République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
Université Akli Mohand Oulhadj – Bouira

Faculté des sciences et des sciences appliquées

Département de Génie Electrique

Mémoire de Master

Option : Technologies des Télécommunications

Réalisé Par :
- LOUIFI Massinissa
- OUABDESSELAM Feriel

Thème :

Reconnaissance automatique de la parole avec intelligence artificielle

Date de soutenance : 23/09/2017

Devant le jury composé de :

- Mouloud AYAD	Président
- Moustapha DIDICHE	Co-promoteur
- Smail MEDJEDOUB	Promoteur
- Mohamed SAIDI	Examinateur
- Mourad BENZIANE	Examinateur

Année universitaire : 2016- 2017

Remerciements

Ce mémoire est le résultat d'un travail épaulé par de nombreuses personnes. En préambule, nous voulons adresser tous nos remerciements aux personnes avec lesquelles nous avons pu échanger et qui nous ont aidés pour la rédaction de ce mémoire.

En commençant par remercier tout d'abord Monsieur Mustapha Didiche et Monsieur Smail Medjedoub, promoteurs de ce mémoire, pour leur aide précieuse et pour le temps qu'ils nous ont consacré.

Un grand merci à Monsieur Mahmoud Sahali pour son implication dans notre recherche et pour l'aide qu'il nous a apporté.

Enfin, nous adressons nos plus sincères remerciements à nos familles : nos parents, frères et sœurs et tous nos proches et amis, qui nous ont accompagnés, soutenus et encouragés tout au long de la réalisation de ce mémoire.

-Résumé :

Le Chapitre 1 est une introduction à la phonétique et présente les systèmes régissant le langage chez l'être humain et le fonctionnement des appareils concernés, introduisant ainsi le système physiologique et le système neurologique et leur anatomie. Le système physiologique se constitue de l'appareil phonatoire qui est le moteur de la production du son des différents phonèmes grâce à l'interaction des trois grands organes (les poumons, le larynx et les cavités bucco-pharyngale) et l'appareil auditif qui a comme organe principal l'oreille est le centre du traitement acoustique et cognitif. Le système neurologique est la partie nerveuse dite le cerveau qui est constitué de neurones assurant ainsi le traitement des différents sons des phonèmes et leur compréhension.

Le chapitre 2 présente le traitement automatique de la parole. Ce dernier comportera les différentes caractéristiques du signal de parole, évoquant les étapes de la numérisation et détaillant les méthodes traditionnellement mises en œuvre pour cette analyse. Ce chapitre sera l'occasion de présenter en profondeur les différentes méthodes du codage LPC et MFCC.

Le chapitre 3 comportera une introduction globale sur l'intelligence artificielle, puis précisément sur les réseaux de neurones, leur évolution durant le siècle dernier citant les différents types des réseaux de neurones. On se focalisera sur un perceptron multicouche MLP afin d'utiliser un nouveau modèle pour l'extraction de caractéristiques le Codage Neuro-Prédictif (NPC, Neural Predictive Coding) qui est une extension au domaine non-linéaire du codage LPC.

Le chapitre 4 sera consacré à une présentation de la langue Amazighe et précisément les lettres Tifinagh puis à l'étude de la mise en forme d'un signal de parole qui sera injecté dans un réseau de neurones MLP (Multi Layer Perceptron), puis la comparaison entre les résultats obtenus par l'utilisation des deux codages : MFCC (Mel Frequency Cepstral Coding) et NPC (Neuronal Predictive Coding).

Mots clés : LPC, MFCC, NPC, MLP, traitement automatique de la parole , langue Amazigh, intelligence artificielle.

-Notations et abréviations :

LPC : Codage Linéaire Prédictif (Linear Predictive Coding)

MFCC: Codage Cepstral (Mel Frequency Cepstral Coefficents)

NPC: Codage Neuro-Prédictif (Neuronal Predictive Coding)

MLP: Perceptron MultiCouches (Multi Layer Perceptron)

PA: Potentiel d'Action

TAP : Traitement Automatique de la Parole

RAP : Reconnaissance Automatique de la parole

f: fréquence en Hertz

T: Période en seconde

Te : Période d'échantillonnage

λ : Longueur d'onde en mètre

c : Célérité de propagation de l'onde en mètre par seconde

n : Nombre de bits

S(t) : Signal original

δ(t) : Signal Dirac

S_e : Signal échantillonné

H(z) : Filtre à réponse impulsionnelle finie

S_a : Filtre pré-accentué

N: Nombre d'échantillons de la parole

M : Nombre d'échantillons qui séparent de trames de parole

S_w : Signal fenêtré

W(n) : fenêtre de Hamming

F_0 : Fréquence fondamentale (Pitch)

m : Fréquence en Mels

FFT : Transformée de Fourier Rapide (Fast Fourier Transform)

DFT : Transformée de Fourier Discrète (Discrete Fourier Transform)

DAP : Décodage Acoustico-Phonétique

DCT : Transformée en Cosinus discrète (Discrete Cosine Transform)

iDCT : Transformée en Cosinus Discrète inverse (inverse Discrete Cosine Transform)

m_M : Nombre de filtres

C_i : Coefficients MFCC

u(n) : signal d'exitation

G : le gain

a_k : Coefficients de prédiction

e(n) : Erreur de prédiction

E : Erreur quadratique totale de prédiction

AP : Acoustico-Phonétique

RF : Reconnaissance de Formes

IA : Intelligence Artificielle

SE : Systèmes Experts

INNS: International Neural Network Society

w: Poids de la première couche

a : Poids de la seconde couche

d_k : Sortie désirée du réseau de neurones

s_k : Sortie obtenue par les réseaux de neurones

p_i : Le potentiel post-synaptique du neurone i.

x_j : L'état du neurone de la couche cachée précédente.

w_{ij} : Le poids de la connexion entre les deux neurones.

G_i : Le gradient

β : Pas de déplacement

Chapitre IV : Résultats et comparaisons

-Liste des Tableaux :

-Liste des figures :

-INTRODUCTION GENERALE:

La parole est le moyen privilégié de communication de l'homme. L'environnement de ce dernier est, en partie, peuplé de machines avec lesquelles il a toujours voulu communiquer. Grâce aux efforts de recherches menés dans différents domaines, de nombreux systèmes de reconnaissance de la parole ou du locuteur sont actuellement proposés. Malgré l'étendue des efforts fournis, le traitement de la parole en temps réelle reste une condition très difficile et qui diminue les performances de ces derniers. Actuellement, les axes de recherche en traitement de la parole sont très divers. On parlera alors de reconnaissance du locuteur ou de la langue ou encore le traitement du signal ou bien la linguistique. Aujourd'hui, l'état d'avancement du développement des systèmes de reconnaissance de la parole, du locuteur ou de la langue est à un degré ou il pourrait être accessible pour le grand public. Cependant, ces applications se heurtent à de nombreuses limites qui ralentissent cette accessibilité. Parmi elles, on peut citer les contraintes d'apprentissage des modèles qui se traduisent par un temps de configuration trop exhaustif pour l'utilisateur. Le riche vocabulaire de certaines langues qui limitent les applications de reconnaissance de la parole. Ainsi que la variété des environnements sonores (bruit de fond, téléphone, musique...) qui limitent considérablement les performances de la reconnaissance. Pour une conception d'un système de reconnaissance optimale il est nécessaire de porter un soin particulier au étapes du traitement du signal (acquisition, filtrage, extraction de caractéristiques, reconnaissance des formes, modèles acoustiques et de la langue...), pour garantir une bonne résistance au bruit de nombreuses méthodes de filtrage ont étaient proposées comme par exemple, le filtrage cepstral. Une avancée très importante est l'utilisation de méthodes d'adaptation statistiques. Ces méthodes modifient les paramètres du système de reconnaissance en vue de meilleures performances, lors du changement du locuteur ou d'environnements sonores (microphones, téléphones…)

Le travail présenté dans ce mémoire concerne une étape importante et commune à la majorité des systèmes de traitement de la parole, il s'agit en effet de l'extraction de caractéristiques. Son objectif est d'obtenir une représentation compacte mais également informative pour les prochaines étapes du système qui différent en fonction de la tâche effectuée (reconnaissance de la parole, du locuteur, de la langue ou de la musique par exemple). L'extraction de caractéristiques est donc une étape fondamentale. De nombreuses méthodes sont couramment utilisées par les chercheurs comme le codage linéaire prédictif (LPC : Linear Predictive Coding) ou le codage cepstral (MFCC : Mel Frequency Cepstral Coding). L'amélioration des systèmes actuels passe, entre autre, par l'optimisation de l'étape d'extraction de caractéristiques. En effet de nombreux auteurs montrent l'importance de cette étape .Ce travail a pour but d'étudier l'étape d'extraction de carac téristiques ainsi présenter, étudier et comparer les différentes méthodes (MFCC : Mel Frequency Cepstral Coding)., (LPC : Linear Predictive Coding), (NPC : Neural Predictive Coding) .Les approches décrites dans ce document concerne le traitement automatique de la parole .Ce domaine est une des voies explorées par de nombreux chercheurs pour l'amélioration des systèmes de reconnaissance. Un des objectifs du traitement automatique de la parole est une meilleure connaissance mais également une meilleure exploitation des différents phonèmes présents durant la production et la perception de la parole. Pour cela nous avons divisé ce mémoire en quatre chapitres :

Le premier chapitre est une introduction à la phonétique et présente les systèmes régissant le langage chez l'être humain et le fonctionnement des appareils concernés, introduisant ainsi le système physiologique et le système neurologique et leur anatomie. Le système physiologique se constitue de l'appareil phonatoire qui est le moteur de la production du son des différents phonèmes grâce à l'interaction des trois grands organes (les poumons, le larynx et les cavités bucco-pharyngale) et l'appareil auditif qui a comme organe principal l'oreille qui est le centre du traitement acoustique et cognitif. Le système neurologique est la partie nerveuse dite le cerveau qui est constitué de neurones assurant ainsi le traitement des différents sons des phonèmes et leur compréhension.

Le deuxième chapitre présente le traitement automatique de la parole. Ce dernier comportera les différentes caractéristiques du signal de parole, évoquant les étapes de la numérisation et détaillant les méthodes traditionnellement mises en œuvre pour cette analyse. Ce chapitre sera l'occasion de présenter en profondeur les différentes méthodes du codage LPC et MFCC.

Le troisième chapitre comportera une introduction globale sur l'intelligence artificielle, puis précisément sur les réseaux de neurones, leur évolution durant le siècle dernier citant les différents types des réseaux de neurones. On se focalisera sur un perceptron multicouche MLP afin d'utiliser un nouveau modèle pour l'extraction de caractéristiques le Codage Neuro-Prédictif (NPC, Neural Predictive Coding) qui est une extension au domaine non-linéaire du codage LPC.

Quant au quatrième chapitre, il sera consacré à une présentation de la langue Amazighe et précisément les lettres Tifinagh puis à l'étude de la mise en forme d'un signal de parole qui sera injecté dans un réseau de neurones MLP (Multi Layer Perceptron), puis la comparaison entre les résultats obtenus par l'utilisation des deux codages : MFCC (Mel Frequency Cepstral Coding) et NPC (Neuronal Predictive Coding). Alors comment se fait la reconnaissance automatique de la parole ? Comment peut-on extraire les caractéristiques de manière à avoir une bonne reconnaissance? Et comment favoriser l'exploitation des phonèmes de la parole ?

I-Introduction :

L'être humain produit et perçoit des sons et peut donc construire des phrases et des mots peu importe la langue parlée c'est la base d'une communication entre deux individus ou plus. Afin de comprendre ce comportement naturel chez l'être humain, la recherche se répartira en deux : le système physiologique qui comportera l'appareil auditif et phonatoire, et le système neurologique viendra compléter l'idée reçue en première partie.

Les sons produits par l'Homme, sont étudiés et transcrits en phonétique et en phonologie ; deux filières du domaine de la linguistique. La phonologie est une branche qui étudie les sons utilisés dans la communication parlée et donc leur production, leur variation plutôt que leur contexte, les unités phonétiques sont les « phones » à la différence de la phonologie qui est une branche qui étudie comment sont agencés « les phonèmes » d'une langue pour former des mots, L'abbé Rousselot (14 octobre 1846 - 16 décembre 1924) est considéré comme le fondateur de la phonétique grâce à ses études sur les sons utilisés dans une communication verbale. [1] [2]

La phonétique se divise en trois branches : Articulatoire, acoustique et auditive. Nous nous pencherons plus à la phonétique auditive qui étudie la façon dont les sons sont perçus par le récepteur.

Par ailleurs, l'assimilation de la naissance et l'acheminement d'un message nerveux est tout aussi important à la production et l'acquisition d'un son. Car le système neurologique apportera non seulement, une interprétation puis une signification au son ou au mot perçu, mais aussi une production d'une phrase sensée, comprise, par d'autres individus afin de pouvoir communiquer.

II-Mécanisme de production de la parole chez l'être humain :

Le processus de production de parole est un mécanisme très complexe qui repose sur une interaction entre le système neurologique et physiologique. Il y a une grande quantité d'organes et de muscles qui entrent dans la production de sons des langues naturelles. [3]

II-1-Système physiologique :

II-1-1 - L'appareil phonatoire :

Le fonctionnement de l'appareil phonatoire humain repose sur l'interaction entre trois grandes classes d'organes: les poumons, le larynx et les cavités bucco-pharyngale. [4]

Dans la phonation, tout débute par l'action des poumons. Ceux-ci libèrent, à un rythme qui est sous le contrôle volontaire du locuteur, un souffle d'air, lequel passe par la trachée et traverse le larynx. Le larynx transforme alors le souffle en son glottique (ou laryngé), C'est, en effet, à l'intérieur du larynx, organe constitué de cartilages réunis entre eux par des ligaments et des muscles, que logent les cordes vocales. Celles-ci sont en fait des fibres musculaires contrôlées par des muscles qui ont comme fonction de les tendre, de les dilater, de les rétrécir ou encore de les allonger. Selon le type d'action des muscles et des cartilages sur les cordes vocales, l'espace entre elles peut être plus ou moins large, ou complètement fermé. C'est dans cet espace que passent les souffles d'air produits par I 'expiration. On appelle glotte cette zone entre les cordes vocales. Lorsque la glotte est largement ouverte, elle permet la respiration et aucun son n'est engendré lorsqu'elle n'est que faiblement ouverte, elle produit le chuchotement et quand elle est complétement fermée, il y a phonation (Figure 1.1). [4]

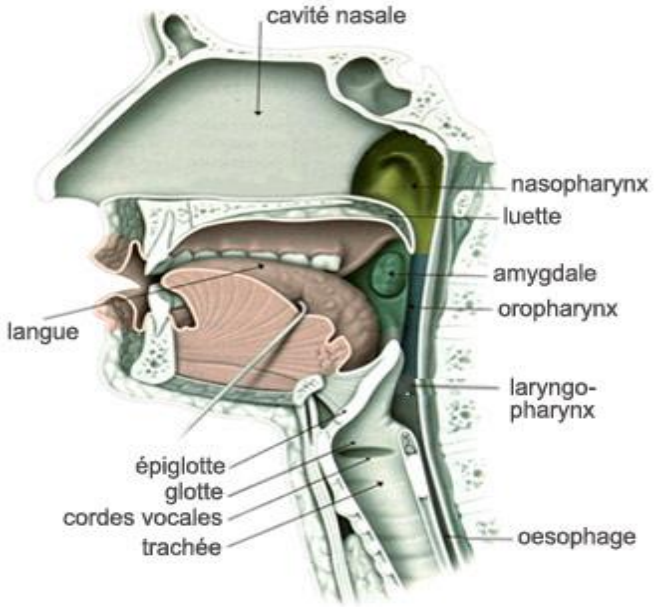

Figure 1.1 Vue de l'appareil phonatoire. [5]

Pendant la respiration calme, les cordes vocales sont écartées et laissent libre passage à l'air. Pour que la phonation puisse se produire et donc pour que les cordes vocales puissent vibrer, elles doivent s'accoler. Si elles ne sont que légèrement accolées de façon à laisser passer une partie du souffle d'air, elles donnent lieu au chuchotement. En se mettant la main devant la bouche et en prononçant une phrase d'abord en chuchotant ensuite à haute voix, on pourra se rendre compte que la pression d'air sur la main est plus forte quand on chuchote que quand on parle à haute voix. [4]

La mise en vibration des cordes vocales qui engendrera le son glottique lorsque l'accolement des cordes vocales constitue un obstacle à l'écoulement normal du souffle d'air venu des poumons, ce qui engendre la création d'une pression sous la glotte qui est fermée. Lorsque cette pression, dite sous-glottique atteint une force suffisamment grande, les cordes vocales cèdent et laissent passer la bouffée d'air. Après le passage de cet air les cordes vocales reprennent rapidement leur position initiale sous l'action de leur masse, de leur tension et de leur élasticité. Ce cycle de pression-relâchement recommence aussitôt que la pression sous-glottique s'est reconstituée. Ces ouvertures et fermetures des cordes vocales sont tellement rapides qu'on dit que ces dernières vibrent et le nombre de vibration a la seconde de ces cordes vocales s'appelle la fréquence d'un son. [4]

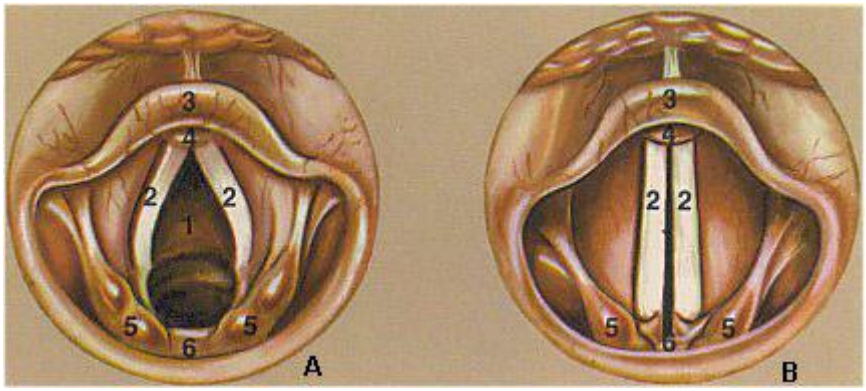

Figure 1.2 Cordes vocales (Abduction (A) et adduction (B)). [5]

Normalement, ce son glottique produit par les vibrations ne peut s'entendre directement. II est immédiatement transforme en pénétrant dans la cavité bucco-pharyngale, car celle-ci, agissant comme un résonateur, le modifie. la cavité bucco-pharyngale se modifie constamment sous l'action d'organes mobiles, telles la langue, les lèvres et l'uvule. La langue peut, en effet, s'abaisser, s'élever, reculer, avancer, les lèvres peuvent être projetées ou écartées, l'uvule peut s'accoler a la paroi pharyngale ou s'en détacher, laissant alors le son glottique entrer dans les fosses nasales. A chacune des configurations que peut adopter la cavité bucco-pharyngale correspond un son diffèrent. [4]

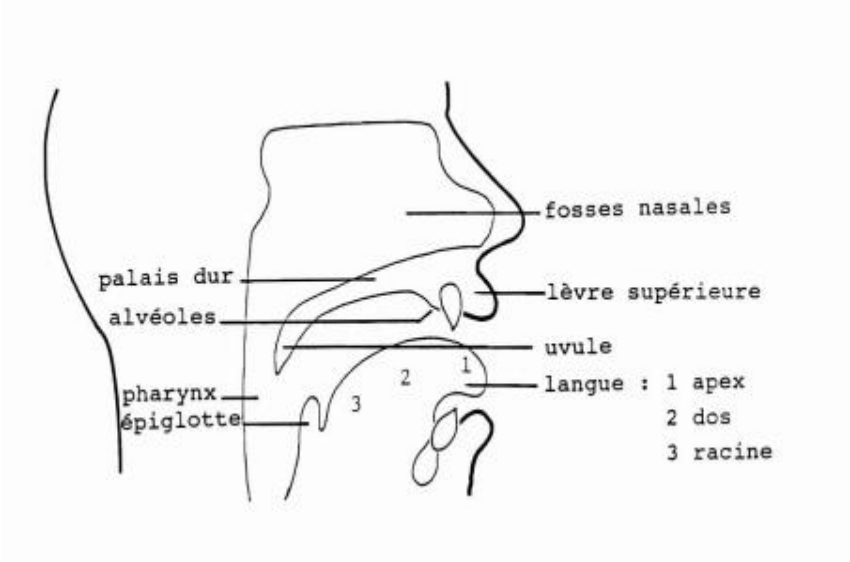

Figure 1.3. Vue de la cavité bucco-pharyngale. [4]

Ainsi, pour produire le son de la voyelle « i », tel qu'on peut l'entendre dans le mot anglais « see » et dans le mot français « ni », la langue doit s'élever et s'avancer, tandis que pour réaliser le son orthographié « ou » en français et « oo » en anglais la langue doit s'élever et reculer; de plus, les lèvres doivent être projetées vers l'avant. De la même manière, pour que soit produite la voyelle française « u », la cavité bucco-pharyngale doit avoir à la fois une configuration identique à la voyelle « i » et une projection labiale, c'est-à-dire que les lèvres doivent être avancées. A la différence lors de la prononciation d'une consonne par exemple la consonne «m », les lèvres sont accolées et le voile du palais (uvule, velum) est abaisse, permettant ainsi à l'air de s'insérer dans les cavités nasales. [4]

II-1-1-1- Les phonèmes :

Le phonème : c'est la plus petite unité distinctive de la chaîne parlée, c'est- à-dire la plus petite unité de son capable de produire un changement de sens par commutation, Le français compte 37 phonèmes, L'anglais 44. L'italien 42. L'allemand 68. L'espagnol 34. [8]

Le graphème : c'est la plus petite unité du système graphique destiné à transcrire les phonèmes. Il est constitué par une ou plusieurs lettres : [o] = o, au, eau (3 graphèmes distincts pour le même phonème), Il est au niveau de la manifestation écrite de la langue ce que le phonème est au niveau de la manifestation orale, Le français compte 130 graphèmes ou plus. [8]

La lettre : c'est une unité de l'alphabet qui en compte 26, seule ou combinée avec d'autres, elle participe à la constitution du graphème. [8]

a- Les phonèmes de la langue Française :

VOYELLES	[i]		*mie, midi, livide*
	[e]	(é fermé)	*des, les, brûlé, chantai*
	[ɛ]	(è ouvert)	*lait, mets, chantais, tête*
	[a]	(a antérieur)	*vache, sac, patte, ta*
	[ɑ]	(a postérieur)	*tas, pâte, âne*
	[£]	(o ouvert)	*Paul, bol, sotte*
	[o]	(o fermé)	*Paule, La Baule, sot*
	[u]	(ou français)	*choux, cour, moule*
	[y]	(u français)	*sur, sûr, j'eus*
	[□]	(eu fermé)	*affreux, meute, heureuse*
	[œ]	(eu ouvert)	*jeune, bonheur, œuvre*
	[…]	(e sourd ou muet)	*cheveux, me*
	[~ɛ]	([ɛ] nasalisé)	*brin, frein, main, faim*
	[Œ]	([œ] nasalisé)	*un, brun, humble*
	[³]	([a] nasalisé)	franc, tante, tente
	[¥]	([£] nasalisé)	rond, mouton, monter
CONSONNES	[p]		*pédicure, appétit*
	[b]		*babouche, aborder, abbé*
	[t]		*tendre, porter, attendre*
	[d]		*dorer, adorer, pardon*
	[k]		*coque, croquer, képi*
	[g]		*gage, naviguer*
	[f]		*fraise, phantasme*
	[v]		*vagabond, wagon*
	[s]		*satin, assez, maçon, cerf, six*
	[z]		*zigzag, prison*
	[Ç]		*louche, chat*
	[ʒ]		*jardin, bourgeon*
	[m]		*marmite, pomme*
	[n]		*nourrir, nonne*
	["gn"]		*rogner, montagne*
	[l]		*laisser, mille*
	[R]		*rare, fourrage*
	[h]		*parking*
SEMI-VOYELLES	[j]		*yeux, œil, renier, fille*
	[□]		*fuir, puits, bruit, duel*
	[w]		*oui, foi, loin, fouet*

-1- Tableau des phonèmes de la langue Française. [10]

Ceci dit il y'a un paramètre très important à prendre en considération, le voisement qui une propriété de certains sons de la parole. Un son est voisé ou dit Sonore si sa production s'accompagne d'une vibration des cordes vocales et sinon, il est non voisé ou dit sourd. [11]

Les voyelles, étant portées par la voix, sont naturellement voisées même si parfois elles subissent un dévoisement avec une atténuation de la vibration laryngale, dans le cas de la parole chuchotée. [11]

Les consonnes, en langue Française sont 21 consonnes dont 3 glissantes, on peut les classées en deux types :

-Consonnes sonores (voisées) :[b], [d], [ɡ], [m], [n], [ɲ], [ŋ], [v], [z], [ʒ], [ʁ], [l] .

-Consonnes sourdes (non-voisées) : [p], [t], [k], [f], [s], [ʃ]. [11]

Et parmi ces consonnes il y'a les consonnes occlusives ou les plosives pour lesquelles le passage de l'air est bloqué par une occlusion momentanée du conduit vocal : [p], [t], [k], [b], [d], [ɡ]. [11]

Puis les consonnes constrictives (fricatives) pour lesquelles le passage de l'air est rétréci par un resserrement du conduit vocal : [f], [v], [s], [z], [ʃ], [ʒ], [ʁ],[l]. [11]

Ces deux catégories de consonnes (occlusives et constrictives) sont des consonnes orales car pendant leur production le voile du palais est relevé et l'air passe donc uniquement par la bouche. [11]

Enfin les consonnes nasales, elles sont produites avec une occlusion momentanée du conduit vocal mais avec un flux d'air continu au niveau des fosses nasales : [m], [n], [ɲ], [ŋ]. [11]

II-2-1- L'appareil auditif :

Le système auditif est en soi une merveille du traitement acoustique et cognitif comportant un capteur qui est l'oreille et un centre de traitement qui est le cerveau, il sert à de multiples tâches comme la maîtrise de l'équilibre (vestibule), la compensation automatique de la pression (par les trompes d'Eustaches) mais aussi l'audition qui est le sujet même de cette partie. [6]

II-2-1-1- Anatomie de l'oreille humaine :

L'oreille est composée de trois grandes parties :

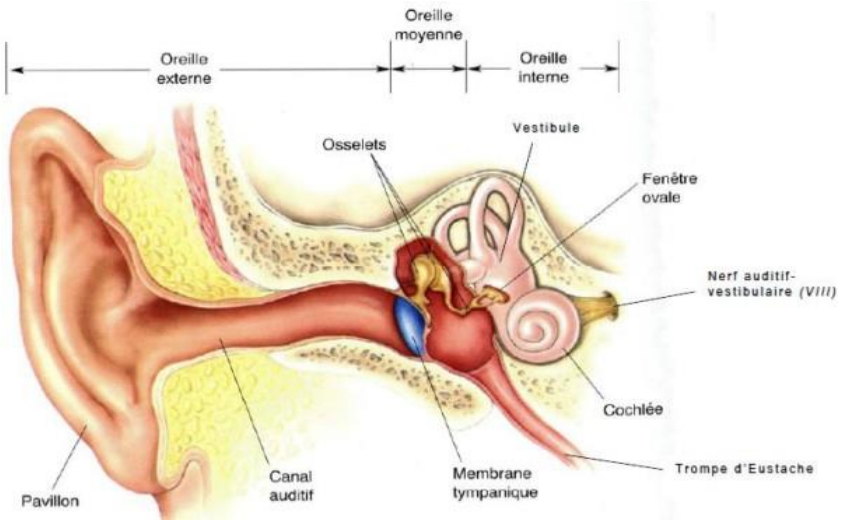

Figure 1.4. Composition de l'oreille humaine. [6]

-L'oreille externe : Elle comporte un pavillon et un conduit qui se termine au fond par le tympan. Le pavillon joue le rôle de " coupe-vent ", et de collecteur d'ondes. Grâce à la présence des deux oreilles, on peut saisir l'effet stéréophonique et, accessoirement repérer la direction où se trouve la source, par estimation de la différence d'intensité ou de phase du même signal perçu par les deux oreilles. [7]

-L'oreille moyenne : Elle comporte un dispositif extrêmement important: la chaîne des osselets. On distingue, dans l'ordre: le marteau, appliqué sur le tympan, l'enclume et l'étrier. Ces osselets sont très petits: l'étrier ne dépasse pas la taille d'un grain de riz. Ils sont articulés les uns aux autres et maintenus élastiquement en place grâce à des ligaments et des muscles. C'est un système articulé déformable. [7]

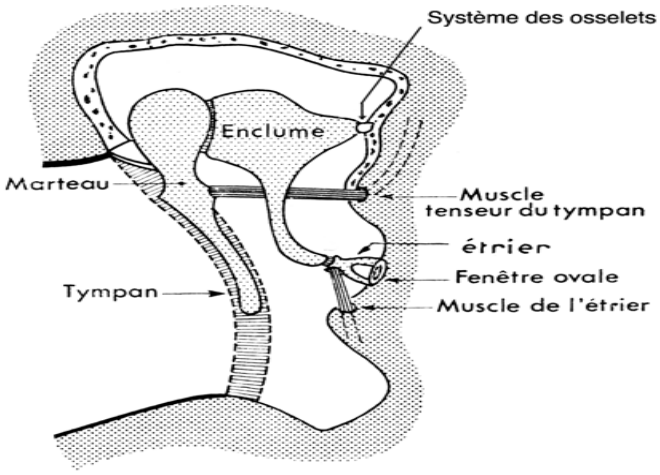

Figure 1.5. L'oreille moyenne. [7]

Ces muscles peuvent freiner les amplitudes, lorsque prévenus d'un signal intense, ils se tendent, en s'adaptant par voie de réflexe, pour protéger le tympan. Inversement, on peut " tendre l'oreille " et favoriser les amplitudes au maximum. Ce mécanisme d'adaptation réflexe montre qu'on ne peut pas mesurer l'intensité perçue d'un son qu'avec des unités physiques. [7]

-L'oreille interne : Elle est en forme de limaçon creusé dans l'os du rocher. Un système de " canaux semi-circulaires " connexe joue un rôle important dans la sensation d'équilibre. Le limaçon est partagé en deux rampes: la rampe vestibulaire et la rampe tympanique qui communiquent à leur extrémité par un petit trou l'hélicotrême. Entre ces rampes, à l'intérieur des deux membranes sont disposées les cellules nerveuses ou cellules de Corti au nombre de 24 000. De chaque cellule de Corti part une fibre nerveuse, dont l'ensemble donne le nerf acoustique qui relie l'oreille au cerveau. [7]

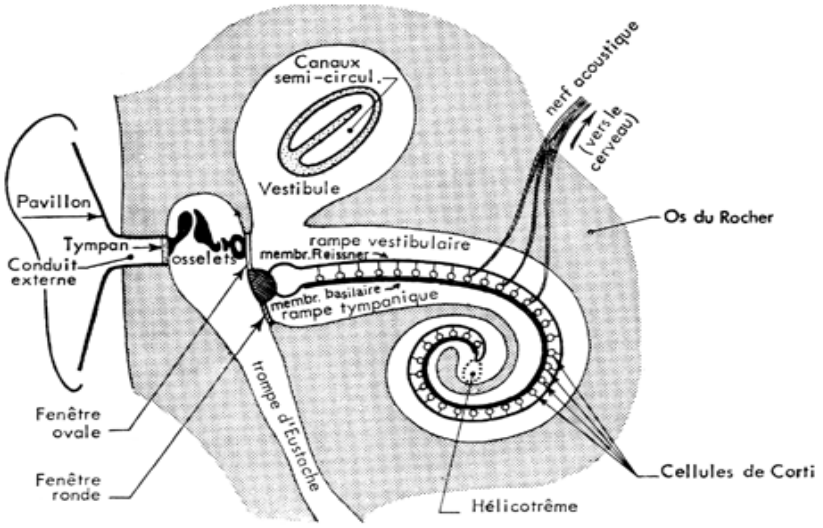

Figure 1.6. L'oreille interne. [7]

II-2-1-2- Fonctionnement de l'appareil auditif humain:

Actuellement nous pensons que le system auditif humain agit comme un prisme acoustique qui peut séparer les fréquences d'un son par plage fréquentielles. [6]

Chaque partie du system auditif a une fonction bien particulière et possède des propriétés bien définies car à l'audition d'un son le pavillon sert à le capter puis le concentrer vers le conduit auditif qui possède une fréquence de résonnance qui oscille entre 1000 Hz et 4000 Hz ou il sera amplifié, une fois arrivé au bout du conduit auditif ce son va faire vibrer le tympan qui fonctionne à la manière d'un transducteur. Il convertit tout d'abord l'énergie acoustique en énergie mécanique avec concentration de cette énergie en son centre. Le tympan possède également une fréquence de résonnance. [6]

Ensuite, les osselets transmettent mécaniquement à la Cochlée le signal acoustique reçu du tympan en réalisant une adaptation de l'impédance et tout ce mécanisme fonctionne dans un milieu aqueux. [6]

Enfin, la Cochlée qui est un organe complexe et précis va traiter chaque plage de fréquence séparément (Les fréquences élevées (aigues) sont captées au début de la Cochlée, et les plus basses (graves) à la fin de la spirale, appelée APEX) et les transformer en flux nerveux dans la partie interne de l'oreille interne. [6]

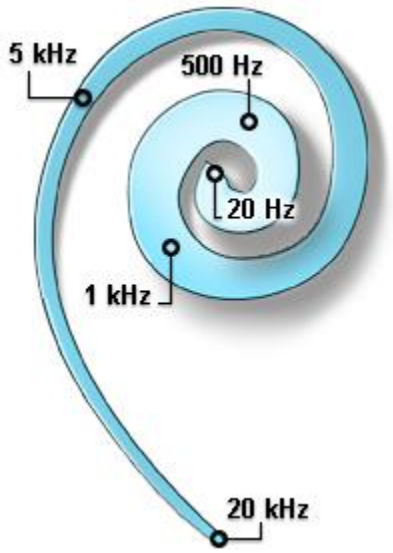

Figure 1.7. Répartition de l'analyse des fréquences.

II-2- Système neurologique :

II-2-1- La Neurophysiologie :

La neurophysiologie, physiologie du cerveau et des cellules nerveuses (neurone et cellule gliale), est la partie de la physiologie qui traite du système nerveux. [9]

II-2-2- Genèse du message nerveux :

Le système nerveux assure la perception de notre environnement et élabore des réponses adaptées permettant le bon fonctionnement de l'organisme comme la coordination des mouvements musculaires ou le contrôle de l'activité cardiorespiratoire. Il est constitué de milliards de neurones reliés par des synapses et organisés en réseau à travers lequel les messages nerveux sont conduits. [9]

Tout changement qui se produit à l'intérieur, ou à l'extérieur, de l'organisme, pouvant être détecté par les fibres sensitives, au niveau du récepteur, s'appelle stimulus. Lorsqu'un stimulus a une intensité suffisante, il excite le récepteur, générant ainsi un message nerveux. [9]

Les cellules nerveuses, comme toutes les cellules de l'organisme, possèdent un potentiel de membrane. La différence de potentiel observée entre le cytoplasme de la cellule et le milieu extérieur, en l'absence de tout stimulus, est le potentiel de repos. Le message nerveux est alors constitué d'une fréquence de potentiels d'action soit une combinaison particulière de signaux nerveux par unité de temps. La fréquence des potentiels d'action dépend de l'intensité du stimulus. Un stimulus de forte intensité déclenche un grand nombre de potentiels d'actions par unité de temps. [9]

II-2-3- Rôle des neurones dans la transmission du message nerveux :

Les neurones sont les cellules fondamentales du système nerveux. Ils sont composés d'un corps cellulaire et de deux types de prolongements : les dendrites qui conduisent le message nerveux jusqu'au corps cellulaire du neurone et un axone qui conduit le message nerveux en direction d'une cellule effectrice de la réponse ou jusqu'à une synapse, zone de communication entre deux neurones. [9]

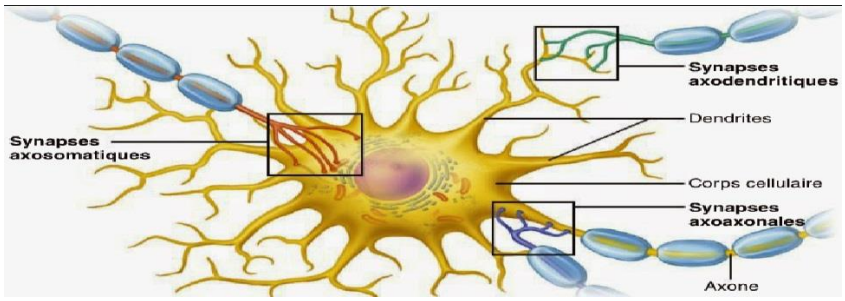

Figure 1.8. Structure d'un neurone.

Le système nerveux central est composé de la moelle épinière et de l'encéphale. C'est là qu'est traitée l'information. Le système nerveux périphérique est constitué des prolongements

ou fibres nerveuses regroupées en nerfs. Un nerf assure la transmission des informations afférentes (de la périphérie vers les centres nerveux) et efférentes (des centres nerveux vers la périphérie). [9]

II-2-4- Le fonctionnement du système nerveux :

Les messages nerveux se propagent sous forme de potentiel d'action de nature électrique, le long des axones des neurones, et par voie chimique, au niveau des synapses. Ces messages sont ensuite intégrés au niveau des centres nerveux, moelle épinière et encéphale, qui produisent une réponse adaptée. [9]

a- L'activité électrique des neurones :

Les neurones sont des cellules qui peuvent être excitées : en réponse à une stimulation, la fibre nerveuse produit un signal de nature électrique, élément constitutif de l'influx nerveux, le potentiel d'action (PA).

Le PA (potentiel d'action) est caractérisé par une durée brève (quelques millisecondes). La membrane du neurone subit une dépolarisation transitoire suivie d'un retour à son potentiel de membrane. Ce phénomène est capable de se propager le long de la fibre. La fibre nerveuse répond à une stimulation selon la loi du tout ou rien : aucun signal n'est émis par le neurone si l'intensité de la stimulation reste en dessous d'une certaine valeur seuil. [9]

b- La transmission chimique du message nerveux au niveau des synapses

La synapse est une structure spécialisée dans la transmission des messages nerveux entre deux neurones, ou entre un neurone et une cellule musculaire cette transmission a lieu par l'intermédiaire de substances chimiques, les neurotransmetteurs. La synapse est constituée de :

- l'extrémité du neurone pré synaptique par lequel le message nerveux arrive.

-une fente synaptique, espace entre les deux neurones, où sont déversés par exocytose les neurotransmetteurs, contenus dans des vésicules.

-la membrane du neurone suivant, dit post synaptique. Les neurotransmetteurs s'y fixent au niveau de récepteurs spécifiques ce qui provoque un changement de l'activité électrique du neurone et la naissance d'un nouveau signal électrique. [9]

c- Le codage des messages nerveux

Au niveau d'une fibre nerveuse, le message nerveux est codé en fréquence de PA : plus l'intensité du stimulus est supérieure à la valeur seuil, plus le nombre de PA se propageant le long de l'axone de la fibre nerveuse augmente, formant ce qu'on nomme des trains de PA successifs.

Au sein d'un nerf, le message nerveux est codé par le nombre de fibres stimulées. Le message se présente sous forme d'un potentiel global dont l'amplitude sera alors plus ou moins importante.

Au niveau d'une synapse, le message est codé en concentration de neurotransmetteur. [9]

II-2-5- L'intégration des messages au niveau des centres nerveux :

A l'échelle cellulaire, l'intégration des messages nerveux a lieu au niveau des motoneurones. Ils effectuent la sommation spatiale et temporelle des nombreux messages nerveux afférents. Si cette sommation est supérieure à leurs seuils d'excitation, ils élaborent de nouveaux messages nerveux efférents.

A l'échelle de l'organisme, le réflexe myotatique par exemple, peut être inhibé sous l'effet de messages nerveux provenant d'un autre type de récepteurs sensoriels, comme ceux qui perçoivent la douleur, ou bien par une commande volontaire provenant de notre cerveau. [9]

III- Conclusion :

L'être humain reste une merveille technologique encore très mal connue. Nous sommes capables, sans aucuns efforts, instantanément et de façon robuste, d'isoler un flux de parole d'un individu donné, à partir d'un paysage sonore complexe et bruité. Le prodige ne s'arrête pas là, nous sommes en plus capable d'authentifier le locuteur, percevoir son état émotionnel et élaborer un sens à cette suite de mots transmis par ce signal.

Nous pourrions évoquer également la formidable dynamique dont est capable notre organe, sans oublier toute la psycho-acoustique qui nous permet de localiser précisément une source sonore (même dans un paysage sonore bruité) ou encore évaluer la densité d'un solide ou d'un liquide par l'écoute du son produit par un impact sur ce dernier.

L'aisance de la production de la parole fait d'elle le moyen de communication le plus répandu dans la société humaine. En effet, il est plus facile de parler que d'écrire ou encore de schématiser. Cette simplicité renferme un traitement très complexe fait par le cerveau humain, ce qui rend la parole difficilement automatisable pour une machine.

I-Introduction :

La parole est l'un des principaux moyens de communication entre les êtres humains. Cette simple apparence à sa production semble être très complexe à l'origine. Ce paradoxe a suscité énormément d'interrogations chez les chercheurs et cela jusqu'à essayer de reproduire le système régissant le langage humain et de créer le dialogue entre l'Homme et la machine.

L'utilisation du clavier et autres périphériques rendent la communication avec la machine très difficile et très lente. L'avancement technologique et surtout de l'informatique a apporté le besoin de nouveaux moyens de dialogue « homme-machine » qui libéreraient l'homme d'un contact constant avec cette dernière. Comme pour l'écriture, le message de la parole doit donc être bien structuré selon des règles connues par tous ceux qui parlent la même langue (grammaire, syntaxe, vocabulaire…). Bien que l'évolution de la langue mue par les tendances actuelles rend de plus en plus difficile le fait de fixer des règles statiques une bonne fois pour toute.

L'être humain grâce à l'ordinateur le plus sophistiqué au monde (son cerveau), pressé de faire passer son message, transgresse le plus souvent ces règles et arrive sans difficultés à le comprendre, ce qui montre le caractère dynamique du cerveau humain qui arrive très bien à s'adapter à de nouvelles situations, caractère qui devrait s'imposer à tout système de traitement de la parole.

Mais d'abord, il est important de comprendre la base du traitement de la parole. Qu'est-ce que le signal de la parole ? Comment se fait le traitement du signal ? Et qu'est-ce que le traitement automatique de la parole ?

II- L'onde sonore :

C'est une onde produite par la vibration mécanique d'un support fluide ou solide et propagée grâce à l'élasticité du milieu environnant sous forme d'ondes longitudinale (Cordes vocales, guitare…). Par extension physiologique, le son désigne la sensation auditive à laquelle cette vibration est susceptible de donner naissance. Par conséquent le son ne peut pas se propager dans le vide.

La science qui étudie les sons s'appelle l'acoustique. La psycho acoustique combine l'acoustique avec la physiologie et la psychologie, pour déterminer la manière dont les sons sont perçus et interprétés par le cerveau.

La propagation d'une onde sonore dans un milieu se traduit par l'existence d'une pression acoustique qui s'ajoute à la pression atmosphérique. [14]

II-1- Caractéristiques d'une onde sonore :

L'onde sonore peut être caractérisée par ces paramètres :

-La Fréquence :

La fréquence est également une caractéristique d'une onde sonore. Celle-ci représente le nombre de périodes en une seule seconde, soit le nombre de fois qu'un phénomène périodique se reproduit par unité de temps. Pour un son, la fréquence représente donc le nombre de vibration par seconde. [14]

La période et la fréquence étant liées directement par la relation suivante :

$$f = 1/T \qquad (2.1)$$

Avec : f (Fréquence) en Hertz
T (période) en seconde

Selon sa fréquence, un son est défini comme grave ou aigu. Les sons graves ont une fréquence comprise entre 20 et 200 Hz. Tandis que les sons aigus ont une fréquence supérieure à 2 000 Hz. Un son dont la fréquence est comprise entre 200 et 2 000 Hz est dit médium. [14]

-L'intensité et l'amplitude :

L'énergie sonore qui traverse une surface de 1m² en 1seconde s'appelle intensité sonore. L'oreille humaine peut entendre des sons dont l'intensité varie entre 10^{-12}J/m2s et 100J/m2s. Le son le plus fort est donc 10^{14} fois plus intense que le plus faible. Les valeurs extrêmes sont trop éloignées l'une de l'autre, c'est pour cela que l'on préfère utiliser une autre unité de mesure que le J/m2s : le décibel (dB). Cette unité varie de 0 à 140 lorsque l'intensité sonore varie entre 10-12J/m2s et 100 J/m2s. Plus l'intensité sonore est élevée, plus le son perçu est fort. A l'inverse, plus l'intensité sonore est basse, plus le son perçu est faible. [14]

-Un son à peine audible (Is= 10^{-12}J/m2s), il correspond un niveau sonore de 0 décibels.

-Si l'intensité sonore est 10 fois plus grande que 10^{-12}J/m2s, on a un niveau sonore de 10dB.

-Si l'intensité sonore est 100 fois plus grande que 10^{-12}J/m2s, on a un niveau de 20dB, etc.

L'intensité est liée à l'amplitude des vibrations sonores. L'amplitude représente les variations de pression de l'onde, plus cet écart est grand plus l'amplitude est grande. Lorsque l'amplitude de l'onde est grande, l'intensité l'est de même et par conséquent, le son est plus fort. [14]

-L*a longueur d'onde :*

La longueur d'onde est la distance séparant deux molécules successives dans le même état vibratoire (même pression) ou encore **la distance parcourue par l'onde pendant une période (notée T).** Plus la longueur d'onde est courte, plus la fréquence est élevée, et donc plus le son est aigu. A l'inverse, plus la longueur d'onde est longue, plus la fréquence est faible, et par conséquent le son est plus grave. La longueur d'onde est notée λ et s'exprime en mètre. [14]

Dans un milieu donné, la fréquence et la longueur d'onde sont liées par la formule :

$$\boldsymbol{\lambda = c/f = c * T} \quad (2.2)$$

Où λ est la longueur d'onde en mètre (m), c la célérité de propagation de l'onde en mètre par seconde (m/s), f la fréquence (Hz) et T la période (s).

-Le timbre :

Le timbre permet à l'oreille de distinguer deux sons qui auraient la même fréquence et la même intensité mais qui seraient produits par deux instruments différents. Grâce à un analyseur de spectre, on peut analyser une note produite par un instrument, on se rend compte que le son résultant n'est pas composé d'un son pur mais d'une multitude de fréquences, appelée son complexe. [14]

En effet, un son pur est très rare car un corps ne peut pas vibrer sans entraîner la vibration d'un corps voisin. Un son naturel est en fait composé de plusieurs sinusoïdes différentes, une plus basse que l'on nomme fondamentale (c'est le son de base), et d'autres plus hautes appelées harmoniques. Un son pur comprend exclusivement la fréquence fondamentale. Les cordes vocales par exemple ne peuvent pas vibrer une par une et produisent donc plusieurs fréquences. Le timbre permet donc également à une personne de reconnaître une voix qui lui est familière. Deux sons peuvent avoir une même fréquence fondamentale et une même intensité, mais ne peuvent donc jamais avoir le même timbre. Cela se nomme l'identité sonore. [14]

II-2- Le signal de la parole :

Le signal de la parole est un phénomène de nature acoustique porteur d'un message. L'information d'un message parlé réside dans les fluctuations de l'air, engendrées, puis émises par l'appareil phonatoire. Ces fluctuations constituent le signal vocal. Elles sont détectées par l'oreille qui procède à une certaine analyse. Les résultats sont transmis au cerveau qui les interprète. [15]

D'autre part, le signal vocal représente la combinaison d'éléments simples et brefs du signal sonore appelés phonèmes, qui permettent de distinguer les différents mots. La parole est un signal réel, continu, d'énergie finie et non stationnaire. Sa structure est complexe et variable avec le temps. [15]

II-3- Caractéristiques du signal de la parole :

Le signal de parole n'est pas un signal ordinaire : il s'inscrit dans le cadre de la communication parlée, un phénomène des plus complexes. Afin de souligner les difficultés du problème, nous faisons ressortir essentiellement quelques caractéristiques notoires de ce signal :

-Un débit intense :

D'un point de vue mathématique, il est ardu de modéliser le signal de parole, car ses propriétés statistiques évoluent au cours du temps. [16]

-Une extrême redondance :

Lorsqu'on a vu une représentation graphique de l'onde sonore on est certainement frappé par le caractère répétitif du signal de parole. En effet, un grossissement visuel permettrait de voir une succession de figures sonores semblant se répéter à l'excès. Un peu de recul laisse apparaître des zones moins stables qu'il convient de qualifier de transitoires. En fait, ce qui semblerait de prime abord superflu, s'avère en réalité fort utile. Les répétitions confèrent à ce signal une robustesse car cette redondance le rend résistant au bruit. . [16]

-Une grande variabilité :

Une même personne ne prononce jamais un mot deux fois de façon identique. La vitesse d'élocution en détermine la durée. Toute affection de l'appareil phonatoire peut altérer la qualité de la production. Un rhume teinte les voyelles de nasalité ; une simple fatigue et l'intensité de l'onde sonore fléchit, l'articulation perd de sa clarté. La diction évolue dans le temps : l'enfance, l'adolescence, l'âge mûr, puis la vieillesse. [16]

La variabilité interlocuteur est encore plus flagrante. La hauteur de la voix, l'intonation et l'accent diffèrent selon le sexe, l'origine sociale, régionale ou nationale. D'ailleurs, la reconnaissance du locuteur est un axe de recherche à part entière. Enfin, toute parole s'inscrit dans un processus de communication où entrent en jeu de nombreux éléments comme le lieu, l'émotion, l'intention, la relation qui s'établit entre les interlocuteurs. Chacun de ces facteurs

détermine la situation de communication, et influe à sa manière sur la forme et le contenu du message. [16]

-Un lieu d'interférences :

La production "parfaite" de chaque son suppose théoriquement un positionnement précis des organes phonatoires. Or, lorsque le débit de la parole s'accélère, le déplacement de ces organes est limité par une certaine inertie mécanique. Les sons émis dans une même chaîne acoustique subissent l'influence de ceux qui les suivent ou les précèdent. Ces effets de coarticulation sont des interférences. Ils entraînent l'altération des formes sonores en fonction des contextes droits ou gauches, selon des règles étudiées par les acousticiens d'un point de vue articulatoire ou perceptif. [16]

II-4- Numérisation d'un signal :

Le signal est une variation (dans le temps de préférence) d'une grandeur physique de nature quelconque porteuse d'information. L'opération de numérisation du signal audio se réalise en théorie en trois étapes (échantillonnage, quantification, codage). [17]

II-4-1- Échantillonnage :

L'échantillonnage consiste à mesurer l'amplitude de l'onde à des intervalles de temps réguliers. Les impulsions représentent les amplitudes instantanées du signal à chaque instant. Plus le nombre d'échantillons est grand et plus le signal sera représenté finement. [17]

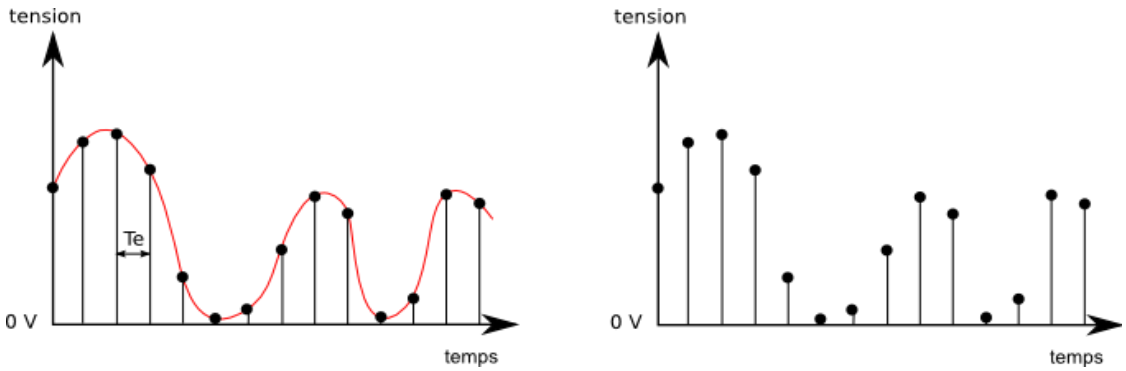

Figure 2.1 Signal échantillonné. [19]

II-4-2- Quantification :

Quantifier un signal consiste à placer les amplitudes des échantillons sur une échelle de valeurs à intervalles fixes. Chaque impulsion correspond donc à un nombre binaire unique. -Une quantification à n bits permet d'utiliser 2^n valeurs différentes.

-Pour 8 bits, on a 256 valeurs et pour 16 bits, on a 65536 valeurs. La transformation d'une valeur physique (en volts) en une valeur binaire introduit donc une distorsion. De même lorsque l'impulsion dépasse la valeur maximale prévue. [17]

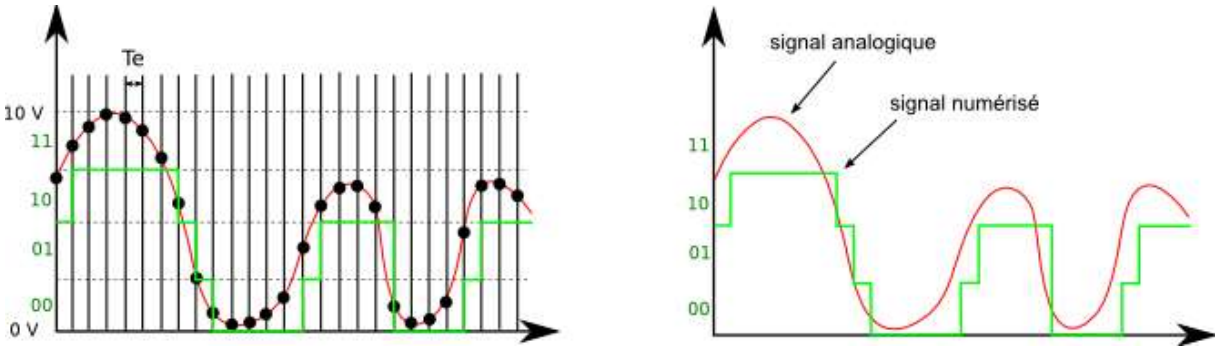

Figure 2.2 Signal quantifié. [19]

II-4-3- Codage :

Dans la littérature technique, ce terme englobe indifféremment toutes les méthodes de compression, les paramétrages d'échantillonnage et de quantification. En principe, le codage désigne le type de correspondance que l'on souhaite établir entre chaque valeur du signal analogique et le nombre binaire qui représentera cette valeur. [17]

III- Reconnaissance de la parole :

La reconnaissance automatique de la parole (RAP) est le processus par lequel la machine tente de « décoder » le signal de la parole qui lui est destiné. Les recherches relatives à la RAP débutèrent dans les années 1950, dans une conjoncture optimiste, car on pensait que les avancées technologiques des ordinateurs rendraient la RAP une tâche aisée. Quelques dizaines d'années plus tard, on se rendait compte que c'était faux, et que la RAP, demeure un problème difficile. Aujourd'hui encore nombre de questions restent posées, les difficultés majeures étant associées à la taille du vocabulaire à reconnaître, la reconnaissance de la parole spontanée, à la reconnaissance indépendamment du locuteur, la parole bruitée, …

La reconnaissance automatique de la parole est très souvent basée sur une représentation paramétrique du signal, son but étant la communication en langue naturel avec une machine. Il s'agit là de deux objectifs différents que l'on peut assigner à un système : la reconnaissance conduisant à une application du type dictée vocale, et la compréhension, qui consiste à accéder à la signification de l'énoncé parlé. [20]

III-1- Concept de base :

La démarche classique suivie lors du processus de reconnaissance automatique de la parole est illustrée par la figure II.1, ce schéma fait ressortir les étapes principales dans un tel processus. [20]

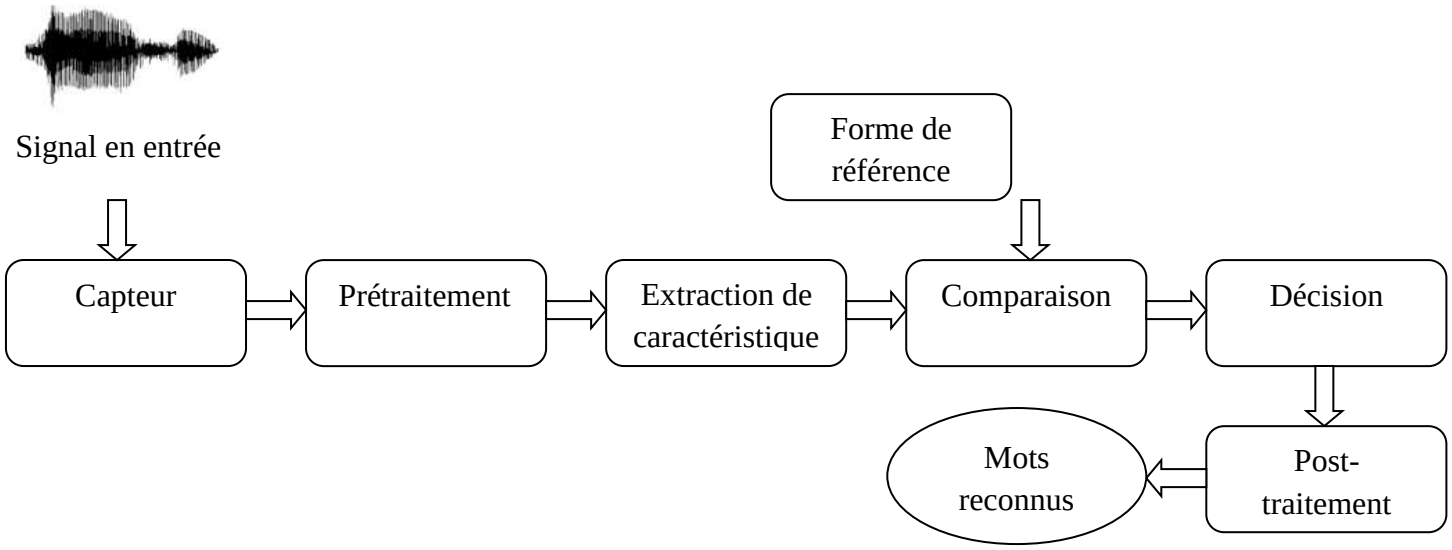

Figure 2.3 Organigramme d'un système de reconnaissance de la parole.

Ainsi, étant donné un signal en entrée du système, celui-ci va subir un prétraitement qui consiste généralement en un filtrage et un échantillonnage qui permet de passer d'un signal continu à des valeurs discrètes, de ces valeurs dont le nombre est important seront extraites des caractéristiques qui permettent de représenter de façon compacte et pertinente le signal originel. Cette étape permet d'avoir une première représentation du signal, ensuite et selon l'approche adoptée par le système de reconnaissance, ce modèle représentatif du signal sera comparé à des formes d'autres signaux que le système « connaît ». Sur la base du résultat de cette comparaison une décision quant au mot reconnu sera prise, celle-ci sera éventuellement validée en considérant les connaissances du domaine. [20]

III-2- Analyse acoustique :

III-2-1- Filtrage et échantillonnage :

Le signal est filtré puis échantillonné. Échantillonner un signal revient à prendre un certain nombre de points régulièrement espacés de ce signal. Cette fonction consiste mathématiquement à multiplier le signal original S(t) avec un signal d'amplitude 1 à chaque instant pour lequel on prend un échantillon (opération répétée à la période T_e) et d'amplitude zéro sinon. Cette fonction est appelée peigne (ou train périodique d'impulsions) de Dirac S(t). [18]

Le choix de la fréquence d'échantillonnage n'est pas aléatoire car une petite fréquence nous donne une présentation pauvre du signal. Par contre une très grande fréquence nous donne des mêmes valeurs, redondance, de certains échantillons voisins, donc il faut prélever suffisamment de valeurs pour ne pas perdre l'information contenue dans **s(t).** [18]

Cette problématique a été résolue par le théorème de Shannon « la fréquence d'échantillonnage assurant un non repliement du spectre doit être supérieure à 2 fois la fréquence haute du spectre du signal analogique », cette fréquence est typiquement de 8 kHz pour la parole de qualité téléphonique et de 16 à 20 kHz pour la parole de bonne qualité. [18]

III-2-2- Préaccentuation :

Préaccentuation : le signal échantillonné S_e est ensuite pré-accentué afin de relever les hautes fréquences qui sont moins énergétiques que les basses fréquences. Cette étape consiste à faire passer le signal S dans un filtre numérique à réponse impulsionnelle finie de premier ordre donné comme suit : [20]

$$\boldsymbol{H(z) = 1 - \alpha z^{-1}} \qquad \text{Avec} \qquad \boldsymbol{0.9 \leq \alpha \leq 1} \qquad (2.3)$$

Ainsi, le signal pré-accentué s_a est lié au signal s_e par la formule suivante :

$$\boldsymbol{S_a(n) = S_e(n) - \alpha S_e(n-1)} \qquad (2.4)$$

III-2-3- Segmentation :

Les méthodes du traitement de signal utilisées dans l'analyse du signal vocal opèrent sur des signaux stationnaires, alors que le signal vocal est un signal non stationnaire. Afin de remédier à ce problème, l'analyse de ce signal est effectuée sur des trames successives de parole, de durée relativement courte sur lesquelles le signal peut en général être considéré comme quasi stationnaire. Dans cette étape de segmentation, le signal pré-accentué est ainsi découpé en trames de N échantillons de parole. En général N est fixé de telle manière à ce que chaque trame corresponde à environ 20 à 30 ms de

parole. Deux trames successives sont séparées de M échantillons correspondant à une période de l'ordre de la centi-seconde. [20]

III-2-4- Fenêtrage :

La segmentation du signal en trames produit des discontinuités aux frontières des trames. Dans le domaine spectral, ces discontinuités se manifestent par des lobes secondaires. Ces effets sont réduits en multipliant les échantillons $\{S_a(n)\}_{n=0\ldots N-1}$de la trame par une fenêtre de pondération $\{w(n)\}_{n=0\ldots N-1}$ telle que la fenêtre Triangulaire, Rectangulaire, Blakman, Hanning et Hamming

$$\boldsymbol{S_w(n) = w(n).s_a(n)} \qquad (2.5)$$

Avec $\boldsymbol{w(n) = 0.54 - 0.46\ \cos(2\pi \frac{n}{N-1})}$ $\boldsymbol{0 \leq \alpha \leq N-1}$ (2.6)

Donc parmi les types de fenêtres citée ci-dessus, on choisit celle de Hamming car elle n'introduit pas de grande perturbation sur le signal (atténuation ou rapport du lobe principal sur le lobe secondaire = 41 dB, avec concentration de l'énergie dans le lobe centrale = 99,96%).Une fois la pondération réalisée, le calcul des coefficients discriminatifs est effectué, pour ce faire il existe des méthodes classiquement répertoriées en trois catégories : les méthodes temporelles, les méthodes fréquentielles et les méthodes cepstrales. [20]

III-3- Extraction des caractéristiques LPC et MFCC :

En acoustique, un son se définit classiquement par son intensité, sa hauteur qui est Fixée par la fréquence de vibration des cordes vocales, appelée fréquence du fondamental ou pitch (F_0). Deux sons de même intensité et de même hauteur se distinguent par le timbre qui est déterminé par les amplitudes relatives des harmoniques du fondamental. Ces amplitudes nommées formants se caractérisent par les maximums de la fonction de transfert (transmittance) du conduit vocal. En général, les trois premiers formants sont essentiels pour caractériser le spectre du signal vocal. [21]

Nous avons précédemment souligné que le signal de la parole présente des particularités telles que la redondance qui justifient tout à fait la recherche d'une représentation plus compacte du signal. Pour cela, il existe différentes techniques d'analyse du signal vocal, que nous regroupons dans ce qui suit en trois catégories : les méthodes temporelles les méthodes spectrales et les méthodes cepstrales. Il y'a d'autres classification de ces techniques en particulier en méthodes spectrales, modèles d'identification et modèles d'audition. [21]

III-3-1- L'échelle Mel :

C'est une modélisation de l'oreille humaine. A noter que le cerveau effectue en quelque sorte une reconnaissance vocale complexe avec filtrage des sons. Prenons l'exemple suivant où vous êtes en compagnie de nombreuses personnes, l'ensemble de ses personnes parle en même temps et vous discutez avec votre voisin. [17]

Malgré le bruit, vous arrivez à discerner clairement ce que vous dit votre voisin, vous ignorez de façon naturelle le bruit de fond et vous amplifier le son qui vous paraît le plus important. [17]

Le cerveau ne se contente non pas seulement de filtrer les sons et de les amplifier mais aussi de prédire. Prenons l'exemple suivant où une personne discute avec vous avec un volume sonore très bas, vous n'avez pas entendue une certaine partie de la phrase mais vous arrivez à la reconstituer et à la comprendre. A partir de l'étude du cerveau nous pouvons nous faire une idée de la complexité de la reconnaissance de la parole et nous pouvons nous rapprocher d'un modèle de plus en plus puissant et parfait.

On considère que l'oreille humaine perçoit linéairement le son jusqu'à 1000 Hz, mais après, elle perçoit moins d'une octave par doublement de fréquence. L'échelle Mel modélise assez fidèlement la perception de l'oreille :

Linéairement jusqu'à 1000 Hz, puis logarithmiquement au-dessus.

La formule donnant la fréquence *m* en Mels, *m* à partir de celle de *f* en Hz, est :

$$m = \frac{1000.\ln(1+\frac{f}{700})}{\ln(1+\frac{1000}{700})} \approx 1127.\ln(1+\frac{f}{700}) \approx 2595.log_{10}(1+\frac{f}{700}) \quad (2.7)$$

L'échelle Mel permet donc de modéliser une perception de l'oreille linéairement. [22]

III-3-2- Mel-scaled Frequency Cepstral Coefficients (MFCC):

Le signal de parole est complexe et redondant et possède une grande variabilité. Pour qu'un système de décodage acoustico-phonétique fonctionne efficacement, des informations caractéristiques et invariantes doivent être extraites du signal de parole. Cette procédure consiste à associer au signal de parole une série de vecteurs de paramètres généralement acoustiques. Le codage MFCC (Mel Frequency Cepstral Coefficients) est une extraction de caractéristiques du signal développée autour de la FFT et de la DCT, ceci sur une échelle fréquentielle non-linéaire (L'échelle Mel). Ce codage paraît être le plus utilisé en DAP vus sa simplicité, sa robustesse au bruit et surtout aux très bons résultats en taux de reconnaissances qu'il a permis d'obtenir par rapport aux autres codages. [17]

La figure ci-dessous, présente les étapes essentielles au calcul des coefficients MFCC

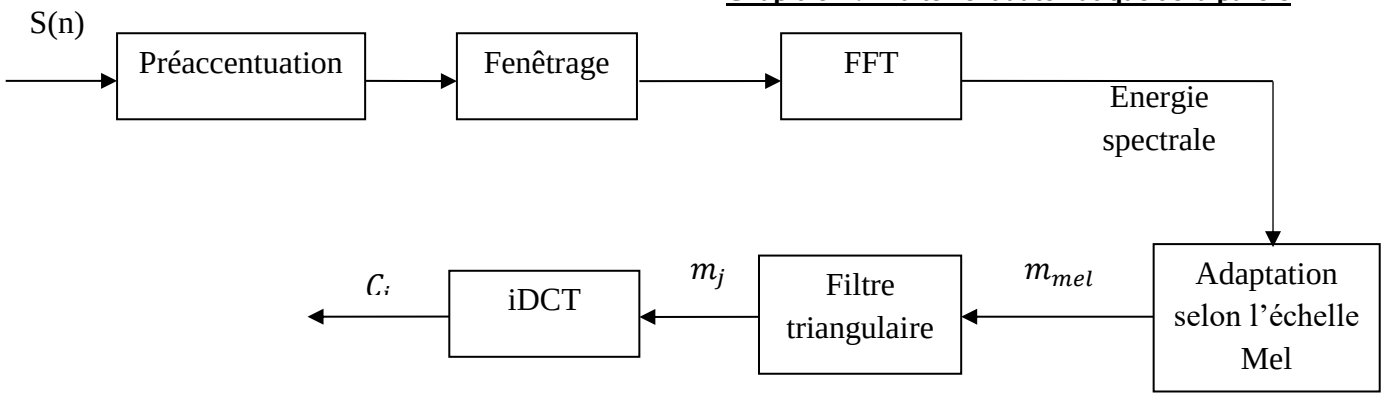

Figure 2.4 Etapes essentielles au calcul des coefficients MFCC.

III-3-2-1- Calcul de la transformée de Fourier (FFT) :

Au cours de cette étape chacune des trames, de N valeurs, est convertie du domaine temporel au domaine fréquentiel. La FFT est un algorithme rapide pour le calcul de la transformée de Fourier discret (DFT) et est définie par la formule.

$$\sum_{n=0}^{N-1} s[n] e^{\frac{-j2\pi}{N}kn} \quad , \quad 0 \leq k \leq N\text{-}1 \qquad (2.8)$$

Les valeurs obtenues sont appelées le spectre. [22]

Seul son module (énergie de la fréquence) est retenu, la phase de la transformée de Fourier numérique du signal de parole ne contient pas d'information suffisamment pertinente pour la reconnaissance de parole. [23]

III-3-2-2- Adaptation selon l'échelle Mel :

Les travaux de Stevens ont permis la mise en évidence de la loi de puissance ou loi de Stevens selon laquelle l'intensité de la perception d'un stimulus n'augmente pas linéairement en fonction de sa puissance mais de façon exponentielle en tenant aussi compte des modalités de l'expérimentation. Les coefficients MFCCs pour Mel-scaled Frequency Cepstral Coefficients, aussi nommés Mel Frequency Cepstral Coefficients dans la littérature, sont donc basés sur une échelle de perception appelée Mel, non linéaire. Celle-ci peut être définie par la relation suivante entre la fréquence en Hertz et sa correspondance en mels :

$$m_{mel} \approx 2595 \times \log_{10}(1 + \frac{f_{Hz}}{700}) \quad (2.9)$$

Pourtant l'utilisation de cette unité n'est pas suffisante. Pour avoir une largeur de bande relative qui reste constante, le banc de filtres Mel est construit à partir de filtres triangulaires positionnés uniformément sur l'échelle Mel donc non uniformément sur l'échelle fréquentielle. [24]

III-3-2-3- Band de filtre Mel :

Pour avoir une largeur de bande relative qui reste constante, le banc de filtres Mel est construit à partir de filtres triangulaires positionnés uniformément sur l'échelle Mel donc non uniformément sur l'échelle fréquentielle.

Cette répartition est illustrée ci-dessous :

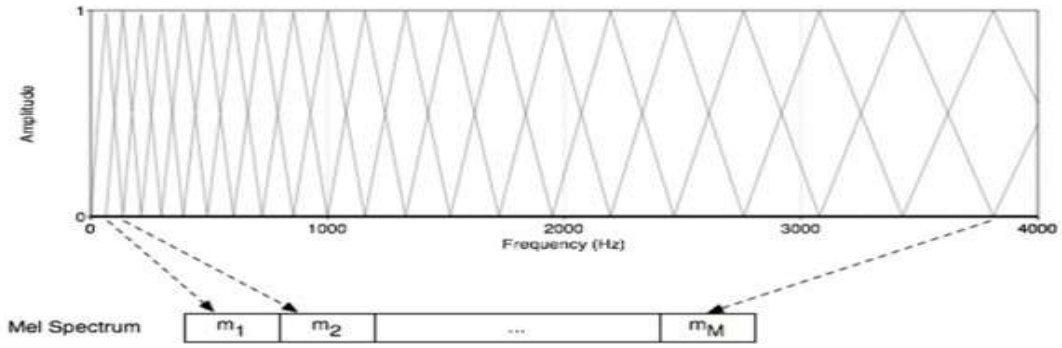

Figure 2.5 Band de filtre Mel.

Sur cette illustration, m_M correspond au nombre de filtres que l'on souhaite. Lorsque ce banc de filtres est en place, il est alors possible de calculer les coefficients MFCCs. [24]

III-3-2-4- Calcul de la transformé en cosinus discrète inverse (iDCT) :

Cette transformée nous fait revenir dans un pseudo-domaine temporel en appliquant la iDCT aux valeurs logarithmiques des m_j (d'où le nom de cepstre, quéfrence, …). La formule deviendra :

$$\boldsymbol{C_{i=}\sqrt{2/M}\sum_{j=1}^{M}\ln(m_j)\cos(\frac{\pi i}{M}(M+0.5))\ ,\quad i=1,2,\dots,M} \qquad (2.10)$$

En général on ne garde que les 13 premiers coefficients. Le premier étant proportionnel au long de l'énergie moyenne, est tout simplement remplacé par l'énergie de la trame. Question de représentativité. On peut voir ça comme une étape de compression. Les coefficients les plus élevés apportent un niveau de détail inutile, voir contre-productif pour la suite. [25]

III-3-3 Analyse par prédiction linéaire (LPC) :

L'étude du mécanisme de phonation montre que la production de chaque unité phonétique (phonème, syllabe, …) dépend de la position articulatoire des organes phonatoires (bouche, langue, lèvres, …). Le conduit vocal est donc considéré comme un filtre soumis à une excitation U(n). Pour les sons voisés ou sonores, cette excitation est un train périodique d'impulsions ; pour les sons non voisés, l'excitation est un bruit blanc. Ce modèle de production est appelé « Autorégressif » (AR).

Dans le domaine temporel, on aura la recursion linéaire suivante :

$$\boldsymbol{S(n) = \sum_{k=1}^{p} a_k S(n-k) + Gu(n)} \qquad (2.11)$$

Cette récurrence exprime le fait qu'un échantillon quelconque S(n) peut être déterminé par une combinaison linéaire des p échantillons qui le précèdent à laquelle il faut ajouter le terme d'excitation (le signal d'excitation u(n) et le gain G).

Les coefficients a_k sont appelés « coefficients de prédiction » ; en effet, si l'excitation était nulle, chaque échantillon S(n) pourrait être prédit exactement à partir des *p* échantillons qui le précédent immédiatement.

On peut donc considérer l'excitation comme étant une erreur de prédiction :

$$\boldsymbol{e(n) = S(n) - \sum_{k=1}^{p} a_k S(n-k)} \qquad (2.12)$$

Il faut alors minimiser l'erreur quadratique totale de prédiction définie par :

$$\boldsymbol{E(p) = \sum_{n=1}^{N} e^2(n)} \quad (2.13)$$

Il existe pour cela des méthodes mathématiques. En particulier, on calcule les dérivées partielles de E(p) par rapport aux coefficients de pondération $a_k$et l'on annule chacune d'entre elles. On obtient ainsi un système d'équation constitué d'un ensemble de *M* équations avec *M* inconnus : $a_1, a_2, \dots, a_M$. [26,19]

$$\begin{bmatrix} \boldsymbol{R(0)} & \boldsymbol{R(1)} & \boldsymbol{R(2)} & \dots & \boldsymbol{R(M-1)} \\ \boldsymbol{R(1)} & \boldsymbol{R(0)} & \boldsymbol{R(1)} & \dots & \boldsymbol{R(M-2)} \\ \boldsymbol{R(2)} & \boldsymbol{R(1)} & \boldsymbol{R(0)} & \cdots & \boldsymbol{R(M-3)} \\ \dots & \dots & \dots & \dots & \dots \\ \boldsymbol{R(M-1)} & \boldsymbol{R(M-2)} & \boldsymbol{R(M-3)} & \dots & \boldsymbol{R(0)} \end{bmatrix} \cdot \begin{bmatrix} \boldsymbol{a_1} \\ \boldsymbol{a_2} \\ \boldsymbol{a_3} \\ \dots \\ \boldsymbol{a_M} \end{bmatrix} = \begin{bmatrix} \boldsymbol{R(1)} \\ \boldsymbol{R(2)} \\ \boldsymbol{R(3)} \\ \dots \\ \boldsymbol{R(M)} \end{bmatrix}$$

Nous n'énumérerons pas tous les types de paramètres employés dans le domaine de la recherche en parole car il y en a énormément et ce n'est pas le propos de notre travail. Pourtant, il est à noter que d'autres approches plus proches de l'audition humaine, telles les modèles d'oreille, ont été étudiées. De plus, le lecteur trouvera des informations sur différents paramètres très largement utilisés pour le codage CELP (Code Exited Linear Predictive) présent dans la norme GSM, pour les PLPs (Perceptual Linear Predictive) et pour les RASTA-PLP, version approfondie des PLP. Cette liste ne se veut pas exhaustive mais permet d'avoir un aperçu des différents paramètres qu'il est possible d'extraire d'un signal de parole. [24]

IV- Les différentes approches de reconnaissance de la parole :

La reconnaissance automatique de la parole est le processus par lequel la machine tente de « décoder » le signal de la parole qui lui est destiné. Pour cela, on recense trois grandes approches, à savoir : l'approche acoustico-phonétique, l'approche reconnaissance de formes et l'approche intelligence artificielle. [19]

IV-1- L'approche acoustico-phonétique :

Historiquement, l'approche acoustico-phonétique (AP) est la première des approches apparues. Elle est basée sur la théorie acoustico-phonétique qui postule que dans un langage parlé, il existe un nombre fini d'unités phonétiques distinctes et que ces unités sont largement caractérisées par un ensemble de propriétés qui se manifestent au travers du signal. Il s'agit alors de définir une relation entre les caractéristiques spectrales du signal et les unités phonétiques du langage. [19]

IV-2- L'approche reconnaissance de formes :

Dans une approche de reconnaissances des formes (RF), la forme du signal (en entend par forme, les coefficients issus de l'analyse) est utilisée directement sans détermination explicite des caractéristiques dans un sens acoustique et phonétique. La plupart des approches de reconnaissance de formes, sont constituées de deux phases : d'abord l'apprentissage et ensuite, la reconnaissance de la forme présentée au système par un processus de comparaison. [19]

IV-3- L'approche intelligence artificielle :

L'approche intelligence artificielle (IA) se définit comme une approche hybride combinant les approches acoustico-phonétique et reconnaissance de formes

Cette approche, qui inclut l'approc²he RF, tente d'automatiser le processus de reconnaissance en s'inspirant de la manière dont l'être humain utilise son intelligence en visualisant, analysant et finalement en décidant sur la base des données acoustiques dont il dispose. L'idée de base étant d'incorporer différentes techniques et d'intégrer diverses sources de connaissances pour la prise en charge d'un problème donné.

Deux concepts clefs sont inhérents à l'intelligence artificielle ce sont : l'apprentissage et l'adaptation, et l'une des voies par lesquelles ces concepts peuvent être implémentés sont les réseaux connexionnistes. D'autre part, les systèmes experts (SE) constituent l'un des outils les plus représentatifs de l'approche IA. [19]

V-Conclusion :

Les codeurs les plus couramment utilisés sont : le codage linéaire prédictif (Linear Predictive Coding LPC) et le codage cepstral (Mel Frequency Cepstral Coding MFCC). Le codage MFCC intègre des connaissances du modèle « auditif » humain, quant au codage LPC, les connaissances recommandées sont celles du modèle « phonatoire » humain. La contrainte des codages cités auparavant, est l'incapacité à traiter la non-linéarité comprise dans les signaux de la parole.

Cependant, un autre codage, qui est une extension non-linéaire du codage LPC, permet d'intégrer les caractéristiques non-linéaires des signaux de la parole, nommé le codage neuronal prédictif (Neuronal Predictive Coding NPC). Ce codage permet de réduire le nombre des coefficients, habituellement élevé dans les structures des prédicteurs non-linéaires, en passant par la phase de paramétrisation afin d'optimiser l'étape de la classification dans la reconnaissance des phonèmes. Ce dernier fera le contenu de notre troisième chapitre.

I-Introduction :

L’intelligence Artificielle souvent abrégée avec le sigle IA, est définie par l’un de ses créateurs, Marvin Lee Minsky, comme : " la construction de programmes informatiques qui s'adonnent à des tâches qui sont pour l’instant, accomplies de façon plus satisfaisante par des êtres humains car elles demandent des processus mentaux de haut niveau tels que : l'apprentissage perceptuel, l'organisation de la mémoire et le raisonnement critiquée. ". C’est une nouvelle technologie en plein développement, cette technologie a vu le jour en 1950 avec Alan Turing qui est l’auteur du fameux Test de Turing qui a été la première conversation entre humain et Intelligence Artificielle. [27]

L’intelligence artificielle évolue en permanence pour essayer d’égaler l’intelligence de l’homme. Les réalisations actuelles de l’intelligence artificielle peuvent être regroupées en différents domaines, tels que Les systèmes experts, L’apprentissage automatique, La reconnaissance des formes, des visages et la vision en général ou encore Le traitement automatique des langues. [27]

Dans ce chapitre nous nous intéressons à un outil qui sera la base de notre travail. Cet outil est l’une des meilleures représentations du system neurologique humain car il en est inspiré, ce sont les réseaux de neurones artificiels.

II-Réseaux de neurones :

Aujourd'hui de nombreux termes sont utilisés dans la littérature pour désigner le domaine des réseaux de neurones artificiels, comme connexionnisme ou neuromimétique. Il semble qu'il faut associer à chacun de ces noms une sémantique précise. Connexionnisme et neuromimétique sont tous deux des domaines de recherche à part entière, qui manipulent chacun des modèles de réseaux de neurones artificiels, mais avec des objectifs différents. L'objectif poursuivi par les ingénieurs et chercheurs connexionnistes est d'améliorer les capacités de l'informatique en utilisant des modèles aux composants fortement connectés. Pour leur part, les neuromiméticiens manipulent des modèles de réseaux de neurones artificiels dans l'unique but de vérifier leurs théories biologiques du fonctionnement du système nerveux central. [28]

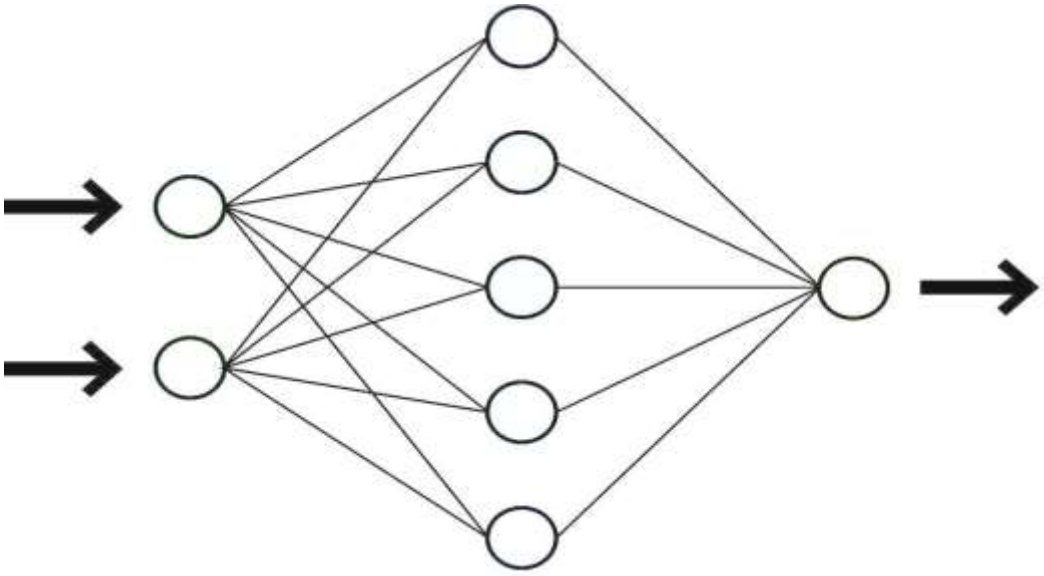

Figure 3.1 Schéma très simplifié d'un réseau neuronal.

Comme on peut le voir dans la (Figure-3-1) un réseau neuronal est un Ensemble de neurones formels interconnectés permettant la résolution de problèmes complexes tels que la reconnaissance des formes ou le traitement du langage naturel, grâce à l'ajustement des coefficients de pondération dans une phase d'apprentissage. Les deux neurones de gauche reçoivent les informations. Le traitement de ces données est déterminé par leurs connexions avec les neurones internes au milieu. Les neurones qui reçoivent une donnée sont activés. L'information finale est envoyée sur le dernier neurone de droite ou sur l'organe effecteur (un moteur par exemple). [29]

Un réseau neuronal s'inspire du fonctionnement des neurones biologiques et prend corps dans un ordinateur sous forme d'un algorithme. Le réseau neuronal peut se modifier lui-même en fonction des résultats de ses actions, ce qui permet l'apprentissage et la résolution de problèmes sans algorithme, donc sans programmation classique. [29]

II-1- Historique :

- 1890 : W. James, célèbre psychologue américain introduit le concept de mémoire associative, et propose ce qui deviendra une loi de fonctionnement pour l'apprentissage sur les réseaux de neurones connue plus tard sous le nom de loi de Hebb.

- 1943 : J. Mc Culloch et W. Pitts laissent leurs noms à une modélisation du neurone biologique (un neurone au comportement binaire). Ceux sont les premiers à montrer que des réseaux de neurones formels simples peuvent réaliser des fonctions logiques, arithmétiques et symboliques complexes (tout au moins au niveau théorique).

- 1949 : D. Hebb, physiologiste américain explique le conditionnement chez l'animal par les propriétés des neurones eux-mêmes. Ainsi, un conditionnement de type pavlovien tel que, nourrir tous les jours à la même heure un chien, entraîne chez cet animal la sécrétion de salive à 7 cette heure précise même en l'absence de nourriture. La loi de modification des propriétés des connexions entre neurones qu'il propose explique en partie ce type de résultats expérimentaux.

Les premiers succès :

- 1957 : F. Rosenblatt développe le modèle du Perceptron. Il construit le premier neuroordinateur basé sur ce modèle et l'applique au domaine de la reconnaissance de formes. Notons qu'à cet époque les moyens à sa disposition sont limités et c'est une prouesse technologique que de réussir à faire fonctionner correctement cette machine plus de quelques minutes.

- 1960 : B. Widrow, un automaticien, développe le modèle Adaline (Adaptative Linear Element). Dans sa structure, le modèle ressemble au Perceptron, cependant la loi d'apprentissage est différente. Celle-ci est à l'origine de l'algorithme de rétropropagation de gradient très utilisé aujourd'hui avec les Perceptrons multicouches. Les réseaux de type Adaline restent utilisés de nos jours pour certaines applications particulières. B. Widrow a créé dès cette époque une des premières firmes proposant neuro-ordinateurs et neuro-composants, la "Memistor Corporation". Il est aujourd'hui le président de l'International Neural Network Society (INNS) sur laquelle nous reviendrons au chapitre Informations pratiques.

- 1969 : M. Minsky et S. Papert publient un ouvrage qui met en exergue les limitations théoriques du perceptron. Limitations alors connues, notamment concernant l'impossibilité de traiter par ce modèle des problèmes non linéaires. Ils étendent implicitement ces limitations à tous modèles de réseaux de neurones artificiels. Leur objectif est atteint, il y a abandon financier des recherches dans le domaine (surtout aux U.S.A.), les chercheurs se tournent principalement vers l'IA et les systèmes à bases de règles.

L'ombre :

- 1967-1982 : Toutes les recherches ne sont, bien sûr, pas interrompues. Elles se poursuivent, mais déguisées, sous le couvert de divers domaines comme : le traitement adaptatif du signal, la reconnaissance de formes, la modélisation en neurobiologie, etc. De grands noms travaillent durant cette période telle : S. Grossberg, T. Kohonen …

Le renouveau

- 1982 : J. J. Hopfield est un physicien reconnu à qui l'on doit le renouveau d'intérêt pour les réseaux de neurones artificiels. A cela plusieurs raisons :

Au travers d'un article court, clair et bien écrit, il présente une théorie du fonctionnement et des possibilités des réseaux de neurones. Il faut remarquer la présentation anticonformiste de son article. Alors que les auteurs s'acharnent jusqu'alors à proposer une structure et une loi d'apprentissage, puis à étudier les propriétés émergentes ; J. J. Hopfield fixe préalablement le comportement à atteindre pour son modèle et construit à partir de là, la structure et la loi 8 d'apprentissage correspondant au résultat escompté. Ce modèle est aujourd'hui encore très utilisé pour des problèmes d'optimisation.

D'autre part, entre les mains de ce physicien distingué, la théorie des réseaux de neurones devient respectable. Elle n'est plus l'apanage d'un certain nombre de psychologues et neurobiologistes hors du coup.

Enfin, une petite phrase, placée en commentaire dans son article initial, met en avant l'isomorphisme de son modèle avec le modèle d'Ising (modèle des verres de spins). Cette idée va drainer un flot de physiciens vers les réseaux de neurones artificiels.

Notons qu'à cette date, l'IA est l'objet d'une certaine désillusion, elle n'a pas répondu à toutes les attentes et s'est même heurtée à de sérieuses limitations. Aussi, bien que les limitations du Perceptron mise en avant par M. Minsky ne soient pas levées par le modèle d'Hopfield, les recherches sont relancées.

La levée des limitations :

- 1983 : La Machine de Boltzmann est le premier modèle connu apte à traiter de manière satisfaisante les limitations recensées dans le cas du perceptron. Mais l'utilisation pratique s'avère difficile, la convergence de l'algorithme étant extrêmement longue (les temps de calcul sont considérables).

- 1985 : La rétropropagation de gradient apparaît. C'est un algorithme d'apprentissage adapté aux réseaux de neurones multicouches (aussi appelés Perceptrons multicouches). Sa découverte réalisée par trois groupes de chercheurs indépendants indique que "la chose était dans l'air". Dès cette découverte, nous avons la possibilité de réaliser une fonction non linéaire d'entrée/sortie sur un réseau en décomposant cette fonction en une suite d'étapes linéairement séparables. De nos jours, les réseaux multicouches et la rétropropagation de gradient reste le modèle le plus étudié et le plus productif au niveau des applications. [30]

II-2- Types d'architectures des réseaux de neurones :

Plusieurs architectures des réseaux existent. On peut en citer :

- **Réseaux monocouche :**

  Un réseau à une seule couche ne permet de séparer que des classes linéairement séparables. On appelle classes linéairement séparables des classes qui peuvent être correctement séparées par un hyperplan ou une fonction linéaire.

- **Réseaux multicouches :**

  Notons que le nombre de couches et le nombre de neurones par couche dépend principalement du problème étudié et des résultats de l'expérimentation sur différentes combinaisons des valeurs de ces paramètres. Les réseaux de neurones à une seule couche sont limités par le fait qu'il ne peut pas séparer des classes non linéairement séparables. [31]

  Pour résoudre un tel problème on utilise les réseaux de neurones multicouches, c.à.d. avec des couches cachées, avec des neurones de fonction d'activation non linéaire. [31]

  Dans la suite de ce chapitre nous avons présenté Le codage Neuro-Prédictif (NPC), mais avant ça nous allons présenter le réseau de neurones perceptron multicouches (MLP) car la performance d'extraction de caractéristiques de la méthode NPC repose sur ce réseau de neurones très performant et très utilisé. [31]

II-3- Réseau de neurones perceptron multicouche (MLP) :

Le perceptron multicouche est un réseau orienté de neurones artificiels organisé en couches et ou l'information voyage dans un seul sens, de la couche d'entrée vers la couche de sortie. La (Figure 3.2) donne l'exemple d'un réseau contenant une couche d'entrée, une couche cachée et une couche de sortie. La couche d'entrée représente toujours une couche virtuelle associée aux entrées du système. Elle ne contient aucun neurone. La couche suivante est une couche de neurones. Dans l'exemple illustré, il y a 4 entrées, 5 neurones sur la couche cachée et un sur la couche de sortie. La sortie du neurone de la dernière couche correspond toujours à la sortie du système. Dans le cas général, un perceptron multicouche peut posséder un nombre de couches quelconque et un nombre de neurones (ou d'entrées) par couche également quelconque. Les neurones sont reliés entre eux par des connexions pondérées. Ce sont les poids de ces connexions qui gouvernent le fonctionnement du réseau et "programment" une application de l'espace des entrées vers l'espace des sorties à l'aide d'une transformation non linéaire. La création d'un perceptron multicouche pour résoudre un problème donné passé donc par l'inférence de la meilleure application possible telle que définie par un ensemble de données d'apprentissage constituées de paires de vecteurs d'entrées et de sorties désirées. Cette inférence peut se faire, entre autre, par l'algorithme dit de rétro propagation. [33]

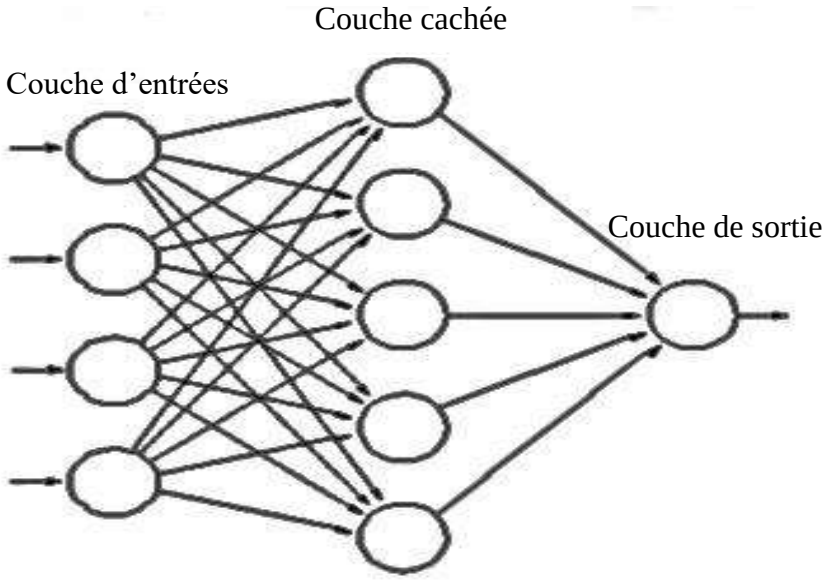

Figure 3.2 Exemple d'un réseau de neurones MLP.

II-3-1- le codage neuro-prédictif :

Le codage neuro-prédictif est une extension du codage **LPC,** donc une méthode de codage temporelle mais contrairement à la méthode **LPC**, le codeur **NPC** extrait les caractéristiques non linéaires d'un phonème. Il est basé sur un **MLP** à une couche cachée suivi d'une couche de sortie à 1 neurone appelé cellule de prédiction. L'étape d'apprentissage consiste à prédire un échantillon (extrait du signal acoustique d'un phonème) à partir des n échantillons précédents grâce à un **MLP**. Tous ces coefficients sont injectés dans un réseau de neurone MLP qui doit déterminer l'erreur quadratique afin de réajuster les poids de la couche d'entrée jusqu'à pouvoir avoir une erreur désirée acceptable, cet algorithme est basé sur la rétro propagation du gradient. [32, 33]

II-3-2- Description du modèle NPC :

Le modèle NPC (Neural Predictive Coding) est un perceptron à une couche cachée. Une des idées importantes dans ce modèle est que le vecteur code ou caractéristique est seulement formé par les poids de la couche de sortie. Cette utilisation des réseaux de neurones est différente des méthodes traditionnellement utilisées. Cependant, elle est à relier à la capacité de représentation interne des réseaux de neurones comme dans le cas de l'analyse non-linéaire des données.

Afin d'atteindre une spécialisation des couches, on considère que la fonction *F* réalisée par le modèle peut être décomposée en deux fonctions : $\boldsymbol{G_w}$ (**w** poids de la première couche) et H_a (**a** poids de la seconde couche) :

$$F_{w,a}(y_k) = G_w * H_a \qquad (3.1)$$

– Les poids de la première couche **w** sont communs à tous les phonèmes, et constituent la partie fixe du système.

– Les poids de la seconde couche **a** sont spécifiques à chaque phonème, et constitue notre vecteur de coefficients. [F]

Le processus d'apprentissage se divise en deux phases :

-La phase de paramétrisation : Cette phase consiste à déterminer tous les poids du réseau en utilisant les exemples des M classes de phonèmes. Les différentes secondes couches créées durant cette phase ne seront plus utilisées. Les poids de la première couche constituent les paramètres du codeur.

-La phase de codage : A la suite de la phase de paramétrisation, on présente un nouveau phonème destiné à être codé. Par minimisation de l'erreur de prédiction à l'aide des poids de la première couche, on détermine les poids de la seconde couche. Ces derniers forment le vecteur code du phonème. [33]

II-3-3- L'algorithme de rétro propagation du gradient :

La rétro propagation est le paradigme des réseaux de neurones artificiels le plus utilisé. Le terme se réfère à un algorithme pour ajuster les poids de connections en un multicouche, ce paradigme a été appliqué avec succès dans différents domaines tel que: le domaine militaire, médical, synthèse de la parole, traitement du signal … [22]

La rétropropagation est basée sur des principes mathématiques dont l'algorithme de descente du gradient et les règles de dérivation des fonctions dérivables (donc peut être appliqué à n'importe quelle structure de fonctions dérivables). [22]

Son objectif est de minimiser l'erreur quadratique pour un couple entrée-sortie, avec d_k la sortie désirée pour le neurone d'indice k et s_k la sortie obtenue par le réseau. L'erreur commise en sortie du réseau sera rétropropagée vers les couches cachées d'où le nom de rétropropagation, elle se traduit par :

$$\boldsymbol{E = \sum(d_k - s_k)^2} \qquad (3.2)$$

II-3-3-1- Les deux phases de l'apprentissage :

-a- Phase de propagation (sens direct) :

Permet le calcul de la sortie du réseau en fonction de son entrée.

Afin d'obtenir la mise à jour du neurone, chaque neurone effectue la somme pondérée de ses entrées et applique la fonction d'activation f, ceci donne :

$$\boldsymbol{s_i = f(p_i)} \qquad \textbf{Où} \qquad \boldsymbol{p_i = \sum_{j=0}^{n} w_{ij}\, x_j} \qquad (3.3)$$

Avec :

p_i : Le potentiel post-synaptique du neurone i.

x_j : L'état du neurone de la couche cachée précédente.

w_{ij} : Le poids de la connexion entre les deux neurones. [22]

-b- Phase de rétropropagation (sens inverse) :

L'algorithme de rétropropagation consiste à effectuer une descente de gradient sur l'erreur quadratique **« E »**.

Le gradient de **E** est calculé par tous les poids de la manière suivante :

$$\frac{\partial \mathrm{E}}{\partial \mathrm{w_{ij}}} = \frac{\partial \mathrm{E}}{\partial \mathrm{p_i}} \frac{\partial \mathrm{p_i}}{\partial \mathrm{w_{ij}}} = \frac{\partial \mathrm{E}}{\partial \mathrm{p_i}}\, x_j \quad (3.4)$$

On notera le gradient **:** $\boldsymbol{G_i = -\frac{\partial \mathrm{E}}{\partial \mathrm{p_i}}}$

On obtiendra deux cas selon la position du neurone ;

-Si le neurone est dans la couche de sortie :

$$\boldsymbol{G_i = -\frac{\partial \mathrm{E}}{\partial \mathrm{p_i}} = -\frac{\partial(\sum_k (\mathrm{d_k - s_k})^2)}{\partial \mathrm{p_i}} = 2(d_i - s_i) f'(p_i)} \quad (3.5)$$

-Si le neurone est dans la couche cachée :

$$\boldsymbol{G_i = -\frac{\partial \mathrm{E}}{\partial \mathrm{p_i}} = -\sum_{k=0}^{n} \frac{\partial \mathrm{E}}{\partial \mathrm{p_i}} \frac{\partial \mathrm{p_k}}{\partial \mathrm{p_i}} = \sum_{k=0}^{n} G_k \frac{\partial \mathrm{p_k}}{\partial \mathrm{p_i}} = \sum_{k=0}^{n} G_k \frac{\partial \mathrm{p_k}}{\partial \mathrm{s_i}} \frac{\partial \mathrm{s_i}}{\partial \mathrm{p_i}}} \qquad (3.6)$$

Donc :

$$\boldsymbol{G_i = f'(p_i) \sum_{k=0}^{n} w_{ki} G_k} \quad (3.7)$$

Les poids seront modifiés tel que :

$$\boldsymbol{w_{ij}^{t+1} = w_{ij}^{t} + \beta G_i S_j} \quad (3.8)$$

Avec :

β : Nombre positif qui représente le pas de déplacement en direction du minimum le plus proche. [22]

IV- Conclusion :

L'intelligence artificielle et spécifiquement les réseaux de neurones arrivent à accomplir d'importantes tâches. Inspirés du cerveau humain, les chercheurs ont réussi à reconstituer l'idée de l'apprentissage afin d'enrichir les performances du réseau et de se rapprocher plus de l'image de l'intelligence humaine.

Comme a dit, un célèbre écrivain québécois, Serge Bouchard : « attention au virus de l'intelligence artificielle. La représentation parfaite endort le cerveau ». Effectivement, l'utilisation de l'IA a trouvé refuge dans de nombreux domaines dont : le domaine militaire, la médecine, la météorologie... Et son exploitation prend un envol vertigineux.

Certes, l'intelligence artificielle est une belle invention, pratique et précise, mais est-elle sans aléas ? Sachant qu'elle repose sur le traitement et la comparaison d'une importante masse de données qu'un simple cerveau humain ne pourrait gérer, ni désigner d'éventuelles erreurs dans ce raisonnement artificiel. Le danger serait donc, que nous arrêtions d'évoluer laissant libre champs aux machines intelligentes.

I-But du projet :

Les systèmes de reconnaissance automatique de la parole se sont étendus à un large périphérique au cours des deux dernières décennies et cela revient à leur forte utilité. La phase de codage, qui permet l'extraction des caractéristiques, est la partie la plus importante dans la chaîne de la reconnaissance, donc exige un travail très minutieux. Cela voudrait dire que, les performances du système, dépendent directement de la qualité de l'extraction de ces caractéristiques. Pour cela, nous avons utilisé deux codages bien distincts : le codage neuronal prédictif (Neuronal Predictive Coding NPC) et le codage cepstral (Mel Frequency Cepstral Coding MFCC) ; deux codages très réputés. Une des branches de la reconnaissance automatique de la parole, est la modélisation neuro-prédictive pour la classification des unités phonétiques. Le travail sur des phonèmes de plusieurs langues, a déjà été mis en œuvre ; entre autre : la langue française, la langue anglaise, la langue arabe… Aujourd'hui notre contribution, ou plutôt le plus qu'on va apporter à ce projet est l'introduction de la langue amazighe en intégrant les lettres Tifinagh.

Finalement notre but, est de permettre à la population Amazighe d'être dans l'air du temps et de profiter de cette technologie comme toute autre langue, proprement dit. Pour ce fait, nous avons construit un logiciel qui permet d'enregistrer et de lire n'importe quel son de la langue Amazighe, puis d'être capable de reconnaître et surtout transcrire en Tifinagh le phonème en question. Après traitement du signal de parole, nous observerons les résultats obtenus par les deux types de codage cités auparavant. La comparaison entre ces deux codages permettra d'étudier leur comportement afin d'optimiser leur utilisation.

II-Tifinagh :

II-1- Présentation :

Le Tifinagh, pluriel de Tafineq qui signifie caractère d'écriture en touarègue, désigne toutes les gravures et les peintures aussi bien que les caractères alphabétiques. D'autres sources tendent à expliquer que le mot « Tifinagh » viendrait du verbe « Fnagh » qui veut dire : J'ai dessiné. Ou encore, une étymologie populaire déclare qu'il s'agirait d'un mot composé de « tifin » qui signifie : trouvaille ou découverte, et de l'adjectif possessif « nagh » qui signifie : notre ; le tout donne : notre découverte. Le Tifinagh est un alphabet consonantique utilisé par les Amazighs en Afrique du nord. L'origine de cette écriture plonge ses racines dans la Préhistoire, à la période Néolithique dite capsienne, comme l'attestent les nombreuses gravures rupestres où les signes de cette écriture, dite « libyque », de formes géométriques simples, se trouvent associés à des scènes de la vie quotidienne, de chasse ou de représentations des animaux de cette époque. Tombé en désuétude depuis l'Antiquité, il fut conservé uniquement dans la sphère linguistique touarègue (Sahara algérien, malien, libyen et nigérien) jusqu'au début du XX^{e} siècle avant d'être réintroduit par les militants kabyles de l'amazighe au Maroc. Cet alphabet a subi des modifications et des variations depuis son origine jusqu'à nos jours, passant du libyque antique jusqu'au néo-Tifinagh utilisé de nos jours. [1]

II-1-1- phonèmes de la langue Amazighe:

Lettres latines (berbère du nord)	Lettres en Tifinagh	prononciation
A	ⵄ	[æ]
B	ⴱ	[b]
C	ⵛ	[ʃ]
Č	�除	[tʃ]
D	ⴷ	[ḏ]
Ḍ	ⴹ	[ðˁ]
E	ⴻ	[e]
Ɛ	ⵄ	[ʕ]
F	ⴼ	[f]
G	ⴳ	[g]
Ǧ	ⵊ	[dʒ]
Γ, ɣ	ⵖ	[ʁ]
H	ⵀ	[h]
Ḥ	ⵃ	[ħ]
I	ⵉ	[i]
J	ⵊ	[ʒ]
K	ⴽ	[k]
L	ⵍ	[l]
M	ⵎ	[m]

N	ⵏ	[n]
Q	ⵇ	[q]
R	ⵔ	[ɾ]
Ṛ	ⵕ	[r̠]
S	ⵙ	[s̻]
Ṣ	ⵚ	[s]
T	ⵜ	[t̻], [θ]
Ṭ	ⵟ	[t̠]
U	ⵓ	[u]
W	ⵡ	[w]
X	ⵅ	[χ]
Y	ⵢ	[j]
Z	ⵣ	[s̺]
Ẓ	ⵥ	[z̺]
Bʷ	ⴱ‍ⵯ	[β]
Gʷ	ⴳ‍ⵯ	[ɟ]
Kʷ	ⴽ‍ⵯ	[ç]
TT		[ts]
ZZ		[dz]

-2- Phonèmes de la langue Amazigh.

La variabilité du signal de la parole fait sa complexité, c.-à-d. qu'un seul modèle est insuffisant pour faire un traitement d'un grand nombre de phonèmes. Pour cela les phonèmes

ont été classifiés en plusieurs groupes phonétiques : Voyelles, semi-voyelles, consonnes : occlusives, affriquées, fricatives, nasales, roulée, spirante (semi-voyelles), comme suit :

-**Les voyelles : [**a], [i] et [u] ([e] : ilem : n'est pas considéré comme une voyelle complète, elle sert uniquement à faciliter la lecture).
-**Spirantes :** [w], [j]…
-**Les consonnes :** le reste des lettres citées ci-dessus.
-**Les occlusives :**
  -Sourdes : [t̪], [k]…
  -Voisées : [d̪], [b]…
-**Les affriquées :**
  -Sourdes : [ts], [tʃ]…
  -Voisées : [dʒ], [dz]…
-**Les fricatives :**
  -Sourdes : [ħ], [θ]…
  -Voisées : [ðˤ], [ʒ]…
-**Les nasales :** [m], [n]…
-**Les roulées :** [ɾ̪], [ɾ]…

Dans la deuxième partie de ce chapitre on va présenter notre logiciel qui est capable de reconnaitre les phonèmes Amazigh grâce aux méthodes d'extraction des caractéristiques cités dans les chapitres précédents et pouvoir les transcrire en lettres tifinagh vues dans le tableau précédent.

III-Présentation de l'interface du logiciel :

Le logiciel est assez facile d'utilisation grâce à sa simple interface qui comporte un oscillogramme (1) qui permet une visualisation temporelle du signal de parole et un spectrogramme (2) permettant une visualisation fréquentielle de la parole ainsi que deux zones de textes (3) qui affichent une transcription phonétique du signal de parole une fois le mécanisme de reconnaissance enclenché, l'une en caractères amazigh et l'autre en caractères occidentaux et enfin un menu de fonctionnalités (4).

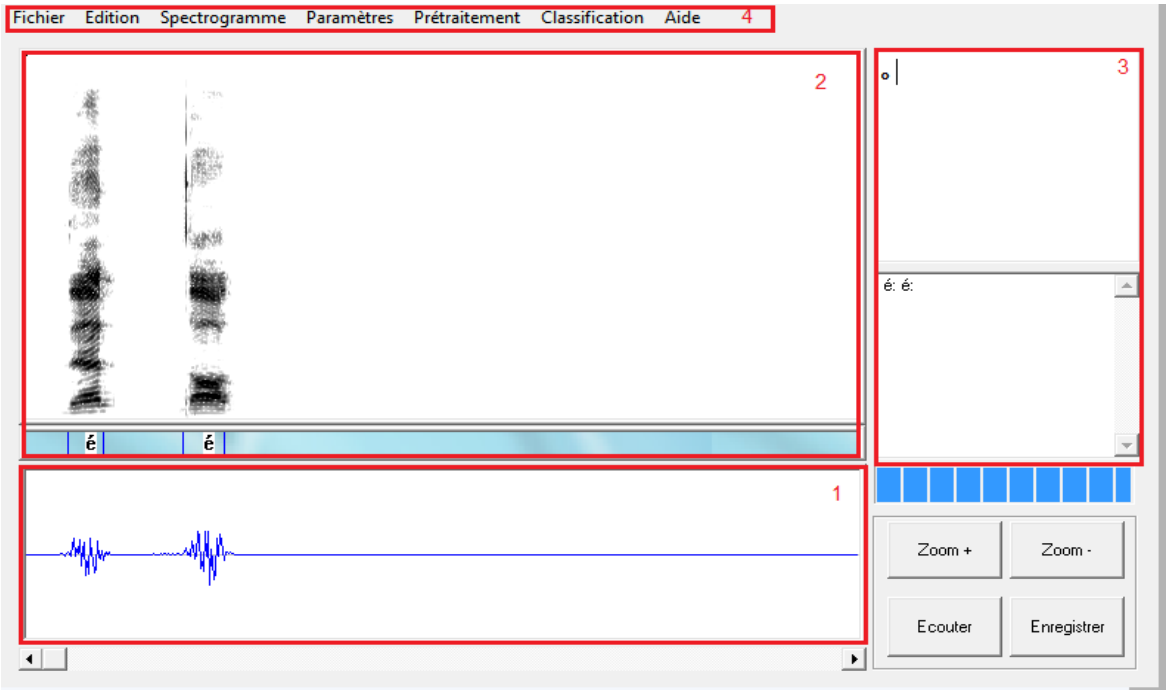

Figure 4.1 Corps du logiciel.

Plusieurs modules forment ce logiciel :

A-Module d'acquisition :

-Fonctionnalités :

-Acquisition ou écoute d'un signal sonore

- Choix de la fréquence d'échantillonnage lors de l'acquisition et la restitution. Elle est fixée par défaut à 16 KHz.

- Lecture d'un fichier de parole.

- Sauvegarde d'une partie ou l'ensemble du signal temporel.

- Affichage du signal temporel sur une fenêtre du menu.

- zoom sur le signal temporel.

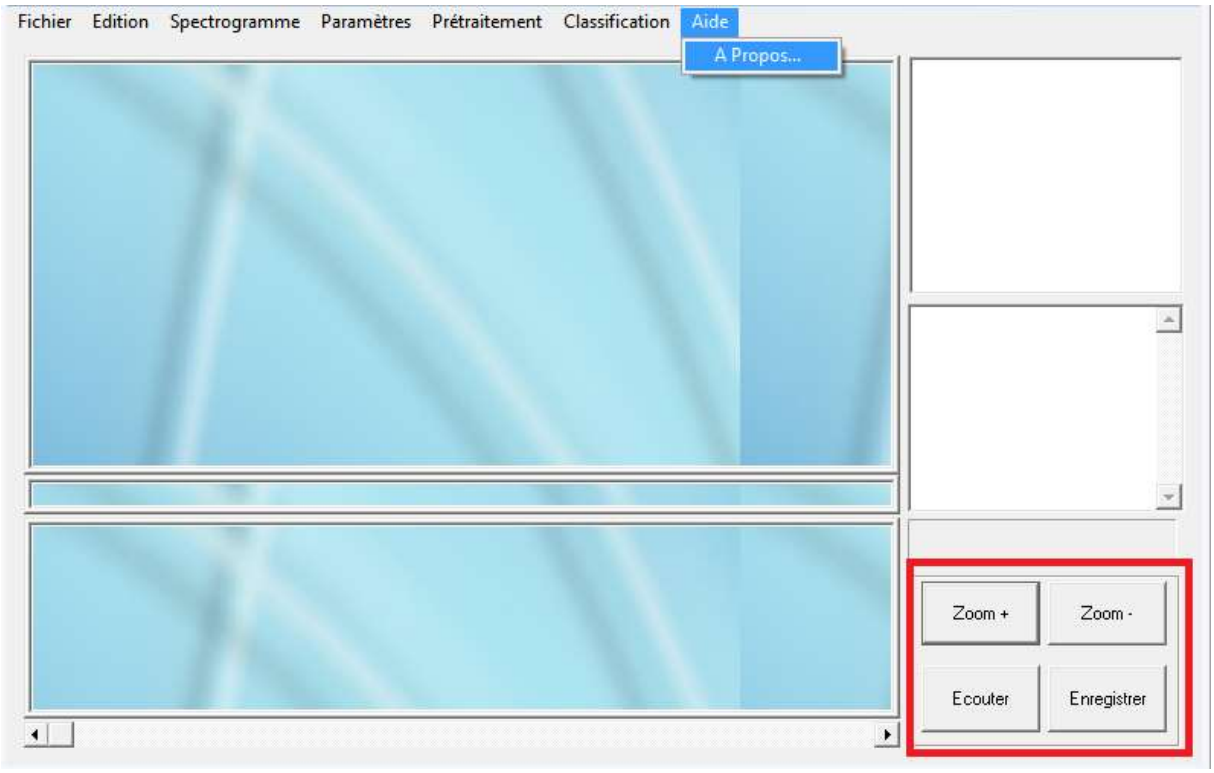

Figure 4.2 Module d'acquisition.

B-Module acoustique :

Ce module permet l'extraction de paramètres acoustique d'un signal temporel.

-Fonctionnalités :

-Calcul et affichage d'un spectrogramme à bande large ainsi qu'à bande étroite pour séparer les harmoniques de la fréquence fondamentale.

- Lissage cepstral de spectrogrammes.

- Calcul d'amplitude du signal.

- Calcule du nombre de passage par zéro du signal temporel pour distinguer entre parole et non parole et permet de différencier entre les sons voisés des sons non voisés.

- Calculer la fréquence fondamentale ou pitch.

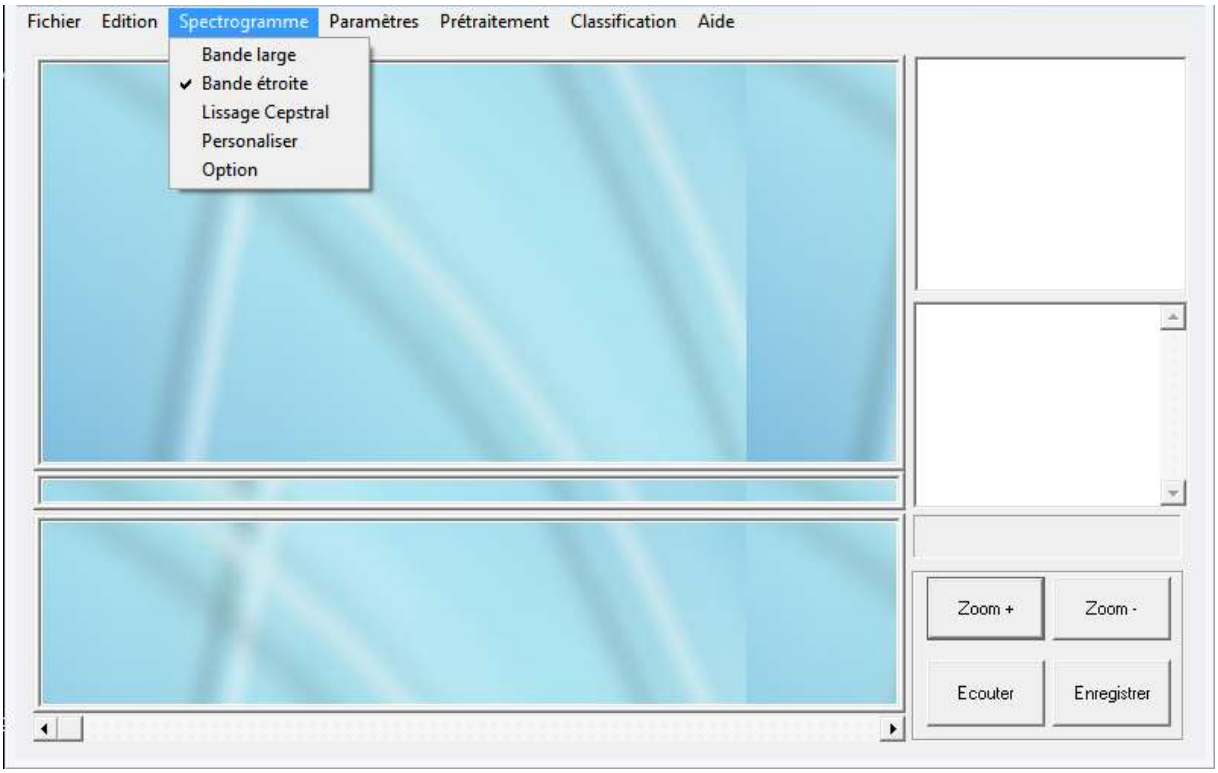

Figure 4.3 Module acoustique (spectrogramme).

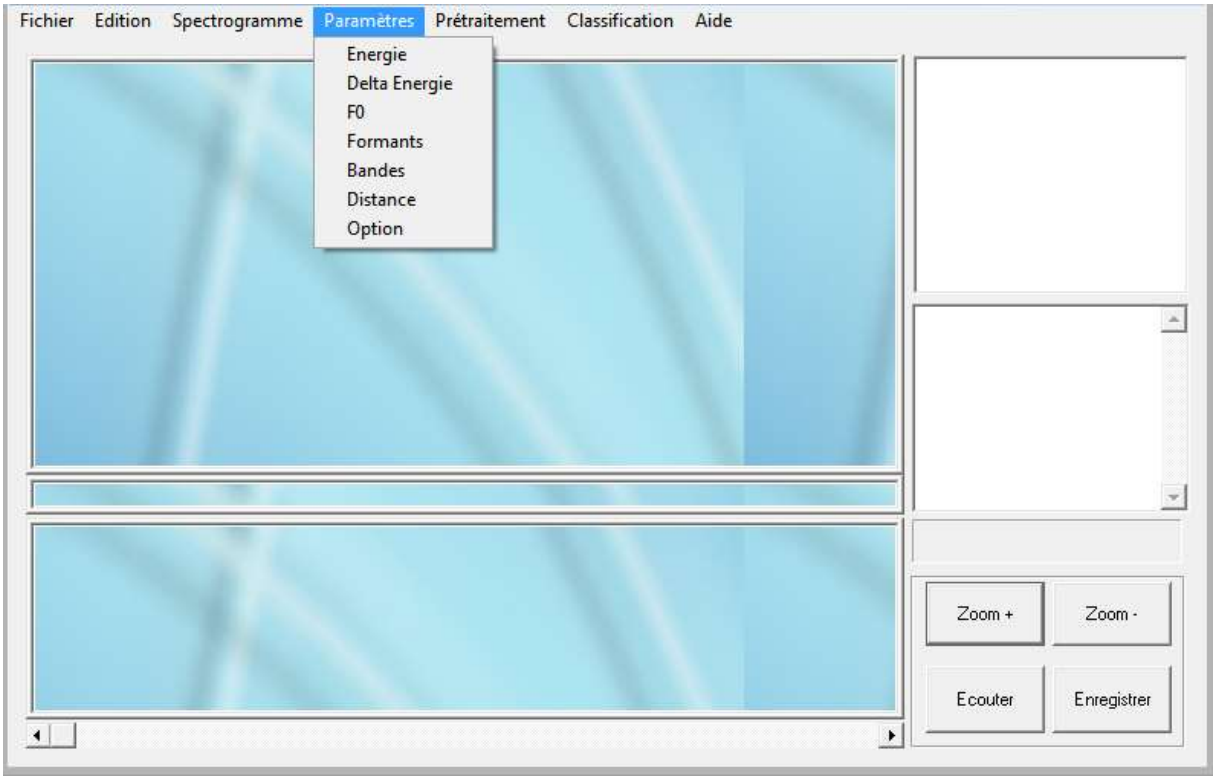

Figure 4.4 Module acoustique (Paramètres).

C-Module de décodage :

Constitué lui-même de deux modules :

-Module de segmentation : son rôle est la segmentation du signal de parole en phonèmes

-Module du calcul d'indice : son rôle est l'extraction des indices phonétiques des signaux de la parole.

D-Module d'étiquetage :

C'est au niveau de ce module que l'opération la plus importante est effectuée, le décodage, car il va tenter de trouver les phonèmes prononcés en utilisant les indices fournis dans les segments fournis par le module segmentation. Puis une étiquette manuelle permet d'attribuer des étiquettes phonétiques aux segments de la parole à partir de la représentation spectrographique du signal de la parole. L'opération d'étiquetage est faisable à partir d'une interface graphique qui permet d'insérer, effacer, changer l'étiquette d'un segment et déplacer la limite de ce dernier ainsi que de calculer et d'afficher un spectrogramme et l'écoute d'un morceau de signal.

Après l'étiquetage manuel il est possible de sauvegarder le résultat dans un fichier consultable et modifiable après et servira en particulier à évaluer les performances du système tant au niveau segmentation qu'au niveau reconnaissance.

E-Module de prétraitement :

Les fichiers WAV qui serviront comme base d'apprentissage ou de test doivent subir un prétraitement avant leur entrée dans le réseau MLP, Après chargement d'un fichier, les données sont sous forme d'un signal temporel. Ce signal va être fenêtré par une fenêtre glissante de 10 ms de type Hamming et qui avancera a pas de 5ms, après découpage du signal en fenêtres de 10ms le calcul des coefficients (NPC/MFCC) sous forme de vecteurs acoustiques sera pour chaque fenêtre.

Ensuite vient l'étape d'identification si la fenêtre existe dans la liste étiquetée alors une information sur l'identifiant du phonème sera ajoutée au vecteur acoustique sinon inconnu, le vecteur connu sera alors ajouté à la liste des vecteurs destinés à l'apprentissage ou au test du réseau MLP.

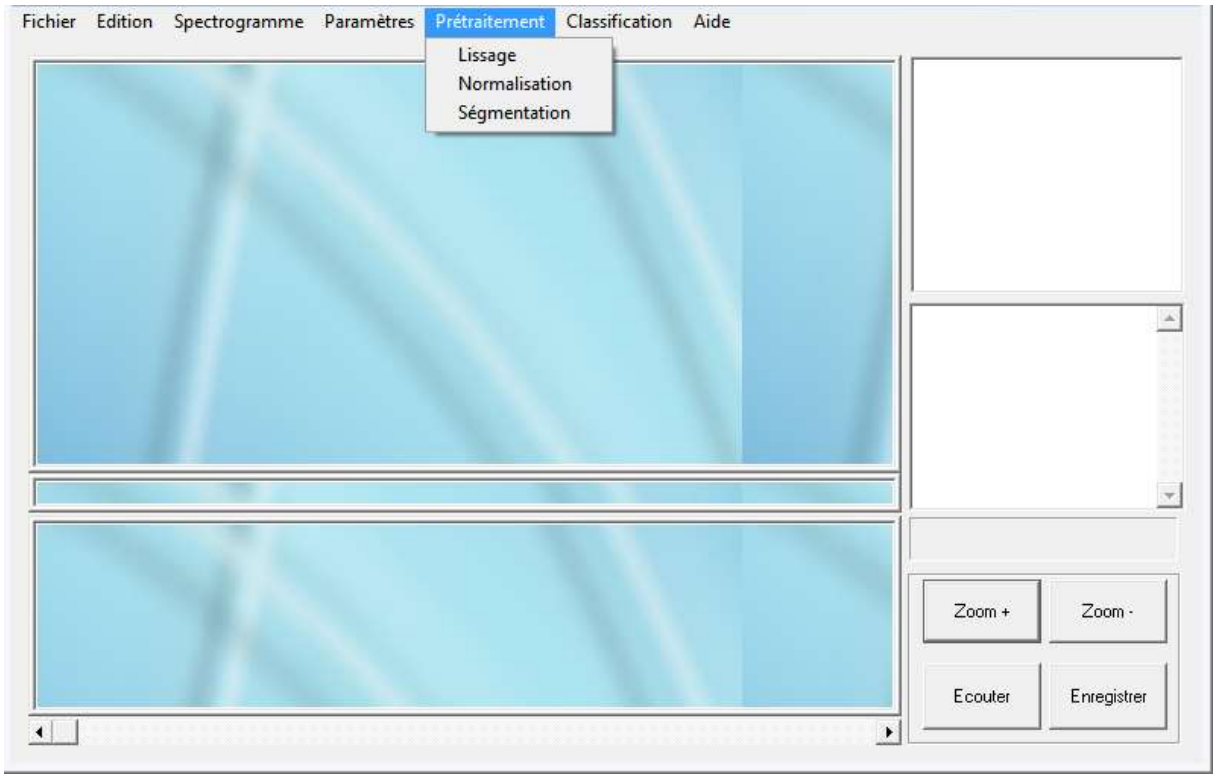

Figure 4.5 Module de prétraitement.

III-1- Test et résultats en reconnaissance phonétique :

Dans notre cas nous avons utilisé un MLP à 3 couches :

-Couche d'entrée.

-Couche cachée.

-Couche de sortie.

Nous avons appliqué notre MLP sur une base de données pour un test, donc nous avons extrait une suite de vecteurs de 16 coefficients MFCC et NPC de nos fichiers audio au format WAV puis lancé le processus d'apprentissage comme expliqué dans la partie précédente et cela pour quelques phonèmes.

Latin	B	S	Ɛ	N
Tifinagh	ⵀ	ⵙ	ⵅ	ⵏ
Taux de reconnaissance MFCC	10%	90%	0%	93%

-3- Taux de reconnaissance des phonèmes (MFCC).

Latin	B	S	Ɛ	N
Tifinagh	ⵀ	ⵙ	ⵅ	ⵏ
Taux de reconnaissance NPC	60%	90%	19%	58%

-4- Taux de reconnaissance des phonèmes (NPC).

-Commentaire :

On peut largement remarquer la supériorité du codage NPC en reconnaissance des plosives car il arrive à mieux reconnaitre les phonèmes rapides dans le temps, quant ’au codage MFCC il arrive à prendre en compte les consonnes de longue durée comme les nasales.

Ce résultat reste encore pour un petit nombre de phonèmes, la langue Amazigh comporte pleins de phonèmes bien difficiles à reconnaitre n’est au moins le codage NPC reste l’un des meilleurs outils à exploiter pour la reconnaissance de la parole.

-CONCLUSION GENERALE :

Dans notre travail le but est de réaliser un logiciel capable de reconnaitre la langue amazighe et le transcrire en lettre tifinagh, l'utilisation des réseaux MLP nous a permis d'évaluer ses performances sur quelques phonèmes seulement le NPC s'avère être une technique intéressante pour l'amélioration des systèmes de reconnaissance dans le cadre du traitement non-linéaire. Epandant il reste beaucoup à faire car ce mémoire n'est qu'une contribution pour le départ de l'intégration de la langue Amazigh dans cette nouvelle technologie.

La recherche dans le domaine de la reconnaissance automatique de la parole ne cesse de s'élargir pour étendre ses différentes applications au niveau industriel et aussi au grand public.

Pour réaliser ces applications, la compréhension des mécanismes de production et de la reconnaissance de la parole ne suffit pas, pour la recherche de grands moyens sont indispensables car pour obtenir un résultat optimal, sachant que la qualité de la reconnaissance dépend directement de la qualité des données vocales qui sont à la base les informations relatives à une voix propre ,plus ces informations seront importantes et connues par le système,meuilleure sera la reconnaissance ,c'est pour cela que les conditions d'enregistrement pour l'établissement des phonèmes doivent s'effectuaient dans des laboratoires.

-PERSPECTIVES :

De multiples perspectives peuvent être envisagé dans ce travail notamment par rapport à l'ajout des lettres tifinagh au grand complet, l'intégration des différents dialectes Amazigh et rendre sa disponibilité au grand public ainsi faire profiter la communauté amazighe de cette technologie, pour ça l'élargissement des corpus d'entrainement est très important afin de mieux entrainer l'apprentissage du réseau et donc mieux reconnaitre les phonèmes.

-Bibliographie :

[1] Laboratoire de phonètique,**«** La parole étudier en laboratoire »**,http://phonetique.uqam.ca/ .** Accédé en juin 2017

[2] Pierre Rousselot , « Autobiographie » **http://www.euskomedia.org/PDFAnlt/riev/16/16554555.pdf .** Accédé en juin 2017

[3] Christian Guilbault. **«** Introduction à la linguistique » **http://www.sfu.ca/fren270/Phonetique/phonetique.htm .** Accédé en juin 2017

[4] Hélène Knoerr**,«** Le mécanisme phonatoire », **http://aix1.uottawa.ca/~hknoerr/Ostiguy1118.pdf .** Accédé en juin 2017

[5] Reconnaissance automatique de la parole

http://outilsrecherche.overblog.com/pages/Notes_131_Lappareil_Phonatoire_Humain-3083095.html . Accédé en juin 2017

[6]http://outilsrecherche.overblog.com/pages/Notes_111_Le_Systeme_Auditif_Humain-3080878.html . Accédé en juin 2017

[7] « Micha VANONY », Le Manuel des Acousmates Juniors,**http://www.studiophebes.com/edu/acoustique/appareilauditifhumain.html .** Accédé en juin 2017

[8] Définitions graphèmes et phonèmes, **http://ww2.acpoitiers.fr/ia16pedagogie/IMG/pdf/graphemes_phonemes_def_liste_frequence.pdf .** Accédé en juin 2017

[9] « Le système nerveux »**http://www.cnrs.fr/cnrs-images/sciencesdelavieaulycee/org_animal/neuro.htm .** Accédé en juin 2017

[10] Graphèmes, phonèmes, définitions ,**http://www.boitaprof.fr/fs/CP_CE1/cpm58-LEXIQUE.pdf .** Accédé en juin 2017

[11] « comprendre le fonctionnement du système oral du français », https://www.projet-pfc.net/le-projet-pfc-ef/le-francais-explique/ . Accédé en juin 2017

[12] « Décodage du signal de la parole »**http://outilsrecherche.over-blog.com/pages/Notes_311_Decodage_du_Signal_de_la_Parole-3082466.html** .Accédé en juillet 2017

[13], Salim**,** « L'autobiographie **» http://www.limag.refer.org/Theses/Salim.pdf** . Accédé en juin 2017

[14] STEVE CHERPILLOD, « ESPACE SONORE » "=**http://archivesma.epfl.ch/2011/016/cherp_enonce/steve_cherpillod_enonce.pdf/** .Accédé aout 2017

[15] Mlle.BELGHITRI KARIMA, «Système sécurisé à base vocale » **http://dspace.univ-tlemcen.dz/bitstream/112/8215/1/Systeme-securise-a-base-vocale.pdf** .Accédé aout 2017

[16] DOUIB OUALID, «RECONNAISSANCE AUTOMATIQUE DE LA PAROLE ARABE PAR CMU SPHINX 4 » http://www.univ-tebessa.dz/fichiers/master/master_1116.pdf **. Accédé en juin 2017**

[17] « Cours de Traitement Automatique de la Parole » **http://www.univ-usto.dz/faculte/fac-mathinfo/Cours/Traitement_Parole2014.pdf** . Accédé en juin 2017

[18] « Numérisation acoustique »,**http://culturesciencesphysique.ens-lyon.fr/ressource/numerisation-acoustique Chareyron2.xml** . Accédé en juin 2017

[19] Lotfi AMIAR, « UN SYSTEME HYBRIDE AG/PMC POUR LA RECONNAISSANCE DE LA PAROLE ARABE » **biblio.univ-annaba.dz/wp-content/uploads/2014/05/Memoire-Amiar.pdf** . Accédé en juin 2017

[20] Hacine Gharbiabdenour**, «** Sélection de paramètres acoustiques pertinents pour la reconnaissance de la parole », http://www.univsetif.dz/Tdoctorat/facultes/facultes1/TEC/2013/HacineGharbiabdenour.pdf . Accédé en juin 2017

[21] Richard Clavel, « Reconnaissance acoustique des émotions » **http://perso.telecom-paristech.fr/~grichard/Publications/livre_ClavelRichard.pdf** . Accédé en juin 2017

[22] Mustapha Kamel Abderrahmane DIDICHE ,« Modélisation neuro-prédictive pour La classification phonétique » **http://thesis.univ-biskra.dz/1299/1/g%C3%A9nie_elect_d6_2014.pdf** . Accédé en aout 2017
[23] Bruno JACOB, « Reconnaissance automatique de la parole » http://www-lium.univ-lemans.fr/~jacob/Recherche/These/bruno.html . Accédé en aout 2017
[24] « Reconnaissance automatique de la parole » http://www-prima.inrialpes.fr/Vaufreydaz/These/Reconnaissance.html . Accédé en aout 2017
[25] « La reconnaissance automatique du locuteur »,**https://blog.groupe-sii.com/ral/** . Accédé en aout 2017

[26] Jean François Frigon and Vladislav Teplitsky **«** Implementation of Linear Predictive Coding (LPC) of Speech » http://www.seas.ucla.edu/~ingrid/ee213a/speech/vlad_present.pdf . Accédé en aout 2017

[27] « L'intelligence artificielle », **http://tpe-intelligence-artificielle-2013.e-monsite.com/pages/definition-de-l-intelligence-artificielle.html**. Accédé en aout 2017

[28] Claude, « les réseaux de neurones artificiels »**http://www.touzet.org/Claude/Web-Fac-Claude/Les_reseaux_de_neurones_artificiels.pdf** . Accédé en aout 2017

[29] « Réseau neuronal »,**http://www.futura-sciences.com/tech/definitions/informatique-reseau-neuronal-601/** . Accédé en aout 2017

[30]Amrani mohamed, « surveillance et diagnostic d'une ligne de production par les réseaux de neurones artificiels »**http://dlibrary.univ-boumerdes.dz:8080/bitstream/123456789/2008/1/Amrani%20Mohamed.pdf** . Accédé en aout 2017

[31] Al Falou Wassim, « Reconnaissance de caractères manuscrits par réseau de neurones » **http://www.lb.refer.org/memoires/215282dea_falou.pdf** . Accédé en aout 2017

[32] Mohamed CHETOUANI, « Codage «neuro-pr´edictif pour l'extraction de caractéristiques de signaux de parole »**http://perso.telecom-paristech.fr/~chollet/Biblio/Theses/TheseChetouani.pdf** . Accédé en aout 2017

[33] Marc Parizeau , « Le perceptron multicouche et son algorithme de retro propagation des erreurs » **https://reussirlem1info.files.wordpress.com/2012/05/mlp.pdf** . Accédé en aout 2017

-Abstract:

The first chapter is an introduction in phonetic and presentation of the systems governing human language with the functioning of the concerned devices, involving physiological and neurological systems and their anatomy. Physiological system consists of vocal apparatus which is the mainspring of the production of sounds of the different phonemes, due to three organs: lungs, larynx and oral pharyngeal cavities; and the auditory system that has, for main organ: ear, the center of acoustic processing. The neurological system is about brain that is considered as connections between neurons, assuring the understanding of meaning of the treated phonemes.

The second chapter introduces the automatic speech processing. It involves the different characteristics of the speech signal explaining digitization steps and the methods of coding: LPC (Linear Predictive Coding) and MFCC (Mel Frequency Cepstral Coefficients).

The third chapter is about artificial intelligence. It contains introduction about neuronal network and their evolution during the last century. We chose one type of neuronal network that is MLP: a Multi-Layer Perceptron in order to use another model of extraction of characteristics which is NPC (Neuronal Predictive Coding), a non-linear extension of LPC.

The forth chapter is dedicated to a presentation of Amzaighe language and Tifinagh letters. Then we have injected a speech signal in a neuronal network MLP so that we could compare the results obtained by both coding: MFCC and NPC.

Keywords: LPC, MFCC, NPC MLP, automatic speech processing, Amazigh language, artificial intelligence.

-Résumé :

Le Chapitre 1 est une introduction à la phonétique et présente les systèmes régissant le langage chez l'être humain et le fonctionnement des appareils concernés, introduisant ainsi le système physiologique et le système neurologique et leur anatomie. Le système physiologique se constitue de l'appareil phonatoire qui est le moteur de la production du son des différents phonèmes grâce à l'interaction des trois grands organes (les poumons, le larynx et les cavités bucco-pharyngale) et l'appareil auditif qui a comme organe principal l'oreille est le centre du traitement acoustique et cognitif. Le système neurologique est la partie nerveuse dite le cerveau qui est constitué de neurones assurant ainsi le traitement des différents sons des phonèmes et leur compréhension.

Le chapitre 2 présente le traitement automatique de la parole. Ce dernier comportera les différentes caractéristiques du signal de parole, évoquant les étapes de la numérisation et détaillant les méthodes traditionnellement mises en œuvre pour cette analyse. Ce chapitre sera l'occasion de présenter en profondeur les différentes méthodes du codage LPC et MFCC.

Le chapitre 3 comportera une introduction globale sur l'intelligence artificielle, puis précisément sur les réseaux de neurones, leur évolution durant le siècle dernier citant les différents types des réseaux de neurones. On se focalisera sur un perceptron multicouche MLP afin d'utiliser un nouveau modèle pour l'extraction de caractéristiques le Codage Neuro-Prédictif (NPC, Neural Predictive Coding) qui est une extension au domaine non-linéaire du codage LPC.

Le chapitre 4 sera consacré à une présentation de la langue Amazighe et précisément les lettres Tifinagh puis à l'étude de la mise en forme d'un signal de parole qui sera injecté dans un réseau de neurones MLP (Multi Layer Perceptron), puis la comparaison entre les résultats obtenus par l'utilisation des deux codages : MFCC (Mel Frequency Cepstral Coding) et NPC (Neuronal Predictive Coding).

Mots clés : LPC, MFCC, NPC, MLP, traitement automatique de la parole, langue Amazigh, intelligence artificielle.