République Algérienne Démocratique et Populaire
Ministère de l'enseignement supérieur et de la recherche scientifique
Université 08 Mai 1945 Guelma
Faculté des sciences et sciences de l'ingénierie
Département d'électronique et télécommunication

Mémoire de fin d'étude

Pour l'obtention du diplôme de Master

Filière : **Télécommunications**
Spécialité **: systèmes des Télécommunications**

Utilisation des réseaux neurones pour la Reconnaissance Automatique de la parole

---

Présenté par :

- **GADOUCHE Salah Eddine**
- **BOUZOUALEGH Ahmed**

Sous la direction de : **Mr .IKNI Samir**

JUIN 2011

Dédicace

Je dédie ce mémoire à mon père et ma mère pour l'éducation

qu'ils ont su me donner et qui m'a permis

avec la grâce de Dieu d'arriver la au je suis.

A mes frères et sœurs Spécialement (mani)

A toute la famille

A tous mes amis Spécialement pour Bilal

Bouzoualegh Ahmed

Dédicace

Je dédie ce mémoire à mon père et ma mère pour l'éducation

qu'ils ont su me donner et qui m'a permis

avec la grâce de Dieu d'arriver la au je suis.

A mes frères et sœurs

A toute la famille

A tous mes amis

Spécialement pour mon frère (Abderrahmane).

Gadouche Salah Edinne

SOMMAIRE

CHAPITRE III : Réseau de Neurones

Introduction générale

Le signal de parole est un signal très redondant, il transporte la même information sous plusieurs formes, ce qui explique son immunité contre les bruits ambiants, et les perturbations extérieures, cette particularité lui permet de protéger son intelligibilité. Seulement pour déchiffrer et décoder ce signal, il nous faut un très puisant calculateur, pour l'appareil auditoire et le cerveau de l'être humain ça ne pose aucun problème, car ces derniers forment un système puissant capable d'effectuer les analyses pragmatiques et sémantiques nécessaires.

Mais pour le traitement automatique sur les machines cela pose beaucoup de problèmes surtout en ce qui concerne le temps d'analyse, l'immense quantité de données à traiter ainsi que l'instabilité du milieu de traitement.

Les êtres humains, révolutionnaire, lancent toujours le défi contre la nature, essaient en utilisant diverses techniques de passer outre ces problèmes ; Ces techniques basées sur des lois tantôt rigoureuses, tantôt approximatives ou même empiriques permettent d'aborder ces problèmes d'un point de vu inspiré de cette nature elle-même (*" il faut s'inspirer de la sagesse de notre créateur dans ses créatures"*), et en fin de compte extraire des résultats satisfaisants dans plusieurs cas. C'est le cas pour la reconnaissance de la parole, cette dernière fera l'objet de ce présent travail.

L'objectif de ce travail est la réalisation d'un système de reconnaissance automatique de la parole avec un vocabulaire limité et multi-locuteurs. Pour l'extraction des caractéristiques, en se base sur les coefficients cepstraux dans l'échelle de Mels (MFCC) qui sont les plus utilisés dans ce domaine. Cette technique s'appuie à la fois sur la théorie cepstrale et la perception de la parole dans l'échelle de Mels.

Dans la phase de classification, le système proposé se base sur les réseaux de neurones. Cette phase constitue une grande catégorie de problèmes actuels qui consiste à attribuer, de façon automatique, un objet à une classe parmi d'autres classes possibles. La résolution de ce type de problèmes demande de représenter les exemples à classifier à l'aide d'un ensemble de caractéristiques. Il s'agit ensuite de concevoir un système capable de classifier ces exemples en se basant sur leur représentation et les réseaux de neurones sont particulièrement bien adaptés à ce type de problème. Ceci est dû à leurs grandes capacités de calcul et à leurs hautes habilités d'apprentissage. De plus, l'estimation de leurs paramètres est

indépendante de la complexité du problème traité ce qui leur permet d'être bien adaptés aux problèmes actuels qui ne cessent d'être de plus en plus complexes.

Le présent mémoire se compose de quarts chapitres :

- Le premier chapitre constitue une généralité sur la reconnaissance automatique de la parole
- Le deuxième chapitre une introduction générale à la reconnaissance de la parole et présente les différentes étapes du traitement de la parole.
- Le troisième chapitre constitue une introduction aux réseaux de neurone et leur fonctionnement, ainsi que leurs algorithmes d'apprentissage.
- Le quatrième chapitre présente les résultats obtenus sur une petite base de données que nous avons élaborée.

Finalement, une conclusion générale conclue ce mémoire.

Chapitre I

Généralité sur Reconnaissances Automatique de la Parole

I.1. Introduction

Parler avec les machines est une vision récurrente de notre imagination collective de l'informatique de futur. Dès 1968 Stanley Kubrick avait imaginé et mis en scène un ordinateur intelligent appelé HAL, capable de résonner et dérouler des opérations logiques mais aussi de communiquer verbalement. Pourtant à cette époque la technologie ne permettrait de reconnaître tout au plus quelques centaines de mots par des systèmes simulés sur des gros ordinateurs. Quarante ans plus tard grâce à l'avènement de l'informatique et aux efforts déployés, les systèmes de reconnaissance vocale sont devenus des produits de consommation destinés à un très large publique. La reconnaissance vocale est devenue une des technologies prépondérantes dans le développement d'interface Homme-Machine avancée. Toutefois, malgré les avancées très importantes de ces dernières années dans ce domaine, les systèmes actuels sont encore en deçà des performances de notre système d'audition. Un des principaux obstacles au déploiement des systèmes de reconnaissance vocale est la robustesse au bruit. Les différences entre les conditions d'utilisation (généralement bruitées) et d'apprentissage (absence de bruit) des modèles acoustiques provoquent une dégradation significative des taux de reconnaissance même si ces dégradations semblent minimes à l'oreille. [1]

I.2. Reconnaissance automatique de la parole (RAP)

Le but ultime poursuivi en reconnaissance de la parole est la communication en langage naturel avec une machine (ordinateur, machine, robot,…). Depuis près de cinquante ans, la reconnaissance de la parole par machine a fait l'objet d'un très grand effort de recherche, mais les performances des systèmes réalisés n'ont pu égaler, encore, celles de l'être humain, notamment dans les conditions réalistes de fonctionnement tel que la parole spontanée ; présence de bruit ambiant, etc.…

Reconnaître et comprendre la parole demeurent parmi les grands défis de l'informatique. Les problèmes à résoudre sont considérables et de natures pluridisciplinaires notamment le traitement du signal, l'intelligence artificielle, la reconnaissance des formes, la phonétique, la linguistique et les neurosciences intervenant à des degrés divers dans les solutions. [2]

I.2.1. Définition

La reconnaissance automatique de la parole est l'un des deux domaines du traitement automatique de la parole, l'autre étant la synthèse vocale. La reconnaissance automatique de la parole permet à la machine de comprendre et de traiter des informations fournies oralement par un utilisateur humain. Elle consiste à employer des techniques d'appariement afin de

comparer une onde sonore à un ensemble d'échantillons, composés généralement de mots mais aussi, plus récemment, de phonèmes (unité sonore minimale). En revanche, le système de synthèse de la parole permet de reproduire d'une manière sonore un texte qui lui est soumis, comme un humain le ferait. [2]

I.3. Classification

Les systèmes de reconnaissance peuvent être classés par ordre de difficulté croissante :

- ❖ Reconnaissance des mots isolés appartenant à un vocabulaire limité (moins de 100 mots) le système est adapté à un locuteur donné : système uni-locuteur.
- ❖ Même spécification mais le vocabulaire est beaucoup plus étendu (quelques milliers de mots).
- ❖ Même spécification mais le système est indépendant du locuteur : système multi-locuteur.
- ❖ Reconnaissance des mots enchainés : les mots appartiennent à un vocabulaire très limité (les dix chiffres par exemple) mais ils sont prononcés sans pause dans un ordre quelconque.
- ❖ Reconnaissance de phrases courtes basées sur un vocabulaire limité : le système est uni-locuteur.
- ❖ Même spécification mais le système est multi-locuteur.
- ❖ Reconnaissance de la parole continue prononcée par un locuteur quelconque. [2]

I.4. Difficultés de la RAP

Le signal de la parole possède des caractéristiques qui compliquent son interprétation et augmentent le nombre de données à traiter.

Il présente un caractère redondant, c'est à dire qu'il renferme plusieurs types d'informations (les sons, la syntaxe et la sémantique de la phrase, l'identité du locuteur et son état émotionnel). Si cette redondance lui confère une bonne résistance au bruit, elle oblige à extraire du signal les informations pertinentes, en essayant de ne pas trop les dégrader.

Le signal est très variable selon le locuteur, c'est la variabilité interlocuteur (timbres différents, différences morphologiques homme ou femme). Mais également pour un même locuteur, on parle alors de variabilité intra-locuteur, due à l'état émotionnel, la voix chantée, parlée, chuchotée. Il s'ajoute aussi les variabilités dues au milieu (le bruit perturbe la prise de son et augmente la variabilité intra locuteur) et à l'acquisition du signal. [2]

Le signal est continu, c'est à dire que lorsqu'on écoute une personne, on perçoit une suite de mots, alors que l'analyse du signal vocal ne permet de déceler aucun séparateur. Le même problème de segmentation se retrouve à l'intérieur du mot. Celui-ci est perçu comme une suite de sons élémentaires (les phonéticiens trouveront le même nombre de phonèmes dans une phrase) que l'analyse ne permet pas d'isoler en segments distincts du signal acoustique. Le signal de la parole est évolutif.

Il y a également le phénomène de coarticulation. C'est l'effet contextuel que produit un phonème sur ses voisins. Il est provoqué par le fait que, lors de la prononciation d'un phonème, l'appareil articulatoire se prépare pour la production du suivant.

Ces caractéristiques compliquent la tâche d'un système de RAP qui doit être capable de « décider qu'un [a] prononcé par un adulte masculin est plus proche d'un [a] prononcé par un enfant, dans un mot différent, dans un environnement différent et avec un autre microphone, que d'un [o] prononcé dans la même phrase par le même adulte masculin ». [2]

I.4.1 Variabilité intra locuteur

La variabilité intra locuteur exprime les différences dans le signal produit par une même personne. Cette variation peut résulter de l'état physique ou moral du locuteur. Une maladie des voies respiratoires peut ainsi dégrader la qualité du signal de parole de manière à ce que celui-ci devienne totalement incompréhensible, même pour un être humain. L'humeur ou l'émotion du locuteur peut également influencer son rythme d'élocution, son intonation ou sa phraséologie. Il existe un autre type de variabilité intra locuteur liée à la phase de production de la parole ou de préparation à la production de parole, due aux phénomènes de coarticulation (voir figure I.1)

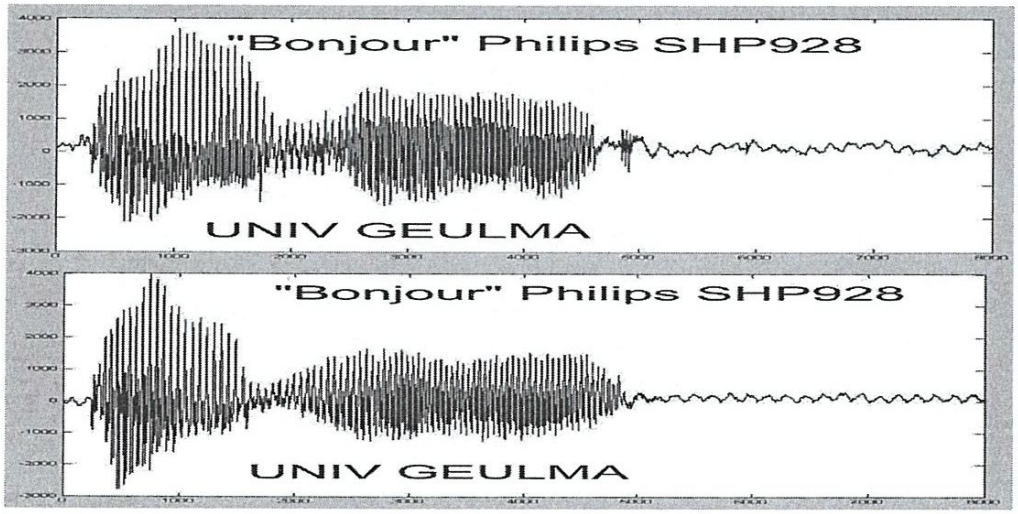


Figure I.1 Variabilité de locuteur

I.4.2 Variabilité interlocuteur

La variabilité interlocuteur est un phénomène majeur en reconnaissance du locuteur. Mais un locuteur reste identifiable par le timbre de sa voix, malgré une variabilité qui peut être parfois importante.

La cause principale des différences interlocuteurs est de nature physiologique. La parole est produite par les vibrations des cordes vocales, qui déterminent l'importance et la forme du flux d'air s'échappant des poumons et amplifiées par les organes respiratoires, cette opération génère un son à une fréquence de base, le fondamental. Cette fréquence de base est différente d'un individu à l'autre et plus généralement d'un genre à l'autre ; une voix d'homme est plus grave qu'une voix de femme, la fréquence du fondamental étant plus faible.

Ce son est ensuite transformé par l'intermédiaire du conduit vocal, délimité à ses extrémités par le larynx et les lèvres. Cette transformation, par convolution, permet de générer des sons différents. Or le conduit vocal est de forme et de longueur variable selon les individus et, plus généralement, selon le genre et l'âge. Ainsi, le conduit vocal féminin adulte est, en moyenne, d'une longueur inférieure de 15% à celui d'un conduit vocal masculin adulte. Le conduit vocal d'un enfant en bas âge est bien sûr inférieur en longueur à celui d'un adulte. Les convolutions possibles seront donc différentes et, le fondamental n'étant pas constant, un même phonème pourra avoir des réalisations acoustiques très différentes.

La variabilité interlocuteur trouve également son origine dans les différences de prononciation qui existent au sein d'une même langue et qui constituent les accents régionaux.

I.4.3 Variabilité due à l'environnement

La variabilité liée à l'environnement peut, parfois, être considérée comme une variabilité intra locuteur mais les distorsions provoquées dans le signal de parole sont communes à toute personne soumise à des conditions particulières. La variabilité due l'environnement peut également provoquer une dégradation du signal de parole sans que le locuteur ait modifié son mode d'élocution. Cette variation, peut être considérée comme du bruit.

La variabilité environnementale due au locuteur peut tout d'abord être de nature physiologique. Ainsi, un système mécanique provocant une déformation du conduit vocal provoquera immanquablement une variation dans le signal de parole produit (voir figure 1.2).

Figure 1.2 Variabilité due a la différence des locuteurs

I.4.4 Variabilité due aux conditions d'enregistrement

Pour appliquer dans le commerce un système de reconnaissance des locuteurs, il est important de connaître les effets de la transmission téléphonique sur un signal sonore.
La transmission de la parole par un canal téléphonique entraîne une limitation dans la gamme de fréquence, de 300 Hz à 3400 Hz de la bande passante.
La caractéristique de transfert n'est pas plate mais change de forme selon la ligne sélectionnée. Les spectres fournis par les lignes téléphoniques sont donc limités par la bande passante et également multipliés par une fonction de transfert de forme inconnue. Dans un premier stade, les études ont montré que la limitation des spectres de, longue durée à la bande passante caractérisant la qualité du téléphone n'affecte pas sensiblement le taux d'identification.
Cependant, la pondération des spectres par des fonctions arbitraires du transfert, détruit la fiabilité de l'identification parce que, dans certains cas, l'effet de la fonction de transfert sur les spectres est plus important que les caractéristiques des voix (voir figure 1.3).

Figure 1.3 Variabilité due a l'enregistrement

I.5. Les méthodes de la RAP

Un système de reconnaissance de la parole reçoit en entrée un signal acoustique et délivre en sortie une suite de symboles discrets. [2]

Il existe deux méthodes différentes :

I.5.1. La reconnaissance globale

L'unité de base considérée est le mot donc elle évite la segmentation. Tous les mots prononcés sont supposés être séparés par des silences de quelques dixièmes de secondes. Les images acoustiques des mots sont isolées les unes des autres à partir de la courbe d'énergie du signal vocal. Après stockage des mots dans un dictionnaire, la comparaison choisit le mot le plus proche [2]

I.5.2. La reconnaissance analytique

Cette approche considère, par contre, comme unité : les phonèmes, les syllabes, les diphones, les allophones, etc. Le problème majeur est le phénomène de coarticulation (les phonèmes ne sont pas isolés mais liés les uns aux autres), la réalisation acoustique d'un phonème change avec celui qui le précède et qui le suit. Cette tâche concerne des informations acoustiques, phonétiques et linguistiques. Cette approche est utilisée dans la reconnaissance de la parole continue. [2]

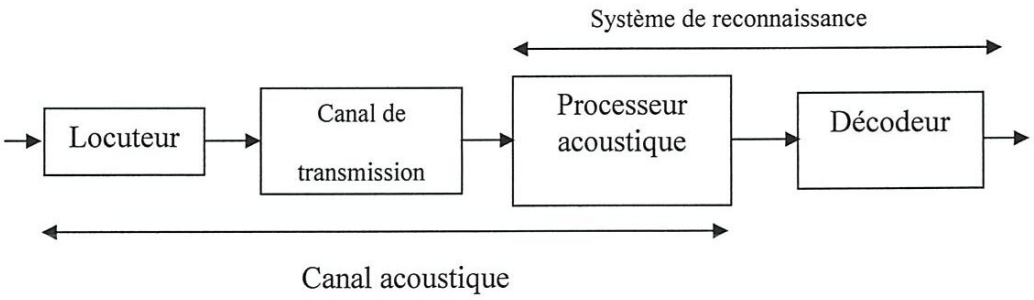


Figure I.4 Eléments intervenants dans la reconnaissance.

I.6. Les approches de la RAP

I.6.1. La comparaison dynamique

Elle permet d'intégrer des séquences de mots ou de phonèmes par l'alignement temporel. Les deux systèmes de comparaison dynamique les plus utilisés sont :

La programmation dynamique (ou en anglais DTW) : Il s'agit de mettre en correspondance deux formes afin d'obtenir une coïncidence optimale au sens d'un critère défini en fonction du domaine étudié (spectre dans la parole).

La DTW, Dynamic time warping, consiste à effectuer une normalisation temporelle au cours de la phase de comparaison entre un mot test et les mots références ; déterminer le début et la fin des formes acoustiques. La DTW comprime les mots dans le temps pour les faire coïncider à un modèle fixe. [2]

I.6.2. Les modèles markoviens

Chaque mot est représenté par un modèle de MARKOV caché. Le critère d'apprentissage le plus utilisé est celui du maximum de vraisemblance. On cherche à déterminer les paramètres des modèles de façon à minimiser la probabilité de générer l'ensemble des mots ou phonèmes du corpus d'apprentissage. Ce modèle autorise l'élision et l'ajout de plusieurs trames ; il peut donc coder l'aspect temporel d'une forme et la variabilité de la parole. [2]

I.6.3. Les systèmes hybrides

Les systèmes hybrides connexionnistes et markoviens sont certainement une voie de recherche prometteuse. D'un point de vue théorique, de nombreux travaux de recherches sont effectivement encore nécessaires pour améliorer les algorithmes d'apprentissage des réseaux et les structures actuelles. Enfin, d'un point de vue expérimental, une validation de cette approche sur une autre base de données devrait permettre d'avoir plus de certitude quant à l'apport des réseaux et des méthodes hybrides dans le domaine de la reconnaissance de la parole. [2]

I.6.4. Les modèles connexionnistes

La modélisation par les réseaux de neurones permet de prendre en compte la variabilité de la parole. La classification des formes est effectuée en utilisant différentes sources d'information de manière auto-organisée. Ce mécanisme qui permet de combiner les connaissances de diverses origines est représenté par les coefficients de pondération du réseau. Cette approche permet donc de passer de l'extraction des caractéristiques du signal à la reconnaissance phonétique en faisant interagir différents indices acoustiques sans avoir à expliciter la stratégie de contrôle à adopter pour combiner au mieux plusieurs attributs phonétiques. [2]

I.7. Conclusion :

De nos jours, la reconnaissance de la parole est un des domaines les plus importants dans la reconnaissance des formes et l'intelligence artificielle. La classification est un outil d'aide à reconnaître la parole, on en trouve les méthodes classiques qui on été largement utilisées mais actuellement nous avons les réseaux de neurones qui ont donné de bons résultats.

Grâce à cette technologie, on peut communiquer oralement avec la machine au lieu d'utiliser les gestes ou les commandes des automatismes, ce qui facilite considérablement l'interaction homme/ machine.

Chapitre II

Traitement Automatique de la parole

II.1. Introduction

Ce chapitre a pour objectif la présentation de quelques notions du signal de parole ainsi que la description des mécanismes de sa production. La parole est constituée de plusieurs éléments appelés phonèmes dépendants les uns des autres. Ces phonèmes peuvent être caractérisés par leurs aspects temporels et fréquentiels. L'analyse de la parole nous permet de déterminer certaines caractéristiques suffisamment pertinentes pour pouvoir ensuite les considérer dans un traitement postérieur. Nos applications porteront sur le signal de parole produit par les hommes. A cet effet, nous exposerons l'appareil phonatoire et auditif de l'être humain. [3]

La parole est un signal réel, continu, d'énergie finie et non stationnaire. Sa structure est complexe et variable avec le temps. [3]

Le système auditif humain est surtout sensible dans une gamme de fréquence située entre 800 Hz à 8 kHz; les limites extrêmes sont respectivement 20 et 20 kHz. [3]

Par contre, le système vocal est encore plus limité, en résumé, pour des sons vocaliques à des fréquences au-dessus de 4 kHz, les hautes fréquences sont plus de 40 kHz en dessous du sommet du spectre. [3]

L'information portée par le signal de parole peut être analysée de bien de façons. On en distingue généralement plusieurs niveaux de description non exclusifs : acoustique, phonétique, phonologique, morphologique, syntaxique et sémantique. [3]

II.2. Etude Phonétique

La parole est le résultat de l'action volontaire et coordonnée des appareils respiratoire et masticatoire. Cette action se déroule sous le contrôle du système nerveux central qui reçoit en permanence des informations par rétroaction auditive et par les sensations cénesthésiques.

L'appareil respiratoire fournit l'énergie nécessaire lorsque l'air est expiré par la trachée artère. Au sommet de celle-ci se trouve le larynx où la pression de l'air est modulé avant d'être appliqué au conduit vocal, qui s'étend du pharynx jusqu'aux lèvres. La figure II.1 représente un schéma simplifié de l'appareil phonatoire qui produit le signal de parole. [4].

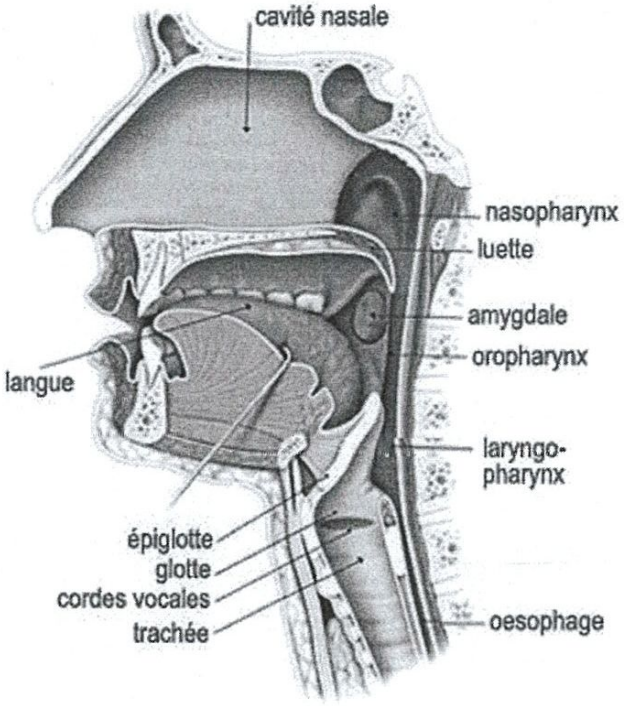


Figure II.1 L'appareil phonatoire

II.3. Les Acteurs de Production

II.3.1. Le Conduit Vocal

Le conduit est un ensemble de cavités situées entre la glotte et les lèvres, il peut être considéré comme un tuyau sonore tridimensionnel d'une longueur approximative de 17cm. Cette cavité est caractérisée par ses fréquences de résonances.

II.3.2. La Source Vocale

Il existe deux sortes de sources d'excitation : la source de bruit ou d'explosion et la source laryngienne. Celle-ci est le larynx qui est situé dans la région moyenne du cou, juste en arrière de la langue, au dessous de l'os hyoïde, au devant de la colonne vertébrale.

II.3.3. Les Cordes Vocales

Les cordes vocales sont des replis musculaires (la muqueuse, le ligament les muscles thyro-arytenoidiens) situés au niveau de la glotte. Les cordes vocales vibrent sous l'effet du passage de l'air à travers la glotte ou précisément sous l'effet de la dépression de part et d'autre de l'espace glottique (pression intra glottique). [4].

II.4. Audition

Dans le cadre du traitement de la parole, une bonne connaissance des mécanismes de l'audition et des propriétés perceptuelles de l'oreille est aussi importante qu'une maîtrise des mécanismes de production. En effet, tout ce qui peut être mesuré acoustiquement ou observé par la phonétique articulatoire n'est pas nécessairement perçu.

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

L'appareil auditif comprend :

- L'oreille externe
- L'oreille moyenne
- L'oreille interne

Le conduit auditif relie le pavillon au tympan : c'est un tube acoustique de section uniforme fermé à une extrémité, son premier mode de résonance est situé vers 3 kHz ; ce qui accroît la sensibilité du système auditif dans cette gamme de fréquences.

La mécanique de l'oreille interne (marteau, étrier, enclume) permet une adaptation d'impédance entre l'air et le milieu liquide de l'oreille interne. Les vibrations de l'étrier sont transmises au liquide de la cochlée (figure II.2). Celle-ci contient la membrane basilaire qui transforme les vibrations mécaniques en impulsions nerveuses.

La membrane s'élargit et s'épaissit au fur et à mesure que l'on se rapproche de l'apex de la cochlée ; elle est le support de l'organe de corti qui est constitué par environ 25000 cellules ciliées raccordées au nerf auditif. La réponse en fréquence du conduit au droit de chaque cellule est esquissée à la figure II.3.

La fréquence de résonance dépend de la position occupée par la cellule sur la membrane au-delà de cette fréquence, la fonction de réponse s'atténue très vite. Les fibres nerveuses aboutissent à une région de l'écorce cérébrale appelée aire de projection auditive et située dans le lobe temporel. En cas de lésion de cette aire, on peut observer des troubles auditives. Les fibres nerveuses auditives afférentes (de l'oreille au cerveau) et efférentes (du cerveau vers l'oreille) sont partiellement croisées : chaque moitié du cerveau est mise en relation avec les deux oreilles internes. [4].

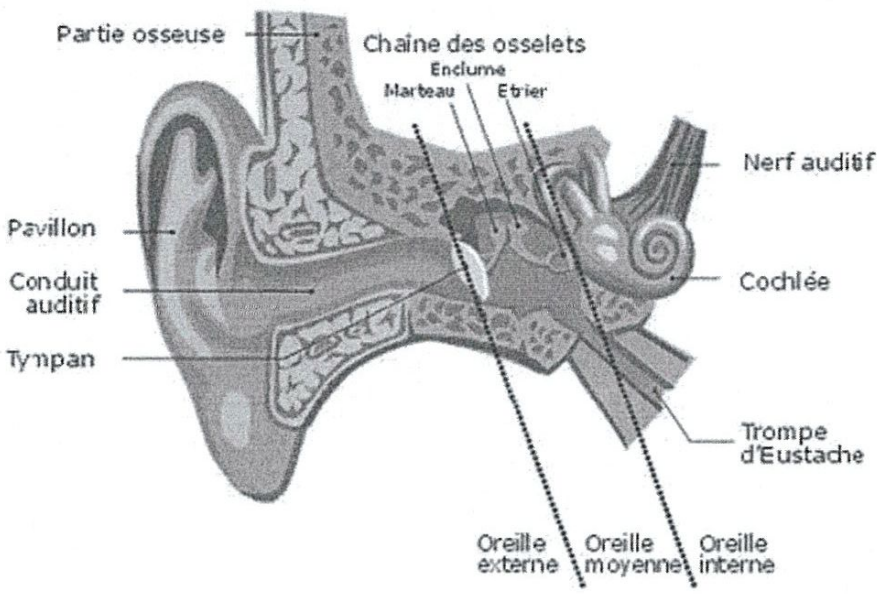


Figure II.2 l'appareil auditif humain.

Ainsi, l'oreille ne répond pas également à toutes les fréquences. La figure II.3 présente le champ auditif humain, délimité par la courbe de seuil de l'audition et celle du seuil de la douleur. Sa limite supérieure en fréquence (16 kHz, variable selon les individus) fixe la fréquence d'échantillonnage maximale utile pour un signal auditif (32 kHz). [2].

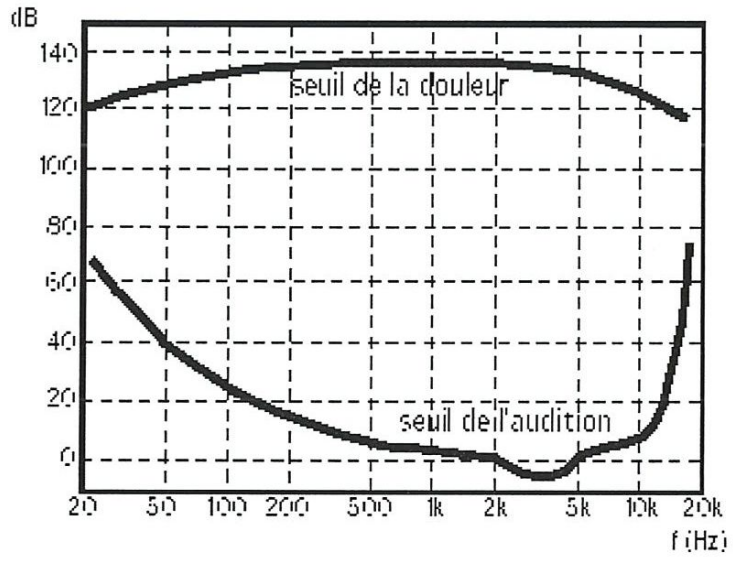


Figure II.3 Le champ auditif humain.

II.4.1. L'échelle des Mels

L'échelle des Mels est une échelle biologique permettant la modélisation de l'oreille humaine. L'échelle des Mels permet de modéliser une perception de l'oreille linéairement avant 1000 Hz puis logarithmiquement.

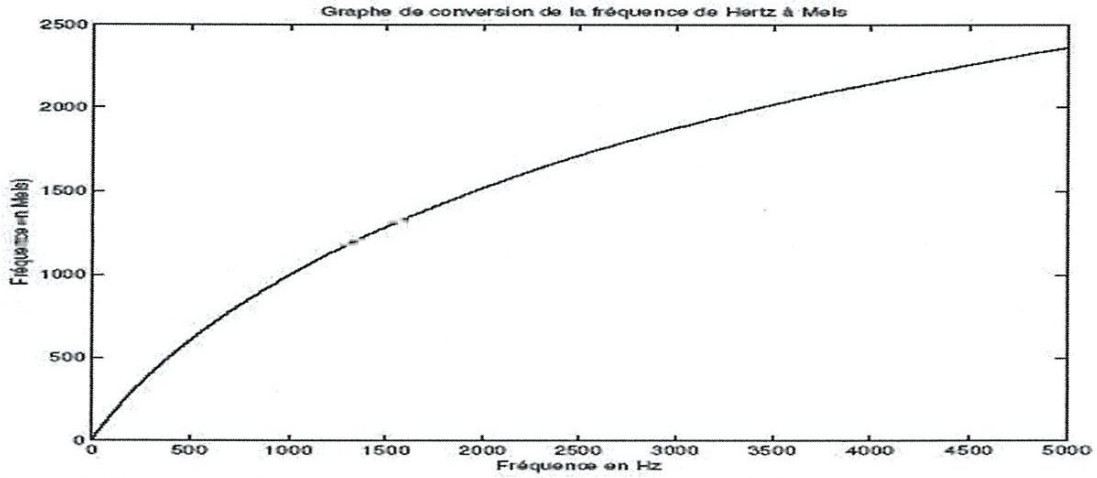


Figure 1.10 : Graphe de conversion de la fréquence d'hertz à Mels

On remarque qu'avant 1000 Hz, la courbe est à peu près droite, ce qui traduit bien l'équivalence entre Hz et Mels à ces fréquences.

On rappel la formule donnant la fréquence en Mels à partir de la fréquence en Hz :

$$m = 2595.\log_{10}\left(1+\frac{f}{700}\right)$$

Ou f est la fréquence en Hertz.

Et m est la fréquence en Mels.

II.5. Le prétraitement de la parole

Le traitement de la parole est aujourd'hui une composante fondamentale des sciences de l'ingénieur. Située au croisement du traitement du signal numérique et du traitement du langage (c'est-à-dire du traitement de données symboliques), cette discipline scientifique a connu depuis les années 60 une expansion fulgurante, liée au développement des moyens et des techniques de télécommunications. [5].

De nature redondante et très complexe, la parole nécessite une représentation et un traitement particuliers, dans le but de minimiser sa redondance d'une part, et de l'adapter au milieu des machines d'autre part. [5].

Un prétraitement d'un signal vocal est réalisé en trois étapes :

- Acquisition.
- Préaccentuation.
- Fenêtrage.

II.5.1. Acquisition

Le traitement de la parole suppose toujours en premier lieu une analyse du signal vocal converti au préalable en signal électrique par un microphone ; Puisque les ordinateurs ne peuvent pas manipuler des sources analogiques, on doit convertir les signaux au format numérique avec un convertisseur A/N. Le processus inverse se refait par le convertisseur N/A. De cette sorte, on va pouvoir travailler sur une représentation spectrale du signal, décomposant ses différentes fréquences avec leurs amplitudes et leurs harmoniques, aboutissant à des «traits » qu'on appelle formants du signal.

La raison d'une analyse numérique c'est qu'elle est plus aisée pour un traitement sophistiqué et qu'elle est beaucoup plus fiable. Le développement rapide des ordinateurs et des circuits intégrés en conjonction avec la croissance des communications numériques a encouragé l'application des techniques numériques au traitement du signal.

La conversion analogique/numérique consiste en l'échantillonnage, la quantification et le codage. L'échantillonnage est le processus de représentation d'un signal continûment variable comme une séquence de valeurs. La quantification conduit à représenter approximativement chaque échantillon dans un ensemble fini de valeurs. Le codage consiste à assigner un numéro réel à chaque valeur.

Avant l'échantillonnage, un filtre passe-bas de fréquence de coupure égale à la moitié de la fréquence d'échantillonnage est inséré pour éviter l'effet dénommé «repliement» ou

«aliasing» postulé par le théorème de Nyquist-Shannon ; Ce filtre est donc appelé filtre « anti-repliement » ou «anti-aliasing».

Il y a deux paramètres qui affectent la qualité du son. Le premier est la fréquence d'échantillonnage (sampling rate) : on la mesure en Hertz (Hz) et des valeurs typiques pour le son sont 4 kHz, 8 kHz, 11.025 kHz, 22.05 kHz, 44.1 kHz et 48 kHz. Cependant d'après le théorème de Shannon, il faut choisir cette fréquence un peu plus grande que la moitié de la bande intéressante parce que les composants électroniques ne sont pas idéaux et qu'il est donc impossible de réaliser un filtre parfait. Le deuxième paramètre qu'affecte la qualité est la quantification en un nombre de bits fixé. Typiquement, ce nombre varie entre 8, 12, 14 ou 16 bits et détermine la dynamique et le rapport signal/bruit. Généralement, il s'agit d'une représentation uniforme mais une amélioration peut être obtenue avec des quantifications non linéaires.

Dans les ordinateurs les fréquences 11025, 22050 et 44100 Hz sont les plus couramment utilisées. [5].

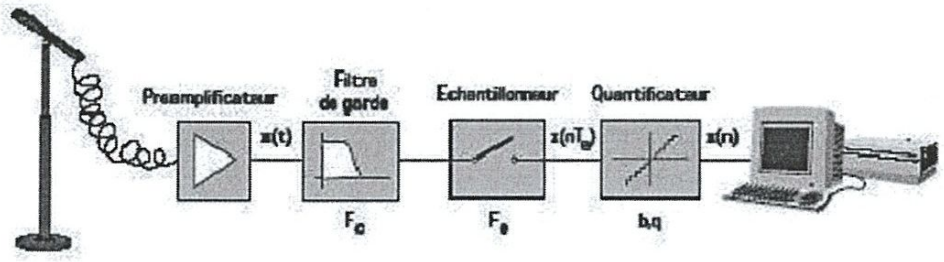


Figure II.4 Acquisition

II.5.2. Classification des sons

Suivant les organes de l'appareil phonatoire mis en jeu et leurs excitations, on peut classer les sons produits dans des différentes classes, parmi ces dernières il y a deux classes de sons importantes : [7]

II.5.2.1. Les sons voisés : les voyelles par exemple

Ce sont des sons ayant une forme quasi-périodique dont la représentation temporelle est illustrée sur la figure II.5.

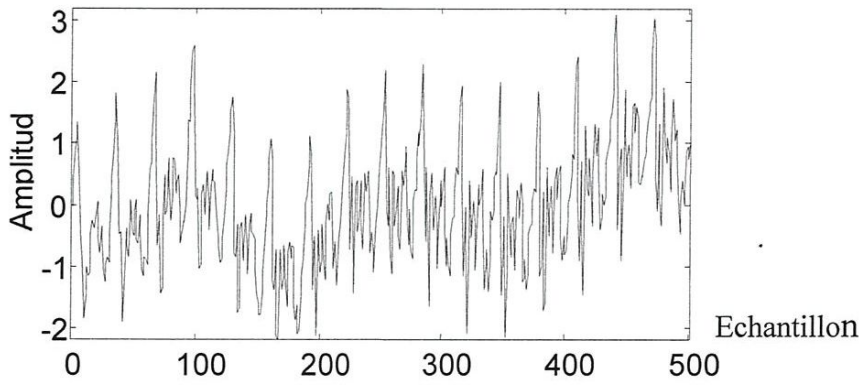


Figure. II.5 La forme d'un son voisé

Le spectre d'un signal voisé (figure II.6) présente la particularité d'avoir une fréquence appelée « *FONDAMENTALE* » ou *PITCH* (la première raie sur la figure II.6). Les autres correspondent aux harmoniques du pitch, dont l'enveloppe de ces raies présente des maxima appelés *FORMANTS*. [7]

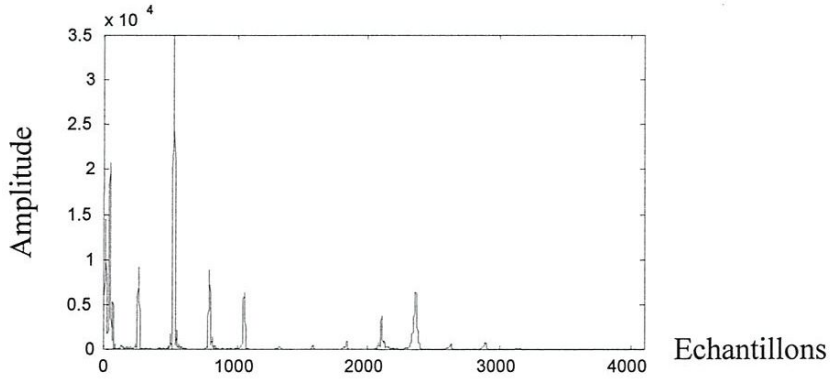


Figure II.6 Le spectre d'un son voisé

II.5.2.2. Les sons non voisés

Ces sons ne présentent pas de structure périodique (figure II.7), ils peuvent être considérés comme un bruit blanc filtré par la transmittance de l'appareil phonatoire. On remarque aussi que leur spectre ne présente pas de structure de pitch. (figure II.8). [7]

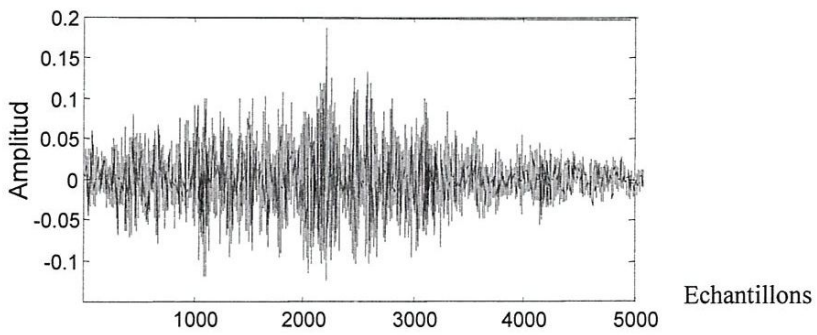


Figure II.7 La forme d'un son non

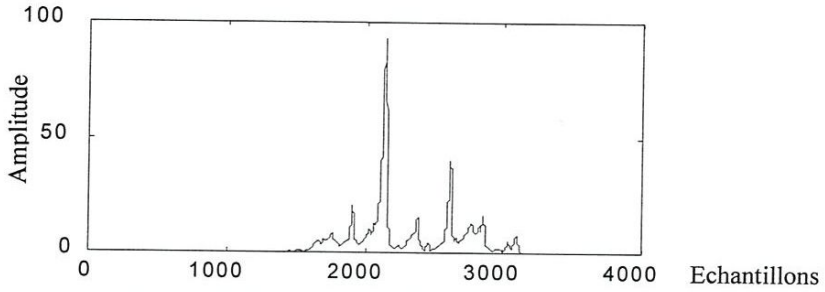


Figure II.8 Le spectre d'un son non voisé

II.5.3. Préaccentuation

En général, le signal vocal se caractérise par une perte de 6 dB/Octave, due à l'influence de la source d'excitation et au rayonnement des lèvres. Une perte de 6 dB/Octave veut dire que les hautes fréquences ont une énergie plus faible que celle des basses fréquences. Pour palier à cet inconvénient la préaccentuation permet d'égaliser les sons aigus avec les sons graves (figure II.9).

Le procédé le plus simple est d'appliquer un filtre de pré-accentuation donné par la fonction de transfert :

$$H(Z)=1-\mu*Z^{-1} \text{ où } 0\leq\mu\leq1.$$

Dans le domaine des signaux discrets (*échantillonnés*) $S(n)$ le problème consiste habituellement à calculer : $Y(n)=X(n)-\mu*X(n-1)$ pour $n\geq0$.

Le facteur de préaccentuation**μ,** est pris entre 0.9 et 1 (souvent **0.95**). Comme conséquence, la préaccentuation introduit une légère distorsion spectrale. [2]

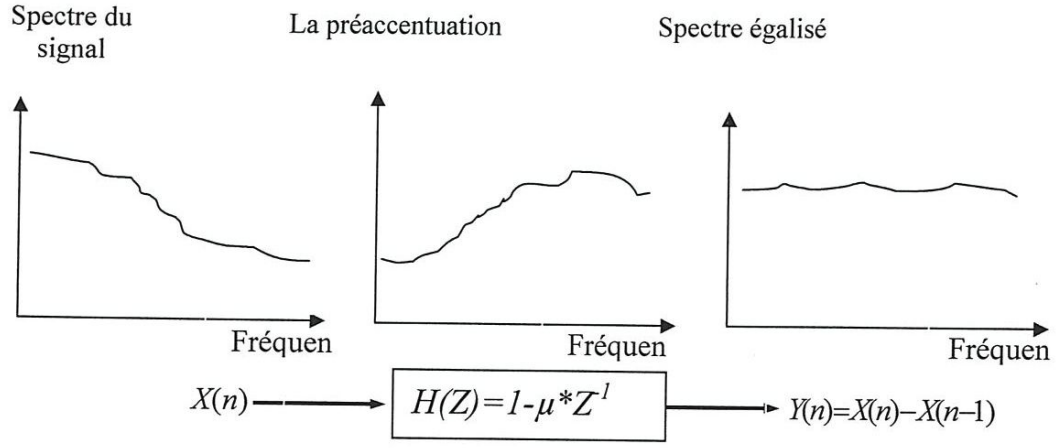


Figure II.9 La Préaccentuation

II.5.4. Fenêtrage

En général, les signaux ne sont pas stationnaires : hauteur, intensité et timbre différents se superposent et/ou se succèdent au cours du temps.

En réalité, le signal vocal n'est pas stationnaire et le conduit vocal se déforme de façon continue.

Lorsque l'on souhaite effectuer un traitement fréquentiel, il est donc nécessaire de ne pas considérer le signal globalement mais sur des fenêtres suffisamment petites (chacune de durée 30 ms environ), pour que le signal soit approximativement stationnaire, localement dans une fenêtre donnée.

Les signaux sur chacun des segments sont ensuite fenêtrés par une fenêtre de pondération *w*. Habituellement on utilise une fenêtre de Hann (ou Hanning), Hamming ou Blackman, qui ont chacune des propriétés différentes. Remarquons que l'absence de fenêtrage revient à utiliser une fenêtre rectangulaire qui présente un lobe principal très fin mais aussi des lobes secondaires d'amplitudes élevées. [6]

Il existe, dans la littérature, différents types de fenêtres ; on citera

- **Fenêtre de Hamming généralisée**

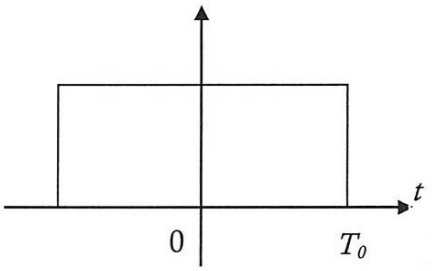
La fenêtre de Hamming généralisée à pour équation :

$$\begin{cases} f_n = \alpha + (1-\alpha).\cos\left(\dfrac{2\pi.n}{N}\right) \ pour \ \ 0 \le n \le N-1 \\ f_n = 0 \ \ ailleurs \end{cases}$$

- **Fenêtre rectangulaire**

On peut également considérer cette fenêtre comme étant le cas particulier de la fenêtre de Hamming généralisée pour α= 1

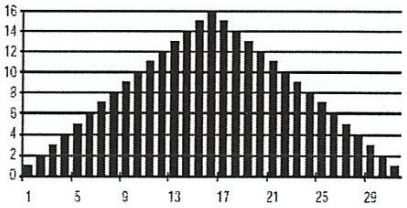
Fenêtre rectangulaire : $h(t) = \begin{cases} 1 \ si \ T \ \ [0 \ T] \\ 0 \ \ \ sinon \end{cases}$



- **Fenêtre triangulaire**

La fenêtre triangulaire peut être vue comme résultat de la convolution de deux fenêtres rectangulaires de longueur N/2.

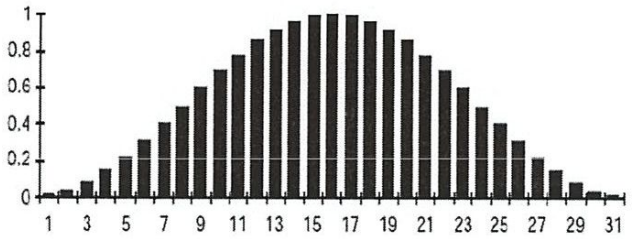
Fenêtre triangulaire (de Bartlett) : $\begin{cases} \frac{2t}{T} & \text{si } t\ [0\ \frac{T}{2}] \\ 2T - \frac{t}{T} & [\frac{T}{2}T] \\ 0 & \text{si non} \end{cases}$



❖ **Fenêtre de Hann**

La fenêtre de Hann (ou Hanning) est le cas particulier de la fenêtre de Hamming généralisée pour $\alpha = 0{,}5$

Fenêtre de Hann $h(t) = \begin{cases} 0.5 - 0.5\cos 2\pi\frac{t}{T} & \text{si } t \in [0, T] \\ 0 & \text{sinon.} \end{cases}$



Parmi ces fenêtres, la fenêtre de Hamming est la plus convenable à la parole, car elle entraîne un minimum de distorsion spectrale du signal de parole, par rapport aux autres fenêtres. (Atténuation du rapport du lobe principal au lobe secondaire est égale a –41 dB, c'est à dire que la concentration de l'énergie dans le lobe principal est égale à 99.96%). [7]

$T_e = 1 / f_e$ est la période d'échantillonnage.

T est la moitié de la longueur de la fenêtre.

II.6. Méthodes d'analyse d'un signal vocal

Plusieurs approches ont été proposées pour l'analyse de la parole, ayant toutes pour but d'extraire le minimum d'information pouvant définir complètement le signal de parole. Parmi ces approches il y a celles qui agissent dans le domaine fréquentiel, celles qui agissent dans le temps, d'autres permettant d'avoir une analyse conjointe temps-fréquence, et finalement les méthodes basées sur la modélisation du système des phonations. [2]

II.6.1. Analyse spectrale

Elles sont fondées sur la décomposition fréquentielle du signal sans connaissance a priori de sa structure fine. La plus utilisée est celle utilisant la transformée de Fourier, appelée

Fast Fourier Transform (FFT). Tout son est la superposition de plusieurs ondes sinusoïdales. Grâce à la FFT, on peut isoler les différentes fréquences qui le composent. Il y a aussi le vocodeur à canaux. [7]

II.6.1.1. La transformation de Fourier

Lorsqu'on veut analyser une entité complexe, une des procédures largement utilisées consiste à la décomposer en une somme d'entités plus simples. L'idée donc est d'exprimer le signal vocal par une combinaison linéaire discrète de fonctions élémentaires de forme simple ; c'est le cas de la transformée de Fourier discrète pour l'estimation du spectre.

La TFD est définie par $S(n) = \sum_{k=0}^{N-1} s(k) \times e^{-j\pi \frac{nk}{N}}$

s(k) : est un signal numérique.

N : est la longueur du support du signal.

Pratiquement la TFD est évaluée par un algorithme rapide appelé FFT (Fast Fourier Trasform). Elle s'opère sur des durées limitées du signal vocal, en prélevant les échantillons de parole à l'aide d'une fenêtre temporelle glissante. En général les fenêtres successives se recouvrent. [13]

Ces fenêtres doivent avoir une largeur si l'on veut que la FFT ait un sens : en général, on prend 256 à 512 points, le recouvrement est par exemple la moitié soit 128 ou 256 respectivement.

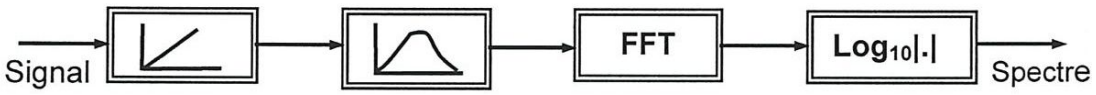


Figure II.10 Traitement par transformé de Fourier

II.6.1.2. Le banc de filtres (vocodeur à canaux)

L'estimation de l'enveloppe spectrale du signal peut se faire à l'aide des filtres en découpant la bande utile en sous-bandes (canaux), dans lesquelles on évalue l'intensité du signal. A la limite, lorsque les sous-bandes sont de largeur nulle, on tend vers une transformée de Fourier si les filtres sont idéaux. Le schéma classique d'un vocodeur à canaux est représenté sur la figure II.11 [7]

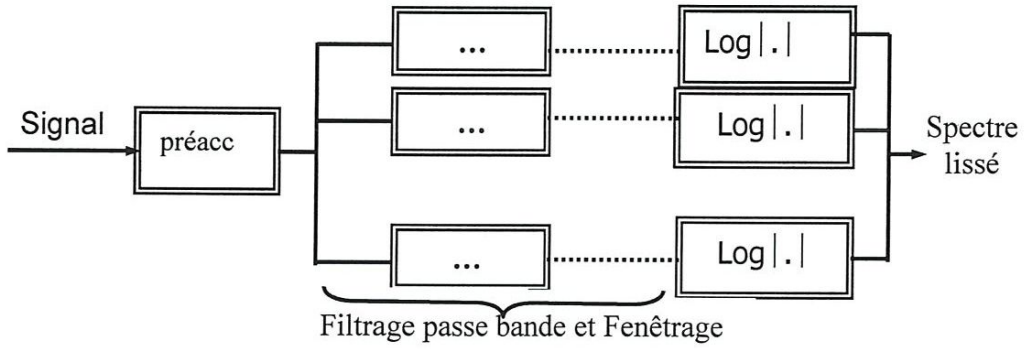


Figure II.11 - Analyse par vocodeur à canaux

II.6.2. Analyse temporelle

II.6.2.1. Energie totale

C'est l'énergie correspondante à la puissance du signal. Elle est souvent évaluée sur plusieurs trames de signal successives pour pouvoir mettre en évidence des variations.

Elle est évaluée par $E = \frac{1}{N}\sum_{k=0}^{N-1} s^2(k)$

N : est la longueur de la fenêtre.

S(k) : est le signal à traité.

Elle joue un rôle important pour délimiter les mots, et la localisation des voyelles qui sont caractérisées par une forte énergie par rapport aux autres phonèmes.

II.6.2.2. La densité de passage par zéro (DPZ)

Ce critère consiste à compter le nombre de passage par zéro pour des trames identiques à celles définies précédemment en commençant du début de l'enregistrement, respectivement à la fin. Si ce nombre dépasse un certain seuil, calculé expérimentalement à partir d'échantillons de silence, alors on est en présence de parole ; début du mot, respectivement fin du mot.

$$DPZ = \frac{1}{2}\sum_{k=0}^{k-1} \left| sign\ (s(k+1)) - sign\ (s(k)) \right|$$

Elle est utilisée pour distinguer le signal de parole du silence et un son voisé d'un son non voisé.

II.6.3. Analyse basée sur la modélisation de la parole

Dans ce domaine on trouve essentiellement les systèmes de prédiction linéaire noté LPC et les systèmes de traitement homomorphique (cepstral) ; Elles sont applicables sur des modèles de parole basés sur une connaissance à priori de la production d'un signal de parole. Avant de présenter ces deux méthodes, il est utile d'examiner l'élaboration d'un modèle de parole. [7]

II.6.3.1. Modèle de la parole

La parole est formée par l'excitation du conduit vocal. On peut considérer le conduit vocal comme un système variant dans le temps, qui impose ses propriétés de transfert selon la forme d'excitation qui lui est appliquée. Si on admet que les excitations du conduit vocal sont relativement indépendantes, la production de la parole peut se résumer dans le modèle de la figure II.12. Dans ce modèle la source d'excitation est soit un générateur d'impulsions périodiques, avec une période dont l'inverse est appelé le fondamental (pour les sons voisés), soit un générateur de bruit blanc (pour les sons non voisés).

Le conduit vocal peut être modélisé par un filtre numérique variant dans le temps, dont les coefficients varient d'une tranche de temps à l'autre.

Un contrôle de gain entre la source et le conduit ajoute une flexibilité supplémentaire pour le niveau sonore. [7]

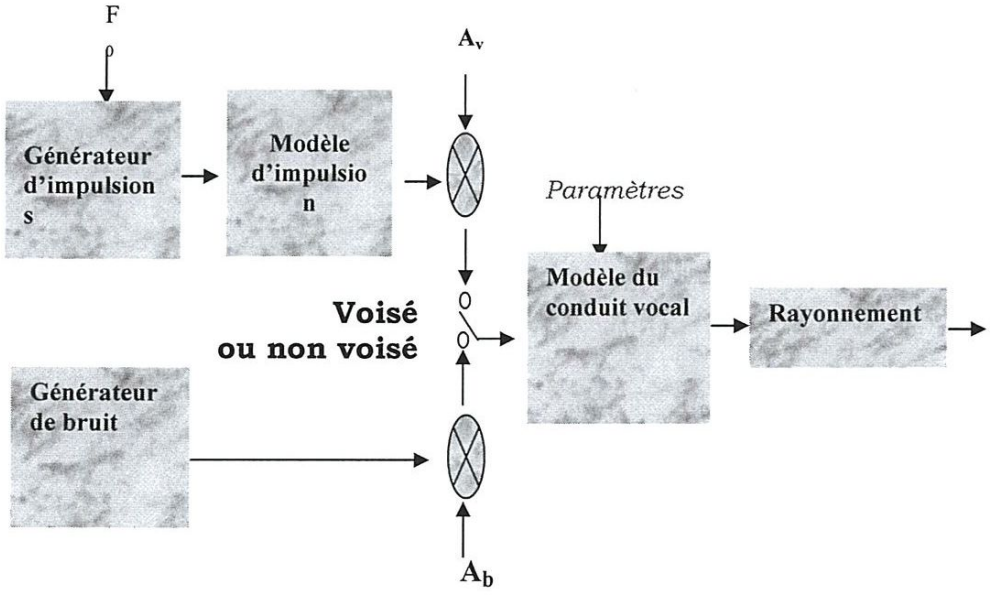


Figure II.12 Modèle de production de la parole

II.6.4. Analyse homomorphique (Cepstrale)

II.6.4.1. Calcul du cepstre complexe

Le défaut majeur de la FFT pour le calcul du spectre vocal, réside dans l'intermodulation source/conduit qui rend difficile la mesure des formants et du fondamental. L'analyse cepstrale est une méthode qui vise à séparer leurs contributions respectives par déconvolution. Pour cela on fait l'hypothèse que le signal vocal x(n) est produit par un signal excitateur g(n) (source glottique) traversant un système linéaire passif de réponse impulsionnelle h(n) (conduit vocal). [7]

D'après ces hypothèses, on aura le système suivant :

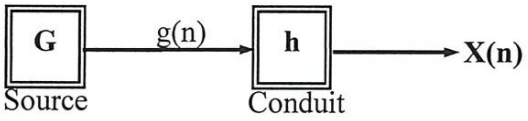


Figure II.13 Modèle source-filtre

Donc on peut écrire pour tout n >0 :

$$x(n) = g(n) * h(n)$$

Pour déconvoluer x(n), c'est à dire pour retrouver les deux composantes g(n) et h(n), il faut se donner une classe de fonctions admissibles pour g(n) (ou pour x(n)), si on suppose que g(n) est une séquence d'impulsions (périodique pour les sons voisés). Il est évident que l'ensemble de ces hypothèses est très limitatif : en toute rigueur cette analyse ne s'applique théoriquement qu'aux parties stables des sons périodiques (voyelles longues, par exemple), dans la pratique, cependant, cette méthode fournie des résultats acceptables sur l'ensemble du signal. [2]

Pour déconvoluer plus aisément x(n), il suffit de transposer le problème par homomorphisme dans un espace où l'opérateur de convolution « * » correspond à un opérateur d'addition « + ». Soit D_*^+ cet homomorphisme.

D_*^+ est un homomorphisme (application) qui applique l'espace vectoriel des signaux d'entrées muni de la loi « * » (convolution), sur l'espace vectoriel.

Des signaux de sortie munie de la loi « + » (addition), donc on est en face de la situation suivante :

$$x(n)=g(n)*h(n) \xrightarrow{D_{*}^{+}} \hat{x}(n)=\hat{g}(n)+\hat{h}(n)$$

Après séparation de $\hat{g}(n)$ et de $\hat{x}(n)$ si la transformation inverse D_{+}^{*} existe, on aura :

$$\hat{g}(n) \xrightarrow{D_{+}^{*}} g(n)$$

$$\hat{h}(n) \xrightarrow{D_{+}^{*}} h(n)$$

L'intérêt de la méthode réside dans le fait que $\hat{g}(n)$ et $\hat{h}(n)$ sont facilement séparables par un filtrage temporel est ceci grâce à l'hypothèse simplificatrice sur g(n).

Les homomorphismes D_{*}^{+} et D_{+}^{*} sont inverses l'un de l'autre, et se définissent par :

$$D_{*}^{+}=TZ(\,)\,0Log(\,)\,0TZ^{-1}(\,)$$

$$D_{+}^{*}=TZ(\,)\,0Exp(\,)\,0TZ^{-1}(\,)$$

Ce qui donne le système schématisé dans la figure II.14

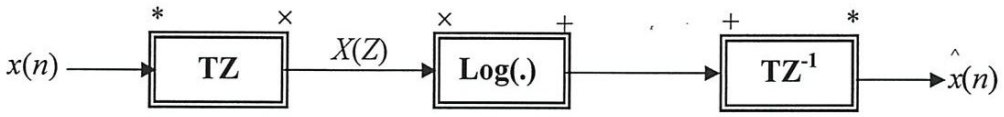


Figure II.14 Calcul du cepstre complexe

Où :

- TZ est la transformée en Z (TZ^{-1} sa transformée inverse).
- La fonction log est utilisée pour le passage du domaine de la loi «.» (La multiplication) au domaine de la loi « + » (l'addition), cette fonction n'est valable que pour les signaux positifs, toutefois, étant donné que la majorité des signaux courants sont bipolaires (positifs et négatifs), donc il faut faire appel à la fonction log complexe, soit :

$$X(Z)=|X(Z)|\times\exp[jArg(X(Z))]$$

Donc :

$$\hat{X}(Z)=\log[X(Z)]=\log|X(Z)|+jArg[X(Z)]$$

- exp : doit être aussi la fonction exponentielle complexe.

D'après le schéma de la figure II.14 on a :

$$X(Z)=TZ[x(n)]$$

$\hat{X}(Z)=\log[X(Z)]$ (La fonction log est complexe)

$$\hat{x}(n)=TZ^{-1}[X(Z)]$$

Tous ceci peuvent être résumés par la notation suivante :

$$\hat{x}(n)=D_{*}^{+}[x(n)]$$

Le signal $\hat{x}(n)$ est appelé cepstre complexe associé au signal x(n).

II.6.4.2. Ambiguïté de la phase

Le problème qui se pose ici est que Arg [X(Z)] n'est défini qu'à 2π près (la valeur principale), c'est à dire que l'on peut ajouter un multiple entier de 2π à la partie imaginaire du log complexe sans changer le résultat.[2]

Ceci montre que l'homomorphisme tel qu'il est défini n'est pas une transformation biunivoque. Pour contourner ce problème, on a introduit la notion du cepstre réel.

II.6.4.3. Définition du cepstre réel

La difficulté du logarithme complexe (à cause de la phase) peut être levée dans le cas de la parole (où l'on ne s'intéresse que rarement à l'information de phase) en prenant un log module (log|.|), ce qui garantie l'inversibilité sans calcul particulier de la phase. [2]

Soit DM_{*}^{+} cet homomorphisme et DM_{+}^{*} son inverse :

$$DM_{*}^{+}=TZ(.)\circ\log|.|\circ TZ^{-1}(.)$$

$$DM_{+}^{*}=TZ(.)\circ\exp|.|\circ TZ^{-1}(.)$$

Or la procédure de déconvolution décrite antérieurement exige que le signal observé soit bien sûr un produit de convolution. C'est la raison pour laquelle on fait souvent l'hypothèse que la fenêtre w(n) recouvre un nombre suffisant M de périodes du fondamental. Dans ce cas, sa variation est faible sur la durée effective de la réponse impulsionnelle ce qui permet d'écrire :

$$x(n) \cong [p(n) * w(n)] \bullet h(n)$$

Soit : $$P_w(n) = p(n) \bullet w(n) = \sum_{k-0}^{M-1} w(kp_0)\ \delta(n - kp_0)$$

Il s'ensuit que $$P(e^{i\theta}) = W(e^{jp_0\theta})$$

Et donc $$\hat{p}_w(n) = \hat{w}(n/p_0)$$

A cause de l'ambiguïté de la phase, on a intérêt à considérer le cepstre réel, qui vaut :

$$\tilde{C}(n) = \tilde{p}_w(n) + \tilde{h}(n)$$

Où :

$\tilde{p}_n(n)$: est une séquence d'impulsions séparée de p_0 échantillons.

$\tilde{h}(n)$: décroît rapidement (en 1/n) avec n, et devient rapidement négligeable du moins pour $n \geq p_0$.

Dans ces conditions, on peut admettre que les premiers coefficients contiennent essentiellement la contribution du conduit vocal et que les pics périodiques ($\tilde{p}_w(n)$) reflètent les impulsions de la source.

On peut séparer la contribution de $\tilde{p}(n)$ par un filtrage temporel en utilisant tout simplement une fenêtre.

Soit F(n) cette fenêtre telle que : $F(n) = \begin{cases} 1 & n<p_0 \\ 0 & \text{autrement} \end{cases}$

Et finalement il ne reste que h(n), définissant le filtre numérique qui modélise le conduit vocal. [7]

II.6.5. Coefficients MFCC: (Mel-scaled Frequency Cepstral Coefficients)

Les paramètres MFCC sont des coefficients cepstraux obtenus à partir des énergies d'un banc de filtre en échelle de fréquence Mel. Il s'agit en faite d'un calcul classique des coefficients cepstraux auquel on a rajouté, avant le logarithme un filtre de Mel. Ces résultats sont intéressants, car le calcul d'une dizaine de coefficients cepstraux est alors suffisant pour des expériences de RAP. [2]

$MFCC_i = \sum_{k=1}^{20} X_k \cos \pi_i \frac{(k-0.5)}{20}$ avec i=1,2,..... p ; 20 est le nombre des filtres et p est le nombre des coefficients.

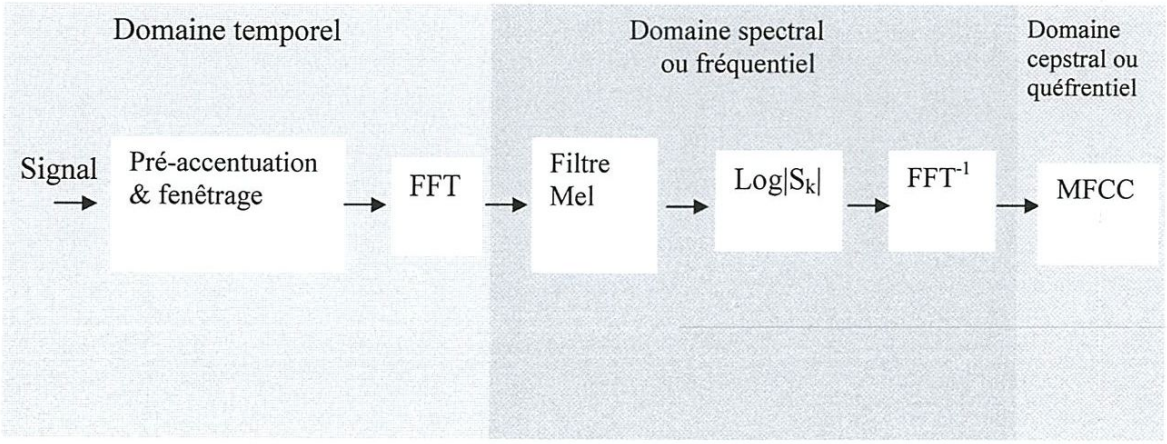


Figure II.16 calcul des coefficients MFCC

II.7. Conclusion

Nous avons présenté brièvement quelques notions sur la production de la parole, les mécanismes de production et la phonation. Le signal de la parole est le résultat de la vibration des cordes vocales après filtrage par le conduit vocal.

Un prétraitement d'un signal vocal est réalisé en trois étapes Acquisition, Préaccentuation, Fenêtrage.

Les paramètres MFCC sont des coefficients cepstraux obtenus à partir des énergies d'un banc de filtre en échelle de fréquence Mel.

Les différentes étapes pour l'obtention MFCC sont : fenêtrage, RFFT, filtre de Mel, calcul du log, $RFFT^{1-}$, MFCC.

Chapitre III

Réseaux de neurones

III.1. Introduction

Les dernières années ont vu un développement technologique puissant dans des domaines divers, et il y a eu un accroissement de besoin pour le contrôle et la gestion des systèmes complexes qui introduisent d'énorme calculs et un nombre de variables important ; d'où la nécessité de chercher de nouvelles méthodes pour une gestion plus souple et moins coûteuse en temps de calculs et en manipulation des variables dont le nombre ne cesse d'augmenter. Pour cela, on s'est intéressé de plus en plus aux systèmes qui apprennent, en utilisant des modélisations des neurones biologiques.

Les modèles de réseaux de neurones ou tout simplement réseaux de neurones, ont été étudiés pendant plusieurs années dans le but d'imiter les performances du cerveau de l'être vivant.

Les réseaux de neurones artificiels sont des réseaux fortement connectés de processeurs élémentaires fonctionnant en parallèle. Chaque neurone calcule une sortie selon les informations reçues .Ils sont adaptés au traitement en parallèle des problèmes complexes comme la reconnaissance vocale et faciale ainsi que la simulation de fonctions de transfert, etc. [8]

III.2. Historique

Les premières recherches sur les réseaux de neurones remonte à la fin du 19e et au début du 20e siècle. Elles consistent en des travaux multidisciplinaires en physique, en psychologie et en neurophysiologie, par des scientifiques tels Hermann Von Helmholtz, Ernst Mach et Ivan Pavlov. A cette époque, IL s'agissait de théorie plutôt générale sans modèle mathématique précis d'un neurone. On s'entend pour dire que la naissance du domaine des réseaux de neurones artificiels remonte aux années 1940 avec les travaux de Warren McCulloch et Walter Pitts qui ont montre qu'avec de tels réseaux on pouvait, en principe, calculer n'importe quelle fonction arithmétique ou logique. Vers la fin des années 1940, Donald Hebb a ensuite propose une théorie fondamentale pour l'apprentissage

La première application concrète des réseaux de neurones artificiels est survenue vers la fin des années 1950 avec l'invention du réseau dit <<perceptron >> par un dénomme Frank Rosenblatt. Malheureusement, il a été démontre par la suite que ce perceptron simple ne pouvait résoudre qu'une classe limitée de problème. Au même moment, Bernard Widrow et

Ted Hoff ont propose un nouvel algorithme d'apprentissage pour entraîner un réseau adaptatif de neurones linéaires, dont la structure et les capacités sont similaires au perceptron.

Vers la fin des années 1960, un livre publie par Marvin Minsky et Seymour Papert est venu jeter beaucoup d'ombre sur le domaine des réseaux de neurones. Entre autres, ces deux auteurs ont démontre les limitations des réseaux développes par Rosenblatt et Windrow-Hoff. Beaucoup de gens ont été influences par cette démonstration qu'ils ont généralement mal interprétée.

Certains chercheurs ont continue en développant de nouvelles architectures et de nouveaux algorithmes plus puissants. E n 1972, Teuvo Kohonen et James Anderson ont développe indépendamment et simultanément de nouveaux réseaux pouvant servir de mémoires associatives. Egalement, Stephen Grossberg a examine ce qu'on appelle les réseaux auto- organisés.

Dans les années 1980, on constate l'invention de l'algorithme de retro-propagation des erreurs, cet algorithme est la réponse aux critiques de Minsky et Papert. C'est ce nouveau développement, généralement attribue a David Rumelhart et James McClellabnd, mais aussi découvert plus ou moins en même temps par Paul Werbos et par Yann LeCun. [12]

III.3. Neurone biologique

Les réseaux de neurones artificiels sont construits selon une architecture semblable, en premier approximation, a celle du cerveau humain. [12]

Les cellules nerveuses, appelées neurones, sont les éléments de base du système nerveux centrale. Celui-ci en posséderait environ 100 milliard. Les neurones possèdent de nombreux points communs dans leur organisation générale et leur système biochimique avec les autres cellules. [14]

Alors Les neurones sont des cellules nerveuses aux sains du cerveau. Elles sont composées de :

- **Corps :**

  (Soma) Ou se déroulent toutes les activités vitales de la cellule, par ailleurs c'est la où se trouve le noyau.

- **Axone et Dendrite :**

Organes spécialisés dans communication avec les autres cellules (neurones).

- **Synapse :**

C'est une jonction entre les m terminaisons axonales et les autres cellules [2]

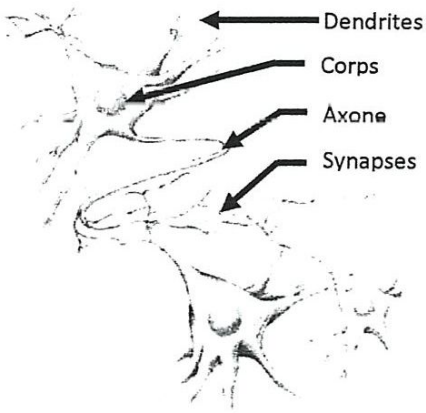


Figure III.1.Le neurone biologique

III.3.1. Mécanisme

L'impulsion nerveuse est une manifestation de la communication intercellulaire. C'est une décharge électrique prenant naissance à l'extrémité de l'axone (segment initiale), cette décharge se propage le long de l'axone pour arriver aux synapses où se déroule des interactions fortement complexes qui transforment la décharge en un signale biochimique et cela par la libération de neurotransmetteurs (acétylcholine, adrénaline,..)

Ces neurotransmetteurs auront pour effets d'exciter les neurones qui les reçoivent, ou au contraire d'inhiber l'influx nerveux. [2]

III.4. Modélisation du problème

C'est à partir de l'hypothèse que le comportement intelligent émerge de la structure et du comportement des éléments de base du cerveau que les réseaux de neurones artificiels se sont développés. Les réseaux de neurones artificiels sont des modèles, à ce titre ils peuvent être décrits par leurs composants, leurs variables descriptives et les interactions des composants. [2]

III.4.1. Neurone formel

Un neurone formel est une représentation mathématique et informatique d'un neurone biologique. Le neurone formel possède généralement plusieurs entrées et une sortie qui correspondent respectivement aux dendrites et au cône d'émergence du neurone biologique (point de départ de l'axone). Les actions excitatrices et inhibitrices des synapses sont représentées la plupart du temps par des coefficients numériques (les poids synaptiques) associés aux entrées. Les valeurs numériques de ces coefficients sont ajustées dans une phase d'apprentissage. Dans sa version la plus simple, un neurone formel calcule la somme pondérée des entrées reçues, puis applique à cette valeur une fonction d'activation, généralement non linéaire. La valeur finale obtenue est la sortie du neurone.

Le neurone formel est l'unité élémentaire des réseaux de neurones artificiels dans lesquels il est associé à ses semblables pour calculer des fonctions arbitrairement complexes, utilisées pour diverses applications en intelligence artificielle.

Mathématiquement, le neurone formel est une fonction à plusieurs variables et à valeurs réelles. [9]

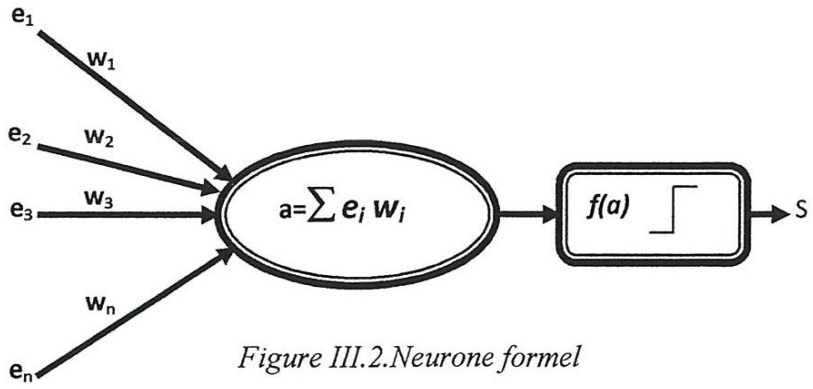


Figure III.2.Neurone formel

III.4.1.1. Le neurone formel de McCulloch et Pitts

Le premier modèle mathématique et informatique du neurone biologique est proposé par Warren McCulloch et Walter Pitts en 1943. En s'appuyant sur les propriétés des neurones biologiques connues à cette époque, issues d'observations neurophysiologiques et anatomiques, McCulloch et Pitts proposent un modèle simple de neurone formel. Il s'agit d'un

neurone binaire, c'est-à-dire dont la sortie vaut 0 ou 1. Pour calculer cette sortie, le neurone effectue une somme pondérée de ses entrées (qui, en tant que sorties d'autres neurones formels, valent aussi 0 ou 1) puis applique une fonction d'activation à seuil : si la somme pondérée dépasse une certaine valeur, la sortie du neurone est 1, sinon elle vaut 0 (cf les sections suivantes).

McCulloch et Pitts étudiaient en fait l'analogie entre le cerveau humain et les machines informatiques universelles. Ils montrèrent en particulier qu'un réseau (bouclé) constitué des neurones formels de leur invention a la même puissance de calcul qu'une machine de Turing.

Malgré la simplicité de cette modélisation, ou peut-être grâce à elle, le neurone formel dit de McCulloch et Pitts reste aujourd'hui un élément de base des réseaux de neurones artificiels. De nombreuses variantes ont été proposées, plus ou moins biologiquement plausibles, mais s'appuyant généralement sur les concepts inventés par les deux auteurs. On sait néanmoins aujourd'hui que ce modèle n'est qu'une approximation des fonctions remplies par le neurone réel et, qu'en aucune façon, il ne peut servir pour une compréhension profonde du système nerveux. . [9]

III.4.1.2. Formulation mathématique

On considère le cas général d'un neurone formel à m entrées, auquel on doit donc soumettre les m grandeurs numériques (ou signaux, ou encore stimuli) notées x_1 à x_m. Un modèle de neurone formel est une règle de calcul qui permet d'associer aux m entrées une sortie : c'est donc une fonction à m variables et à valeurs réelles.

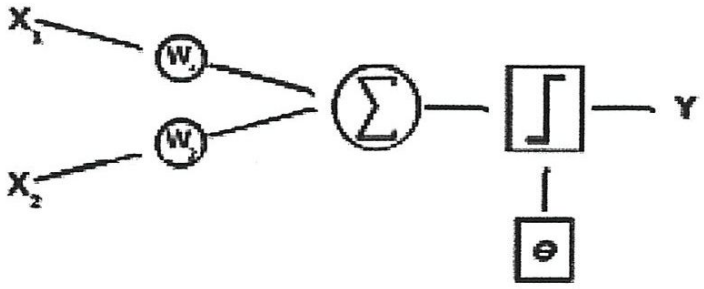


Figure III.3.Neurone formel

Neurone formel avec 2 entrées et une fonction d'activation à seuil.

Dans le modèle de McCulloch et Pitts, à chaque entrée est associé un poids synaptique, c'est-à-dire une valeur numérique notée de w_1 pour l'entrée 1 jusqu'à w_m pour l'entrée m. La première opération réalisée par le neurone formel consiste en une somme des grandeurs reçues en entrées, pondérées par les coefficients synaptiques, c'est-à-dire la somme.

$$W_1X_1 + \ldots + W_mX_m = \sum_{j=1}^{m} W_jX_j$$

A cette grandeur s'ajoute un seuil w_0. Le résultat est alors transformé par une fonction d'activation non linéaire (parfois appelée fonction de sortie) φ .La sortie associée aux entrées x_1 à x_m est ainsi donnée par :

$$\varphi\left(W_0 + \sum_{j=1}^{m} W_jX_j\right)$$

Qu'on peut écrire plus simplement :

$$\varphi\left(\sum_{j=0}^{m} W_jX_j\right)$$

En ajoutant au neurone une entrée fictive x_0 fixée à la valeur 1.

Dans la formulation d'origine de McCulloch et Pitts, la fonction d'activation est la fonction de Heaviside (fonction en *marche d'escalier*), dont la valeur est 0 ou 1. Dans ce cas, on préfère parfois définir la sortie par la formule suivante

$$\varphi\left(\sum_{j=1}^{m} W_jX_j - w_0\right)$$

Qui justifie le nom de seuil donné à la valeur w_0. En effet, si la somme $\sum_{j=1}^{m} w_j x_j$ dépasse w_0 la sortie du neurone est 1, alors qu'elle vaut 0 dans le cas contraire : w_0 est donc le seuil d'activation du neurone, si on considère que la sortie 0 correspond à un neurone « éteint ». [9]

"Un neurone formel fait une somme pondérée des potentiels d'actions qui lui proviennent des autres neurones, puis s'active suivant la valeur de cette sommation pondérée. Si cette sommation dépasse un certain seuil, le neurone est activé et transmet une réponse dont la valeur est celle de son activation. Si le neurone n'est pas activé, il ne transmet rien. [2]

III.4.1.3. Fonction d'activation

La plupart des neurones formels utilisés actuellement sont des variantes du neurone de McCulloch et Pitts dans lesquelles la fonction de Heaviside est remplacée par une autre fonction d'activation. Les fonctions les plus utilisées sont :

- la fonction sigmoïde.
- la fonction tangente hyperbolique.
- la fonction identité.
- dans une moindre mesure, certaines fonctions linéaires par morceaux.

Ces choix sont motivés par des considérations théoriques et pratiques issues de la combinaison des neurones formels en un réseau de neurones formels.

Il est claire que la fonction d'activation joue un rôle très important dans le comportement du neurone .Elle retourne une valeur représentatif de l'activation du neurone, cette fonction a comme paramètre la somme pondéré des entrés ainsi que le seuil d'activation.

La nature de cette fonction diffère selon le réseau. On en compte divers types, parmi elles : [2]

Nom de la fonction	Relation d'entrée/sortie	Icône	Nom Matlab
seuil	$a = 0$ si $n < 0$ $a - 1$ si $n \geq 0$		hardlim
seuil symétrique	$a = -1$ si $n < 0$ $a - 1$ si $n \geq 0$		hardlims
linéaire	$a = n$		purelin
linéaire saturée	$a = 0$ si $n < 0$ $a = n$ si $0 \leq n \leq 1$ $a = 1$ si $n > 1$		satlin
linéaire saturée symétrique	$a = -1$ si $n < -1$ $a = n$ si $-1 \leq n \leq 1$ $a = 1$ si $n > 1$		satlins
linéaire positive	$a = 0$ si $n < 0$ $a = n$ si $n \geq 0$		poslin
sigmoïde	$a = \frac{1}{1+\exp^{-n}}$		logsig
tangente hyperbolique	$a = \frac{e^n - e^{-n}}{e^n + e^{-n}}$		tansig
compétitive	$a = 1$ si n maximum $a = 0$ autrement		compet

III.5. Les réseaux de neurones formels

III.5.1. Définition

Un réseau de neurones est un outil d'analyse statistique permettant de construire un modèle de comportement à partir de données qui sont des exemples de ce comportement.

Un réseau de neurones est un maillage de plusieurs neurones, généralement organise en couches. Pour construire une couche de S neurones, il s'agit simplement de les assembler comme à la figure III.4. Les S neurones d'une même couche sont tous branchés aux R entrées. On dit alors que la couche est totalement connectée. Un poids wi,j est associé a chacune des connexions. Nous noterons toujours le premier indice par i et le deuxième par j (jamais l'inverse). Le premier indice (rangée) désigne toujours le numéro de neurone sur la couche, alors que le deuxième indice (colonne) spécifie le numéro de l'entrée. Ainsi, wi,j désigne le poids de la connexion qui relie le neurone i a son entrée j. L'ensemble des poids d'une couche forme donc une matrice W de dimension S × R : [12]

$$\begin{pmatrix} W_{11}\ W_{12}\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots W_{1m} \\ W_{21}\ W_{22}\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots W_{2m} \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ \ldots\ldots\ldots W_{ij} \ldots\ldots\ldots \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ W_{n1}\ W_{n2}\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots W_{nm} \end{pmatrix}$$

W_{ij} : poids de la liaison du neurone "j" vers le neurone "i".

L'activation totale du réseau est décrite par un vecteur d'activation :

A= (a1, a2,..., an) ai : activation du neurone "i" i=1, 2,3,..., n

- ❖ Les réseaux de neurones sont organisés en couches:

- ➢ **Une couche d'entrée** : qui reçoit les informations provenant de l'extérieur. Les neurones de cette couche correspondent par exemple aux neurones sensoriels présents dans la rétine. Ils sont d'un type particulier car ils se contentent de transmettre l'information qui leur est présentée sans traitement
- ➢ **Une ou plusieurs couches intermédiaires** : encore appelées couches cachées car elles ne sont pas directement en contact avec le monde extérieur.
- ➢ **Une couche de sortie** : correspondant aux neurones moteurs qui actionnent les muscles [10].

Les neurones étant les nœuds, qui seront connectes par des liens appeler liens synaptiques ou synapses. Ces liens synaptiques sont pondérés par des poids judicieusement choisis.

La propagation de l'activation à travers les liens synaptiques des neurones en avale vont influencer les autres neurones en amont, cette activation sera pondérée par le lien qu'elle prendra ainsi on appellera poids synaptique le poids de chaque liaison synaptique.

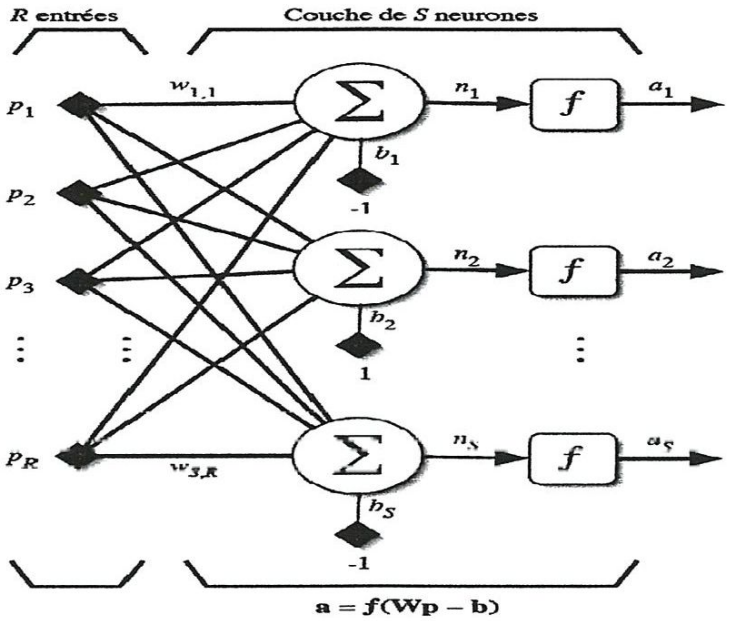


Figure III.4.Couche de S neurones.

L 'architecture d'un réseau de neurone influence considérablement sur son comportement globale. On distingue deux grandes architectures de réseaux de neurones :

III.5.2. Les réseaux non bouclés

Forme générale

Un réseau de neurones non bouclé est donc représenté graphiquement par un ensemble de neurones «connectés» entre eux, l'information circulant des entrées vers les sorties sans «retour en arrière» : si l'on représente le réseau comme un graphe dont les nœuds sont les neurones et les arêtes les « connexions » entre ceux-ci, le graphe d'un réseau non bouclé est Acyclique : si l'on se déplace dans le réseau, à partir d'un neurone quelconque, en suivant les connexions, on ne peut pas revenir au neurone de départ. La représentation de la topologie d'un réseau par un graphe est très utile, notamment pour les réseaux bouclés. Les neurones qui effectuent le dernier calcul de la composition de fonctions sont les neurones de sortie ; ceux qui effectuent des calculs intermédiaires sont les neurones cachés. La figure III.5 représente un réseau de neurones non bouclé qui a une structure particulière, très fréquemment utilisée : il comprend des entrées, une couche de neurones "cachés" et des neurones de sortie. Les neurones de la couche cachée ne sont pas connectés entre eux. Cette structure est appelée Perceptron multicouche. [11]

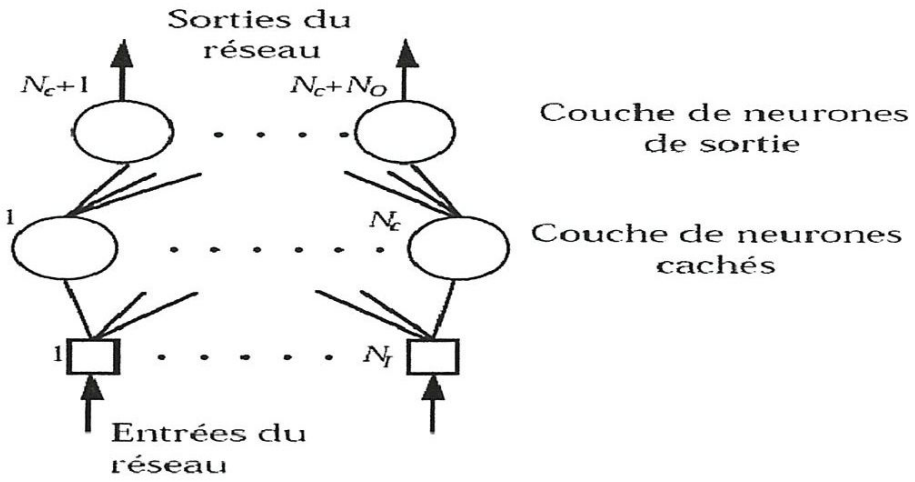


Figure III.5. Un réseau de neurones non bouclé.

III.5.3. Les réseaux bouclés

Forme générale

Nous introduisons ici l'architecture la plus générale pour un réseau de neurones, les « réseaux bouclés », dont le graphe des connexions est cyclique : lorsqu'on se déplace dans le réseau en suivant le sens des connexions, il est possible de trouver au moins un chemin qui revient à son point de départ (un tel chemin est désigné sous le terme de « cycle »). La sortie d'un neurone du réseau peut donc être fonction d'elle-même ; cela n'est évidemment concevable que si la notion de temps est explicitement prise en considération.

Ainsi, à chaque connexion d'un réseau de neurones bouclé (ou à chaque arête de son graphe) est attaché, outre un poids comme pour les réseaux non bouclés, un retard, multiple entier (éventuellement nul) de l'unité de temps choisie. Une grandeur, à un instant donné, ne pouvant pas être fonction de sa propre valeur au même instant, tout cycle du graphe du réseau doit avoir un retard non nul. [11]

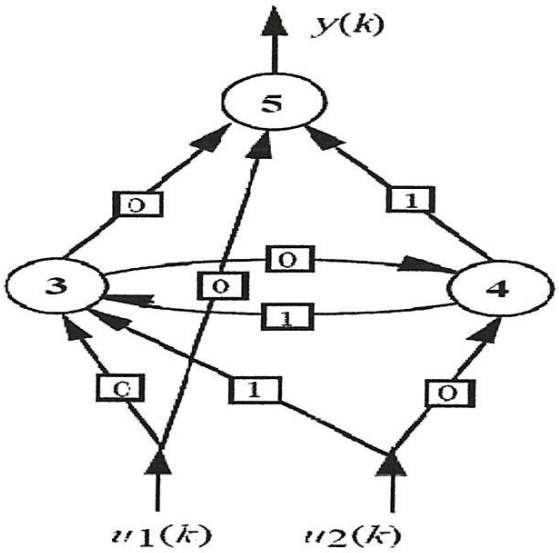


Figure III.6. Un réseau de neurones bouclé

III.6. Structure d'interconnexion

Les connexions entre les neurones qui composent le réseau décrivent la topologie du modèle. Elle peut être quelconque, mais le plus souvent il est possible de distinguer une certaine régularité.

III.6.1. Réseau multicouche (au singulier) :

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 avalent. 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.

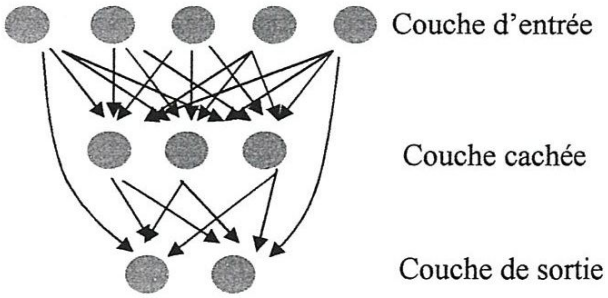


Figure III.7.Réseau multicouche

III.6.2. Réseau 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 III.8). Les connexions sont donc moins nombreuses que dans le cas d'un réseau multicouche classique.

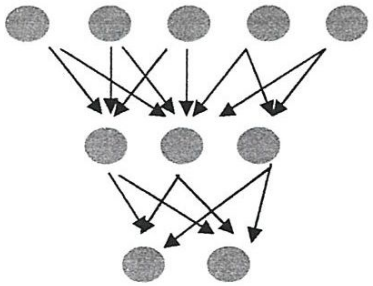


Figure III.8.Réseau à connexions locales

III.6.3. Réseau 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

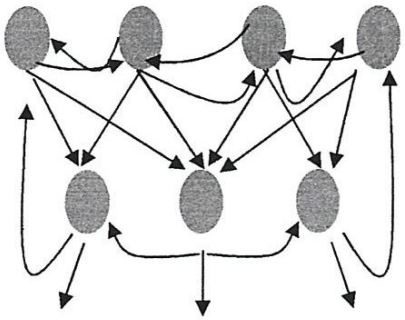


Figure III.9.Réseau à connexions récurrentes

III.6.4. Réseau à connexion complète

C'est la structure d'interconnexion la plus générale (Figure III.10). Chaque neurone est connecté à tous les neurones du réseau.

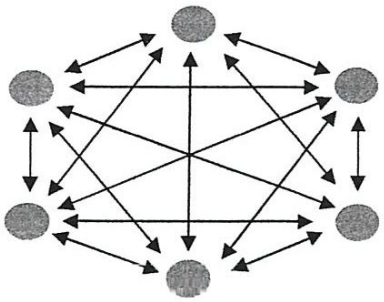


Figure III.10.Réseau à connexion complète

Il existe de nombreuses autres topologies possibles, mais elles n'ont pas eu à ce jour la notoriété des quelques unes que nous avons décrites ici.

III.7. Apprentissage

III.7.1. Définition

Parmi les propriétés désirables pour un réseau de neurones, la plus fondamentale est sûrement la capacité d'apprendre de son environnement, d'améliorer sa performance à travers un processus d'apprentissage. Mais qu'est-ce donc que l'apprentissage ?

Malheureusement, il n'existe pas de définition générale, universellement acceptée, car ce concept touche à trop de notions distinctes qui dépendent du point de vue que l'on adopte.

Dans le contexte des réseaux de neurones artificiels, nous adopterons un point de vue pragmatique en proposant la définition suivante :

« L'apprentissage est un processus dynamique et itératif permettant de modifier les paramètres d'un réseau en réaction avec les stimuli qu'il reçoit de son environnement.

Le type d'apprentissage est détermine par la manière dont les changements de paramètre surviennent ».

Cette définition implique qu'un réseau se doit d'être stimulé par un environnement, qu'il subisse des changements en réaction avec cette stimulation, et que ceux-ci provoquent dans le futur une réponse nouvelle vis-à-vis de l'environnement. Ainsi, le réseau peut s'améliorer avec le temps.

Dans la plupart des architectures, l'apprentissage se traduit par une modification de l'efficacité synaptique, c'est-à-dire par un changement dans la valeur des poids qui relient les neurones d'une couche à l'autre. Soit le poids $w_{i,j}$ reliant le neurone *i* à son entrée *j*. Au temps t, un changement $\Delta w_{i,j}(t)$ de poids peut s'exprimer simplement de la façon suivante : [12]

$$w_{i,j}(t) = w_{i,j}(t+1) - w_{i,j}(t)$$

et, par conséquent, $wi,j(t+1) = wi,j(t) + wi,j(t)$, avec $w_{i,j}(t+1)$ et $w_{i,j}(t)$ représentant respectivement les nouvelle et ancienne valeurs du poids $w_{i,j}$.

Un ensemble de règles bien définies permettant de réaliser un tel processus d'adaptation des poids constitue ce qu'on appelle l'algorithme1 d'apprentissage du réseau.

III.7.2. Protocoles d'apprentissages

Presque la totalité des réseaux de neurones ont en commun un même protocole d'apprentissage celui-ci comporte quatre étapes :

- **Etape 1 :** Initialisation des poids synaptiques avec des petites valeurs aléatoires.
- **Etape 2 :** Présentation du patron d'entré et propagation de l'activation des neurones.
- **Etape 3 :** Calcule de l'erreur, dans le cas d'un apprentissage supervisé cette erreur dépend de la différence entre l'activation des neurones et le patron de référence
- **Etape 4 :** Calcule du vecteur de correction à partir des valeurs des erreurs, avec lequel on effectue la correction des poids synaptiques

  Les étapes 2-3-4 sont répétées jusqu'à la fin de l'apprentissage. [2]

III.7.3. Les types d'apprentissage

Les procédures d'apprentissage peuvent se subdiviser, elles aussi, en trois grandes catégories

- Apprentissage supervisé.
- Apprentissage non supervisé.
- Apprentissage semi-supervisé.

III.7.3.1. Apprentissage supervisé

Dans ce type d'apprentissage l'utilisateur dispose d'un comportement de référence qu'il désir inculquer au réseau. Le réseau est donc capable de mesurer la différence entre son comportement actuel et le comportement de référence, et de corriger ses poids de façon à réduire cette erreur.

III.7.3.2. Apprentissage semi - supervisé

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

III.7.3.3. Apprentissage non supervisé (appelé aussi auto organisation)

Ici la procédure consiste à modifier les poids du réseau en fonction des critères internes comme coactivation des neurones. Les comportements résultant de ces apprentissages sont en général comparables à des techniques d'analyse de données.

Dans la majorité des réseaux de neurones actuels, l'apprentissage du réseau s'effectue lors d'une période d'apprentissage préliminaire à son utilisation subséquente, les poids synaptiques sont figés. [16]

III.7.4. Règles d'apprentissage

III.7.4.1. La loi de Hebb, un exemple d'apprentissage non supervisé

La loi de Hebb (1949) s'applique aux connexions entre neurones, Elle s'exprime de la façon suivante :

Dans un contexte neurobiologique, Hebb cherchait à établir une forme d'apprentissage associatif au niveau cellulaire. Dans le contexte des réseaux artificiels, on peut reformuler l'énonce de Hebb sous la forme d'une règle d'apprentissage en deux parties :

1. Si deux neurones de part et d'autre d'une synapse (connexion) sont activés simultanément (D'une manière synchrone), alors la force de cette synapse doit être augmentée.

2. Si les mêmes deux neurones sont activés d'une manière asynchrone, alors la synapse correspondant doit être affaibli ou carrément élimine. [12]

xi	xj	∂ wij
0	0	0
0	1	0
1	0	0
1	1	+

Table III.2 La loi de Hebb

xi et xj sont respectivement les valeurs d'activation des neurones i et j, ∂w_{ij} (dérivée partielle du poids) correspond à la modification de poids réalisée.

Une telle synapse est dite «synapse hebbien». Il utilise un mécanisme interactif, dépendant du temps et de l'espace, pour augmenter l'efficacité synaptique d'une manière proportionnelle à la corrélation des activités pré- et post-synaptiques. De cette définition ressort les propriétés suivantes :

1. Dépendance temporelle. Les modifications d'une synapse hebbien dépendent du moment exact des activités pré- et post-synaptiques.

2. Dépendance spatiale. Etant donne la nature même du synapse qui constitue un lieu de transmission d'information, l'apprentissage hebbien se doit de posséder une contiguite spatiale.

C'est cette propriété qui, entre autres, permet l'apprentissage dit non supervise sur lequel nous reviendrons bientôt ;

3. Interaction. L'apprentissage hebbien dépend d'une interaction entre les activités de part et d'autre de la synapse.

4. Conjonction ou corrélation. Une interprétation de l'énonce de Hebb est que la condition permettant un changement dans l'efficacité synaptique est une conjonction des activités pré et post-synaptiques. C'est la co-occurrence des activités de part et d'autre de la synapse qui engendre une modification de celui-ci. Une interprétation plus statistique refere `a la corrélation de ces activités. Deux activités positives simultanées (corrélation positive) engendrent une augmentation de l'efficacité synaptique, alors que l'absence d'une telle corrélation engendre une baisse de cette efficacité. [12]

Mathématiquement, on peut exprimer la règle de Hebb sous sa forme la plus simple par la formule suivante :

$$w_{ij}(t+1) = w_{ij}(t) + \partial w_{ij}(t)$$

Où : $w_{ij}(t+1)$ est le nouveau poids, $w_{ij}(t)$ l'ancien

$$\partial w_{ij}(t) = x_i . x_j$$

L'algorithme d'apprentissage modifie de façon itérative (petit à petit) les poids pour adapter la réponse obtenue à la réponse désirée. Il s'agit en fait de modifier les poids lorsqu'il y a erreur seulement. [15]

1/ Initialisation des poids et du seuil S à des valeurs (petites) choisies au hasard.

2/ Présentation d'une entrée $E_l = (e_1, \ldots e_n)$ de la base d'apprentissage.

3/ Calcul de la sortie obtenue x pour cette entrée :

$a = \sum (w_i . e_i) - S$ (La valeur de seuil est introduite ici dans le calcul de la somme pondérée)

x = signe (a) (si a > 0 alors x = +1 sinon a = 0 alors x = -1)

4/ Si la sortie x est différente de la sortie désirée d_l pour cet exemple d'entrée E_l alors modification des poids (μ est une constante positive, qui spécifie le pas de modification des poids) :

$$wij(t+1) = wij(t) + \mu.(xi.xj)$$

5/ Tant que tous les exemples de la base d'apprentissage ne sont pas traités correctement (i.e. modification des poids), retour à l'étape 2.

Exemple: Nous allons réaliser l'apprentissage sur un problème très simple. La base d'apprentissage est décrite par la table 2 :

e_1	e_2	x
1	1	1
1	-1	1
-1	1	-1
-1	-1	-1

Table III.3. Base d'exemples d'apprentissage pour la loi de Hebb.

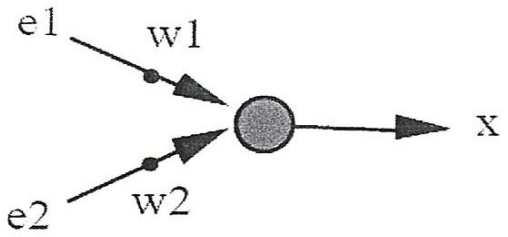


Figure III.11. Réseau de 3 neurones

1/ Conditions initiales : $\mu = +1$, les poids et le seuil sont nuls.

2/ Calculons la valeur de x pour l'exemple (1) :

3/ $a = w_1.e_1 + w_2.e_2 - S = 0 . 1 + 0 . 1 - 0 = 0 \quad a \leq 0 \Rightarrow x = -1$

4/ La sortie est fausse, il faut donc modifier les poids en applicant :

$w_1 = w_1 + e_1.x = 0.0 + 1.1 = 1$

$w_2 = w_2 + e_2.x = 0.0 + 1.1 = 1$

2/ On passe à l'exemple suivant (2) :

3/ $a = 1.1 + 1.-1 -0.0 = 0 \qquad a \leq 0 \Rightarrow x = -1$

4/ La sortie est fausse, il faut donc modifier les poids en appliquant :

$w1 = 1 + 1.1 = 2$

$w2 = 1 + 1.-1 = 0$

L'exemple suivant (3) est correctement traité : $a = -2$ et $x = -1$ (la sortie est bonne).

On passe directement, sans modification des poids à l'exemple (4). Celui-ci aussi est correctement traité. On revient alors au début de la base d'apprentissage : l'exemple (1). Il est correctement traité, ainsi que le second (2). L'algorithme d'apprentissage est alors terminé : toute la base d'apprentissage a été passée en revue sans modification des poids. [15]

III.7.4. 2. La règle d'apprentissage du Perceptron un exemple d'apprentissage supervisé

La règle de Hebb ne s'applique pas dans certain cas, bien qu'une solution existe. Un autre algorithme d'apprentissage a donc été proposé, qui tient compte de l'erreur observée en sortie.

L'algorithme d'apprentissage du Perceptron est semblable à celui utilisé pour la loi de Hebb. Les différences se situent au niveau de la modification des poids.

1/ Initialisation des poids et du seuil S à des valeurs (petites) choisies au hasard.

2/ Présentation d'une entrée $E_l = (e_1, \ldots e_n)$ de la base d'apprentissage.

3/ Calcul de la sortie obtenue x pour cette entrée :

$$a = \sum (w_i . e_i) - S$$

x = signe (a) (si a > 0 alors x = +1 sinon a = 0 alors x = -1)

4/ Si la sortie x du Perceptron est différente de la sortie désirée d_l pour cet exemple d'entrée E_l alors modification des poids (μ le pas de modification) :

$$wij(t+1) = wij(t) + \mu.((d - x).e_i)$$

5/ Tant que tous les exemples de la base d'apprentissage ne sont pas traités correctement (i.e. modification des poids), retour à l'étape 2. [15]

Exemple de fonctionnement de l'algorithme d'apprentissage du Perceptron:

Base d'exemples d'apprentissage :

e_1	e_2	d	(EX)
1	1	1	(1)
-1	1	-1	(2)
-1	-1	-1	(3)
1	-1	-1	(4)

1/ Conditions initials: $w_1 = -0.2$, $w_2 = +0.1$, S = 0, ($\mu = +0.1$)

2/ a (1) = -0.2 + 0.1 -0.2 = -0.3

3/ x (1) = -1 (la sortie désirée d (1) = +1, d'où modification des poids)

4/ $w_1 = -0.2 + 0.1 . (1 + 1) . (+1) = 0$

w2 = +0.1 + 0.1 . (1 + 1) . (+1) = +0.3

2/ a (2) = +0.3 - 0.2 = +0.1

3/ x (2) = +1 Faux

4/ $w_1 = 0 + 0.1 . (-1 - 1) . (-1) = +0.2$

$w_2 = +0.3 + 0.1 . (-1 - 1) . (+1) = +0.1$

2-3/ a (3) = -0.2 - 0.1 - 0.2 = -0.5 Ok

2-3/ a (4) = +0.2 - 0.1 - 0.2 = -0.1 Ok

2-3/ a (1) = +0.2 + 0.1 - 0.2 = +0.1 Ok

2-3/ a (2) = -0.2 + 0.1 - 0.2 = -0.1 Ok

5/ Tous les exemples de la base ont été correctement traités, l'apprentissage est terminé.

III.8. Le Perceptron

Ce modèle fut le premier proposé en 1958 par "Frank Rosenblatt". Le perceptron était constitué de trois couches appelées :

- Rétine (neurones d'entrées)
- Aire d'association (neurones cachés)
- Neurones réponse (neurones de sortie)

III.8.1. Le Perceptron Multi Couches (PMC)

Le PMC est le type de réseaux le plus répandu et le plus utilisé, vu la simplicité de sa structure et la rapidité de son apprentissage. Dans le passé, les PMC étaient peu utilisés, à cause du manque de règles rigoureuses et d'algorithmes efficaces pour gérer la phase d'apprentissage. On utilisait les perceptrons à deux couches seulement, ces derniers peuvent être entraînés avec des règles d'apprentissage relativement simples. En contre partie les réseaux à deux couches sont limités aux simples applications. Cela a changé récemment avec le développement de nouvelles règles pour le PMC ; ce qui a permis d'aborder des applications de plus en plus complexes, et de combler les lacunes des réseaux à deux couches.

III.8.1.1. Structure

Un PMC est un réseau à couches où ses nœuds (neurones formels) sont arrangés dans des couches; Il peut comprendre une ou plusieurs couches entre les deux couches d'entrée et de sortie. Sur la Figure III.12 on trouve un exemple d'un PMC à trois couches [2]

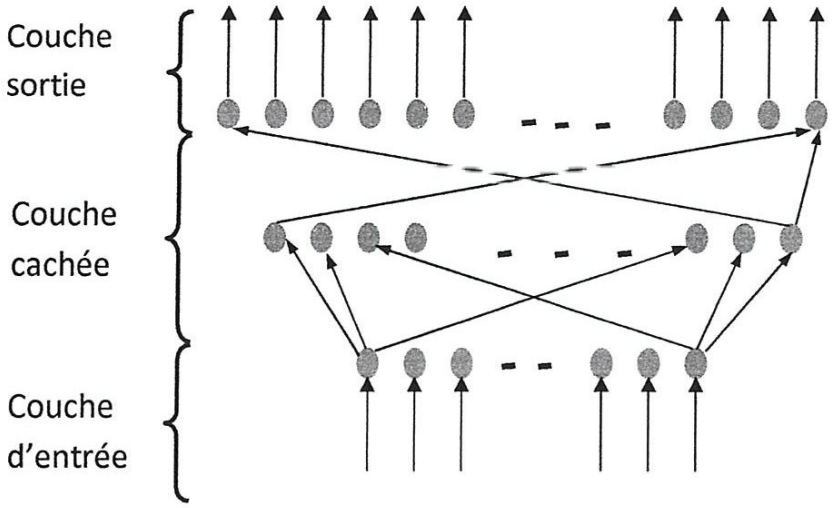


Figure III.12.Architecture d'un PMC à trois couches

Les fonctions d'activation utilisées dans ce types de réseaux sont principalement les fonctions a seuil ou sigmoïdes. Il peut résoudre des problèmes non-linéairement séparables et des problèmes logiques plus compliquent [2]

III.9. Conclusion

L'objectif de ce chapitre et la présentation de quelques réseaux artificiels, est la mise en exergue d'une organisation structurelle des neurones. Chaque structure est dotée d'une fonction particulière et ces structures adaptent leur comportement par des mécanismes d'apprentissage, qu'est Le grand avantage des réseaux de neurones réside dans leur capacité d'apprentissage automatique, ce qui permet de résoudre des problèmes sans nécessiter l'écriture de règles complexes, tout en étant tolérant aux erreurs.

Les réseaux de neurones sont des outils statistiques, qui permettent d'ajuster des fonctions non linéaires très générales à des ensembles de points ; comme toute méthode statistique, l'utilisation de réseaux de neurones nécessite que l'on dispose de données suffisamment nombreuses et représentatives

Chapitre IV

Applications

IV.1. Introduction

L'objectif de ce travail est la réalisation d'un système de reconnaissance automatique de la parole. Le système qu'on propose est à un vocabulaire limité et multi-locuteurs.
Pour l'extraction des caractéristiques, on se base sur les coefficients cepstraux dans l'échelle de Mels (MFCC) qui sont les plus utilisé dans ce domaine. Dans la phase de classification, le système proposé se base sur les réseaux de neurones.

Nous réalisons une petite base de données avec 4 mots (numéros) : Dix(10), Vingt(20), Trente(30) et Quarante(40) prononcés par deux locuteurs.

IV.2. Les démarche de l'application

Les différentes démarches qui constituent l'application sont les suivantes :

- Acquisition des mots isolés pour la construction du corpus d'apprentissage
- Paramétrage des mots (extraction des formants)
- Construction des corpus d'apprentissages (d'entrée et de test).
- Réalisation de la phase de l'apprentissage pour multi-locuteur.
- La généralisation du réseau neurones et le calcul du pourcentage de reconnaissance et d'apprentissage.
- Réalisation logiciel : la réalisation est implémentée sur MATLAB.

VI.2.1. Acquisition

L'acquisition a été faite sur PC à l'aide d'une carte soundblaster. la fréquence d'échantillonnage a été fixée à 8KHz ; c'est à dire de travailler sur une bande de largeur 4KHz (bande téléphonique).

L'acquisition est suivie d'une isolation, permettant ainsi l'élimination des parties non désirées des mots. Ce sont les parties qui existent au début et à la fin du mot prononcé, et qui seront très gênantes par la suite dans l'extraction des formants d'une part, et d'augmenter le temps de calcul d'autre part

VII.2.2. Paramétrage

Une fois l'acquisition est faite, le mot va subir un traitement comprenant les tâches suivantes
La préaccentuation et Extraction des caractéristiques(MFCC).

- **La préaccentuation** : cette tache permet d'accentuer la partie haute du spectre vocal.

- **Extraction des caractéristiques**

Les paramètres MFCC sont des coefficients cepstraux obtenus à partir des énergies d'un banc de filtre en échelle de fréquence Mel. Il s'agit d'un calcul classique des coefficients cepstraux auquel on a rajouté, avant le logarithme un filtre de Mel.

Les différentes étapes pour l'obtention des MFCC sont :

1. Fenêtrage
2. RFFT
3. Filtre de MEL
4. Calcul du Log

5 $RFFT^{-1}$

6. MFCC

IV.3. Phase de classification

Pour la classification nous utilisons un PMC qui est le réseau de neurones le plus utilisé pour la classification. Le PMC utilisé comporte un nombre de neurones d'entrée égale au nombre de caractéristique (MFCC) donc N=12. Ce PMC comporte un nombre de neurones de sortie égale aux nombres de classe donc **J**=4. Pour le nombre de neurones dans la couche cachée, nous avons choisie (par expérience) M=8.

Quand on applique un réseau neurones pour la classification, le nombre de neurones de sortie égale au nombre de classes, la $i^{ème}$ neurone correspond à la $ième$ classe, ce qui nous a mené dans le calcul de la sortie désirée de forcé à 1 le neurone qui représentent la classe du mot propager dans le réseau et des 0 ailleurs,dans le but à apprendre au réseau que le mot propager appartient à cette classe.

Mots	Classe
Dix(10)	**1**
Vingt (20)	**2**
Trente (30)	**3**
Quarante(40)	**4**

Tableau IV.1

IV.4. Résultats du test

Nous avons établie une petite base de données avec 4 mots (numéros) : Dix(10), Vingt(20), Trente(30) et Quarante(40) prononcés par deux locuteurs. Chaque a été répété 8 fois pours chaque locuteurs. Nous avons réalisé des tests de classification, les résultats sont donnés par le tableau suivant :

Mots	Taux d'apprentissage	Taux de reconnaissance	Classe (résultat)
Dix(10)	93.5484%	100%	1
Vingt (20)	87.0968%	100%	2
Trente (30)	93.5484%	100%	3
Quarante(40)	93.5484%	100%	4

Tableau IV.2

IV.5. Conclusions

Les résultats obtenus dans les différents tests présidant montrent que le PMC a permet de bien classifier tous les exemples d'apprentissage dans les quatre (4) cas.

Conclusion générale

Dans ce travail nous avons présenté une nouvelle technique de traitement en réseaux de neurones appliquée à la reconnaissance des mots isolés. Elle utilise les formants comme paramètres, ces derniers ont été obtenus par l'analyse spectrale basée sur les coefficients cepstraux dans l'échelle de Mels (MFCC)

Nous avons réalisés d'un système de reconnaissance automatique de la parole avec un vocabulaire limité et deux-locuteurs. Pour l'extraction des caractéristiques, en se base sur les coefficients cepstraux dans l'échelle de Mels (MFCC) qui sont les plus utilisés dans ce domaine. Cette technique s'appuie à la fois sur la théorie cepstrale et la perception de la parole dans l'échelle de Mels.

Dans la phase de classification, le système proposé se base sur les réseaux de neurones. Cette phase constitue une grande catégorie de problèmes actuels qui consiste à attribuer, de façon automatique, un objet à une classe parmi d'autres classes possibles. La résolution de ce type de problèmes demande de représenter les exemples à classifier à l'aide d'un ensemble de caractéristiques. Il s'agit ensuite de concevoir un système capable de classifier ces exemples en se basant sur leur représentation et les réseaux de neurones sont particulièrement bien adaptés à ce type de problème. Ceci est dû à leurs grandes capacités de calcul et à leurs hautes habilités d'apprentissage. De plus, l'estimation de leurs paramètres est indépendante de la complexité du problème traité ce qui leur permet d'être bien adaptés aux problèmes actuels qui ne cessent d'être de plus en plus complexes

[1] Contributions à la reconnaissance automatique de la parole avec données manquantes- Doctorat de l'université Henri Poincaré {Nancy 1 (spécialité informatique) par Sébastien Demange- novembre 2007

[2]Utilisation Des Réseaux Neuro-Flous Pour La Reconnaissance Automatique De La Parole PFE ITO juin 2003

[3] Traitement de la parole par René Boite et Murat Kunt. Ed Presse Polytechniques Romande « L/621.103 »

[4] Compression de la parole dans la communication numérique.

PFE unv-Guelma juin 2007

[5] Traitement de la parole par Guy Almouzni (cours & TP).

[6] Les fenêtres de pondérations Exposé 1ere année Master sous la direction de Mr : Tabaa Mohamed Tahar -2010-

[7] Utilisation des réseaux Neuro-Flous en reconnaissance de La Parole PFE ITO juin 2001

[8] Application du Réseau de Neurones Gamma à la Reconnaissance de la Parole

4th International Conference: Sciences of Electronic, Technologies of Information and Telecommunications March 25-29, 2007 – TUNISIA

Site web

[9] www.wikipedia.com

[10] RESEAUX de NEURONES en TRAITEMENT d'IMAGES: Des Modèles Théoriques aux Applications Industrielles- doctorat de l'Université de Bretagne Occidentale par Gilles BUREL- décembre 1991-


[11] Réseau de Neurones : Méthodologies et applications G-Dreyfus, J-M-Martinez , M-Samuelides sous la direction de G-Dreyfu Ed Eyrolles-2004 « L/004.958 »


[12] RESEAUX DE NEURONES- GIF-21140 et GIF-64326- par Marc Parizeau- Université de Laval- 2006


[13] La parole et son traitement automatique par Calliope. Ed Masson


[14] livre Réseaux de neurones par Davalo. EYROLLES. 1993.


[15] Livre Réseaux de neurones artificiels par François Blayo .Puf. 1996


[16] Livre Les réseaux neuromimétiques par Jean François Jodoin