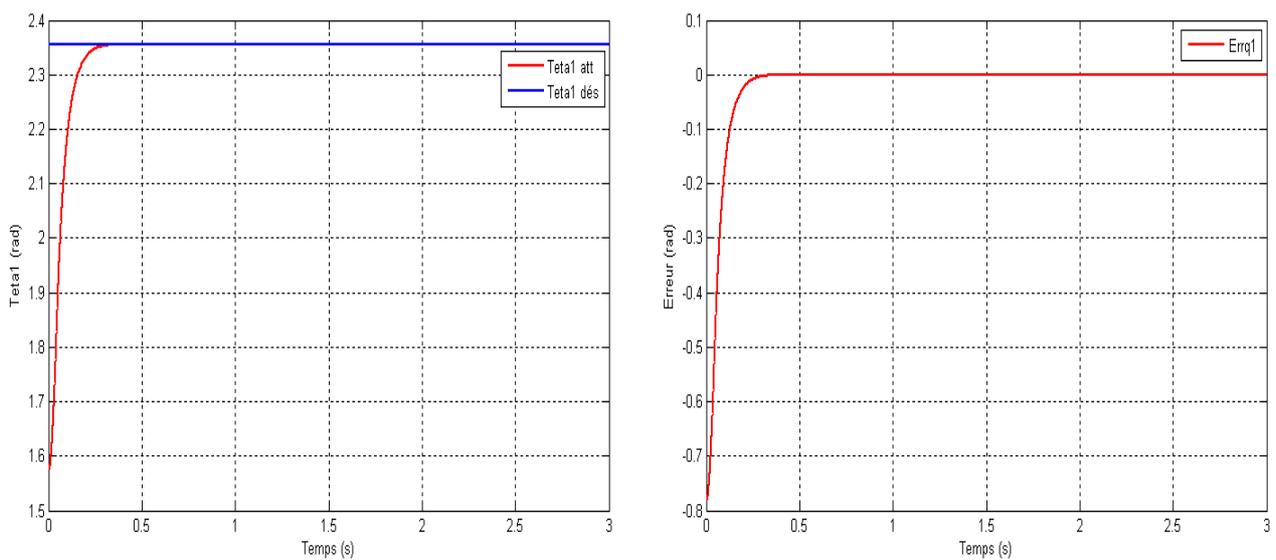


Figure IV-33 l'évolution de l'articulation θ_1 et son erreur pour le mot « Gauche »

➤ **le mot Droite :**

Les états initiaux :

$$\theta_1 = \frac{\pi}{2}; \quad \theta_2 = \frac{\pi}{4}; \quad \theta_3 = -\frac{\pi}{2}$$

Le déplacement désiré

$$\theta_1 = \frac{\pi}{2} - \frac{\pi}{4}; \quad \theta_2 = \frac{\pi}{4}; \quad \theta_3 = -\frac{\pi}{2}$$

Résultats obtenus

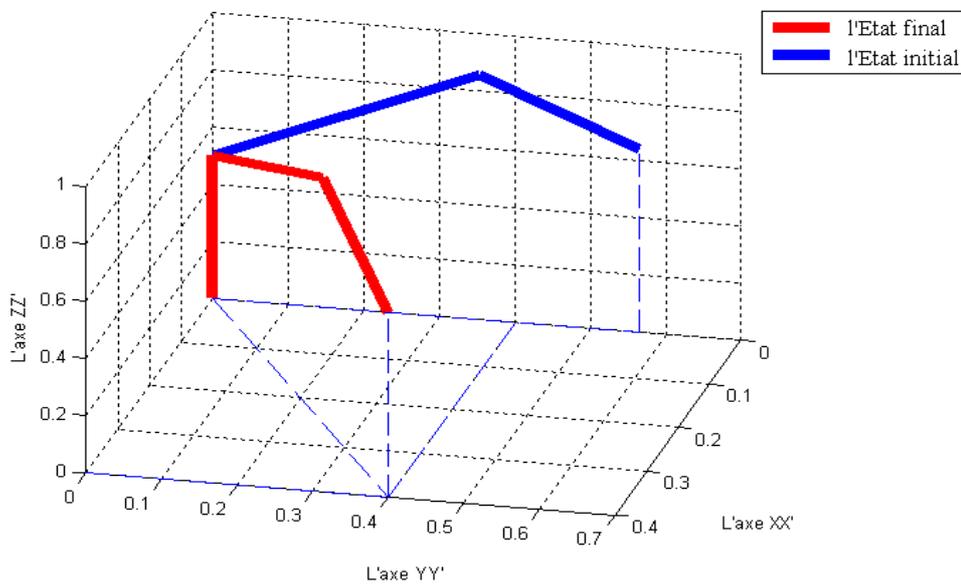


Figure IV-34 présentation de rotation du bras à droite

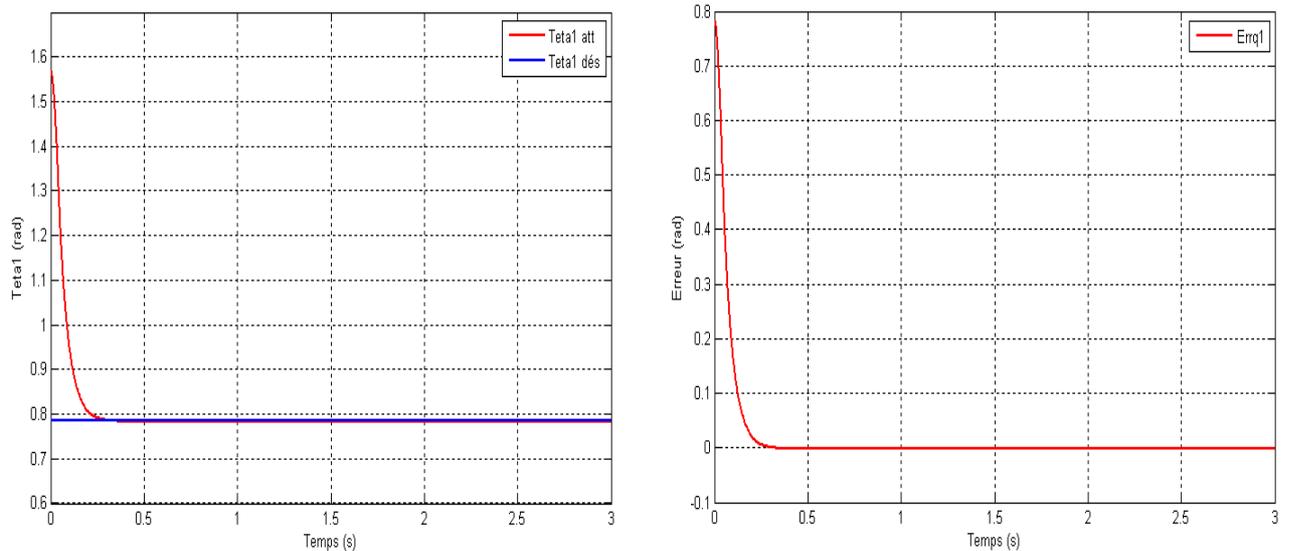


Figure IV-35 l'évolution de l'articulation θ_1 et son erreur pour le mot «Droite »

IV-6 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).

Les erreurs de poursuite enregistrées sont différentes d'un cas à l'autre, mais elles sont globalement proches de zéro ce qui montre l'efficacité de la commande retenue.

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 chaîne 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 été retenus pour cet objectif.

La reconnaissance proprement dite est opérée par utilisation d'une approche statistique basé sur le modèle de Markov caché. Dans une première phase, on opère un entraînement des paramètres du modèle 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 liberté commandé en mode glissant est utilisé. Un ensemble de mots ordres lui sont envoyés et les résultats des simulations sont encourageants.

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 maître dont il a fait la reconnaissance.

Bibliographie

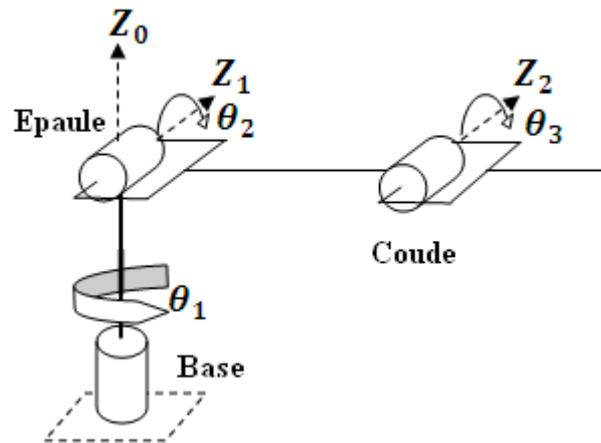
- [1] Khiair NAIT CHABANE ,"*Exploitation de la redondance pour la commande coordonnée d'un manipulateur mobile d'assistance aux personnes handicapées*" Université d'every 2006.
- [2] Y. BAAZI, "*Etude d'un Manipulateur à 3 Degrés de Liberté Application à la Fonction de Préhenseur*", thèse de magister, Institut d'électronique, Université de Batna, 2000
- [3] N.Azoui, "*Commande non linéaire d'un bras manipulateur*", Thèse de Magister, université de Batna 2009.de Batna, 2000.
- [4] C. VIBET, "*Robots Principes et Contrôle*", Editions Ellipses, 1988.
- [5] Bernard BAYLE, "*Introduction à la Robotique*", Université Louis Pasteur de Strasbourg 2004
- [6] Gilles Léothaud, "*Théorie de la Phonation*"2004-2005;
- [7] Thomas HUEBER, "*Reconstitution de la parole par imagerie ultrasonore et vidéo de l'appareil vocal : vers une communication parlée silencieuse*", thèse de doctorat Université Pierre et Marie Curie 2009.
- [8] Thomas Hueber, "*Synthèse de la parole à partir d'imagerie ultrasonore et optique de l'appareil vocal*", thèse de doctorat Ecole Supérieure de Chimie Physique Electronique de Lyon
- [9] Deller, J.R., Hansen, J.H. L., Proakis J. G ,"*Discrete Time Processing of Speech Signals*" IEEE Press 1999.
- [10] Calliope , "*La parole et son traitement automatique*" Paris, Masson 1989.
- [11] René Boite, Hervé Boulard "*Traitement de la parole*" presses polytechniques et Universitaires de Romandes 2000

-
- [12] Laurent BUNIET "*Traitement automatique de la parole en milieu bruité : étude de modèles connexionnistes statiques et dynamiques*", thèse doctorat Université Henri Poincaré - Nancy 1 1997
- [13] J. Koreman, B. Andreeva, et H. Strik. "*Acoustic parameters versus phonetic features in ASR*", In Proceedings of the International Congress of Phonetic Sciences, 1999.
- [14] J. P. Haton, J. M. Pierrel, G. Perennou, J. Caelen, et J. L. Gauvain, "*Reconnaissance automatique de la Parole*" dunod, 1991.
- [15] E. Zwicker et R. Feldtkeller. Psychoacoustique , "*L'oreille, récepteur d'information*", Masson, 1981.
- [16] 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.
- [17] F. Jelinek, "*Continuous speech recognition by statistical methods*", Proceeding of IEEE, 1976.
- [18] J. Ueberla, "*Analysing a simple language model : some general conclusions for language models for speech recognition*", Computer, Speech and Language, 1994.
- [19] Mark W. Spong, Seth Hutchinson, and M. Vidyasagar "*Robot Modeling and Control*"
- [20] Salma Jamoussi "*Methodes statistiques pour la comprehension automatique de la parole*" thèse Doctorat de l'universite Henri Poincare-Nancy decembre 2004
- [21] L. Rabiner et B. H. Juang, "*A tutorial on hidden Markov models and selected application in speech recognition*", Proceeding of IEEE, 1989.
- [22] Yassine BEN AYED, "*Détection de mots clés dans un flux de parole* ", Doctorat de l'Ecole Nationale Supérieure des Télécommunications décembre 2003

-
- [23] J. L. Gauvain et C. Lee, "*Maximum a-posteriori estimation for multivariate gaussian mixture observations of Markov chains*", IEEE Transactions on Speech and Audio Processing, 1994.
- [24] R. Cardin, Y. Normandin, et R. De Mori, " *High performance connected digit recognition using maximal mutual information estimation*", In Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, , 1991.
- [25] S. Kapadia, V. Valtchev, et S. J. Young, "*MMI training for continuous phoneme recognition on the TIMIT database*", In Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, , 1993.
- [26] 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
- [27] J.J.Slotine, W.Li "*Applied nonlinear control*", Englewood Cliffs, NJ: prentice Hall, 1991.
- [28] M. Hacene "*contribution à l'étude de la programmation et de la commande d'un robot manipulateur* " thèse du doctorat (1981)

Annexe

➤ **Modèle dynamique de bras manipulateur utilisé**



$$\tau = M(q)\ddot{q} + N(q, \dot{q})\dot{q} + G(q)$$

$$M(1,1) = 2.57 + 1.38\cos(\theta_2)^2 + 0.3\sin(\theta_2)\sin(\theta_3) + 0.744\cos(\theta_2)\sin(\theta_2 + \theta_3);$$

$$M(1,2) = 0.69\sin(\theta_2) - 0.134\cos(\theta_2 + \theta_3) + 0.0238\cos(\theta_2);$$

$$M(2,1) = M(1,2)$$

$$M(1,3) = -0.134\cos(\theta_2 + \theta_3) - 3.97 \times 10^{-3}\sin(\theta_2 + \theta_2);$$

$$M(3,1) = M(3,1)$$

$$M(2,2) = 6.79 + 0.744 * \sin(\theta_3);$$

$$M(2,3) = 0.333 + 0.372 * \sin(\theta_2) - 0.011 * \cos(\theta_3);$$

$$M(3,2) = M(2,3)$$

$$M(1,1) = 1.16;$$

$$N(1,1) = (-2.76\sin(\theta_2)\cos(\theta_2) + 0.744 * \cos(2\theta_2 + \theta_3) + 0.6\sin(\theta_2)\cos(\theta_3) - 0.0213(1 - 2\sin(\theta_2)\sin(\theta_3)))\dot{\theta}_2$$

$$N(1,2) = (0.69\cos(\theta_2) + 0.134\sin(\theta_2 + \theta_3) - 0.0238\sin(\theta_2))\dot{\theta}_2 + (0.267\sin(\theta_2 + \theta_3) - 7.58 \times 10^{-3}\cos(\theta_2 + \theta_3))\dot{\theta}_3;$$

$$N(1,3) = (0.744\cos(\theta_2)\cos(\theta_2 + \theta_3) + 0.6\sin(\theta_2)\cos(\theta_3) + 0.022\cos(\theta_2)\sin(\theta_2 + \theta_3) - 0.0213(1 - 2\sin(\theta_2)\sin(\theta_3)))\dot{\theta}_1 + 0.5(0.267\sin(\theta_2 + \theta_3) - 7.58 \times 10^{-3}\cos(\theta_2 + \theta_3))\dot{\theta}_3;$$

$$N(2,1) = -0.5(-2.76\sin(\theta_2)\cos(\theta_2) + 0.744\cos(2\theta_2 + \theta_3) + 0.6\sin(\theta_2)\cos(\theta_3) - 0.0213(1 - 2\sin(\theta_2)\sin(\theta_3)))\dot{\theta}_1;$$

$$N(2,2) = (0.022\sin(\theta_3) + 0.744\cos(\theta_3))\dot{\theta}_3;$$

$$N(2,3) = 0.5(0.022\sin(\theta_3) + 0.744\cos(\theta_3))\dot{\theta}_3;$$

$$N(3,1) = -0.5(0.744\cos(\theta_2)\cos(\theta_2 + \theta_3) + 0.6\sin(\theta_2)\cos(\theta_3) + 0.022\cos(\theta_2)\sin(\theta_2 + \theta_3) - 0.0213(1 - 2\sin(\theta_2)\sin(\theta_3)))\dot{\theta}_1;$$

$$N(3,2) = -0.5(0.022\sin(\theta_3) + 0.744\cos(\theta_3))\dot{\theta}_2;$$

$$N(3,3) = 0;$$

$$g_1 = 0;$$

$$g_2 = -37.2\cos(\theta_2) - 8.4\sin(\theta_2 + \theta_3) + 1.02\sin(\theta_2);$$

$$g_3 = -8.4\sin(\theta_2 + \theta_3) + 0.25\cos(\theta_2 + \theta_3);$$