République Algérienne Démocratique et Populaire

Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

Université Ibn Khaldoun Tiaret

جامعة ابن خلدون تيارت

Faculté des Mathématiques et de l'informatique

Thème :

Reconnaissance Automatique de la Parole Appliquée à Sur Quelques Mots Arabes à Consonnes Pharyngales

POUR L'OBTENTION DU DIPLOME DE MASTER

Spécialité : **R**éseaux et **T**élécom

Dirigé par :

Mr. K, MEZZOUG

Réalisé par :

Mr. CHAKER Hocine

Mr. DALIA Noureddine

Année Universiraire 2012/2013

Dédicaces

- *A mes parents pour m'avoir transmis la soif du savoir et l'amour des études dés mon plus jeune âge. Pour leurs encouragements et leur soutien lors des moments les plus difficiles.*
- *A mes frères qui sont toujours là pour moi.*
- *A mes amis si nombreux que je ne vais pas nommer car la liste est longue et je ne veux pas faire des jaloux ni prendre le risque d'oublier quelques uns.*
- *Et enfin et non pas des moindres, a toute personne me connaissant et ayant de prés ou de loin apporté son aide a la réalisation de ce travail*

Remerciements

Avant tout on remercie dieu ALLAH le tout puissant Qui nous a dotés de toutes les facultés et nous avoir donné la force le courage, la santé et les moyens afin de réaliser ce modeste travail.

A notre encadreur ***Mr. K, MEZZOUG***

Pour avoir accepté de nous assister, afin de réaliser notre mémoire de fin d'étude pour ses précieux conseils avec gentillesse et disponibilité pour son bon sens et son esprit vif, pour la confiance qui nous a témoigné et

La patience qu'il nous a accordée .On remercie également nos membres de Jury. Nous Adressons également nos vifs Remerciements à l'ensemble des enseignants de la Faculté. A tout le personnel administratif.

Résumé

La Reconnaissance Automatique de la Parole (RAP) est affectée par les nombreuses Variabilités présentes dans le signal de parole. En dépit de l'utilisation de techniques sophistiquées, un système de RAP seul n'est généralement pas en mesure de prendre en compte l'ensemble de ces variabilités. Nous proposons l'utilisation de diverses sources d'information acoustique pour augmenter la robustesse des systèmes de reconnaissance.

Notre objectif est le développement et le test d'une méthode RAP (Reconnaissance Automatique de La Parole) particulièrement de la langue arabe en prenant en compte le phénomène des consonnes pharyngales. Le système global est représenté par les RNA (Réseau de Neurone Artificiel) et plus particulièrement les MLP (Multi Layer Perceptrons) pour reconnaissance automatique de la parole. Les résultats montrent un taux de reconnaissance important et variant et ceci d'une façon générale selon plusieurs paramètres de l'algorithme d'apprentissage et selon les scenarios de reconnaissance préposés dans ce modeste travail.

Abstract

Automatic Speech Recognition (ASR) is affected by much variability present in the speech signal. Despite sophisticated techniques, a single ASR system is usually incapable of considering all these variability's. We propose to use various sources of acoustic information in order to increase precision and robustness of recognition systems.

Our objective is the development and the test of an ASR method (Automatic Recognition of The Speech) particularly of the Arabic language taking in account the phenomenon of pharyngeal's consonants. The global system is represented by the ANN (Artificial Neuron Network) and more especially the MLP (Multi Layer Perceptrons) for automatic speech recognition. The results show an important and various recognition rates according to several parameters of training algorithm and according to scripts of recognition employees in this modest work.

Mots clé -Key words

ANN (Artificial Neural Network), MLP (Multi Layer Perceptron) ASR (Automatic Speech recognition), MFCC (Mel-frequency cepstrum coefficients).

Sommaire

Chapitre II: La reconnaissance des formes

Chapitre III: La reconnaissance automatique de la parole

Chapitre IV: Implémentation et tests

Sommaire

Liste des tableaux

Liste des figures

Liste des figures

Introduction générale

La parole est certainement le moyen le plus direct et le plus naturel utilisé par l'homme pour échanger l'information. Le progrès enregistré dans le domaine du traitement du signal, le développement des moyens informatiques (matériels et logiciels) et l'apport de l'intelligence artificielle permettent d'envisager l'utilisation de la parole pour communiquer et dialoguer avec une machine.

La Reconnaissance Automatique de la Parole (RAP) est une technologie informatique permettant à un logiciel d'interpréter une langue naturelle humaine. Elle permet à une machine d'extraire le message oral contenu dans un signal de parole.

Les caractéristiques phonétiques et linguistiques de la langue sont largement impliquées dans le processus.

L'Arabe est une langue sémitique. Elle comprend un standard compris par l'ensemble de la communauté arabophone et une multitude de dialectes différents les uns des autres.

La langue arabe a fait l'objet de plusieurs études anciennes et récentes, mais jusqu'à présent peu de travaux ont été effectués concernant la reconnaissance automatique de la parole ou RAP, Avec l'apparition de la nouvelle génération des smart phones les utilisateurs peuvent parler avec leurs téléphones avec des langues spécifiques.

L'absence de la langue arabe parmi ces langues reflète la pauvreté des recherches sur la parole arabe.

Sur le plan phonétique, l'Arabe standard présente la particularité d'être une langue essentiellement consonantique qui se caractérise par la présence des consonnes pharyngales, glottales et emphatiques.

Ce projet de fin d'étude représente une initiative à la proposition d'un modèle neuronale destiné à la reconnaissance automatique de la parole appliquée sur quelques mots arabes à consonnes pharyngales, ce modèle repose essentiellement sur le codage cepstrale MFCC (Mel Frequency Ceptrum Coefficients) pour automatiser la reconnaissance de la parole et d'extraire les paramètres cepstraux pertinents.

Dans le premier chapitre on va présenter des généralités sur le traitement de signal Puis on va présenter d'une manière plus détaillée, le traitement de signale de parole.
Dans le deuxième chapitre, nous présenterons la reconnaissance des formes et les réseaux de neurones artificiels comme outil de reconnaissance et de validation de notre travail.

Le troisième chapitre est consacré à la description de la reconnaissance automatique de la parole, et en particulier la langue arabe et ces différents aspects acoustiques et phonétiques concernant le caractère pharyngale de quelques consonnes arabes, puis nous récapitulons notre chapitre par une description de quelques travaux scientifiques en matière de RAP de la langue arabe.

Le quatrième et dernier chapitre sera consacré à l'implémentation et les expériences en plus des observations et différentes constatations, enfin nous terminant notre étude par une conclusion générale tout en évaluant les performances de telle approche.

Chapitre I
Traitement de Signal de la Parole

I. traitement de signal de parole

1. Introduction

Le signal de parole n'est pas un signal ordinaire, il est le vecteur d'un phénomène extrêmement complexe, Il est difficile de le modéliser car ses propriétés statistiques varient au cours du temps.

Dans ce chapitre on va présenter des généralités sur le traitement de signal, la représentation et la classification des signaux et aussi on va parler comment échantillonner, quantifier et coder un signal…etc. Puis on va présenter d'une manière plus détaillée, le traitement de signale de parole qui est aujourd'hui une composante fondamentale des sciences de l'ingénieur.

2. traitement de signal

2.1. Définition

Le traitement du signal (T.S) est une discipline technique qui a pour objet l'élaboration, la détection et l'interprétation des signaux porteurs d'informations. Son but est donc de réussir à extraire un maximum d'information utile sur un signal perturbé par du bruit en s'appuyant sur les ressources de l'électronique et de l'informatique.

2.2. Domaines d'applications

Le traitement de la parole : analyse, synthèse, codage, reconnaissance…etc.

Les télécommunications : téléphone, radio, mobile, vidéoconférence.

Le biomédical : échographie, imagerie.

Défense : systèmes d'armes, surveillance, guidage, navigation.

L'astronomie : imagerie optique et radio.

Le radar et le sonar, l'imagerie satellitaire …etc.

2.3. Le signal

Le signal est la représentation physique de l'information qui est convoyée d'une source vers un destinataire et qui évolue dans le temps ou dans l'espace. *[1.1]*

2.4. Le bruit

Le bruit est défini comme tout phénomène perturbateur gênant la perception ou l'interprétation d'un signal.

- Rapport signal sur bruit

Le rapport signal sur bruit mesure la quantité de bruit contenue dans le signal. Il s'exprime par le rapport des puissances du signal (p_S) et du bruit (p_N). Il est souvent donné en décibels (dB)

$$\left(\frac{s}{n}\right)_{dB} = 10 \log \frac{p_S}{P_N}$$

2.5. Système

Un système est un appareil où l'on peut distinguer des signaux d'entrée et des signaux de sortie (téléphone, modem)

2.6. Signal analogique

a)Définition

Le signal analogique est un signal continu qui par définition contient un nombre infini d'éléments, (domaine des temps et des amplitudes continus).

b) Représentation des Signaux

Un signal physique est représenté par des fonctions S(t) à valeurs réelles d'une variable réelle t. Par conséquent, le signal possède les caractéristiques suivantes :

– énergie bornée ;

– amplitude bornée ;

– continu temporellement ;

– causal (**s(t)** = **0** pour **t** < **0**) **;**

– spectre du signal borné (tend vers 0 lorsque f tend vers ∞) (Figure1.1).

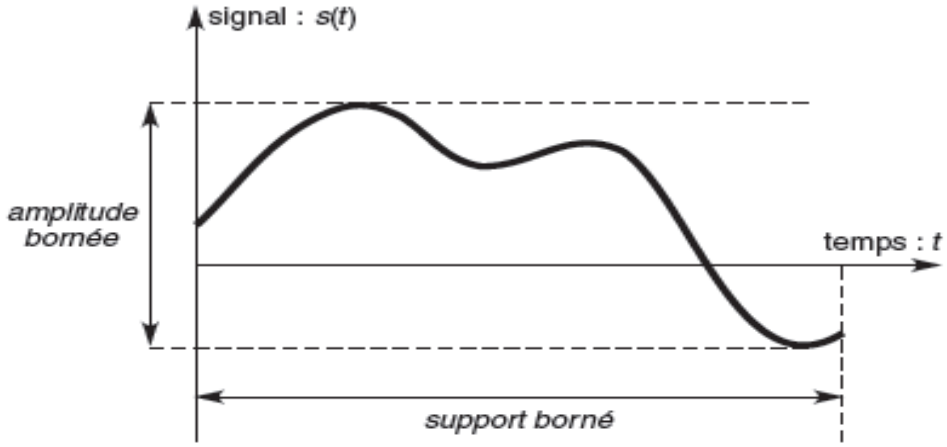


Figure 1. 1 : *Représentation d'un signal physique.*

c) Classification des signaux

Pour faciliter l'étude des signaux, différents modes de classification peuvent être envisagés :

- Classification temporelle des signaux ;
- Classification spectrale

➢ Classification temporelle des signaux

La première classification, basée sur l'évolution du signal en fonction du temps, fait apparaître deux types fondamentaux :

- Les signaux certains (ou déterministes) dont l'évolution en fonction du temps peut être parfaitement décrite par un modèle mathématique. Ces signaux proviennent de phénomènes pour lesquels on connaît les lois physiques correspondantes et les conditions initiales, permettant ainsi de prévoir le résultat ;

- les signaux aléatoires (ou probabilistes) dont le comportement temporel est imprévisible et pour la description desquels il faut se contenter d'observations statistiques.

➢ Classification spectrale

Un signal peut être classé suivant la distribution de son amplitude, sa puissance ou son énergie en fonction de la fréquence (spectre du signal). Le domaine des fréquences occupé par son spectre est aussi appelé la largeur de bande spectrale du signal ΔF (figure 1.2): *[1.2]*

$$\Delta F = F_{max} - F_{min}$$

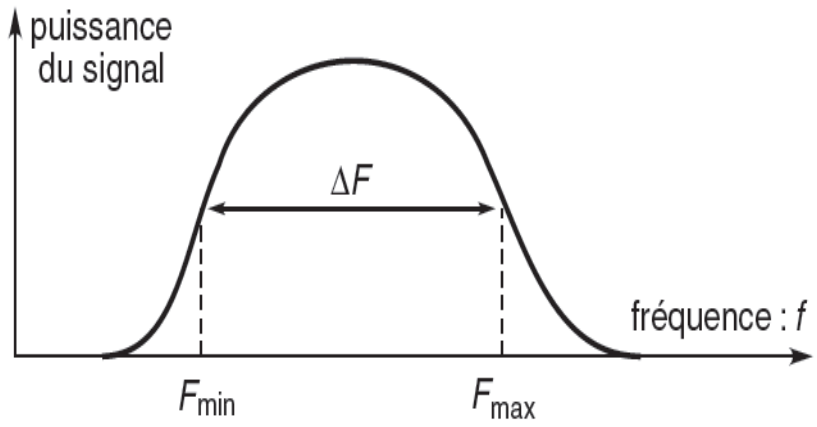


Figure 1. 2 : *Distribution spectrale d'un signal.*

2.7. Numérisation des signaux

a) Définition

Numérisation d'un signal est l'opération qui consiste à faire passer un signal de la représentation dans le domaine des temps et des amplitudes continus au domaine des temps et des amplitudes discrets.

b) Échantillonnage

L'échantillonnage consiste à prélever un nombre déterminé d'éléments (échantillons) qui seront suffisants pour reconstituer à l'arrivée d'un signal analogique de qualité.

Les différentes études ont montré qu'il suffit d'échantillonner à deux fois la fréquence supérieure contenu dans le signal. Ainsi, pour un signal de la parole où l'information est contenue dans une bande de 4000 Hz (0-4000), un échantillonnage à 8000 Hz suffit (c'est à dire toutes les 125 µs) (Figure 1. 3) *[1.3]*.

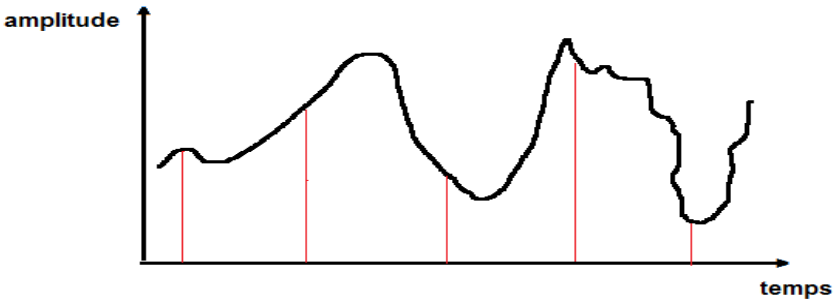


Figure 1. 3 : *Echantillonnage d'un signal analogique.*

- Théorème de Shannon

Lorsqu'un signal x(t) a un spectre à support borné [X(f) = 0 pour |f| >F_{max}], il est possible d'échantillonner ce signal sans perdre d'information : il suffit pour cela de choisir une fréquence D'échantillonnage $f_e \geq 2.F_{max}$. On pourra alors reconstruire x(t) parfaitement à partir des échantillons x(nT_e), avec $T_e = \frac{1}{F_e}$, représenté (figure 1.4) *[1.4]*

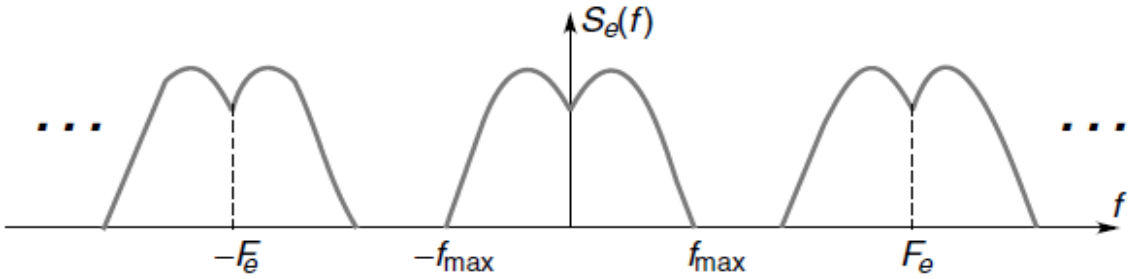


Figure 1.4 : *Périodisation du spectre du signal échantillonné.*

c) Quantification

Elle consiste à représenter un échantillon par une valeur numérique au moyen d'une loi de correspondance. La loi la plus simple consiste à diviser l'ordonnée en segments égaux. Le nombre de segments dépend du codage. Par exemple un codage sur 8 bits engendre 2^8 segments. L'erreur effectuée dans l'approximation est appelée bruit de numérisation.

La figure 1.5 représente une quantification d'un signal avec un codage sur 2 bits *[1.5]*

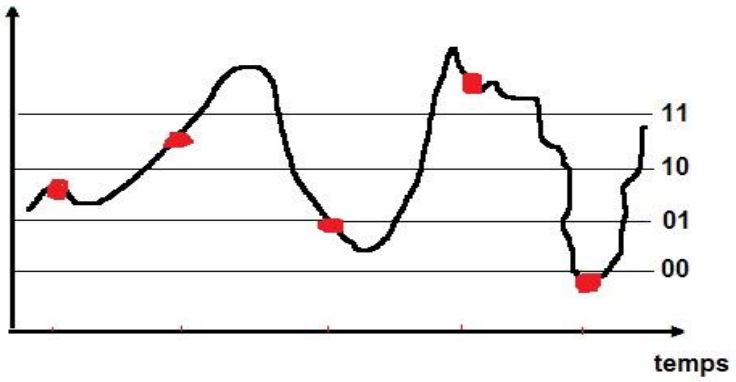


Figure 1. 5 : *Quantification d'un signal analogique*

d) Codage

Le codage consiste à associer à un ensemble de valeurs discrètes un code composé d'éléments binaires.

Les codes les plus connus : code binaire naturel, code binaire décalé, code Manchester , code DCB, code Gray. Figure 1.6 montre le codage sur 2 bits d'un signal analogique *[1.5]*

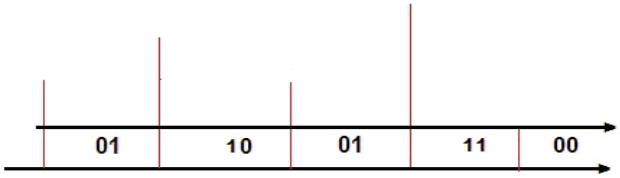


Figure 1. 6 : *codage d'un signal analogique*

2.8. Les Différentes représentations Fréquentiels

L'analyse fréquentielle des signaux apporte une information supplémentaire importante. Pour les différents cas de signaux, classés selon les caractéristiques continu ou discret et périodique ou transitoire. De plus les méthodes, utilisées pour calculer ces représentations spectrales, ne sont pas les mêmes selon ces différents types de signaux (tableau 1.1).

	Spectre	
Signal	**Méthode de calcul**	**Caractéristiques**
1 - continu et périodique	Série de Fourier	Discret et non périodique
2 - continu et non périodique	Intégrale de Fourier	Continu et non périodique
3- discret et non périodique	Intégrale de Fourier	Continu et périodique
4 - discret et périodique	Transformée de Fourier discrète (TFD)	Discret et périodique

Tableau 1. 1 : *Correspondance entre les signaux et leurs spectres associés*

2 .9. Transformée De Fourier

La transformée de Fourier est une des méthodes pour représenter un phénomène temporel dans le domaine fréquentiel. Il existe d'autres méthodes plus complexes mais plus efficaces.

a. Transformée De Fourier Discrète

On appelle transformée de Fourier discrète d'une suite de termes (0) (1)

(−1), la suite de termes (0) (1) (−1), définis par :

$$X(k) = \sum_{n=0}^{n-1} x(n) e^{-j2\pi \frac{nk}{N}}$$

b. Transformée De Fourier Rapide

La transformée de Fourier rapide TFR ou FFT (Fast Fourier Transforme) est simplement un algorithme permettant de réduire le nombre d'opérations, en particulier le nombre de multiplications, pour calculer la TFD. Ce temps de calcul est en effet primordial pour réaliser des systèmes numériques en « temps réel » [1.6].

2.10. Convolution

Le produit de convolution d'un signal s(t) par un autre h(t) est donné par :

$$s(t) * h(t) = \int_{-\infty}^{+\infty} s(k)\, h(t-k)\, dk$$

2.11 Notion de Filtrage

Le filtrage est une forme de traitement de signal qui modifie le spectre de fréquence et/ou la phase du signal présent en entrée du filtre et donc par conséquent sa forme temporelle. Il peut s'agir soit :

- d'éliminer ou d'affaiblir des fréquences parasites indésirables
- d'isoler dans un signal complexe la ou les bandes de fréquences utiles.

On classe les filtres en deux grandes familles :

- les filtres numériques réalisés à partir de structure intégrée micro programmable (DSP).

- les filtres analogiques réalisés à partir de composants passifs (résistance, Inductance, condensateur).

II. Signal de la parole

1. définition

Le traitement de la parole est une discipline scientifique localisé au croisement du traitement du signal numérique et du traitement du langage.

2. La production de la parole

2.1. Définition

La production da la parole est une action volontaire et coordonnée d'un certain nombre de muscles du système articulatoire.

L'appareil respiratoire fournit l'énergie nécessaire à la production des sons, en poussant l'air à travers l'appareil phonatoire, celui-ci est composé essentiellement du larynx qui contient les cordes vocales qui sont l'élément important, différentes cavités (la bouche, le pharynx, le nez) et différents muscles ou mécanismes qui contrôlent la forme et l'occlusion de ces cavités (la langue, la mâchoire, la luette, les lèvres).

2.2 Appareil phonatoire

L'appareil phonatoire est composé principalement de trois éléments qui contribuent ensemble à la production de la parole. Ces éléments dont le contrôle et la coordination sont assurés par le système nerveux central, sont :

Les poumons, Le larynx Le conduit vocal

a) Les poumons

Ils fournissent l'énergie (l'air) nécessaire à la production du son

b) Le larynx

Son rôle est la production des sons. C'est un ensemble de cartilages articules comprenant les deux "cordes vocale ". ces dernières sont des organes vibratoires constituées de tissu musculaire et de tissu conjonctif résistant. *[1.7]* (Figure 1.8)

c) Le conduit vocal

C'est le conduit entre le larynx et les lèvres, il est composé de plusieurs cavités reliées entre elles. On retrouve la cavité pharyngale (le pharynx), la cavité nasale (les fosses nasales), la cavité buccale (la bouche) et la cavité labiale (les lèvres).(figure 1.7)

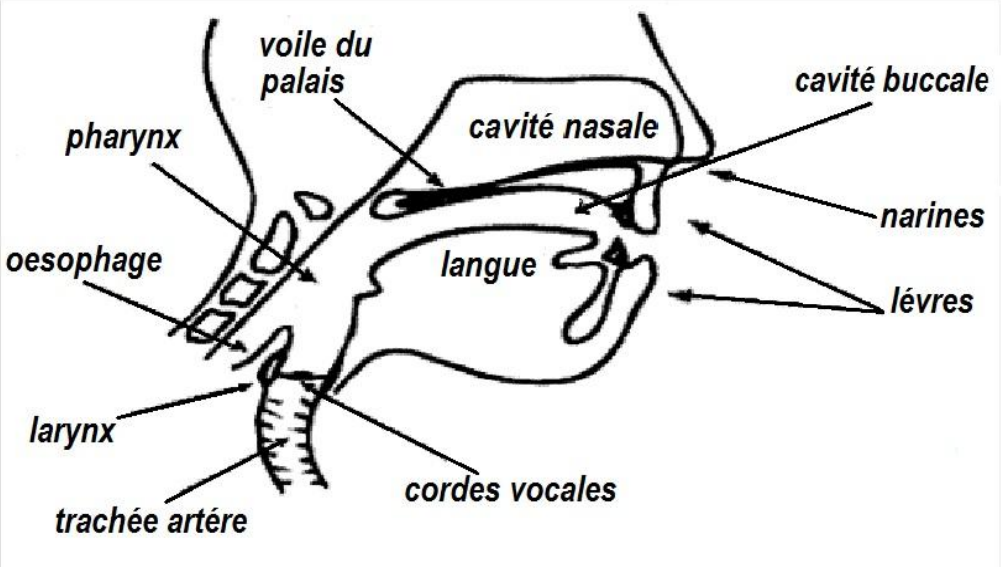


Figure 1.7: *l'appareille phonatoire*

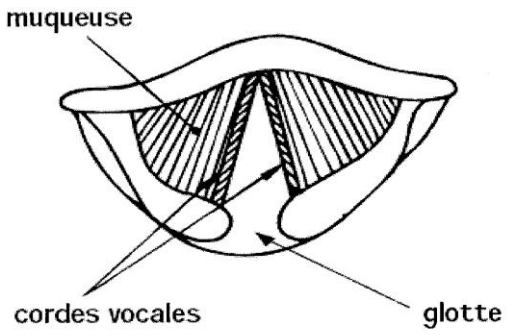


Figure 1.8 *: le larynx*

3. Sons voises et non-voises

Les différents sons de la parole sont classés en deux catégories principales selon que les cordes vocales vibrent ou ne vibrent pas :

3.1. Sons voisés

Pendant certains sons, la glotte s'ouvre brusquement libérant ainsi la pression accumulée en amont sous forme d'impulsions périodiques, ces impulsions mettent les cordes vocales en vibration quasi-périodique.

Le spectre d'un son voisé présente des raies correspondantes à l'harmonique du fondamental (structure de pitch) c'est le cas des voyelles, l'enveloppe de ces raies présente des maximums appelés les formants, les trois ou quatre premiers formants sont essentiels pour caractériser le spectre vocal

3.2. Sons non-voisés :

Si les cordes vocales sont écartées; soit une turbulence quasi aléatoire d'air est produite dans le conduit vocal par diminution de sa section, soit le conduit est momentanément fermé complètement pour augmenter la pression et rouvert Instantanément produisant une transitoire décroissante. Les sons ainsi produits sont appelés sons non voisés. Le son non voisé ne présente pas une structure périodique, il peut être considéré comme un bruit blanc, ainsi son spectre ne présente pas une structure de pitch *[1.8]*

4. Le phonème

Le phonème est la plus petite unité présente dans la parole et susceptible par sa présence de changer la signification d'un mot. Le nombre de phonème est toujours très limité, La notion de phonème ne tint compte que des caractéristiques acoustiques qui permettent une distinction entre des mots, elle ne tint pas compte des phénomènes physiques de la production du son *[1.7], [1.9]*

5. Perception de la parole

Le signal de parole est un vecteur acoustique porteur d'informations d'une grande complexité, variabilité et redondance. Les caractéristiques de ce signal sont appelées traits acoustiques. Chaque trait acoustique a une signification sur le plan perceptuel.

Le premier trait est la fréquence fondamentale f0, c'est la fréquence de vibration des cordes vocales. La connaissance de la fréquence f0 est nécessaire pour faire la différence entre la voix d'un homme, d'une femme ou d'un enfant.

Le deuxième trait est le spectre fréquentiel qui dépend principalement le timbre de la voix. Le timbre est une caractéristique permettant d'identifier une personne à la simple écoute de sa voix.

Le timbre dépend de la corrélation entre la fréquence fondamentale et les harmoniques qui sont les multiples de cette fréquence.

Le dernier trait acoustique est l'énergie correspondant à l'intensité sonore. Elle est habituellement plus forte pour les segments voisés de la parole que pour les segments non voisés

6. Le système auditif

Les vibrations mécaniques du signal sont converties en impulsion nerveuse du nerf auditif par les cellules ciliées au niveau de la cochlée.

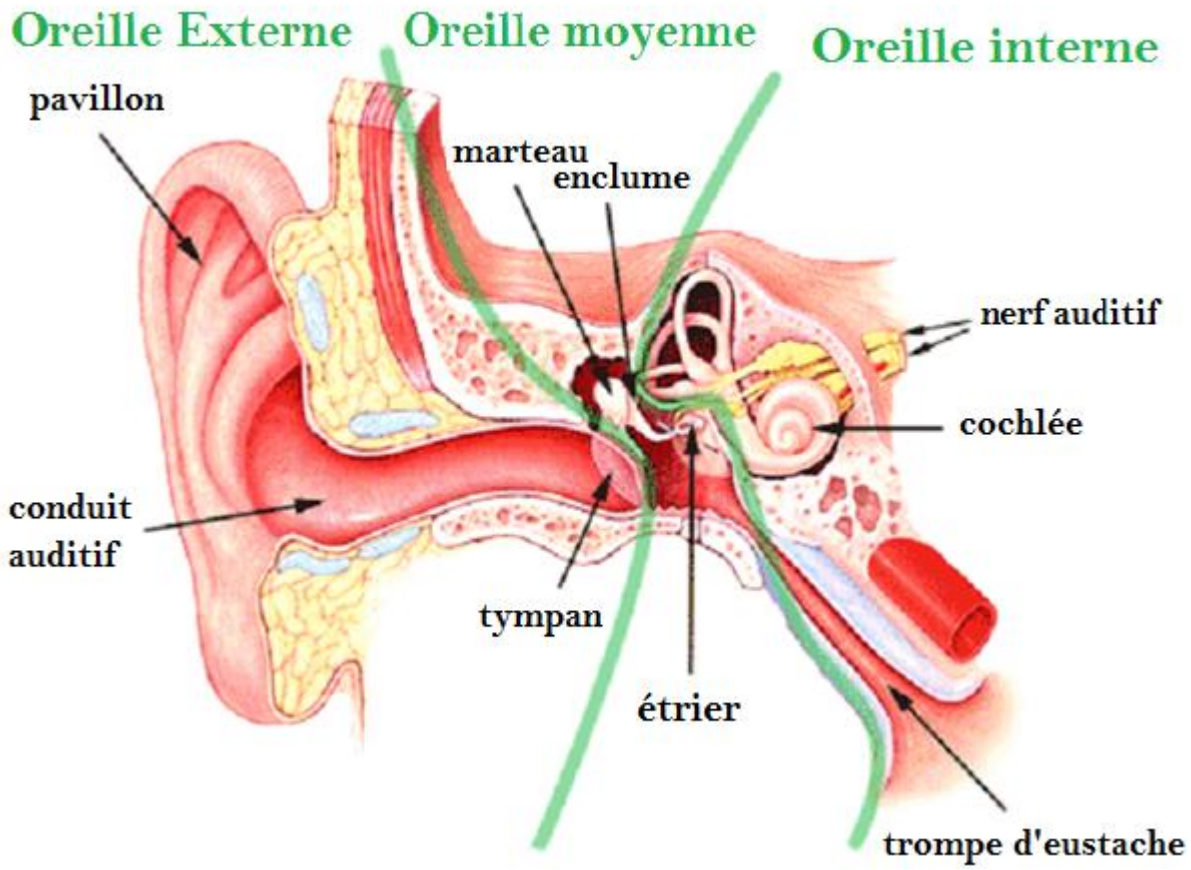


Figure 1.9 : *Le système auditif*

6.1. Anatomie de l'oreille humaine

L'oreille est composée de trois grandes parties :

a)L'oreille externe

Composée d'un pavillon et du conduit auditif. (vibrations aériennes)

Pavillon : Protection et Amplification du signal acoustique

Conduit auditif : possède une fréquence de résonnance dans une plage de fréquence correspondant à la voix humaine

b) L'oreille moyenne

Composée du tympan et des trois petits osselets (marteau, enclume et étrier). (vibrations mécaniques)

Toutes les composantes de l'oreille moyenne ont une fonction d'Adaptation d'impédance.

c) L'oreille interne

Composée de la cochlée (ou limaçon), départ du nerf auditif vers le cerveau. (vibrations liquidiennes)

Cochlée : Filtrage / Analyse du signal et conversion en flux nerveux.

Nerf auditif **:** transport du flux nerveux vers les neurones du cortex auditif pour procéder au traitement de l'information. (Figure 1.9)

6.2. Analyse fréquentielle

Il y a environ 25000 cellules ciliées qui sont réparties au niveau de la cochlée,
Une cellule ciliée «vibre» à une certaine fréquence dite de résonance qui dépend de la position sur la cochlée.

L'oreille effectue donc une sorte d'analyse en fréquence du signal acoustique et
La transformation en impulsion nerveuse est sensible à la fréquence mais est insensible à la phase.

6.3. Réponse en fréquence de l'oreille

• L'oreille ne répond pas de manière égale à toutes les fréquences.

• La limite supérieure en fréquence est d'environ 16000-20000 Hz

• Fréquence d'échantillonnage Fe= 2 Fmax= 40000 Hz

7. Conclusion

À travers ce chapitre on a réalisé une présentation générale sur le traitement de signal, on a étudié la parole comme un type du signal, son analyse et sa modélisation. On a présenté les différents aspects physiques du signal de la parole.

Dans le chapitre suivant on va traiter la reconnaissance des formes et les réseaux de neurones artificiels comme un modèle de calcul de classification.

Chapitre II
La reconnaissance des formes

I .Reconnaissance Des Formes

1. Introduction

La reconnaissance de formes (RdF) est un domaine majeur de l'informatique dans lequel les recherches sont particulièrement actives. Il existe en effet un très grand nombre d'applications qui peuvent nécessiter un module de reconnaissance notamment dans les systèmes de traitement visant à automatiser certaines tâches que l'homme fait manuellement. Dans ce chapitre, on va présenter en premier lieu des notions de base sur la reconnaissance de formes (généralité et processus de RdF) et les différentes méthodes appliquées pour réaliser un système de reconnaissance de formes.

2. Définition

La reconnaissance des formes (ou parfois reconnaissance de motifs) est un ensemble de techniques et méthodes visant à identifier des motifs à partir de données brutes afin de prendre une décision dépendant de la catégorie attribuée à ce motif. On considère que c'est une branche de l'intelligence artificielle qui fait largement appel aux techniques d'apprentissage automatique et aux statistiques.

Les formes ou motifs à reconnaître peuvent être de nature très variée. Il peut s'agir de contenu visuel (code barre, visage, empreinte digitale...) ou sonore (reconnaissance de parole), d'images médicales (rayon X,...) ou multi spectrales (images satellitaires) et bien d'autres. *[2.1]*

3. Domaines d'application

Nous pouvons citer certains de ces domaines d'application :

- Reconnaissance des formes sur signaux temporels

- Signal de parole : reconnaissance de la parole (qu'à t on dit ?), reconnaissance du locuteur (qui a parlé ?)

- Signaux biomédicaux : électrocardiogramme, ...

- Surveillance d'instruments, diagnostic de panne.

- Reconnaissance des formes dans les images numériques

- Lecture automatique de caractères

- Reconnaissance d'empreintes digitales

- Analyse de scènes, interprétation d'images. *[2.2]*

4. Le principe de la RDF

La RDF est la réduction méthodique d'information. A partir d'une donnée très riche, par exemple un signale numérisée, on veut obtenir une information pertinent qui tient en quelque bits, on considère donc souvent la reconnaissance des formes comme un problème de classification.

4.1.Prérequis

Le principe de reconnaissance des formes utilise des termes précis pour définir les éléments formels .Voici la signification de ces principaux termes

- **Observation :** En reconnaissance des formes, elle va se représenter par un vecteur de caractéristique.
- **Caractéristique** : tout se qui permet de rendre un objet distinctif (signes)
- **Classe** : Ensemble d'entités possédant des caractères communs.
- **Classer** : Classer consiste à reconnaitre un nombre d'une classe. *[2.3]*

5. Schéma général de la reconnaissance des formes

La RDF, au sens large du terme, est une discipline vieille. Elle pose la question de l'intelligence humaine et des possibilités de l'intelligence artificielle, la reconnaissance automatique de la parole et plus généralement la RDF sont des domaines de recherche actifs depuis la fin des années soixante.

La RDF consiste à attribuer automatiquement une étiquette à une forme présente dans un signal. D'une autre manière, la reconnaissance automatique de la parole a pour objet de convertir des signaux qui sont compréhensibles par l'homme, en un code interprétable par un ordinateur. *[2.4]*

Le schéma général de RDF est présenté sur la figure 2.1.

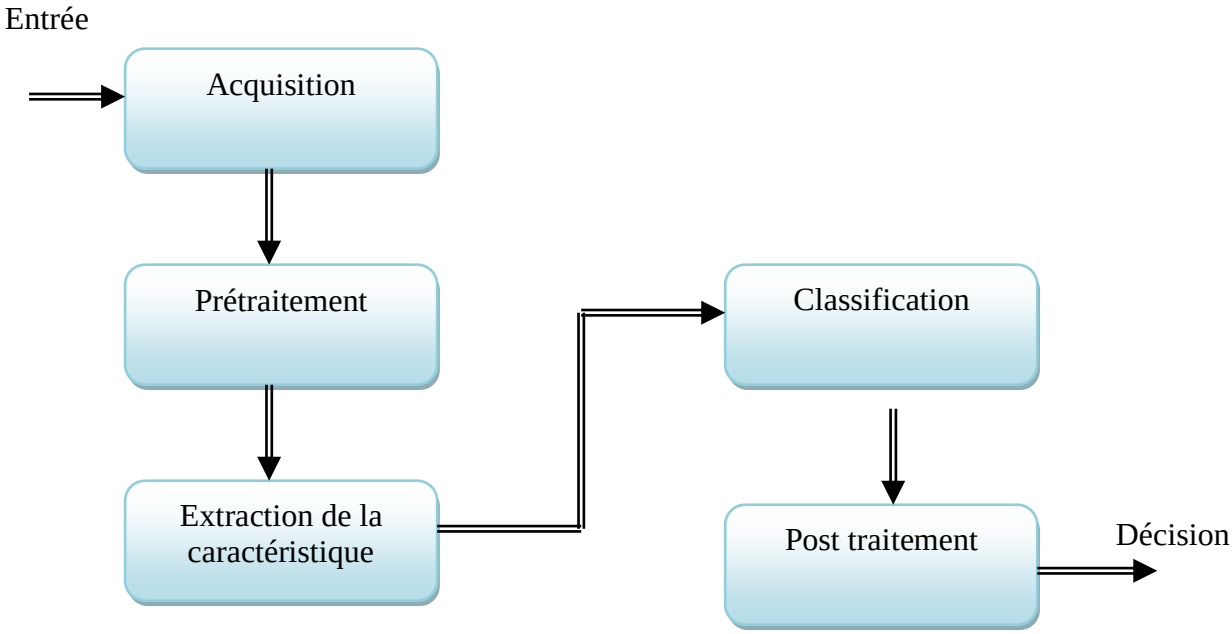


Figure 2.1 : *Schéma globale d'un processus de RDF*

5.1. L'acquisition ou la détection

Il faut d'abord acquérir l'information originale (la forme à reconnaître) à l'aide des capteurs physiques, et de la convertir à des grandeurs numériques, pour en mettre dans une chaîne de traitements informatisés.

Dans le cas où l'information est disponible, le capteur physique c'est le microphone.

5.2. Le prétraitement

L'objectif des prétraitements est de faciliter la caractérisation de la forme. C'est l'étape de la préparation des données pour celle d'extraction des caractéristiques, elle agit sur deux axes, l'un est de réduire le bruit et les déformations ou de l'éliminer carrément s'il est possible et l'autre est de ne maintenir que les caractéristiques significatives de la forme.

5.3. Extraction des caractéristiques

L'objectif de l'extraction des caractéristiques dans le domaine de la reconnaissance consiste à exprimer les primitives sous une forme numérique ou symbolique appelée codage. Selon le cas, les valeurs de ces primitives peuvent être réelles, entières ou binaires.

La difficulté ici est de trouver de bonnes caractéristiques, permettent aux classifier de reconnaître facilement les différentes classes d'objets, on dit alors qu'elles sont discriminantes.

Elles doivent aussi être invariantes à certaines transformations (la lettre a appartient à la même classe quelle que soit sa taille)

5.4. Classification

La classification est l'élaboration d'une règle de décision qui transforme les attributs caractérisant les formes en appartenance à une classe (passage de l'espace de codage vers l'espace de décision).

Le type d'une méthode de classification se décline généralement en deux familles : le mode supervisé et le mode non supervisé. Si l'on dispose d'un ensemble de points étiquetés, on parlera de classification supervisée.

Dans le cas contraire, on effectue une classification non supervisée appelée également classification automatique. La RDF, ainsi définie, est l'apprentissage ou la découverte de structures appelées classe dans un ensemble de données éventuellement perturbées.

5.5. Post-traitement

L'étape ultime du processus de RdF, appelée post-traitement, regroupe toutes les actions à prendre lors de la classification ou non de l'objet à classer (évaluation d'un seuil de confiance, décision de classer ou de rejeter un objet), en utilisant d'autres informations de haut niveau (lexicales, syntaxiques ou sémantiques) pour sélectionner une solution.

6. Les Approches de la Reconnaissance des Formes

Il n'y a pas une théorie unifiée de RdF cependant, il existe deux approches principales: L'approche basée sur la théorie statistique de la décision et l'approche structurelle.

6.1 Approche statistique

Une approche classique en RdF est fondée sur l'étude statistique des mesures que l'on a effectuées sur les objets à reconnaitre. L'étude de leur répartition dans un espace métrique et la caractérisation statistique des classes permet de prendre une décision de reconnaissance du type « plus forte probabilité d'appartenance à une classe ». Ces méthodes s'appuient en général sur des hypothèses concernant la description statistique des familles d'objets analogues dans l'espace de représentation.

6.2 Approche structurelle

Si l'approche statistique permet de se placer dans un cadre mathématique solide et général, elle présente néanmoins le défaut d'oublier la nature des mesures qui sont faites sur les formes et de les traiter de façon abstraite. On conçoit cependant qu'il est plus simple et plus riche d'utiliser

des paramètres descriptifs liés à la nature même des formes étudiées. Si l'on possède une technique de suivi de contour (ce qu'il est facile d'imaginer dans les images à deux niveaux de gris) et un détecteur de segment, on voit que tous les triangles par exemple peuvent se décrire par une caractéristique du type : « la frontière est composée d'un segment horizontal de longueur l, suivi d'un segment oblique de longueur à peu près l, qui se termine au point de départ de premier».

7. Comparaison des méthodes

Méthodes statistique	Méthodes structurelle
- approche quantitative - problème d'optimisation - appliquées aux formes élémentaires - moyennement sensible au bruit	- approche qualitative - validation de propriétés - appliquées aux formes complexes - pas du tout ou très sensible au bruit

Table 2.1 : *Comparaison entre les méthodes statique et méthodes structurelles*

8 Les difficultés de la RDF

8.1 La diversité des formes

Une des principales difficultés de la reconnaissance des formes est la variabilité des formes. En effet, si nous prenons le cas d'une forme de signal, il existe de nombreuses formes différentes qui correspondent pourtant à un même signal, voir la figure 2.2. *[2.5]*

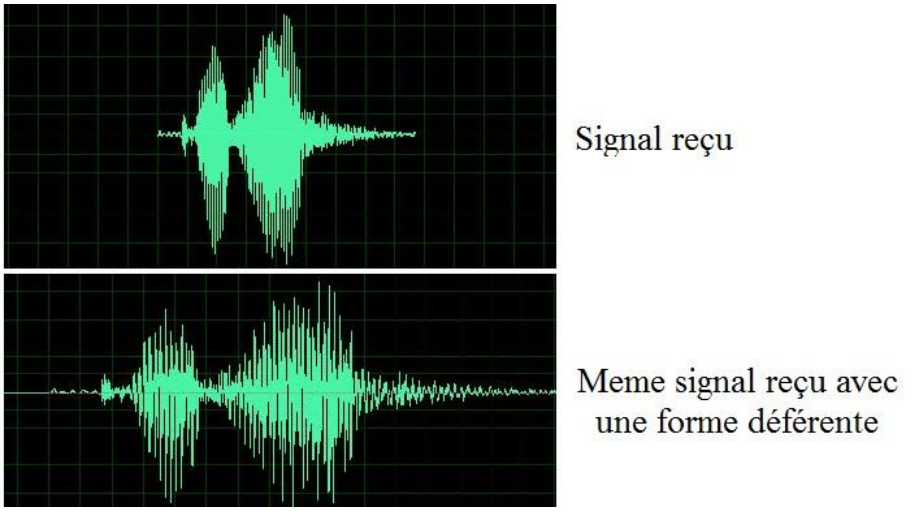


Figure 2.2 : *La diversité des formes*

8.2 La déformation et le bruit:

La seconde difficulté de la RdF est la distorsion due au bruit et à la déformation.

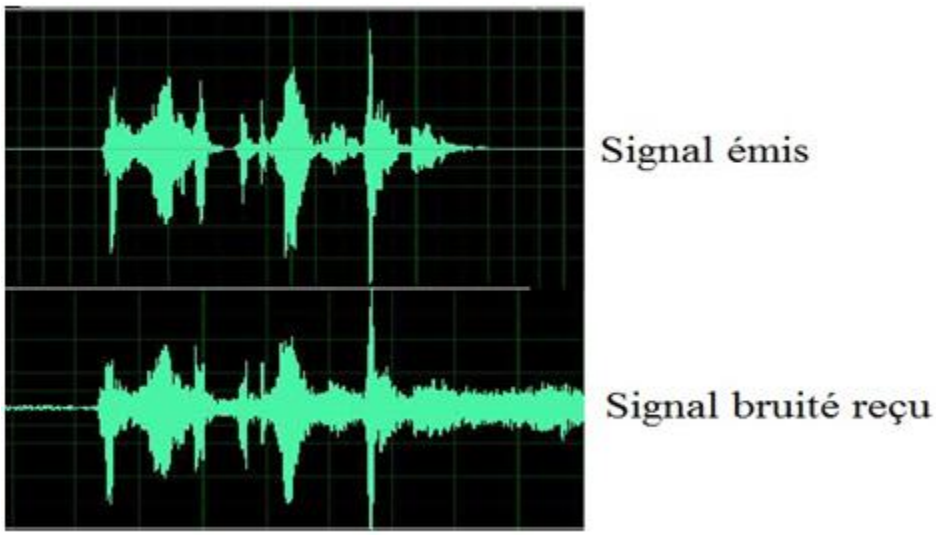


Figure 2.3 : *La déformation et le bruit d'un signal*

II les réseaux de neurones artificiels

1. introduction

Les réseaux de neurones artificiels ou réseaux connexionnistes sont fondés sur des modèles qui tentent d'expliquer comment les cellules du cerveau et leurs interconnexions parviennent, d'un point de vue global, à exécuter des calculs complexes.

Ces systèmes qui stockent et retrouvent l'information de manière "similaire" au cerveau sont particulièrement adaptés aux traitements en parallèle de problèmes complexes comme la reconnaissance automatique de la parole, la reconnaissance de visages ou bien la simulation de fonctions de transfert. Ils offrent donc un nouveau moyen de traitement de l'information utilisé en reconnaissance de formes (vision, image, parole, etc.).

2. neurone biologique

2.1. Structure d'un neurone biologique

On pense que le système nerveux compte plus de 1000 milliards de neurones interconnectés. Bien que les neurones ne soient pas tous identiques, leur forme et certaines caractéristiques permettent de les répartir en quelques grandes classes. En effet, il est aussi important de savoir, que les neurones n'ont pas tous un comportement similaire en fonction de leur position dans le cerveau. Avant de rentrer plus en avant dans les détails, examinons un neurone. *[2.6], [2.7]*

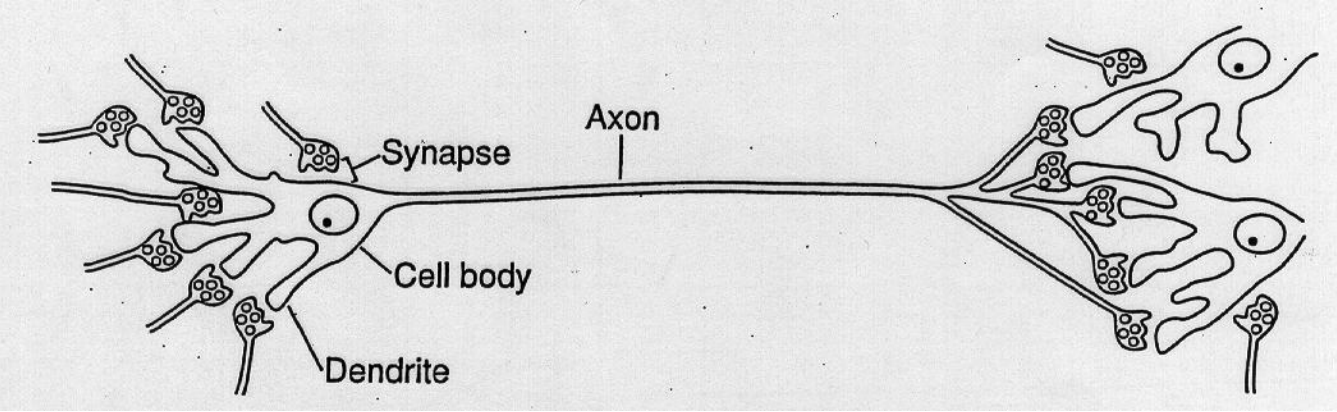


Figure2.4 : *Le* comportement biologique du système neuronal

3. Les neurones artificiels

Formellement un réseau de neurones artificiels est un graphe dont les nœuds sont des unités de calcul appelés neurones formels, et les arêtes représentent les liens synaptiques. *[2.8], [2.9]*

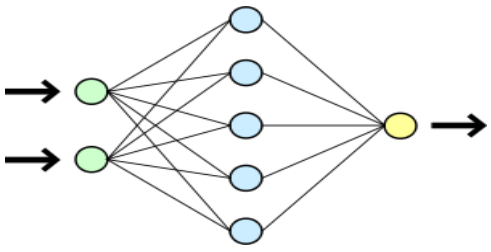


Figure2.5 : un réseau multi couches

3 .1. Structure du neurone artificiel

La figure 2.6 montre la structure d'un neurone artificiel. Chaque neurone artificiel est un processeur élémentaire. Il reçoit un nombre variable d'entrées en provenance de neurones amonts. A chacune de ces entrées est associé un poids w abréviation de poids (weight en anglais) représentatif de la force de la connexion. Chaque processeur élémentaire est doté d'une sortie unique, qui se ramifie ensuite pour alimenter un nombre variable de neurones avals. A chaque connexion est associé un poids.

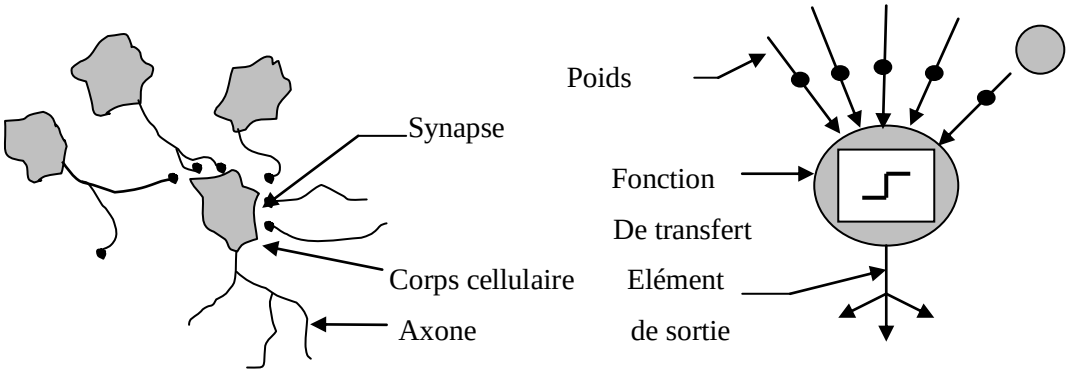


Figure 2.6 : *Mise en correspondance neurone biologique / neurone artificiel*

On donne les notations que nous utilisons par:

Pour le neurone d'indice i, les entrées sur celui-ci sont de poids wij alors que les connexions avals sont de poids wki.

Un neurone artificiel est une fonction non linéaire paramétrée.

Un neurone est considéré comme un élément élémentaire de traitement et comme la figure 2.7 nous démontre qu'il reçoit les entrées et produit un résultat à la sortie :

$$s = \sum_{i=1}^{N} w_i x_i + \theta \;=\; W'X + \theta$$

$$y = f(s)$$

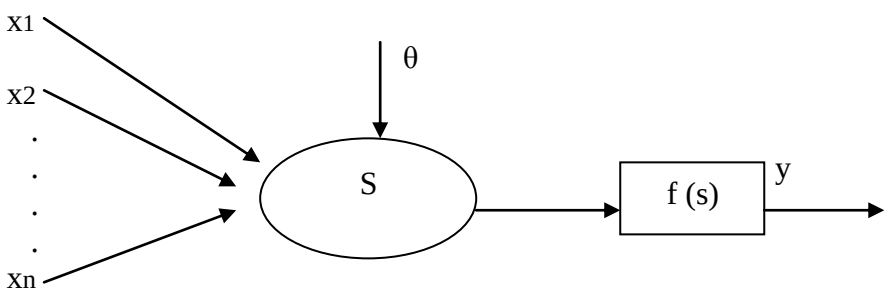


Figure 2.7 : Représentation d'un réseau de neurone artificiel

x1, x2, ..., xn sont les entrées qui proviennent de l'environnement externe au réseau ou d'autres neurones. y est la sortie et w1, w2, ..., wn sont les poids associés à chaque connexion. X est le vecteur d'entrée, W' est le vecteur poids, $\boldsymbol{\theta}$ est appelé le biais. La fonction f est appelée fonction d'activation.

Il existe plusieurs fonctions d'activation:

- Fonction à seuil f(s) = sign(s) = +1 si s > 0

  f(s)=-1si s≤ 0.

- $$f(s) = \frac{1}{1- e^{x}}$$
- **Fonction sigmoïde**
- **Function tanh(*s*)**
- **...**

Un réseau réalise une ou plusieurs fonctions algébriques de ses entrées, par composition des fonctions réalisées par chacun des neurones. La capacité de traitement de ce réseau est stockée sous forme de poids d'interconnexions obtenus par un processus d'apprentissage à partir d'un ensemble d'exemples d'apprentissage.

On distingue deux phases. La première est habituellement le calcul de la somme pondérée des entrées (a) selon l'expression suivante :

$$a = \sum(w_i e_i)$$

À partir de cette valeur, une fonction de transfert calcule la valeur de l'état du neurone. C'est cette valeur qui sera transmise aux neurones avals. Il existe de nombreuses formes possibles pour la fonction de transfert présenté sur la figure 2.8. On remarquera qu'à la différence des neurones biologiques dont l'état est binaire, la plupart des fonctions de transfert sont continues, offrant une infinité de valeurs possibles comprissent dans l'intervalle ([0, +1] ou [-1, +1]). *[2.10]*

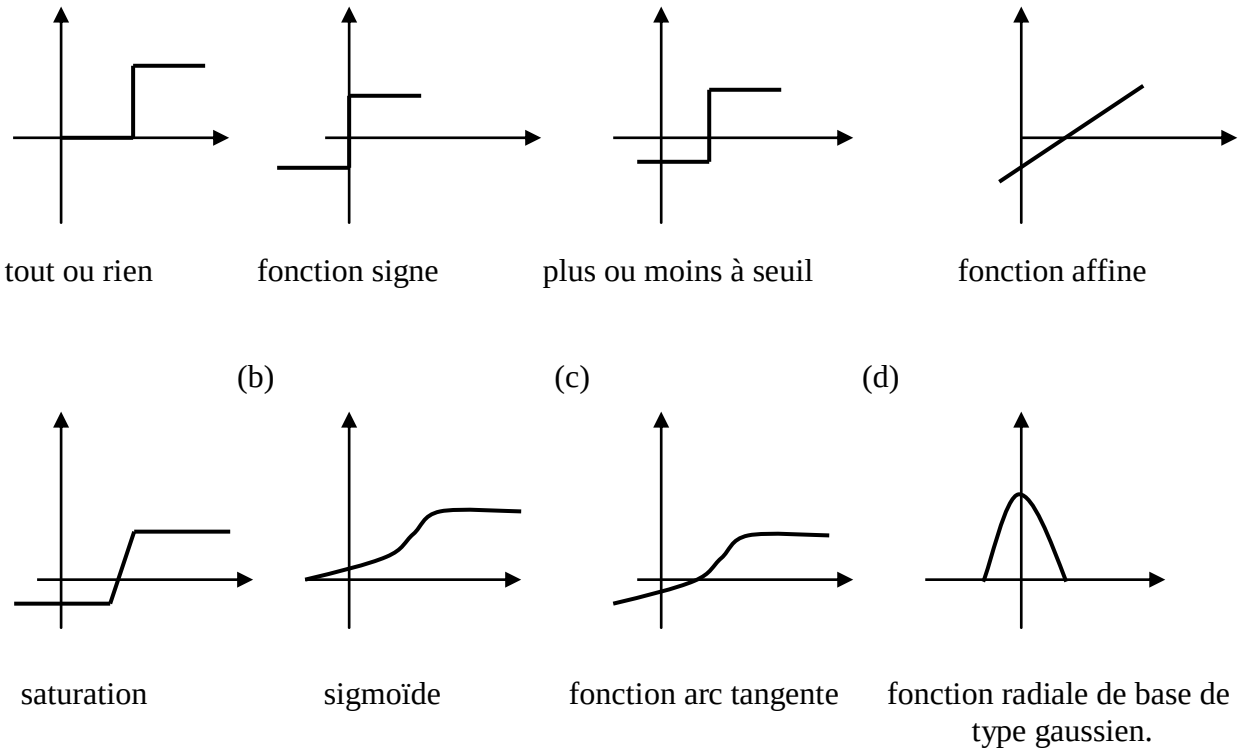


Figure 2.8 : *Différents types de fonctions de transfert pour le neurone artificiel*

Nous constatons que les équations décrivant le comportement des neurones artificiels n'introduisent pas la notion de temps. En effet, et c'est le cas pour la plupart des modèles actuels de réseaux de neurones, nous avons affaire à des modèles à temps discret, synchrone, dont le comportement des composants ne varie pas dans le temps

3.2. L'apprentissage

L'apprentissage est une phase du développement d'un réseau de neurones durant laquelle le comportement du réseau est modifié jusqu'à l'obtention du comportement désiré. L'apprentissage neuronal fait appel à des exemples de comportement.

L'apprentissage peut être considéré comme le problème de la mise à jour des poids des connexions au sein du réseau afin de réussir la tâche qui lui est demandée. L'apprentissage est la caractéristique principale de RNA, il peut se faire de différentes manières et selon de différentes règles.

On distingue trois grandes classes d'algorithmes d'apprentissage :

- L'apprentissage supervisé.
- L'apprentissage semi supervisé.

- L'apprentissage non supervisé.

3.2.1. Types d'apprentissage :

a) L'apprentissage supervisé :

Dans ce type d'apprentissage, le réseau s'adapte par comparaison entre le résultat qu'il a calculé, en fonction des entrées fournies, et la réponse attendue en sortie. Ainsi, le réseau va se modifier jusqu'a ce qu'il trouve la bonne sortie, c'est-à-dire celle attendue, correspondant à une entrée donnée.

b) L'apprentissage semi supervisé

L'utilisateur ne possède que des indications imprécises (par exemple, échec /succès de réseau) sur le comportement finale désiré.

c) L'apprentissage non supervisé

Dans ce cas, l'apprentissage est basé sur des probabilités. Le réseau va se modifier en fonction des régularités statistiques de l'entrée et établir des catégories, en attribuant et en optimisant une valeur de qualité, aux catégories reconnues.

3.2.2. Principales règles d'apprentissages

Les stratégies de modification des poids synaptiques sont dérivées des règles générales suivantes :

- La règle de poids des connexions entre deux processeurs élémentaires est renforcée si les deux processeurs élémentaires sont activés simultanément.
- La règle delta, règle ou le poids synaptique est adapté pour obtenir la diminution de l'erreur entre la sortie réelle du processeur élémentaires et la sortie désirée

4. L'algorithme de la descente du gradient

L'algorithme d'optimisation le plus simple est la descente de gradient. Dont le principe est de partir d'un point aléatoire puis de se déplacer dans la direction de la plus forte pente. En appliquant un certain nombre d'itérations converge vers une solution qui est un minimum local de f (Figure 2.9).

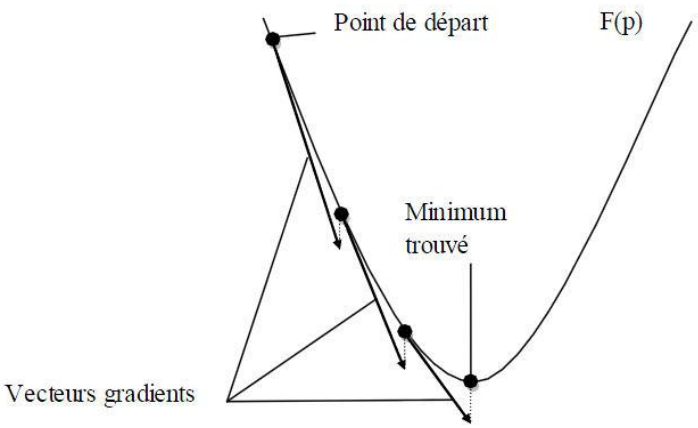


Figure 2.9 : *illustration du principe de la descente de gradient*

4.1. Description générale de l'algorithme

Soient le vecteur des entrées x et le vecteur des coefficients synaptiques w.

La sortie vaut alors : o = ϭ (x.w) = ϭ (x0.w0 + … + xn.wn),

ϭ étant une fonction de transfert continue et dérivable.

Posons : y = x.w

Soit S la base d'apprentissage composée de couples (x, c), où c est la sortie attendue pour x.

On définit ainsi l'erreur sur le réseau pour la base d'apprentissage S :

$E(w) = 1/2 \sum[(x,c) \text{ dans } S] (c - o)^2$

Problème : trouver w qui minimise E(w).

Algorithme :

. Initialiser aléatoirement les coefficients wi.

. Répéter :

. Pout tout i :

. Δwi = 0

.Fin Pour

.Pour tout exemple (x, c) dans S

. Calculer la sortie o du réseau pour l'entrée x

. Pout tout i :

. Δwi = Δwi +ε*(c - o)* xi * ϭ(x.w)

. Fin Pour

.Fin Pour

.Pour tout i :

. wi = wi +Δ wi

.Fin Pour

.Fin Répéter

5. Propriété fondamentale des Réseaux de Neurones

La propriété fondamentale des RdN remarquable qui est à l'origine de leur intérêt pratique dans des domaines très divers est « l'approximation parcimonieuse ». Cette expression traduit deux propriétés distinctes. D'une part, les RdN sont des approximateurs universels, et, d'autre part, une approximation à l'aide de RdN nécessite, en général, moins de paramètres ajustables (les poids des connexions) que les approximateurs usuels. Plus précisément, le nombre de poids varie linéairement avec le nombre de variables de la fonction à approcher, alors qu'il varie exponentiellement (beaucoup plus rapidement) avec la dimension de l'espace des entrées dans le cas de la plupart des autres approximateurs usuels.

En vertu de cette propriété « l'approximation parcimonieuse », le comportement de tout système statique peut être approché par un RdN non bouclé, et celui de tout système dynamique par un réseau bouclé. La propriété d'approximation parcimonieuse RdN d'excellents outils pour la résolution des problèmes (il faut s'assurer d'avoir bien posé le problème) de modélisation et de classification.

6. Topologies de Réseaux de Neurones Artificiels

La topologie des réseaux de neurones artificielle peut aller d'une connectivité totale à une connectivité locale ou les neurones ne sont liés qu'à leur plus proche voisin. Il est courant d'utiliser des réseaux à structure régulière pour faciliter leur utilisation.

6.1. Les Réseaux à Connexions Complètes

C'est la structure d'interconnexion la plus générale. Chaque neurone est connecté à tous les neurones du réseau (et à lui-même).

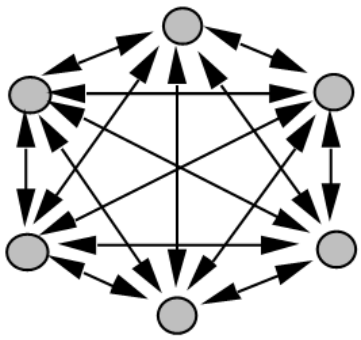


Figure 2.10 : Réseaux à Connexions Complètes

6.2. Les Réseaux à Connexions Récurrentes

Les connexions récurrentes ramènent l'information en arrière par rapport au sens de propagation défini dans un réseau multicouche. Ces connexions sont le plus souvent locales.

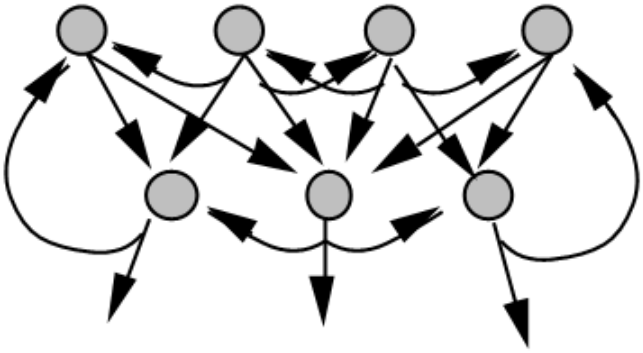


Figure 2.11 : *Réseaux à Connexions Récurrentes*

6.3. Les Réseaux à Connexions Locales

Il s'agit d'une structure multicouche, mais qui à l'image de la rétine, conserve une certaine topologie. Chaque neurone entretien des relations avec un nombre réduit et localisé de neurones de la couche avale (***Figure 2.12***). Les connexions sont donc moins nombreuses que dans le cas d'un réseau multicouche classique

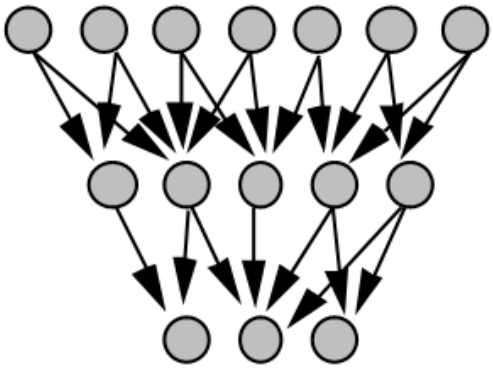


Figure 2.12 : *Les Réseaux à Connexions Locales*

6.4. Les Réseaux Multicouches ou MLP

Les neurones sont arrangés par couche. Il n'y a pas de connexion entre neurones d'une même couche et les connexions ne se font qu'avec les neurones des couches avales ***(fig. 2.13).*** Habituellement, chaque neurone d'une couche est connecté à tous les neurones de la couche suivante et celle-ci seulement. Ceci nous permet d'introduire la notion de sens de parcours de

l'information (de l'activation) au sein d'un réseau et donc définir les concepts de neurone d'entrée, neurone de sortie. Par extension, on appelle couche d'entrée l'ensemble des neurones d'entrée,

couche de sortie l'ensemble des neurones de sortie. Les couches intermédiaires n'ayant aucun contact avec l'extérieur sont appelés couches cachées.

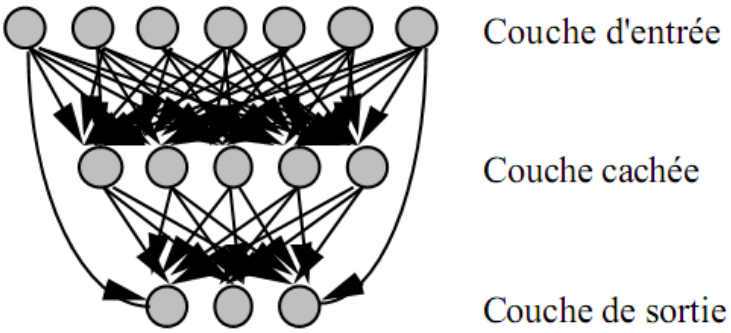


Figure 2.13 : *Réseaux Multicouches ou MLP*

7. Modèles des réseaux de neurones

7.1. Les réseaux de Hopfield

Les réseaux de Hopfield sont des réseaux récurrents et entièrement connectés, dans ce type de réseau, chaque neurone est connecté à chaque autre neurone et il n'y a aucune différenciation entre les neurones d'entrée et de sortie. Ils fonctionnent comme une mémoire associative non-linéaire et sont capables de trouver un objet stocké en fonction de représentations partielles ou bruitées.

L'application principale des réseaux de Hopfield est l'entrepôt de connaissances mais aussi la résolution de problèmes d'optimisation. Le mode d'apprentissage utilisé ici est le mode non-supervisé.

7.2. Les réseaux de Kohonen

Ce sont des réseaux à apprentissage non-supervisé qui établissent une carte discrète, ordonnée typologiquement, en fonction de patterns d'entrée. Le réseau forme ainsi une sorte de treillis dont chaque nœud est un neurone associé à un vecteur de poids. La correspondance entre chaque vecteur de poids est calculée pour chaque entrée. Par la suite, le vecteur de poids ayant la meilleure corrélation, ainsi que certains de ses voisins, vont être modifiés afin d'augmenter encore cette corrélation.

8. conclusion

Les réseaux de neurones sont depuis quelque temps un point de focalisation des médias, du public et des scientifiques. Les travaux menés dans le domaine des sciences de la cognition artificielle ont été marquées par quelques apports non négligeables mais surtout par beaucoup d'optimisme.

Les années qui viennent concrétiseront cet optimisme ou bien relègueront cette technique parmi les nombreuses " recettes " informatiques.

Dans ce chapitre, nous avons présenté l'essential sur la reconnaissance des formes et les réseaux de neurones et on a expliqué le fonctionnement d'un algorithme d'apprentissage adapté aux réseaux multicouches.

Chapitre III
la reconnaissance automatique de la parole

I. la reconnaissance automatique de la parole ou RAP

1. Introduction

La reconnaissance de la parole consiste à faire reconnaître par une machine les mots ou phrases prononcés par un locuteur. C'est une discipline quasi contemporaine de l'informatique.

La reconnaissance automatique de la parole pose de nombreux problèmes d'un point de vue théorique. Leur complexité fait que seuls des sous-problèmes ont pu être à ce jour résolus.

Ces solutions partielles correspondent à des contraintes plus ou moins fortes, et les systèmes existants supposent une coopération plus ou moins grande des utilisateurs. *[3.1]*

2. Domaines d'applications des systèmes RAP

Téléphonie: Téléphonie, Assistance et Services

Embarqué : Automobile, Maison intelligente

Multimédia : Dictée vocale, Logiciels pédagogiques, Jeux vidéos

Médical : Aide aux personnes handicapées, rééducation assistée

***Industriel* :** Biométrie, Contrôle vocal de machines

3. Les techniques de reconnaissance

Pour réaliser des systèmes de reconnaissance, deux approches principales sont utilisées, l'une consiste à reconnaître globalement des mots séparés par des intervalles de silence, et l'autre permet en revanche d'aborder le problème de la reconnaissance de la parole continue éventuellement dans un contexte multilocuteur. Nous allons décrire ces deux approches.

3.1. Méthode globale de RAP

Dans cette approche, le mot est considéré comme une entité globale. Cette méthode est caractérisée par le fait qu'une phase d'apprentissage est nécessaire pendant laquelle l'utilisateur prononce la liste des mots du lexique de son application.

Lors de la phase de reconnaissance, le mot à reconnaître est comparé à tous les mots de références du lexique. Le mot ressemblant le plus au mot prononcé est alors reconnu. Les avantages de cette approche sont d'une part l'indépendance vis-à-vis des particularités de la langue du fait de la

phase d'apprentissage, et d'autre part, l'excellente capacité de reconnaissance pouvant atteindre les 99%. Néanmoins, le vocabulaire est assez limité et les systèmes sont le plus souvent monolocuteur, de plus la prononciation en mots isolés est peu naturelle.

3.2. La méthode analytique

Cette approche tente de détecter et d'identifier des unités élémentaires (phonèmes, syllabes…) puis de reconnaître la phrase effectivement prononcée. La méthode fait apparaître plusieurs modules qui communiquent entre eux. Le module acoustique a pour rôle d'extraire les caractéristiques du signal de parole destinées aux modules phonétiques et prosodiques.

Le module prosodique sert à trouver les informations sur le rythme et l'intonation de la phrase et le module phonétique traduit la liste des indices en une suite d'unité phonétiques. Le module phonologique porte sur les phénomènes de la langue dont le contenu phonétique est modifié par les articulations rapides, les liaisons et les variétés dialectales.

Au niveau du lexique, interviennent les informations sur les mots qui composent la langue

Le module syntaxique renferme les règles de la grammaire qui permettent de décrire et d'analyser la langage en termes grammatical et fonctionnel et permet donc de définir toutes les séquences de mots acceptables.

Le niveau sémantique permet de donner la signification de l'énoncé et le rejet des phrases syntaxiquement correctes n'ayant aucune interprétation. Le module pragmatique, utilisé en dialogue, permet de déterminer le sens de la phrase dans le contexte de l'application et de gérer l'historique du dialogue *[3.2]*

4. Décodage acoustico-phonétique

De toutes les opérations décrites par les différents modules de l'approche analytique, la transformation du signal vocal en une suite d'étiquettes phonétiques est la plus fondamentale. Toute erreur à ce niveau augmente considérablement l'indéterminisme des traitements ultérieurs. Le décodage acoustico-phonétique est lié à deux aspects importants en reconnaissance de la parole: la représentation paramétrique et l'identification phonétique.

4.1. La représentation paramétrique

Elle consiste en la conversion du signal de parole en une représentation linguistique structurée. Son rôle est la réduction des données redondantes et le calcul des paramètres qui permettent de distinguer les phonèmes *[3.3]*. Les principales représentations sont:

4.1.1. La représentation temporelle

Le signal sonore qui constitue le message est préalablement converti par un microphone en un signal électrique, le résultat est une tension alternative dont l'amplitude varie de façon continue en fonction du temps (signal analogique). Pour être exploité par ordinateur, le signal passe par une phase de digitalisation qui le transforme en une suite de nombres mesurant son amplitude à des intervalles de temps successifs très brefs (on parle d'échantillonnage du signal) à partir de quoi on calcule les paramètres: énergie, nombre de passage par zéro, fréquence fondamentale, etc. C'est une représentation sensible au bruit et dépend de la fréquence fondamentale et du déphasage de la voie de transmission.

4.1.2. La représentation fréquentielle

Elle est obtenue en mesurant pendant des intervalles de temps plus larges, les éléments des différentes fréquences qui composent le signal et leurs amplitudes. Elle est assurée en exprimant le signal (de forme compliqué) en une combinaison linéaire de fonctions de base (sinusoïdes ou exponentielles) dont les propriétés sont bien connues et qui sont facilement manipulables.

La représentation fréquentielle la plus utilisée est obtenu par l'application d'une transformée de Fourier sur le signal, et elle se calcule par un algorithme de transformée de Fourier rapide (FFT)

Un spectrogramme est un graphique de ce type de représentation, appelée parfois représentation temps-fréquence-amplitude :

-Abscisse : temps.

–Ordonnée : fréquence.

–Niveau de gris : amplitude.

La représentation fréquentielle permet de réduire d'un facteur dix environ le flux d'information représenté par le signal vocal et d'éliminer les redondances présentes dans celui-ci, elle est d'ailleurs utilisée par le système auditif humain. *[3.4]*

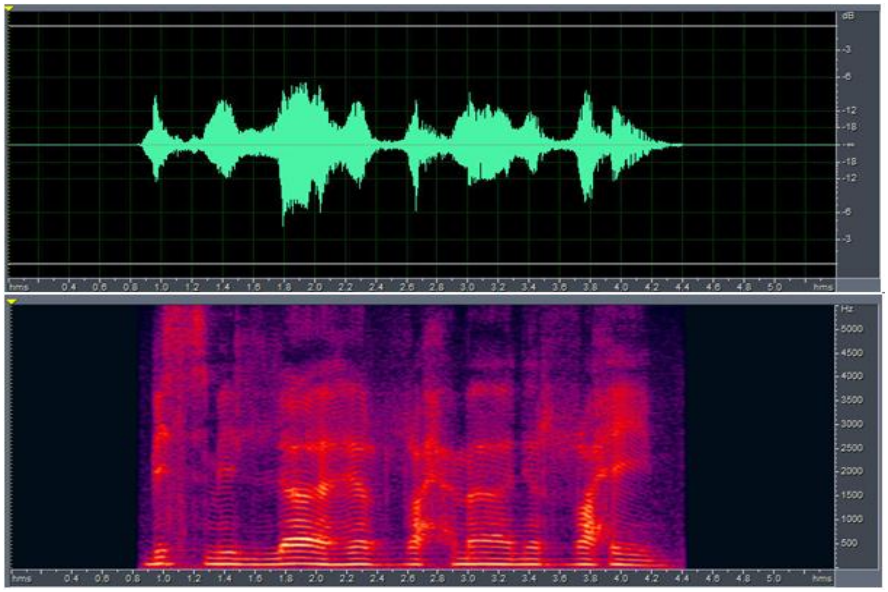


Figure3.1 : *Exemple de spectrogramme*

4.1.3. L'analyse par prédiction linéaire LPC

Abrégée en LPC, comme "Linear Prediction Coding", la méthode a été utilisée pour la transmission de la parole par compression importante de données. Elle fait intervenir le modèle de production de la parole, elle est donc bien adaptée à sa représentation paramétrique. L'extraction des paramètres se fait de manière très rapide dans le domaine temporel. L'idée est d'exprimer le signal à un instant donné comme combinaison linéaire de son passé, tout en essayant de régler les coefficients de combinaisons pour que l'énergie de l'erreur entre le signal réel et le signal prédit soit minimale. Une transformée de Fourier à partir de coefficients LPC permet de rendre plus précises la largeur de bande et la position des formants, mais la LPC présente l'inconvénient de représenter mal les creux dans le spectre et ne convient donc pas bien pour les sons nasalisés.

4.1.4. Étapes de calcul du vecteur caractéristique de types MFCC :

La MFCC (Mel Frequency Cepstral Coefficients) est une extraction de caractéristique du signal développée autour de la FFT et de la DCT sur une échelle de Mel.

La fonction de la MFCC est à peu près identique à celle du LPC. Néanmoins, la MFCC est plus robuste au bruit que le LPC.

Les différentes phases de d'algorithmes sont :

- Phase 1 : Découper le signal en plusieurs fenêtres qui se recoupent entre elles et on applique la MFCC à chaque fenêtre.

- **Phase 2** : Appliquer une fenêtre de Hamming au signal:

$$w(n) = 0.54 + 0.46 . Cos(\frac{2 .\pi.n}{N+1})$$

Par la suite on multiplie cette fonction par le signal à transformer.

- Phase 3 : Appliquer ensuite la FFT à la fenêtre pour en ressortir la magnitude, on obtient donc le spectre.

- Phase 4 : On passe à l'échelle de Mel. En effet, après des études sur l'oreille humaine, il a été montré que l'homme se base sur une échelle fréquentielle spécifique. Pour simuler l'oreille humaine, il faut passer par un Banc Filtre, un filtre pour chaque fréquence que l'on cherche. Ces filtres ont une réponse de bande passante triangulaire. Pour connaitre l'intervalle entre chaque filtre, on utilise une constante: Mel-Frequency interval. .

- Phase 5 : Pour finir, on travaille avec le Cepstre, on convertit le spectre logarithmique de Mel en temps au moyen de la DCT (Discret Cosinus Transform) La formule de cette transformation est donner par la formule suivante :

$$\boldsymbol{C_i = \sum_{j=0}^{M} S(j) cos(i(j - \frac{1}{2})\frac{\pi}{N_f})} \textbf{ pour i= 0,1,..., M} - \textbf{1}$$

M correspond au nombre de coefficients cepestraux .

.$\boldsymbol{N_f}$ désigne le nombre de filtre.

Ainsi, on réduit le nombre de données caractérisant le signal. Voir figure 3.2

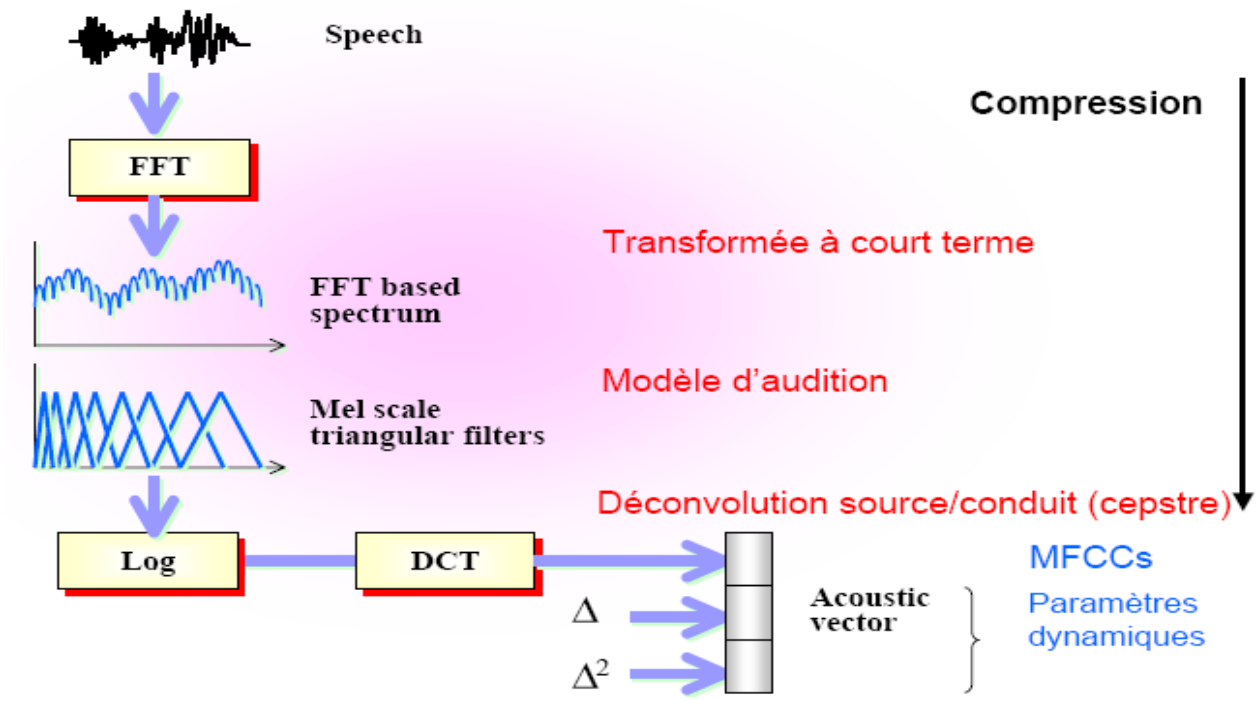


***Figure3.2**: méthode de calcul Mel-Frequency Cepstral Coefficients (MFCCs)*

4.2. Les techniques de décodage phonétique

Les techniques utilisées en décodage phonétique peuvent être classées selon trois approches:

- Une approche de classification automatique avec ou sans apprentissage utilisée pour définir des prototypes.

- Une approche de reconnaissance des formes qui consiste à affecter des étiquettes à des segments grâce à des critères de proximité.

- Une approche basée sur la reconnaissance de traits où l'on retrouve les approches système expert.

4.2.1. La quantification vectorielle

Elle consiste à utiliser les probabilités statistiques des sons dans leur espace de représentation (spectre d'amplitude, coefficients cepstraux ou de prédiction). Cette technique rentre dans la problématique plus générale de la classification automatique. Elle part du postulat que deux formes proches dans leur espace de représentation sont aussi proches en soi. L'utilisation de cette technique pose plusieurs difficultés:

-La représentation correcte des formes de références.

-Le découpage de l'espace en classes pertinentes.

-Le choix de la métrique dans cet espace.

Cette technique nécessite une phase d'apprentissage pour constituer le dictionnaire de référence. L'autre inconvénient de cette méthode est qu'elle permet de faire une série de décisions locales mais sans référence aux échantillons passés sans tenir compte des phénomènes de coarticulation. Son principal avantage est de réduire considérablement le débit par référence à une liste de prototypes connue à priori.

4.2.2. La comparaison dynamique

En utilisant le principe de mise en correspondance optimale, la programmation dynamique permet de tenir compte des distorsions temporelles entre deux formes à comparer.

Chaque mot du lexique est représenté par une suite de vecteurs r(l), ... , r(j), chaque forme à reconnaître est représentée par t(l), ...t(i). II faut trouver un chemin de recalage qui à chaque vecteur de T, fait correspondre un vecteur R. Ce chemin devra prendre en compte les contraintes naturelles de la parole. Ensuite, parmi tous les chemins de recalages possibles, il faut choisir celui dont la somme des distances le long du chemin est minimale.

L'algorithme permet d'éliminer rapidement des références lorsque des différences notables apparaissent au cours d'une étape quelconque de l'examen de la référence à reconnaître.

Les avantages de la méthode sont d'une part son excellente capacité de reconnaissance et son indépendance vis à vis des particularités de la langue. Mais dans ce cas le problème de l'apprentissage et de la segmentation se posent de manière plus aiguë que pour la quantification vectorielle, c'est pourquoi elle est souvent utilisée en aval de la quantification vectorielle et sert à la reconnaissance des mots isolés ou enchaînés. *[3.5]*

5. Les modèles stochastiques

Dans les modèles stochastiques, les formes acoustiques sont représentées par un graphe sous forme d'une chaîne de Markov ou plus précisément par des modèles de Markov cachés HMM (Hidden Markov Model). Le graphe est composé d'un nombre fini d'états représentant les segments stables du signal, tandis que les variations spectrales sont modélisées par les arcs de transition. *[3.6]*.

Si nous considérons un signal acoustique S, le principe de la reconnaissance peut être explique comme le calcule de la probabilité P(W|S) d'une suite de mots (ou phrase) W qui correspond au signal acoustique S, et de déterminer la suite de mots que peut maximiser cette probabilité.

En utilisant la formuler de Bayes, P(W|S) peut s'écrire :

Avec :

$$\mathbf{P\ (W|S) = p(w).\ P(S|W)\ /P(S)}$$

P(w) est la probabilité de la suite de mot W.

P(S|W) est la probabilité du signal acoustique S, étant donné la suite de mot W.

Un modèle HMM est défini par l'ensemble de donnée suivantes :

Une matrice A qui indique la probabilité de transition d'un état q_i vers un autre état ou vers lui-même, soit $P(q_j \mid q_i) = a_{ij}$.

Une matrice B qui indique la probabilité d'émission de l'observation dans chaque état.

Cette probabilité est de type multi gaussienne, définie par les vecteur moyenne, les matrice de covariance et de poids associe a chaque gaussienne.

Une matrice C donne la distripions de départ des états, c'est-à-dire pour chaque état la probabilité d'être atteint à partir de l'état initial q_i .

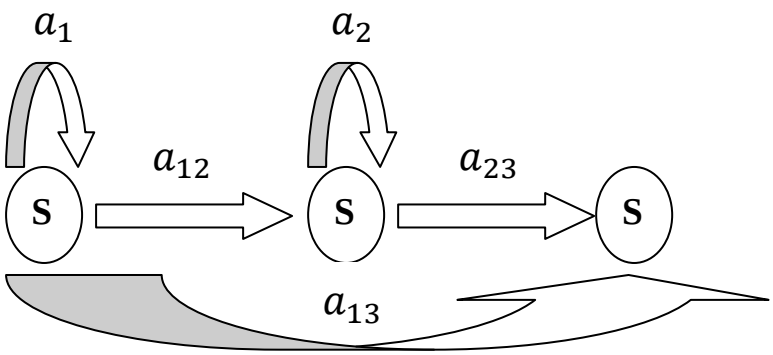


Figure 3.3 : *modèle HMM gauche-droite d'ordre 1 à 3 états*

Cette structure rend possible la représentation des changements dus a la prononciation. En effet pour l'articulation lente, il y a répétition d'état. Il sera représente dans un modèle HMM par une transition d'état sur lui-même (la transition a_{ii} sur la figure précédente).

Alors que pour les articulations rapides, le modèle admet le saut à l'état suivante (transition a_{13}).

6. Les modèles connexionnistes

Ils sont fondés sur une modélisation des réseaux de neurones. Ces derniers possèdent des avantages forts intéressants tels que le parallélisme, le raisonnement à partir de données incomplètes et la capacité de généralisation. Nous assistons actuellement à un regain d'intérêt pour l'utilisation

des modèles connexionnistes en reconnaissance automatique de la parole même s'ils n'ont pas encore prouvé leur supériorité par rapport aux autres méthodes.

Les méthodes décrites ci-dessus sont plutôt des techniques de reconnaissance de formes qui s'avèrent mal adaptées à la reconnaissance d'unités phonétiques instables. L'identification phonétique nécessite la prise en compte du contexte. Un problème qu'on ne peut résoudre avec les méthodes de reconnaissance de formes.

7. La reconnaissance de traits

Cette approche met en jeu plusieurs types de connaissances et cherche à représenter les entités phonétiques (phonèmes) en termes de traits et contexte. La reconnaissance s'effectue en général en deux étapes:

1-La segmentation du signal en grandes classes phonétiques. Cette phase a pour but de délimiter des segments sur le signal de parole,

2- l'étiquetage des segments en utilisant les traits et les contextes. Cette phase consiste à identifier les segments obtenus lors de la segmentation en affectant à chaque segment une suite d'étiquettes phonétiques. Le résultat est le plus souvent un treillis phonétique utilisé comme entrée aux modules du niveau supérieur.

8. Types d'erreurs lors de la reconnaissance

Il est clair qu'aucun système n'est parfait. Cela veut dire qu'il y a toujours des erreurs de suppression, d'insertion ou de substitution des mots. Ces notions sont utilisées pour classer les différents systèmes existant dans le marché selon leur Word Accuracy et Word Error Rate. Dans le cas d'une substitution, ce qui est dit par l'utilisateur est mal reconnu: "c'est" →"ces", ceci peut être du à une mauvaise prononciation, articulation du mot. Dans celui d'une insertion, par exemple, l'utilisateur dit "je pense que" et c'est "je ne pense pas que" qui est reconnu, ici les deux mots ''ne'' et ''pas'' sont insérés en changeant fondamentalement le sens. Et en dernier pour la notion de suppression, l'utilisateur dit "je pense que" et c'est "je pense" qui est reconnu, le mot que est effacé.

II. la langue arabe

1. Présentation de la langue arabe

La langue arabe est une langue sémitique, elle est parmi les langues les plus anciennes dans le monde.

L'arabe classique standard à trente quatre (34) phonèmes parmi lesquels six (6) sont voyelles et vingt huit (28) sont des consonnes.

Les 28 consonnes arabes ont été divisées en deux groupes:

- 14 consonnes solaires qui assimilent le ل de l'article.
- 14 consonnes lunaires qui n'assimilent pas le ل de l'article.

Sur le plan phonétique, l'arabe standard présente la particularité d'être une langue essentiellement consonantique qui se caractérise par la présence des consonnes pharyngales, glottales et emphatiques et l'existence d'une opposition temporelle brève-longue des voyelles.

Au niveau morphologique, l'Arabe possède un système complet basé sur la notion de racine qui constitue le pilier du dictionnaire. *[3.7]*

2. La Phonétique de la langue arabe

La phonétique est le domaine de la linguistique qui a pour objet l'étude des langues naturelles dans leurs dimensions sonores. Le phonème est la plus petite unité discrète ou que l'on puisse isoler par segmentation dans la chaîne parlée. Un phonème est en réalité une entité abstraite, qui peut correspondre à plusieurs sons. Il est en effet susceptible d'être prononcé de façon différente selon les locuteurs ou selon sa position et son environnement au sein du mot. Les phones sont d'ailleurs les différentes réalisations d'un phonème.

Les phonèmes arabes se distinguent par la présence de deux classes qui sont appelées pharyngales et emphatiques. Ces deux classes sont caractéristiques des langues sémitiques comme l'hébreu

Les syllabes permises dans la langue arabe sont : CV, CVC, CVV, CVVC et CVCC

Où le V désigne une voyelle courte ou longue et le C représente une consonne

La langue arabe comporte cinq types de syllabes classées selon les trais ouvert/fermé et court/long. Une syllabe est dite ouverte (respectivement fermée) si elle se termine par une voyelle (respectivement une consonne). Toutes les syllabes commencent par une consonne suivie d'une

voyelle et elles comportent une seule voyelle. La syllabe CV peut se trouver au début, au milieu ou à la fin du mot *[3.5] , [3.8]*

Le tableau (tableau 3.1) représente quelques exemples de mots arabes avec leur prononciation en Alphabet Phonétique Internationale.

Mots en Arabe	Prononciation	Signification	Représentation syllabique
كَتَبَ	Kataba	Il a écrit	CV CV CV
يَكْتُبُ	i :aktobo	Il écrit	CVC CV CV
كَاتِبٌ	Ka :tibon	Ecrivain	CV CV CVC
جَمِيلٌ	Jami :lon	Beau	CV CV CVC
صَبْرٌ	Sabr	Patience	CVCC

Tableau 3.1 : *exemple d'analyse syllabique de quelques mots arabes*

3. Modes d'articulation des consonnes arabe

3.1. Les fricatives

Les fricatives sont produites dans la cavité vocale par une constriction étroite qui rend la circulation d'air turbulente. Acoustiquement, les fricatives non voisées possèdent en général un haut bruit aléatoire et les fricatives voisées possèdent des structures de résonance faibles qui apparaissent comme des ombres de formants faibles avec un léger bruit.

Exemple ف /fa / : est une fricative labiodentale non voisée.

ز /zin / : est une fricative dentale voisée.

ت , س , ش …etc. (voir tableau 3.2)

3.2. Les plosives

Une plosive est caractérisée par:

- la formation d'une fermeture à l'intérieur de la cavité vocale par un ou plusieurs articulateurs à l'endroit où le conduit de pression est bloqué et qui apparaît comme un vide sur le spectrogramme.
- la brusque libération de cette pression qui apparaît comme une barre d'explosion ou burst sur le spectrogramme. Par exemple ض ,ك , ط, …etc (tableau 3.2)

3.3. Les nasales

La nasalité est définie en terme physiologique comme étant la formation d'une ou plusieurs fermetures orales et le passage de l'air à travers le nez. Au cours de la production des nasales les deux cavités orale et nasale sont donc normalement utilisées.

En Arabe, il ya deux consonnes nasales le م /mim/ et le ن /non/.

Latéral ل /lam/ et des vibrant ر /ra/ glottales…etc. (tableau 3.2).

3.4. Les emphatiques

Le système phonétique de l'Arabe tire son originalité de la présence des consonnes emphatiques: la langue arabe est souvent appelée la langue du, ض, Nad/. Un phonème qui n'existe qu'en Arabe et qui est d'ailleurs difficile à prononcer. Les quatre articulations emphatiques sont : t' ط , d' ض , s' ص , dh' ظ .

		labiale	Dentale	Inter-dentale	Emphatique dentale	(Alveo-) Palatale	Vélaire	Uvulaire	Pharyn-gal	Glottale
plosives	Sans voisée			t ت	t' ط		k ك	q ق		a -ء
	voisée	b ب		d د	d' ض	dZ[1] ج				
Fricatives	Sans voisée	f ف	T ت	s س	s' ص	Sh ش	kh خ		X\ ح	h ه
	voisée		dh ذ	z ز	dh' ظ		gh غ		?\ ع	
Nasales		m م		n ن						
latéral				l[2] ل						
Vibrant				r ر						
Semi-voyelles		w و			j ي					

Tableau3.2 : *Arabe phonétique*

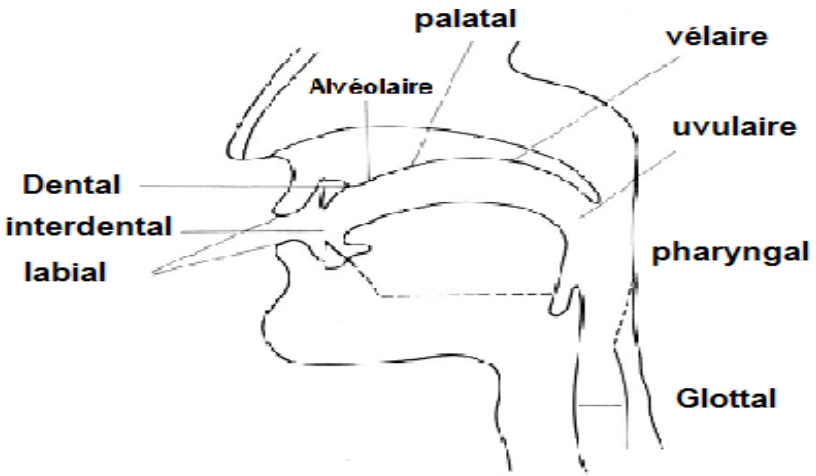


Figure3.4: *Le système phonétique de la langue arabe*

4. La pharyngalisation

Traditionnellement, la pharyngalisation est appelée ‘emphase’ mais le premier terme est plus précis et correspond à une réalité linguistique typologiquement bien décrite.

La pharyngalisation est un trait d'articulation secondaire des sons d'une langue. Ce trait, comme beaucoup d'autres articulations secondaires, est considéré comme un trait accompagnant les voyelles. On le trouve fréquemment associé aux consonnes, dans des langues telles que l'arabe .

Acoustiquement, il s’agit d’une articulation secondaire qui combine une articulation d’arrière à une articulation primaire d’avant. Concrètement, pendant qu’une consonne, normalement alvéolaire, est réalisée, la racine de la langue recule vers la partie postérieure du pharynx et le dos de la langue se creuse l’os hyoïde et le pharynx sont légèrement rehaussés, tout cela implique une diminution de la cavité pharyngale, et de fait une augmentation de la cavité orale.

Pour des raisons ayant certainement à voir avec des contraintes articulatoires ou perceptives

(Barkas-Defradas & Embarki, 2009 : 22), la pharyngalisation ne concerne que les consonnes coronales, autant en berbère qu’en arabe. En effet, il semble que des consonnes pharyngalisées palatales ou vélaires soient irréalisables, puisque le dos de la langue qui doit se rapprocher du palais pour ces deux articulations devrait en même temps être abaissé pour l’articulation pharyngalisée. Toutefois, les consonnes labiales peuvent apparaître pharyngalisées, de même que les post-alvéolaires, mais autant en berbère qu’en arabe, ces dernières n’accèdent généralement pas au statut de phonème.

Notons que Mc Carthy (1989, 1991, 1994), qui a rédigé les articles qui sont peut-être les plus élaborés sur la question des consonnes pharyngalisées et des consonnes d’arrière en général,

distingue les consonnes d'arrière qui possèdent un trait pharyngal de celles qui ont un caractère plutôt uvulaire. Cette dernière distinction permet entre autre d'expliquer la différence de comportement qui peut exister entre les consonnes pharyngalisées et uvulaires d'une part, et les consonnes laryngales et pharyngales d'autre part. Cette distinction nous sera utile quand nous traiterons des voyelles, mais, pour les consonnes, nous nous bornerons à utiliser le trait strictement acoustique, le trait pharyngal, qui nous suffira pour l'analyse.

5. Les consonnes pharyngales

5.1. Le pharynx

1. Le pharynx est un conduit faisant communiquer la bouche et l'œsophage d'une part, les fosses nasales et le larynx d'autre part.
   C'est donc le carrefour des voies aériennes et des voies digestives qui se croisent à ce niveau, il est composé de 3 parties

- Rhino-pharynx.
- Oro-pharynx.
- Laryngo-pharynx.

a) la consonne "Ha" ح

Est la sixième lettre l'alphabet arabe, utilise une friction un peu plus importante, mais sonne un peu comme un courant d'air entre les mains ou comme un papier de verre très fin.

/ ح / est une fricative pharyngale non voisée de durée 100-150 msec qui devient voisée en milieu intervocalique. Lors de la production du / ح / une constriction est formée par le dorsum de la langue contre la paroi postérieure du pharynx. Acoustiquement, le / ح / apparaît comme un bruit plus fort que celui du /h/.

Exemple **بسم الله الرحمان الرحيم**

b) la consonne "ayn" ع

Est le dix-huitième lettre l'alphabet arabe Le son le plus difficile à entendre en arabe est le ain (ع), il sonne comme un "ai". Il est un peu comme le son qu'un docteur vous demande lorsqu'il regarde au fond de votre gorge. Quand vous dites "aaah", rentrez un peu le dos de votre langue dans votre gorge.

(ع) est décrit comme une fricative pharyngale voisée dont la structure est très dépendante du contexte de production. En position initiale, le (ع) apparaît sur le spectrogramme comme une sorte de "burst" - durée 40-50 msec - dont l'intensité est quelque part entre 1450 et 1.550 Hz.

Exemple ربيــع، عمل، نعم

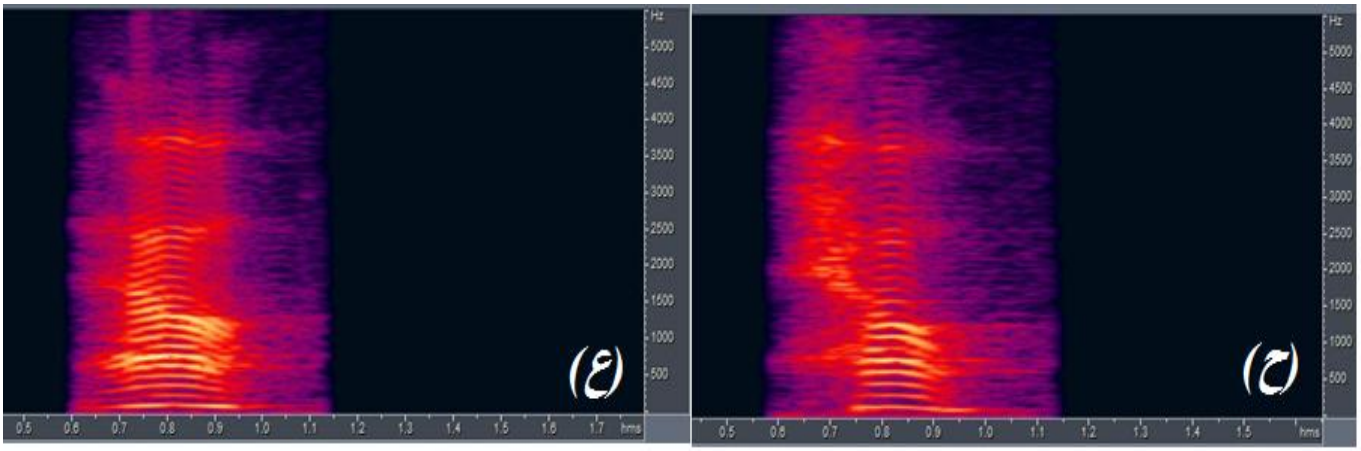


***Figure3.5** : Spectrogramme de consonnes (ع) et (ح)*

6. Travaux Scientifiques Récents

En matière de RAP de la langue arabe nous pouvons cités quelques travaux récents dans ce domaine :

Daneil. L. Newman a enquêté sur le statut phonétique d'arabe, spécifiquement sur l'originalité de phonèmes arabes spéciaux, un groupe de recherche (Kirchhoff, K., Bilmes, J., Das, S., Duta, N., Egan, M., Bande J.,FENG H., HENDERSON, J., DABEN L., NOAMANY, M., SCHONE, P.,,Schwartz, R., et Vergyri, D) ont effectué des recherches sur la nouvelle reconnaissance de la Parole arabe, aussi un autre groupe a fait des expériences sur adaptation et l'accent arabe Non-indigène dans la reconnaissance automatique de la parole, cette équipe est composée de (Yousef Alotaibi et Sid-Ahmed Selouani, et Douglas O'Shaughnessy)

D'autres études ont été faites par :

Selouani, S. et Caelen, J., sur la Reconnaissance traits phonétiques arabes qui utilise des Architectures connexionniste modulaires, Mansour Alghamdi qui s'intitulé sur l'analyse, synthèse et Perception de la parole arabe, et Yousif A. El – Imam que a fait Un vocabulaire sans restriction de la Parole arabe ».

7. Conclusion

Nous avons présenté dans ce chapitre les notions les plus importantes de la reconnaissance automatique de la parole sur lesquelles reposent l'essentiel de notre travail de l'aspect pharyngal de la langue Arabe standard en plus d'une brève description des travaux parmi les plus importants à ce niveau d'étude.

Chapitre IV
Implémentation et tests

1. Introduction

Notre objectif dans ce chapitre est de développer sous MATLAB, un système de reconnaissance automatique de la parole Mono-locuteur.

Premièrement en construire une base de données de fichiers de format **.WAV** ensuite d'appliquer le codage MFCC sur cette base pour avoir des vecteurs d'apprentissage ensuite ces vecteurs subissent une phase de normalisation par interpolation de l'image cepstrale pour pouvoir injecter ces vecteur dans un MLP.

L'apprentissage et le test sont réalisés via les réseaux de neurones artificiels (ANN) et plus précisément par un MLP avec l'algorithme de correction des poids synaptique selon la descente du gradient (voir chapitre RdF), afin d'estimer le taux de reconnaissance globale pour chaque scenario à part, (figure 4.1).

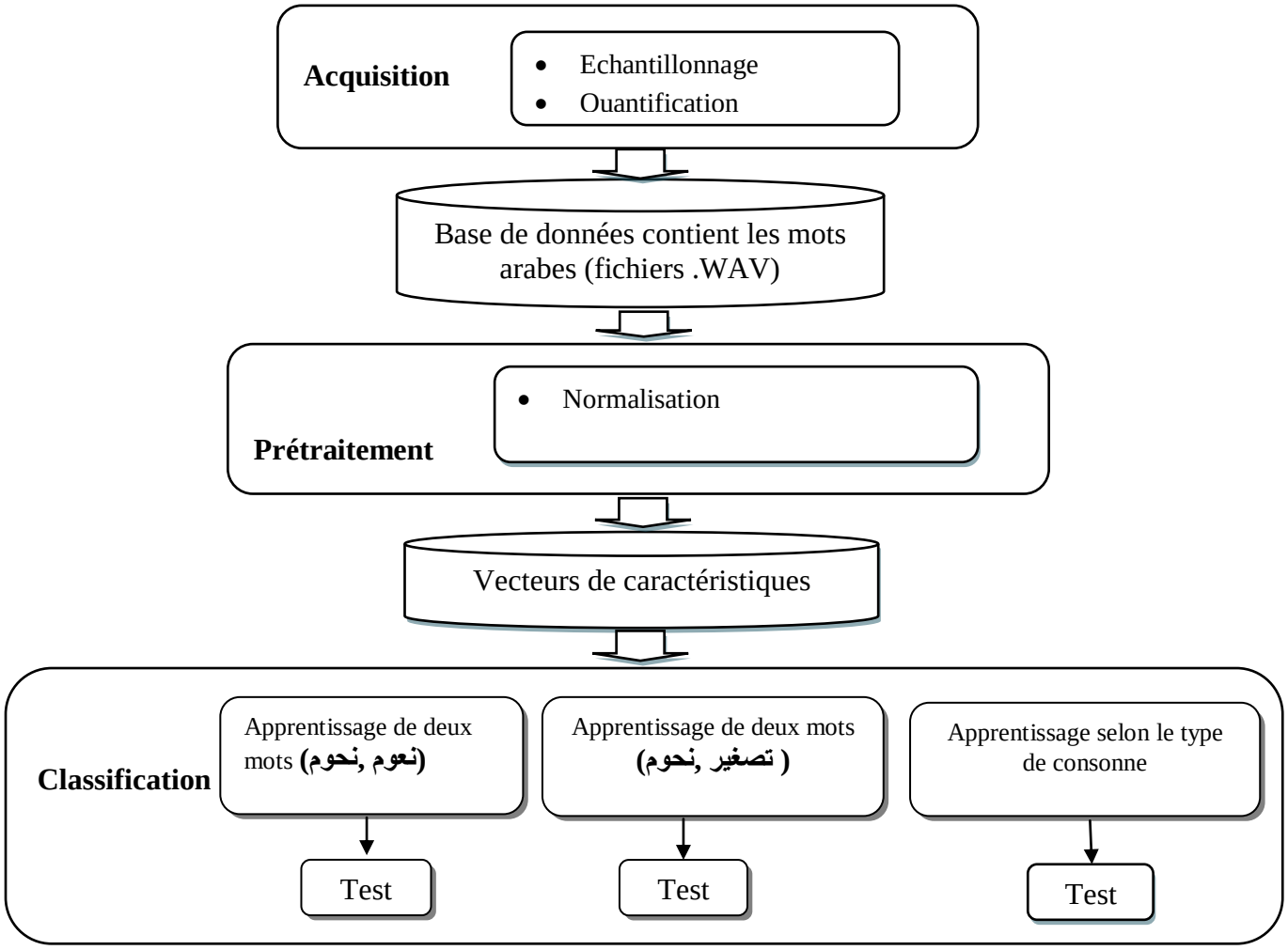


Figure 4.1 *Schéma Général des différentes étapes de l'apprentissage*

2. Environnement de programmation : MATLAB

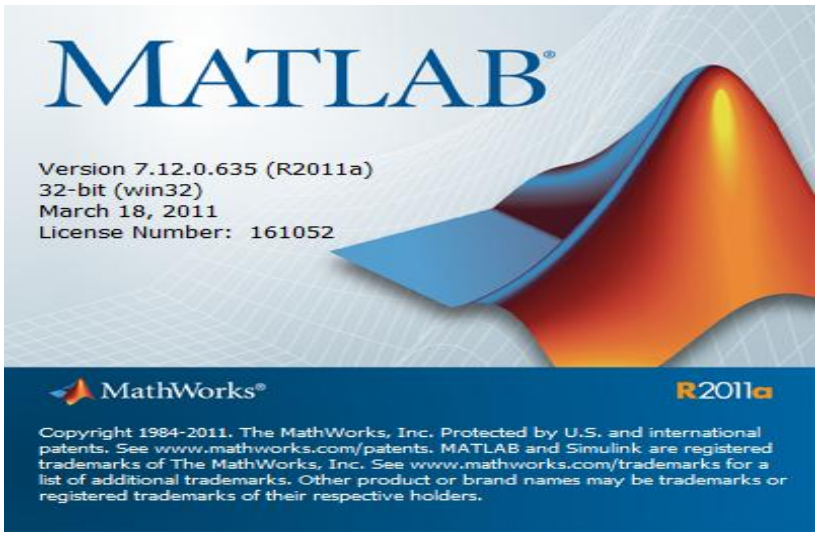


Figure 4. 2 : *Logiciel MATLAB.*

2.1. MATLAB (MATrix LAboratory)

Est un logiciel pour effectuer des calculs numériques. Il a été conçu pour faciliter le traitement des matrices.

2.2. Pourquoi MATLAB ?

Le succès actuel de MATLAB vient de sa simplicité de prise en main et d'utilisation. De plus, il existe des boîtes à outils (toolbox) optionnelles mais très utiles dans certains domaines comme l'optimisation, le traitement du signal et de l'image, l'apprentissage (réseaux de neurones. . .), l'automatique (Simulink), etc. Ce logiciel est de plus très utilisé tant dans le monde industriel que dans le monde universitaire.

MATLAB possède une riche bibliothèque de fonction prédéfinies qui simplifié grandement l'élaboration de programmation plus complexes.

2.3. Intérêts

- Programmation infiniment rapide pour le calcul et pour l'affichage.
- Une librairie très riche.
- Possibilité d'inclure une programmation en C/C++.
- Langage interprété : Pas de compilation donc pas d'attente pour compiler.
- Possibilité d'exécuter du code en dehors du programme.
- Code facile à comprendre et très lisible.

- Une aide très bien faite.

2.4 Inconvénients

- Vitesse de calcul moine rapide qu'en C/C++.
- Application auto-exécutable peu pratique.

MATLAB nécessite une configuration spécial de matériel et logiciel, pour ce la on a exploites les ressources suivant :

- Système d'exploitation : Microsoft Windows 7 Professionnel.
- Version : Pack 1.
- Ordinateur : HP 620.
- Fabricant : Hewlett-Packard.
- Mémoire physique totale : 2Go.
- Processeur intel dual-core : 2.3 GHz.

3. Interface de L'application Implémentée

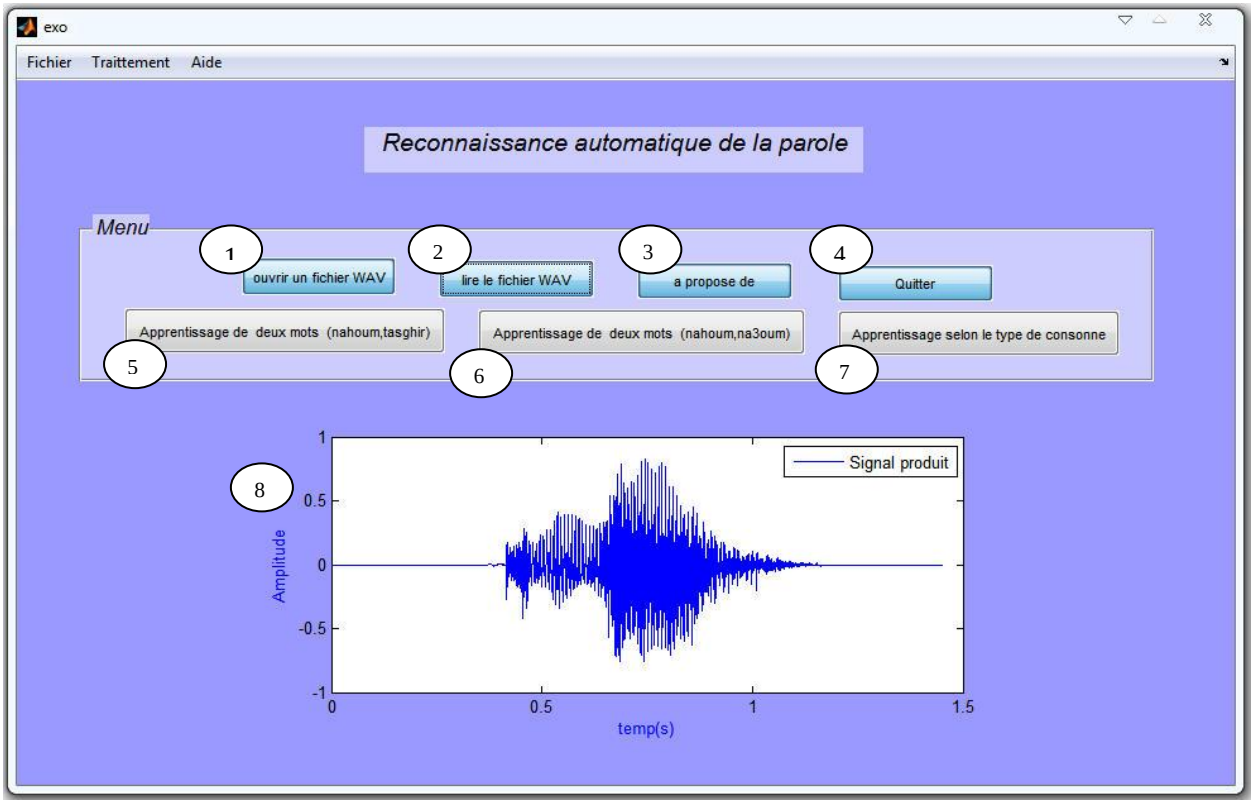


Figure 4.3 : *Fenêtre de menu principal*

1. Ouvrir un fichier .WAV .
2. Lire le fichier .WAV sélectionné .
3. A propose de développeur de l'application.
4. Quitter l'application.
5. Permet de faire l'apprentissage de deux mots (نحوم, تصغير).
6. Permet de faire l'apprentissage de deux mots (نحوم, نعوم).
7. Apprentissage selon le type de consonne.
8. Espace pour afficher le signal temporelle.
9. Aide sur l'application et sur l'environnement MATLAB.
10. Accéder a la fenêtre de modification les paramètres d'apprentissage.

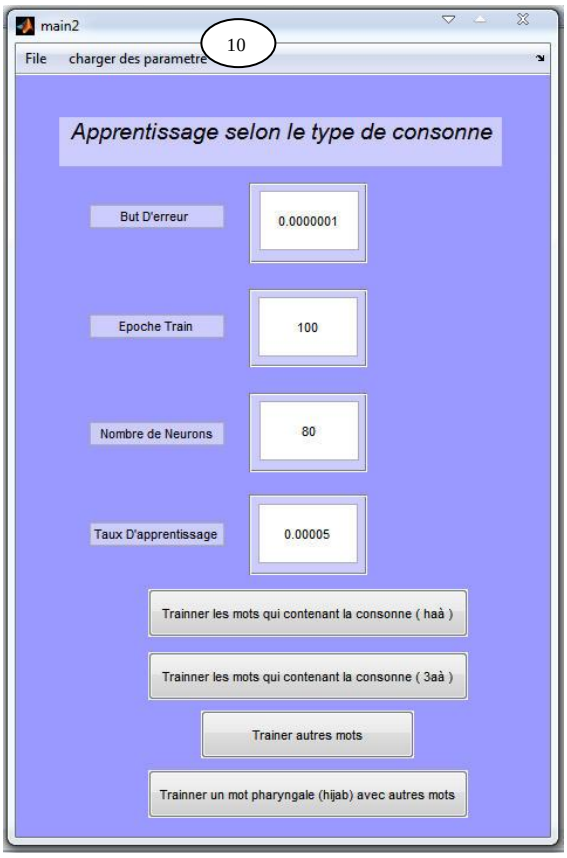


Figure 4.4 : *Fenêtre de l'apprentissage selon le type de consonne*

Figure 4.5 : *Fenêtre de l'apprentissage de deux mots (نحوم, تصغير).*

Figure 4.6: Fenêtre de l'apprentissage de deux mots (نحوم, نعوم)

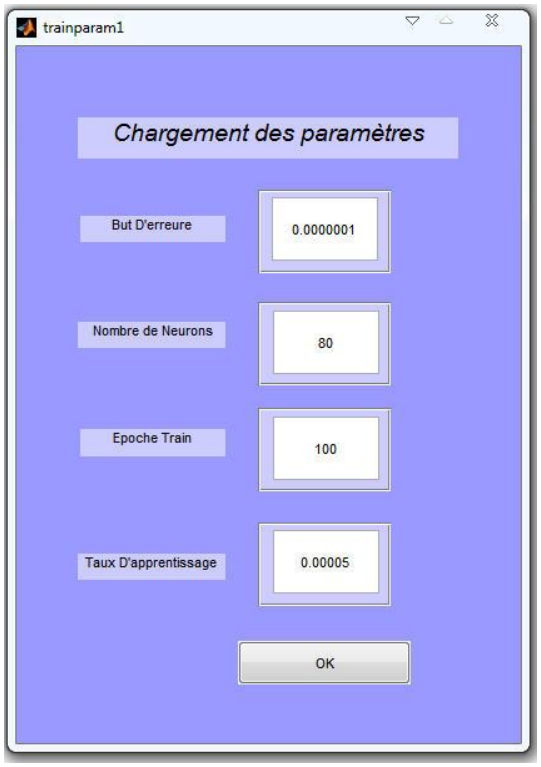
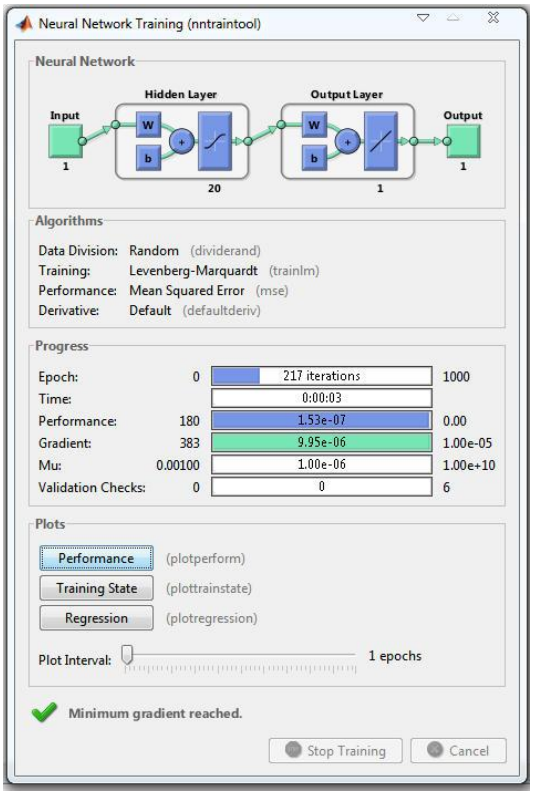


***Figure.** 4.7 : Fenêtre pour modifier les paramètres d'apprentissage*



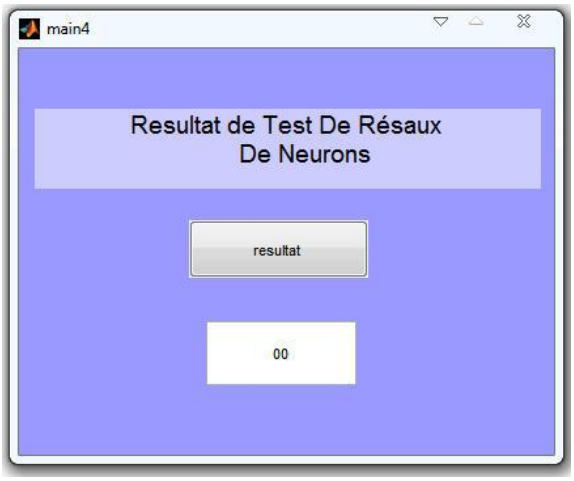


***Figure 4.8:** Apprentissage du MLP et affichage du résultat de test*

4. Expériences et évaluations

Á travers cette partie, nous allons procéder à une succession d'expériences selon Le paramètre de :

1- nombre de neurones cachés.

2- Le type de consonne pharyngale.

Les résultats sont présentés sous formes de tableaux.

Notations :

Nbr $_{NC}$: nombre de nuerons cachés.

Nbr $_{EP}$: nombre d'epochs d'apprentissage.

4.1. Apprentissage selon le type de consonne

rythme D'apprentissage	Nbr_{NC}	Nbr_{EP}	Mots qui contenant la consonne (ح)	Mots qui contenant la consonne (ع)	Autres mots	Un mot pharyngale avec 03 non-pharyngales
			Taux (%)	Taux (%)	Taux (%)	Taux (%)
0 .00005	80	100	**65,57**	**57 ,14**	**70,13**	**61.53**
0 .00005	100	100	**65,95**	**58,29**	**75,47**	**62.5**
0 .00005	200	100	**67.21**	**58,57**	**76,92**	**62,55**
0 .00005	300	100	**68,85**	**60,03**	**78,42**	**64,06**

Tableau 4.1 : *Expériences selon type de consonne*

4.2. Apprentissage de deux mots (نحوم, تصغير)

rythme D'apprentissage	Nbr_{NC}	Nbr_{EP}	Taux d'apprentissage de deux mots (نحوم, تصغير)
			Taux (%)
0 .00005	80	100	**57,23**
0 .00005	100	100	**70,69**
0 .00005	200	100	**85,10**
0 .00005	300	100	**65,86**

Tableau 4.2 : *Expériences de deux mots (تصغير ,نحوم)*

4.3. Apprentissage de deux mots (نحوم, نعوم)

rythme D'apprentissage	Nbr_{NC}	Nbr_{EP}	Taux d'apprentissage de deux mots (نحوم, نعوم)
			Taux (%)
0 .00005	80	100	**55**,29
0 .00005	100	100	**49**,**53**
0 .00005	200	100	**57**,**08**
0 .00005	300	100	**57**,**97**

Tableau 4.3 : *Expériences de deux mots (نعوم ,نحوم)*

5. Observations et Argumentations

5.1. Apprentissage selon le type de consonne

On remarque que :

- **Pour les mots qui contenant la consonne (ح) :**

Le taux de reconnaissance est maximal de **68,85%** pour un nombre de neurones cachés égale à **300.**

- **Pour les mots qui contenant la consonne (ع) :**

Le taux de reconnaissance est maximal de **60,03%** pour un nombre de neurones cachés égale à **300.**

- **Pour les autres mots :**

Le taux de reconnaissance est maximal de **78,42%** pour un nombre de neurones cachés égale à **300.**

- **Un mot pharyngale avec 03autre mots non-pharyngale**

- Le taux de reconnaissance est maximal de **64,06%** pour un nombre de neurones cachés égale à **300.**

5.2. l'apprentissage de deux mots (نحوم, تصغير)

Le taux de reconnaissance est maximal de **85,10%** pour un nombre de neurones cachés égale à **200.**

5.3. l'apprentissage de deux mots (نحوم , نعوم)

Le taux de reconnaissance est maximal de **57,97%** pour un nombre de neurones cachés égale à **300.**

- ✓ On remarque que plus le nombre des neurones caché augmente ,plus le taux d'apprentissage s'améliore

6. Argumentations

1- Le taux d'apprentissage de mots pharyngale est inferieur que le taux d'apprentissage de mots non-pharyngale ce signifier que les mots qui contiennent les consonnes pharyngale (ح) et (ع) sont des mots un peu difficile à reconnaitre par rapport aux autres mots.

2- Quand on considère un scénario composé d'un mot pharyngale et 03 mots non-pharyngale le taux diminué sa-ve-dire que les mots pharyngaux ont un effet négatif sur le taux de reconnaissance.

3- le taux de test des deux mots **(نحوم, تصغير)** est supérieur que le taux de test de deux mots **(نحوم , نعوم),** se qui implique qu'il existe une ressemblance cepstrale entre les deux mots **(نحوم , نعوم)** .

7. Conclusion

À travers ce chapitre nous avons procédé a une successions d'expériences dans le but de tester et valider notre modèle neuronal proposé, et c'est dans ce contexte qu'on peut dire d'une manière générale que la préparation et la prononciation des échantillons d'apprentissages est très importante surtout dans la langue arabe. Enfin nous pouvons d'un coté dire que pour avoir un bon résultat satisfaisant il faut menée plusieurs expériences selon le nombre de neurones cachées et ceci dans le but de déterminer avec précision la configuration optimale. Et de l'autre coté les expériences ont montrées l'influence des consonnes pharyngales sur le taux globale de reconnaissance, ainsi nous recommandons les futur PFE de bien prendre en considération la contrainte du codage du signal car une bonne analyse du spectre de ce genre de consonnes et suivant les résultats obtenus on peut facilement modifier la configuration des vecteurs MFCC ou bien utiliser un autre type de codage tel que les coefficients LPC (Linear Prediction Coding).

On conclu que la variation de nombre de neurones cachés affecte beaucoup sur la qualité de reconnaissance.

D'une manière générale les résultats étaient entre modeste et encourageante de telle façon que ce model puisse être pris en considération pour de futurs travaux.

Conclusion générale

Dans ce travail nous avons abordé un domaine en cours d'expansion cette dernière décennie : c'est la reconnaissance automatique de la parole et particulièrement en arabe.

La RAP (Reconnaissance Automatique de la Parole) permet d'améliorer les performances des systèmes, et c'est dans ce contexte qu'un certain nombre de méthodes ont été proposées pour extraire des caractéristiques adaptées à la reconnaissance de la parole, tel que le codage MFCC utilisé déjà dans notre approche dans le but d'extraire des paramètres statistiques a base de l'information cepstrale, etc…

Nous avons construit un système de reconnaissance automatique de la parole appliquée sur quelques mots arabes à consonnes pharyngales. Ce système se compose d'une base d'apprentissage et d'une base de test.

L'approche la plus utilisé dans ce domaine est bien la représentation cepstrale du signal de la parole. Ce type de codage permet de compresser le signal de la parole tout en conservant l'ensemble des informations Spectro-Temporelles utiles pour les différentes tâches de reconnaissance.

Enfin nous espérons d'un côté, que ce modeste travail servira comme une base solide à d'autres futurs travaux en prenant en compte d'autres paramètres ou contraintes acoustiques.

Et de l'autre côté, représentera un outil performant pour l'élaboration automatique de bases de données de codage MFCC destinées aux systèmes de reconnaissance automatique de la parole arabe.

Glossaire

RAP **:** Reconnaissance Automatique de la Parole

T.S : traitement du signal

P_s : puissances du signal

P_n **:** puissances bruit

S(t) : fonction de signal physique

ΔF : largeur de bande spectrale du signal

TFD : Transformée de Fourier discrète

FFT : Fast Fourier Transform

DSP : structure intégrée micro programmable

RdF : reconnaissance de formes

LAD : Lecture Automatique de Document

RNA : réseaux de neurones artificiels

HMM : Hidden Markov Model

ZCR : Taux de passage par zéro

MFCC : Mel Filter Cepstral Coefficients

LPC : Linear Prédictive Coding

PLP : Perceptual Linear Predictive

LDA : Linear Discriminant Analysis

DAP : décodage acoustico-phonétique

CV : consonne, voyelle

Bibliographie

[1.1] F. de Coulon : " Théorie et traitement des signaux " ; Ed. Dunod M. Bellanger : " Traitement numérique du signal " ; Ed. Masson

[1.2] M .BELLANGER , Traitement numérique du signal , Dunod ,1998

[1.3] Gérard-Michel Cochard Technologies des réseaux de communication Université Virtuelle de Tunis 2007

[1.4] A. J. Jerri, "The Shannon sampling theorem - its various extensions and applications : a tutorial review" ,1977, pp. 1565-1596.)

[1.5] cour reseau M chadli 2010

[1.6]: G .BOUDOIN et J.BERCHER, *Eléments du Traitement du Signal* ,1998.

[1.7] T.Dutoit Introduction au traitement automatique de la parole faculté polytechnique de mons.

[1.8] EliasNemer, Rafik Goubran «The fourth order cumulant of speech signal with application to voice activity detection » IEEE Trans. On signal Proc Vol. 42, N° 1, pp. 222-224, Jan. 1994

[1.9] A.DECHEVEIGNE et H.KAWAHARA. a fundamental frequency estimator for speech and music page 1917-1930, 2002

[2.1] Damien Poisson et Sami Mahjoub. « La reconnaissance de Forme Comment améliorer les techniques de reconnaissance de forme 3D ? » 06/06/2010 , Paris.

[2.2] BOUGAMOUZA Fateh .Contribution à la reconnaissance automatique de l'écriture Manuscrite arabe, application sur les montants littéraux des chèques, Université de Constantine. 2008.

*[2.3]*DERDOUR Khedidja et all .Reconnaissance de formes du chiffre arabe imprimé Application au code à barre d'un produit, Université de Batna.2009.

[2.4]. derdour khedidja. « Reconnaissance de formes du chiffre arabe imprimé : Application au code à barre d'un produit ». Université hadjlakhdar –batna faculté des sciences de l'ingénieur .

[2.5] BEN KROSE et PATRICK van der Smagt, « An Introduction to Neural Network » Ecole Nationale d'Ingénieurs de Tunis. 1996.

[2.6] Le Cerveau, Bibliothèque de la revue Pour la Science, 1982.

[2.7] R. Masland, L'architecture fonctionnelle de la rétine, 1987.

[2.8] Eric Davals, patricknaim des réseaux de neurones 2eme Edition eyrolles 1992

*[2.9]*claude tauzet, les réseaux de neurone artificiel, introduction au connexionnisme « Ecole supérieure d'Électricités, France » 1992.

[2.10] Claude Touzet les réseaux de neurones artificiel Introduction au connexionnisme 1992 .

[3.1] Jean-Paul Haton «La reconnaissance automatique de la parole »

[3.2] J.M. Pierrel. Le Dialogue Oral homme-machine: Connaissance linguistiques, stratégies et architectures des systèmes. Collection Hermès, Paris,1987.

[3.3] Y. Gong. Introduction au traitement du signal pour la représentation paramétrique en reconnaissance automatique de la parole, Département Mathématique Appliquée et Informatique, Université de Nancy l, 1985.

[3.4] Y. Gong. Conception et réalisation d'un système de transformée de Fourier Rapide (FFT). Département d'Electronique, Université de Pierre et Marie CURIE (ParisVI), 1983.

[3.5] H. Sakoe et S. Chiba. Dynamic programming algorithm optimisation for spoken word recognition. IEEE Trans. Acoust., Speech, Signal Processing, February 1978.

[3.6] Baker. Stochastic Modelling for Automatic Understanding. Speech Recognition, pages 51-58. R. Reddy editor, New York, Academie Press, 1975.

[3.7] A. Muhammad, "Alaswaat Alaghawaiyah," Daar Alfalah, Jordan, 1990 (in Arabic).

[3.8] M. Elshafei. Toward an arabic text-to-speech system. 1991. vol. 4B no. 16,pp. 565–583