

**REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE
MINISTERE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA RECHERCHE
SCIENTIFIQUE
UNIVERSITE MOULOUD MAMMERI DE TIZI-OUZOU**



**FACULTE DE GENIE ELECTRIQUE ET D'INFORMATIQUE
DEPARTEMENT D'ELECTRONIQUE**

**Mémoire de fin d'études
Présenté en vue d'obtention
du Diplôme d'Ingénieur d'Etat en Electronique**

Option : communication et instrumentation

Thème :

**Traitement de la parole :
Application des réseaux
de neurones à la
reconnaissance du locuteur**

Proposé et dirigé par :

M^r : S .HADDAB

Etudié par :

M^{elle} : ZIANI NORA

M^{elle} : TALEB LYNDA

Année universitaire 2009/2010

sommaire

Introduction générale	1
------------------------------------	---

Chapitre I Signal de parole

I-1Système de production de la parole	3
I-1-1-L'appareil phonatoire	3
I-1-2-La voix humaine	4
I-2-définitions des classes phonétiques	5
I-2-1-Les voyelles	5
I-2-2-Les consonnes	6
I-2-2-1-Les occlusives	6
I-2-2-2-Les fricatives	6
I-2-2-3-Les sonnantes	6
I-3- L'alphabet phonétique international	7
I-4-principe d'acquisition du signal de parole	8
I-4-1-Transducteurs et capteurs (microphone)	8
I-4-2-L'amplificateur	9
I-4-3-Le filtre anti-repliement (filtre de garde)	9
I-4-4-Numérisation du signal	9
I-5-Caractéristiques du signal de parole	11
I-5-1-caractéristiques physique	11
I-5-1-1-La redondance	11
I-5-1-2-La coarticulation	11
I-5-1-3-La variabilité du signal parole	12
I-5-1-4-Les interférences	12
I-5-2-Caractéristiques électriques	12
I-5-2-1-Fréquence fondamentale	13
I-5-2-2-La bande spectrale	13
I-6-Conclusion	14

Chapitre II L'analyse et la reconnaissance de signal vocal

II-1-les différentes méthodes d'analyse de la parole	15
II-1-1-La méthode directe	15
II-1-1-1-La glottographie photoélectrique	15
II-1-1-2-La glottographie électrique	15
II-1-1-3-La glottographie ultrasonore	15
II-1-2-La méthode indirecte	16
II-1-2-1-L'analyse temporelle	16
II-1-2-1-1-Le taux de passage par zéro (T P Z)	16
II-1-2-1-2-Analyse par l'énergie	16
II-1-2-2-Analyse fréquentielle	17
II-1-2-2-1- analyse fréquentiel par la transformée de Fourier à court terme(TFCT).....	17
II-2-La reconnaissance vocale	23

II-2-1- Modes de fonctionnement des systèmes de reconnaissance.....	23
II-2-1-1-Dépendant du locuteur (mono locuteur)	23
II-2-1-2-Pluri locuteur (multi locuteur).....	24
II-2-1-3-Indépendant du locuteur	24
II-2-2-Le mode d'élocution.....	24
II-3-la reconnaissance du locuteur	25
II-3-1- Identification et vérification du locuteur	25
II-3-1-1-La vérification du locuteur (ou l'authentification).....	25
II-3-1-2- l'identification du locuteur	26
II-4-Système de reconnaissance de locuteur	26
II-4-1-Paramétrisation	27
II-4-2-Modélisation des locuteurs	27
II-4-2-1-L'approche vectorielle	28
II-4-2-2-L'approche statistique	28
II-4-2-3-L'approche connexionniste	28
II-4-2-4-L'approche relative	28
II-4-3-Décision et mesure des performances	28
II-5-Evaluation des performances en reconnaissance du locuteur	29
II-6-Les limites de la reconnaissance du locuteur (difficultés)	29
II-7-Conclusion	30

Chapitre III Les réseaux de neurone

III-1-Le neurone biologique	31
III-1-2-Fonctionnement d'un neurone	31
III-1-3-Anatomie des neurones	31
III-2-Le neurone formel (artificiel)	32
III-2-1-Modélisation d'un neurone formel	33
III-3-Réseaux de neurones	33
III-3-1-La couche d'entrée	34
III-3-2-La couche cachée	34
III-3-3-La couche de sortie	34
III-4-La fonction d'activation	35
III-5-L'apprentissage	36
III-6-Le perceptron	37
III-7-Les étapes de fonctionnement d'un réseau de neurone	37
III-7-1-Collecte des données	37
III-7-2-Analyse des données	37
III-7-3-Séparation des données	37
III-7-4-Choix du réseau de neurones	38
III-8-Algorithmes de la retro propagation	38
III-9-Les applications des réseaux de neurones	39
III-9-1- L'application à la reconnaissance des formes	39
III-9-2-L'application à la planification	39
III-9-3-L'application à l'identification.....	39
III-9-4-L'application à la commande	40
III-9-5-L'application au diagnostic	40
III -10-Conclusion	40

Chapitre IV

Application à la reconnaissance du locuteur

IV-1-Description et acquisition de la base de données	41
IV-2-Procédure d'enregistrement	42
IV-3-Extraction des paramètres caractéristiques des locuteurs	43
IV-4-Choix de l'architecture du réseau de neurones	45
IV-4-1-Choix des variables d'entrée	45
IV-4-2-Choix de la couche intermédiaire	46
IV-4-3-Choix des sorties	46
IV-5-La base d'apprentissage	46
IV-6-Normalisation des données	47
IV-7-Apprentissage du réseau	47
IV-8-Evaluation des performances du réseau	49
IV-8-1-Diagramme de l'erreur	49
IV-8-2-L'influence du nombre de neurone de la couche cachée	49
IV-8-3-Influence du coefficient d'apprentissage	51
IV-8-4-Influence de nombre d'itération	53
IV-8-5-Influence l'erreur désirée	55
IV-9-Résultats de l'application du réseau de neurones	57
IV-10-Conclusion	59
Conclusion générale	60

Remerciements

Nous tenons à remercier :

✚ Le bon Dieu qui nous a donné la bonne santé et la persévérance durant notre cursus.

✚ Nos parents et tous ceux qui nous ont aidé pour mener à terme ce travail.

✚ Notre promoteur Mr Haddab. S.

✚ Les membres de jury.

✚ Toutes les personnes qui ont participé aux enregistrements

✚ Et tous les lecteurs

Dédicaces

Je dédié ce modeste travail à :

✚ A ma très chère mère source de tendresse

✚ A mon très cher père, qui m'encourage

Dans les instants délicats

✚ Mes frères : Tahar, Nassim

✚ Mes sœurs : Karima, Salima, Mounia et ma petite
adorable Imen

✚ A ma très chère amie Nora

✚ Tous mes amis qui ont partagé mes bons souvenirs

✚ Tous ceux qui me sont très cher(e)s

Lynda

Dédicaces

Je dédié ce modeste travail à :

✚ A ma très chère mère source de tendresse

✚ A mon très cher père, qui m'encourage

Dans les instants délicats

✚ Mes frères : Ali, Samir

✚ Mes sœurs : Sabrina et ma petite

adorable Cilia

✚ Tous mes oncles, tontes, et cousins, cousines

✚ A ma très chère amie Lynda

✚ Tous mes amis qui ont partagé mes bons souvenirs

✚ Tous ceux qui me sont très cher(e)s

Nora

Chapitre I

Signal de parole

Le processus de production de la parole est complexe et entraîne l'activation de nombreux muscles et articulateurs dont les commandes sont parfaitement synchronisées et contrôlées par le système nerveux. Dans ce chapitre, nous proposons d'étudier ce système afin de mieux comprendre le processus de production de la parole. Nous allons présenter d'abord les différents organes biologiques constituant le système vocal puis nous décrirons brièvement le système de phonation et les classes phonétiques qui en découlent et nous terminerons par l'explication du principe d'acquisition du signal de parole et ses diverses caractéristiques.

I-1- Système de production de la parole :

I-1-1- L'appareil phonatoire :

L'appareil phonatoire nous permet de produire des sons très variés dans un espace fréquentiel et énergétique donné (**figure I-1**). L'appareil phonatoire humain a été la base de recherches visant à simuler mécaniquement ses capacités, ces recherches ayant permis, en retour, de mieux comprendre son fonctionnement

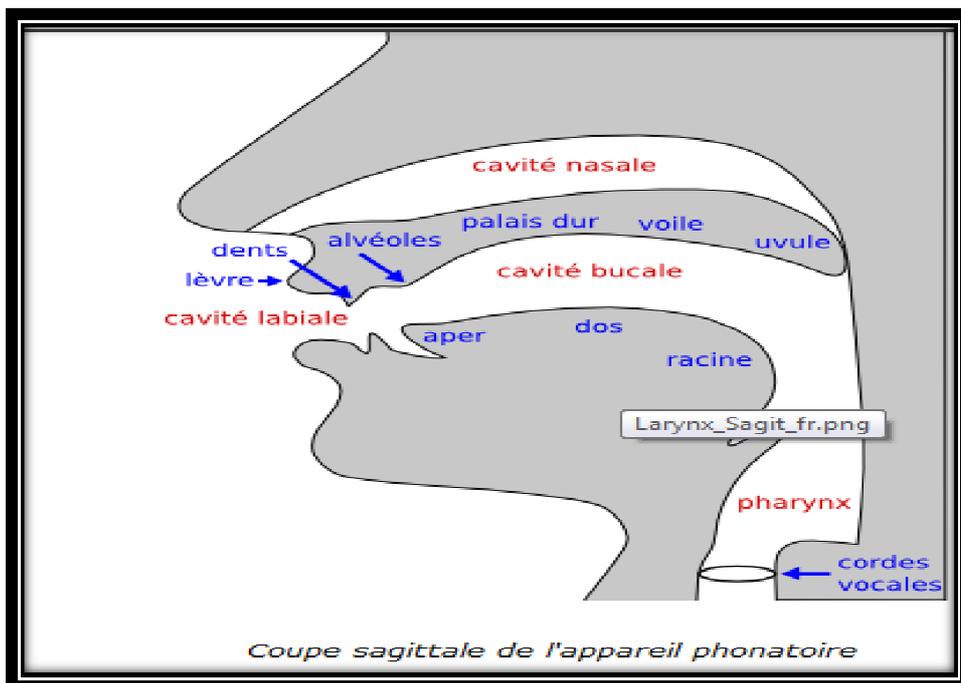


Figure I-1- L'appareil phonatoire.

I-1-2- La voix humaine :

La parole peut être décrite comme le résultat de l'action volontaire et coordonnée d'un certain nombre de muscles. Cette action se déroule sous le contrôle du système nerveux central qui reçoit en permanence des informations par rétroaction auditive et par les sensations kinesthésiques.

L'appareil respiratoire fournit l'énergie nécessaire à la production de sons, en poussant de l'air à travers la trachée-artère. Au sommet de celle-ci se trouve le larynx où la pression de l'air est modulée avant d'être appliquée au conduit vocal. Le larynx est un ensemble de muscles et de cartilages mobiles qui entourent une cavité située à la partie supérieure de la trachée (**Figure I-2**).

Les cordes vocales sont, en fait, deux lèvres symétriques placées en travers du larynx. Ces lèvres peuvent fermer complètement le larynx et, en s'écartant progressivement, déterminer une ouverture triangulaire appelée glotte. L'air y passe librement pendant la respiration et la voix chuchotée.

Le larynx est d'abord complètement fermé, ce qui accroît la pression en amont des cordes vocales et les force à s'ouvrir, ce qui fait tomber la pression, et permet aux cordes vocales de se fermer. Des impulsions périodiques de pression sont ainsi appliquées au conduit vocal, composé des cavités pharyngienne et buccale. Lorsque la luette est en position basse, la cavité nasale vient s'y ajouter en dérivation. Notons, pour terminer, le rôle prépondérant de la langue dans le processus phonatoire.

Sa hauteur détermine la hauteur du pharynx : Plus la langue est basse, plus le pharynx est court.

Elle détermine aussi le lieu d'articulation, région de rétrécissement maximal du canal buccal, ainsi que l'aperture, écartement des organes au point d'articulation.

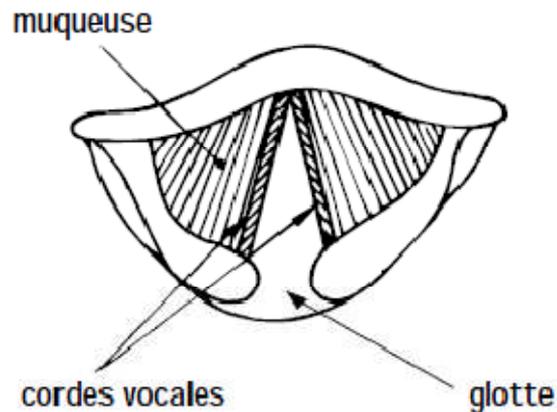


Figure I-2- Section du larynx, vu de haut.

I-2-Définitions des classes phonétiques :

La phonétique peut être définie comme l'étude de l'articulation du phonème (un phonème est la plus petite unité présente dans la parole). Elle peut être considérée descriptive, associant chaque son de la langue à un symbole et à une classe.

Il est intéressant de grouper les sons de parole en classes phonétiques, en fonction de leur mode articulaire (caractérise de la forme et de la façon dont la construction est produite par le conduit vocal). On distingue, généralement, deux classes principales : les voyelles et les consonnes.

I-2-1-Les voyelles :

Les voyelles sont des sons qui ont une durée assez longue, caractérisés par un libre écoulement et une grande constance de leurs propriétés fréquentielles dans le temps.

Il y a deux types de voyelles : les voyelles orales (i, e, u, ...) qui sont émises sans intervention de la cavité nasale et les voyelles nasales (ã, e~, ...) qui font intervenir la cavité nasale. La langue française comprend douze voyelles orales émises seulement par la bouche, ainsi que quatre voyelles nasales correspondant à la mise en parallèle des cavités nasales sur la cavité buccale par abaissement du voile du palais.

Chaque voyelle se caractérise par les résonances du conduit vocal qu'on appelle "**les formants**". En général, les trois premiers formants sont suffisants pour caractériser toutes les voyelles.

I-2-2-Les consonnes :

Les consonnes sont des phonèmes qui, lors de leur articulation, rencontrent un obstacle (lèvres pour les voyelles labiales, dents pour les dentales, fermeture du palais pour [k], etc.). Elles sont généralement beaucoup plus courtes que les voyelles et beaucoup plus variables dans le temps. Les consonnes peuvent être rangées en trois catégories : les occlusives, les fricatives et les sonantes.

I-2-2-1-Les occlusives :

Les occlusives sont caractérisées par un silence provenant de la fermeture complète du conduit vocal (occlusion) en un point précis. La durée du silence, est influencée par l'entourage phonétique et par le débit de parole. Mais ce silence peut ne pas être total (dans le cas des occlusives sonores) car il peut se former une " barre de voisement ", qui est produite par une vibration des cordes vocales. Cette barre, de faible énergie, est concentrée dans les basses fréquences (100 à 300 Hz). La fin de l'occlusion provoque une perturbation acoustique, sous la forme d'une onde de pression due au relâchement de l'air qui était comprimé par l'occlusion. Cette perturbation est de courte durée.

I-2-2-2-Les fricatives :

Les fricatives (ou constrictives), sont des bruits produits par l'écoulement turbulent de l'air. Lorsque cet écoulement rencontre un rétrécissement «un lieu de constriction» il se produit un bruit de haute fréquence.

I-2-2-3-Les sonantes :

Les consonnes sonantes (c'est à dire les liquides, les nasales et les semi-consonnes) présentent la particularité de posséder des formants et peu ou pas de bruit.

I-3-L'alphabet phonétique international :

L'alphabet phonétique international (IPA) associe des symboles phonétiques aux sons, de façon à permettre l'écriture universelle des prononciations (**Tableau I-1** pour le français) :

IPA	EXEMPLES	IPA	EXEMPLES
i	idée, ami	p	patte, repas, cap
e	ému, ôté	t	tête, ôter, net
ɛ	perdu, modèle	k	carte, écaille, bec
a	alarme, patte	b	bête, habile, robe
ɑ	bâton, pâte	d	dire, rondeur, chaud
ɔ	Obstacle, corps	g	gauche, égal, bague
o	auditeur, beau	f	feu, affiche, chef
u	coupable, loup	s	sœur, assez, passe
y	punir, élu	ʃ	chanter, machine, poche
ø	creuser, deux	v	vent, inventer, rêve
œ	malheureux, peur	z	zéro, raisonner, rose
ə	petite, fortement	ʒ	jardin, manger, piège
ɛ̃	peinture, matin	l	long, élire, bal
ɑ̃	vantardise, temps	ʀ	rond, chariot, sentir
ɔ̃	rondeur, bon	m	madame, aimer, pomme
œ̃	lundi, brun	n	nous, punir, bonne
j	piétiner, briller		agneau, peigner, règne
w	oui, fouine	ŋ	jumping, smoking
ɥ	huile, nuire	h	halte, hop (exclamations)

Table I-1 : Alphabet phonétique internationale (API). La liste des symboles est restreinte aux phonèmes du français.

Après avoir expliqué le processus de production de signal de parole on passe maintenant à la description des différentes étapes nécessaire pour l'acquérir afin d'extraire les caractéristique qu'on va utiliser.

I-4-Principe d'acquisition du signal de parole :

L'acquisition de données consiste à recueillir les signaux de leurs sources et les numériser pour les stocker, les analyser et les présenter sur un PC. Différentes composantes sont à prendre en compte lors de la construction d'un système d'acquisition du signal de parole (Figure I-2) :

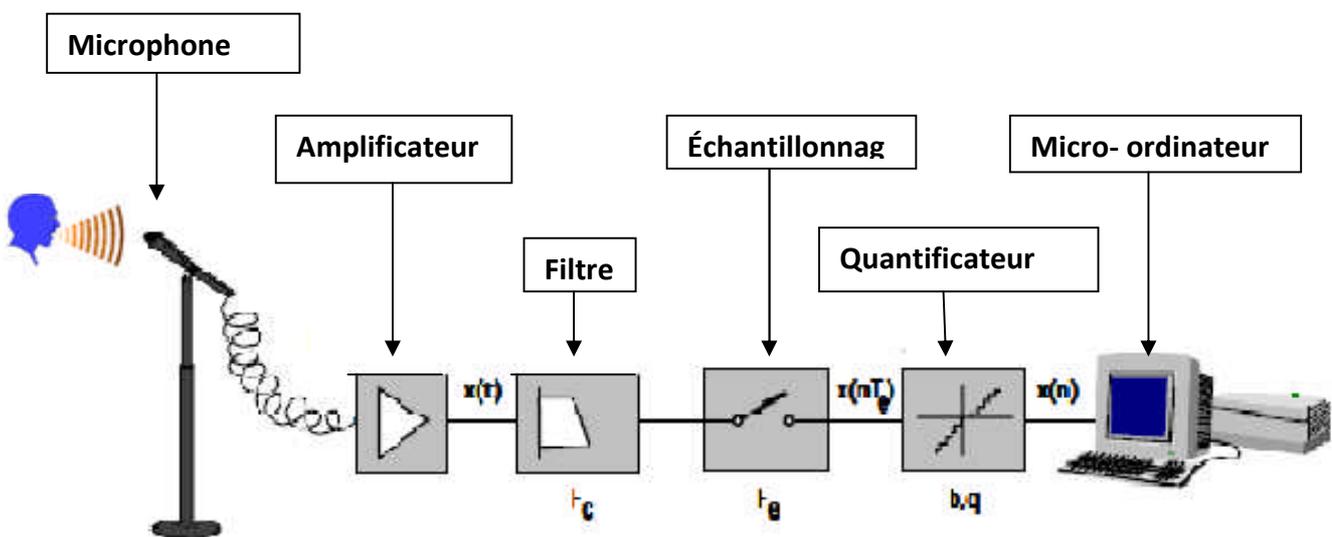


Figure I-3 : Le schéma synoptique d'enregistrement numérique du signal de parole

I-4-1-Transducteurs et capteurs (microphone) :

Une fois que le son a été émis par le locuteur, il est capté par un microphone qui transforme les vibrations sonores en courant électrique (l'énergie mécanique en énergie électrique).

Une bobine mobile est solidaire de la membrane. Cette dernière sous l'effet des variations de pression acoustique fait osciller la bobine dans un champ magnétique annulaire produit par un aimant permanent. La bobine coupe les lignes de force du champ magnétique. En raison de cette oscillation périodique de la membrane, et donc de la bobine dans le champ, il y a induction d'un courant électrique dans la bobine mobile-courant qui peut être rendu utilisable par une amplification appropriée (Figure I-3).

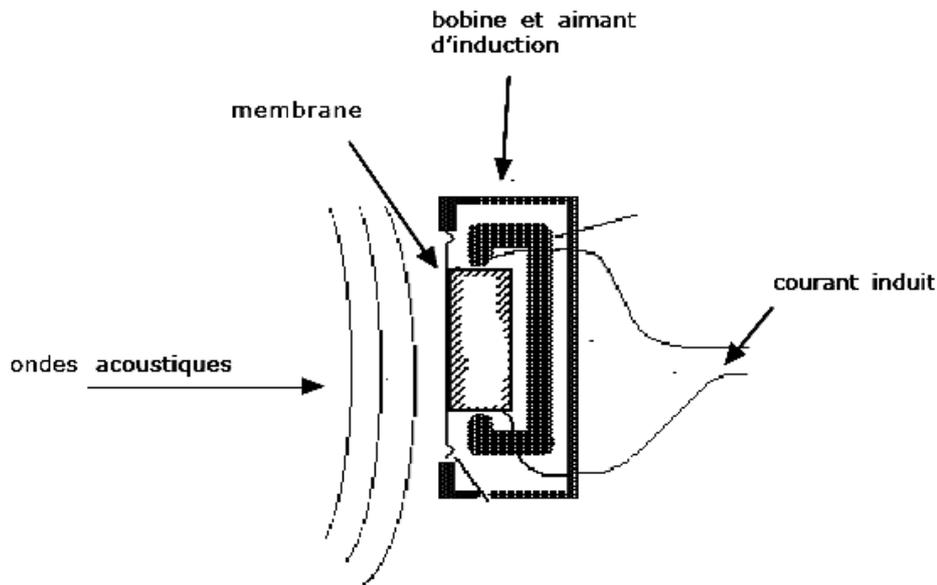


Figure I-4 : Microphone dynamique

I-4-2-L'amplificateur :

Il est placé juste à la sortie du capteur, son rôle est d'amplifier le signal issu du microphone qui est de faible énergie, il assure également la protection contre une éventuelle surcharge, court circuit et échauffements, de manière à ce qu'il soit traitable par les autres étages.

I-4-3-Le filtre d'anti-repliement (filtre de garde) :

Le signal électrique est ensuite appliqué à l'entrée d'un filtre passe bas dit filtre d'anti-repliement dont l'utilité est de centrer le signal de parole.

I-4-4-Numérisation du signal :

La numérisation du signal est une opération indispensable afin de pouvoir le transmettre vers le PC. On utilise pour cela, un convertisseur analogique-numérique (CAN) qui assure les opérations d'échantillonnage et de quantification.

➤ **Première opération : Echantillonnage**

L'échantillonnage consiste à passer d'un signal à temps continu en une suite discrète de valeur (valeur mesurée à intervalle régulier). On appelle période d'échantillonnage, notée T_e la durée entre deux échantillons successifs, son inverse est appelée «fréquence d'échantillonnage», qui est notée : $F_e = 1/T_e$.

Dans le cas ou on souhaite reconstituer le signal analogique, il faut poser une contrainte dans le choix de ' T_e ', cette contrainte découle du théorème «Shannon» qui dit que. La fréquence $F_e = 1/T_e$ doit valoir au moins le double de la fréquence maximal ' F_{max} ' du signal c'est-à-dire : $F_e \geq 2.F_{max}$

Les fréquences inférieurs aux demi-fréquences d'échantillonnage ne seront pas affectées par l'effet de repliement, donc si on souhaite représenter correctement des fréquences jusqu'à 20Khz, il est nécessaire d'échantillonné le signal a une fréquence supérieure ou égale à 40Khz. Par une répétition de spectre d'amplitude trace les F_e .

Le filtre anti-repliement permet, justement, de supprimer les répétitions du spectre pour ne garde que l'allure centrale qui correspond au signal d'origine continu.

(La Figure I-5) suivante représente le mot« parenthèse » échantillonné à $F_e=44.1$ KHz

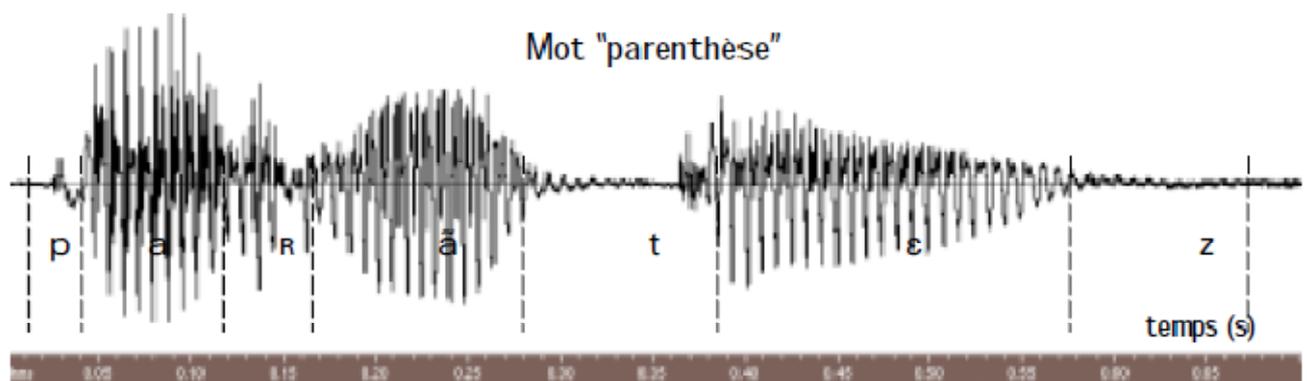


Figure I-5 : Audiogramme du signal de parole

➤ **Deuxième opération : Quantification**

Quand on prélève des valeurs du signal à des instants multiples de T_e , la valeur obtenue, $x(kT_e)$, n'est pas forcément entière et peut correspondre à une infinité de valeurs après la virgule (exemple: 1.26394...). Pour remédier à cela, on définit une précision, donnée par le pas de quantification q : toute valeur comprise dans un intervalle donnée de largeur q est assimilée à la valeur au début de l'intervalle, ce qui revient à définir une précision des amplitudes échantillonnées correspondantes (exemple : 1.26394... devient 1.26).

L'amplitude mesurée du signal est alors codée sur 'b' bits ce qui donne 2^b classe. L'approximation de la valeur analogique par une valeur numérique discrète provoque une distorsion du son dit bruit de quantification. Pour limiter ce bruit on choisit un nombre de classe élevé. L'erreur relative de quantification vaut à la moyenne $(1/2^b)$ tel que b est le nombre de bits utilisé pour le codage des échantillons.

I-5-Caractéristiques du signal de parole

I-5-1-Caractéristique physique

Le signal de la parole n'est pas un signal ordinaire, il s'inscrit dans le cadre de la communication parlée et il est difficile de le modéliser compte tenu de sa variabilité. Nous allons ici, tenter de mettre en évidence quelques caractéristiques importantes du signal non stationnaire afin de faire ressortir les problèmes posés lors de son traitement.

I-5-1-1-La redondance :

Les répétitions confèrent à ce signal une robustesse. La redondance le rend résistant au bruit. Dans une certaine mesure, elle fonctionne comme un code correcteur d'erreur, puisqu'un interlocuteur humain sait décrypter un message même s'il est entaché de bruits dus à de possibles interférences.

I-5-1-2-La coarticulation

La production parfaite d'un son suppose un positionnement précis des organes phonatoires. Le déplacement de ces organes est limité par une certaine inertie mécanique. Les sons émis dans une même chaîne acoustique subissent alors l'influence de ceux qui les précèdent ou les suivent. Ces effets de coarticulation constituent un facteur de variabilité supplémentaire important du signal de parole.

I-5-1-3-La variabilité du signal parole :

Les caractéristiques des sons de parole portent des informations aussi bien sur les catégories linguistiques que sur l'identité des locuteurs et sur leurs états émotionnels.

➤ La variabilité intra-locuteur :

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 variabilité interlocuteur :

Elle est plus évidente. La hauteur de la voix, l'intonation l'accent diffèrent selon le sexe, l'origine sociale, régionale ou nationale.

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.

I-5-1-4-Les interférences

Le signal de parole peut être perturbé pour certains phénomènes due à l'environnement tels que :

- l'acoustique du lieu (milieu protégé ou environnement bruyant)
- la qualité du microphone et sa position par rapport à la bouche
- les bruits de bouche.

I-5-2 Caractéristiques électriques :

Elles sont décrites par : la fréquence fondamentale, la bande spectrale et le taux de passage par zéro.

I-5-2-1 Fréquence fondamentale :

Dans le domaine fréquentiel, un signal est caractérisé par sa bande spectrale qui regroupe toutes les composantes contenues dans ce signal, l'une parmi ces composantes est caractérisée par une amplitude maximale, dont sa fréquence représente la fréquence fondamentale du signal.

I-5-2-2-La bande spectrale :

L'oreille est l'organe de l'audition qui effectue la transformation des signaux physiques (vibrations acoustiques) en signaux nerveux électrique qui seront décodés par le cerveau.

L'oreille humaine perçoit des fréquences comprises entre 20 Hz (fréquence la plus grave) et 20000 Hz (fréquence la plus aiguë).

Toute fréquence inférieure à 20 Hz est qualifiée d'infrason, même si l'oreille d'autres animaux (taupe par exemple) peut capter des "sons" de quelques Hertz.

Toute fréquence supérieure à 20 kHz est qualifiée d'ultrason, alors qu'un chien entend jusqu'à 40 kHz et une chauve-souris jusqu'à 160 kHz.

L'appareil auditif comprend l'oreille externe, l'oreille moyenne, et l'oreille interne Figure I-5. Le conduit auditif relie le pavillon au tympan : c'est un tube acoustique de section uniforme fermé à une extrémité, son premier mode de résonance est situé vers 3000 Hz, ce qui accroît la sensibilité du système auditif dans cette gamme de fréquences. Le mécanisme de l'oreille interne (marteau, étrier, enclume) permet une adaptation d'impédance entre l'air et le milieu liquide de l'oreille interne. Les vibrations de l'étrier sont transmises au liquide de la cochlée. Celle-ci contient la membrane basilaire qui transforme les vibrations mécaniques en impulsions nerveuse.

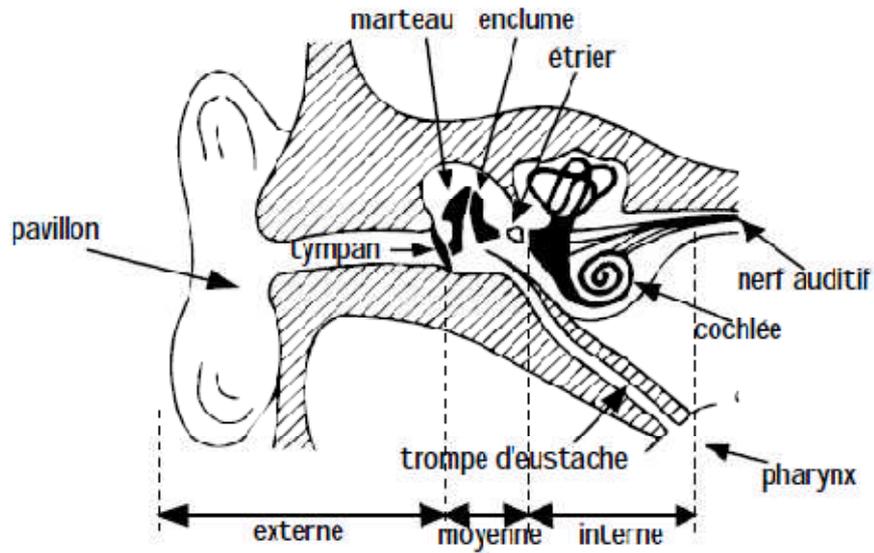


Figure I-6 : Le système auditif

I-6-Conclusion

La parole est un moyen de communication où de nombreux éléments entrent en jeu, tels que le lieu, l'émotion du locuteur, la relation qui s'établit entre les locuteurs (stressante ou amicale). Ces facteurs influencent sur la forme et le contenu du message. L'acoustique du lieu (milieu protégé ou environnement bruyant), la qualité du microphone, les bruits de bouche, les hésitations, les mots hors vocabulaire sont autant d'interférences supplémentaires sur le signal de parole qui rendent, parfois, son interprétation complexe

Chapitre II

L'analyse et la reconnaissance de signal vocal

L'analyse est une étape indispensable à toute application de synthèse, de codage et de reconnaissance de la parole. Dans ce chapitre on basant d'abord sur L'analyse du signal fournissant une description du signal acoustique et l'extraction des paramètres significatifs tel que la fréquence fondamentale, les formants, l'énergie et pour cela il existe plusieurs méthodes. Puis à la reconnaissance automatique du locuteur dont on va décrire les divers systèmes, les modes de la reconnaissance. Puis les limites et les performances d'un système de reconnaissance du locuteur, ainsi que ses domaines d'application.

II-1-Les différentes méthodes d'analyse de la parole :

Il existe deux méthodes différentes, selon la technique utilisée. La méthode est dite directe lorsqu'on agit sur les organes vocaux. Elle est indirecte lorsqu'on exploite le signal électrique traduisant la parole.

II-1-1-La méthode directe :

On distingue plusieurs cas selon le principe utilisé :

II-1-1-1-La glottographie photoélectrique :

Elle utilise un principe optique qui consiste à mesurer les variations continues de la surface de la glotte par l'intermédiaire d'un phototransistor inséré par le nez dans le pharynx, et qui mesure l'intensité lumineuse à travers la peau du cou au niveau de la glotte. Le signal électrique délivré par le phototransistor représente la variation de l'air de la glotte. C'est la photographie photoélectrique.

II-1-1-2-La glottographie électrique :

Elle s'effectue au moyen de deux électrodes fixées sur le cou sans modifier les conditions normales de phonation est permet de mesurer les variations de l'impédance du tissu musculaire du larynx.

II-1-1-3-La glottographie ultrasonore :

A travers la réflexion des ondes ultrasonores sur des cristaux piezo électriques placés autour du cou, on peut mesurer le temps de fermeture et d'ouverture des cordes vocales.

II-1-2-La méthode indirecte :

C'est cette méthode qui nous intéresse car elle traite le signal électrique délivré par un microphone. On peut traiter ou analyser le signal dans le domaine temporel ou bien dans le domaine fréquentiel selon les paramètres à extraire.

II-1-2-1-L'analyse temporelle :

Elle fournit des paramètres à partir du traitement effectué sur la représentation temporelle du signal. De nombreuses mesures peuvent être effectuées, à partir de cette représentation telle que le taux de passage par zéro que l'on utilise dans notre application

II-1-2-1-1-Le taux de passage par zéro (T P Z) :

Pour un signal échantillonné, il y a passage par zéro lorsque deux échantillons successifs sont de signe opposé. Le taux de passage par zéro à court terme peut être estimé par la formule suivante, dans l'intervalle du temps compris entre les échantillons 0 et N

$$(\text{TPZ}) = -\sum_{n=0}^{N-1} a_n \quad \text{Avec } a_n = 1 \text{ si } x_{n+1} \cdot x_n < 0 \text{ (0 sinon).}$$

n : Numéro d'échantillon du signal de parole.

N : Nombre d'échantillons par fenêtre.

x_n : Echantillon du signal.

x_{n+1} : Echantillon suivant du signal.

Le **TPZ** est un indice très simple à calculer. Il représente le nombre de fois que le signal, dans sa représentation amplitude/temps, passe par la valeur centrale (zéro).

II-1-2-1-2-Analyse par l'énergie :

L'énergie à court terme sert à détecter les silences, elle est élevée s'il y a présence d'un signal (son), et petite s'il n'y a pas de son c'est-à-dire présence d'un silence.

L'énergie E en décibel (dB) est définie par

$$E = 10 \log - \sum ()$$

N : nombre d'échantillons par fenêtre.

s(n) : échantillon de signal vocal

II -1-2-2-Analyse fréquentielle :

L'analyse spectrale a pour but de caractériser la répartition d'énergie ou de puissance d'un signal en fonction de la fréquence.

II-1-2-2-1- Analyse fréquentielle par la transformée de Fourier à court terme (TFCT) :

La transformée de Fourier (**TF**) présente un intérêt majeur et reste un outil de base en traitement du signal. Cependant, si cet opérateur permet de connaître les différentes fréquences contenues dans un signal c'est-à-dire son spectre de fréquence il ne permet pas de savoir à quels instants ces fréquences ont été émises. Autrement dit, la **TF** donne une information globale et non locale des composantes fréquentielles. Cette perte de localité est un inconvénient pour analyser des signaux non stationnaires ou quasi-stationnaire, notamment le signal parole pour lesquelles la connaissance temporelle de répartition spectrale est primordiale pour remédier à cela, c'est-à-dire aux problèmes par l'inadéquation de la TF dans le cas des signaux non stationnaire on définit la TFCT

Pour illustrer ceci, considérons un signal réel triangulaire et calculons son spectre d'amplitude. On remarque, sur la répartition fréquentielle du signal, que l'allure de S(f) (**Figure II-1-a**) ne permet pas de déduire à quel instant se trouve l'amplitude maximale de s(t)

(**Figure II-1-b**), c'est-à-dire à quel instant apparaît le triangle.

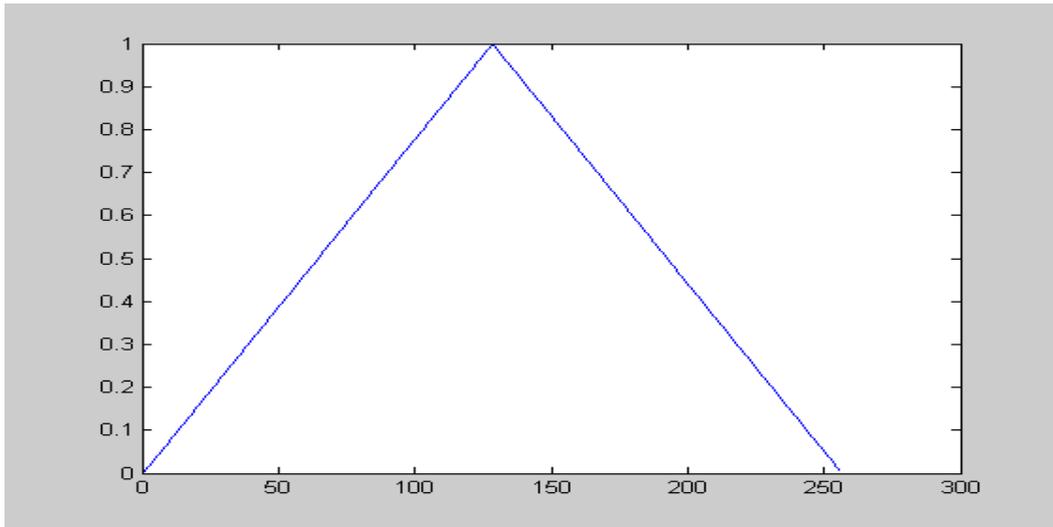


Figure II-1-a : Le signal $s(t)$

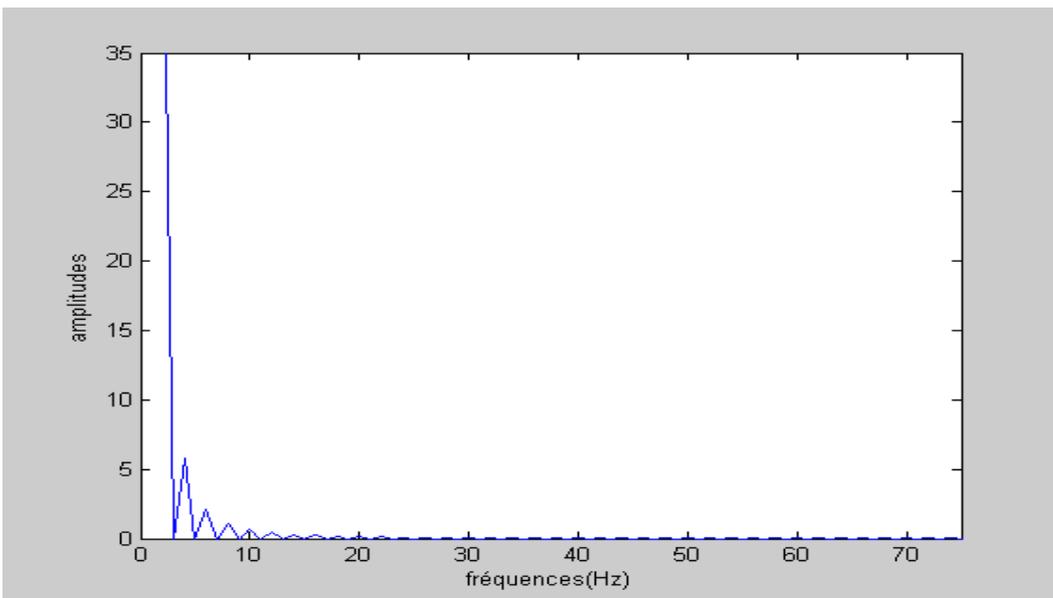


Figure II-1-b : La TF de $s(t)$

La première réflexion des traiteurs de signaux est l'introduction de la notion de stationnarité locale, qui consiste à découper le signal à traiter en segments dans lesquels le signal est stationnaire. On définit alors la **transformée de Fourier à court terme (TFCT)**.

La **TFCT** consiste alors, à appliquer au signal $s(t)$ une fenêtre d'observation glissante de sorte que les hypothèses de stationnarité soient localement satisfaites. On applique alors la TF non pas au signal global $s(t)$, mais à chacune des tranches ainsi découpées. On obtient ainsi le

spectre local du signal. La totalité des spectres locaux indique alors comment le spectre varie au cours du temps.

Sur l'axe horizontal on porte le numéro de la fenêtre (c'est-à-dire le temps) et sur l'axe vertical la fréquence. En effet on suppose que le signal est stationnaire sur toute la longueur de chaque fenêtre.

Pour un signal $x(t)$, la TFCT est définie par :

$$\text{TFCT} = \int_{-\infty}^{\infty} () w(t-\tau) \exp(-2j\pi ft) dt$$

Limites de la TFCT : Résolution temps – fréquence :

La TFCT ne peut pas décrire le contenu du signal strictement à l'instant « t » et à la fréquence « f ». la taille de la fenêtre d'analyse est assujettie au principe d'incertitude d'Heisenberg qui stipule que la résolution conjointe temps-fréquence est minorée :

$$\Delta t \Delta f > 1/4\pi$$

Si on souhaite une bonne localisation temporelle on a besoin d'une fenêtre étroite $h(t)$, et si on souhaite une bonne localisation fréquentielle la fenêtre $H(f)$ doit être étroite.

➤ Application au signal de parole :

Comme le signal de parole est un signal non stationnaire, l'utilisation de la TFCT, donc du fenêtrage, impose.

Les fenêtres les plus utilisées en le traitement de signal sont les suivants :

- La fenêtre Hamming (à bande étroite) (**Figure II-2-a**)
- La fenêtre rectangulaire (**Figure II-2-b**)
- La fenêtre Hanning (à bande large). (**Figure II-2-c**)
- Fenêtre blackman. (**Figure II-2-d**)

Si l'on considère la fenêtre de Hamming, ses expressions fréquentielle, dans le domaine numérique, s'écrit

$$W[n] = 0.54 + 0.46 \cos((2n\pi) / (2M+1)).$$

n: numero de l'échantillon

$W(n)$: TFCT de la fenêtre

Cette expression se traduit par une bande de transition étroite (inférieure à 2Khz) et une atténuation stagnante de -75 db dans la bande rejetée (**Figure II-2**)

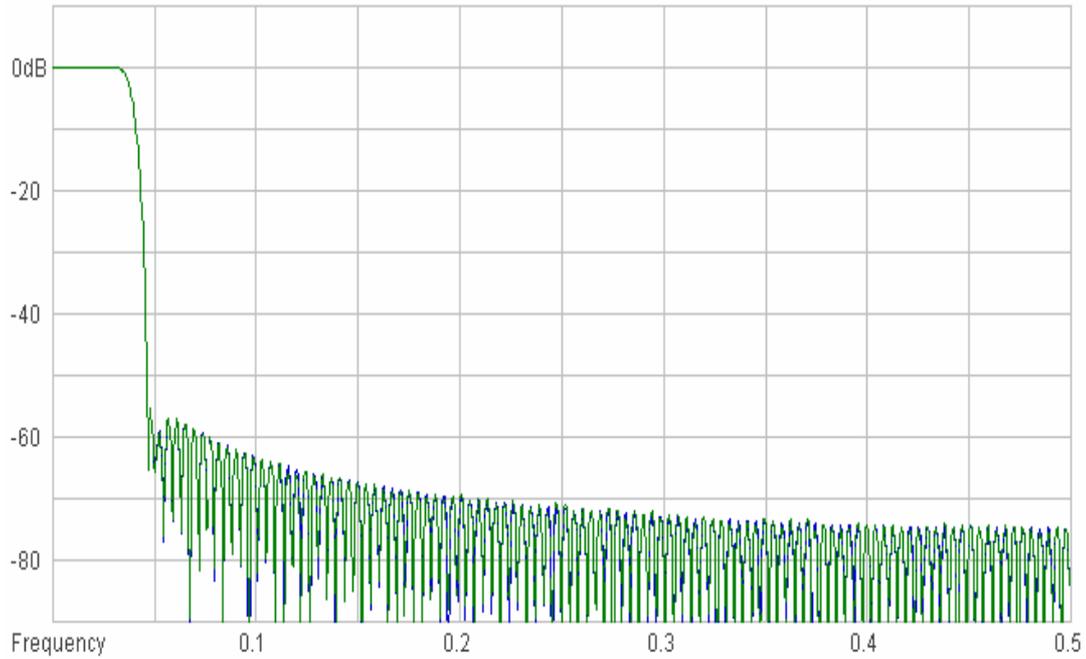


Figure II-2-a : Spectre d'amplitude de la Fenêtre Hamming

De la même façon, les représentations fréquentiels des autres fenêtres sont données par les figures suivantes

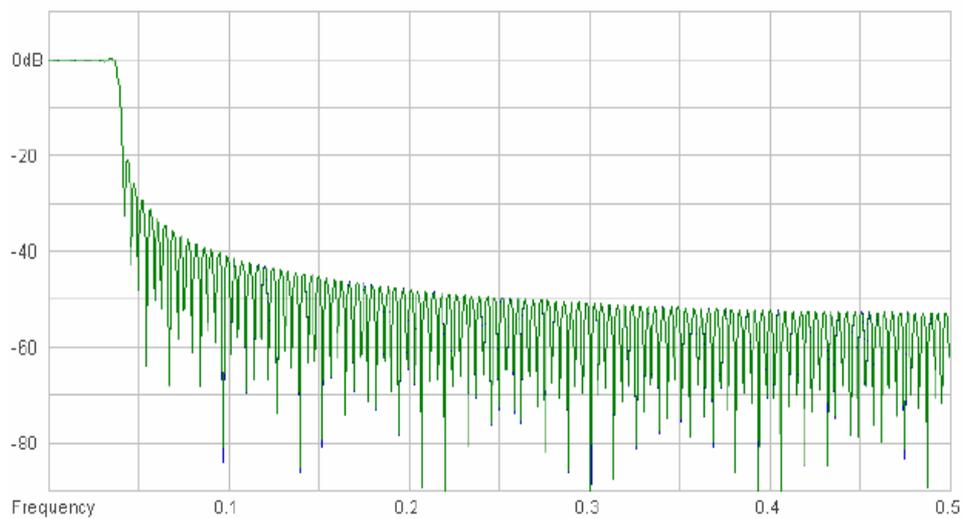


Figure II-2-b : Fenêtre Rectangulaire

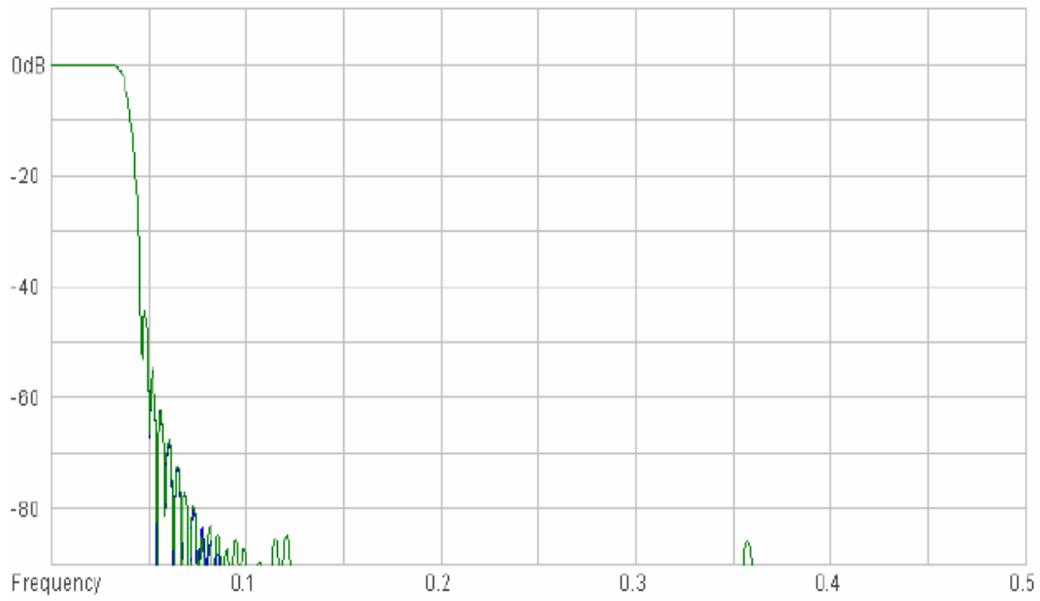


Figure II-2-c : Fenêtre Hanni

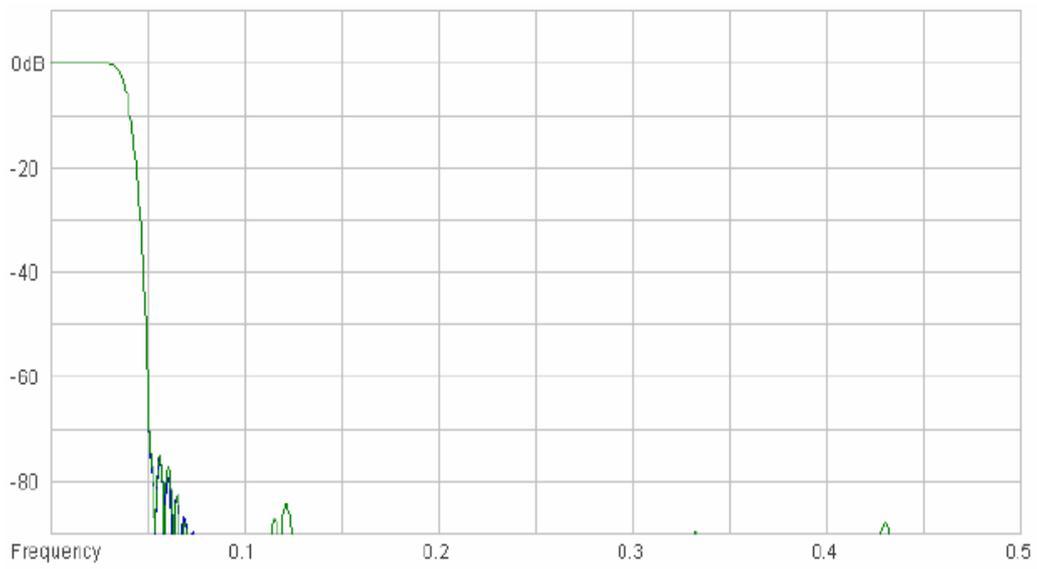


Figure II-2-d : Fenêtre Blackman

Exemple :

Si on applique la TFCT au signal de parole correspondant a la phrase « merci bon dieu », on obtient l'allure suivante (**Figure II-3**)

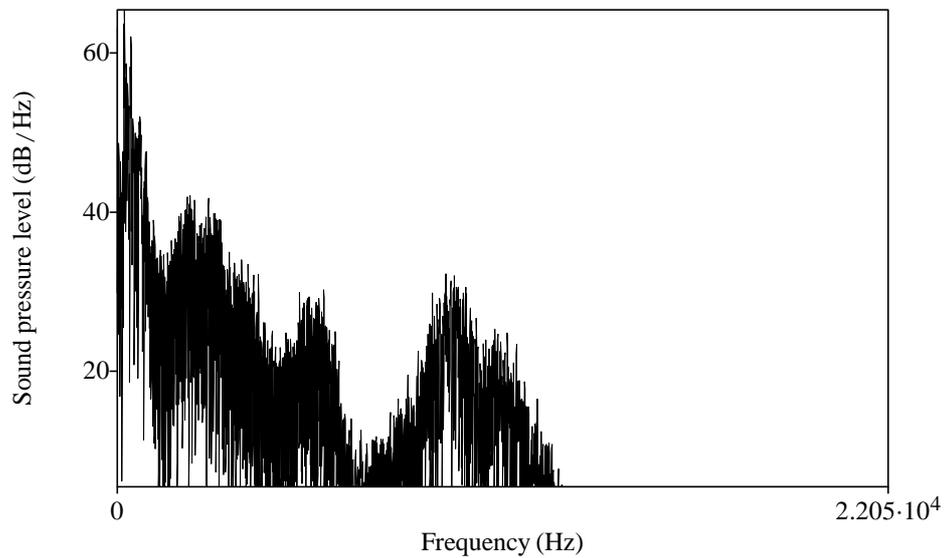
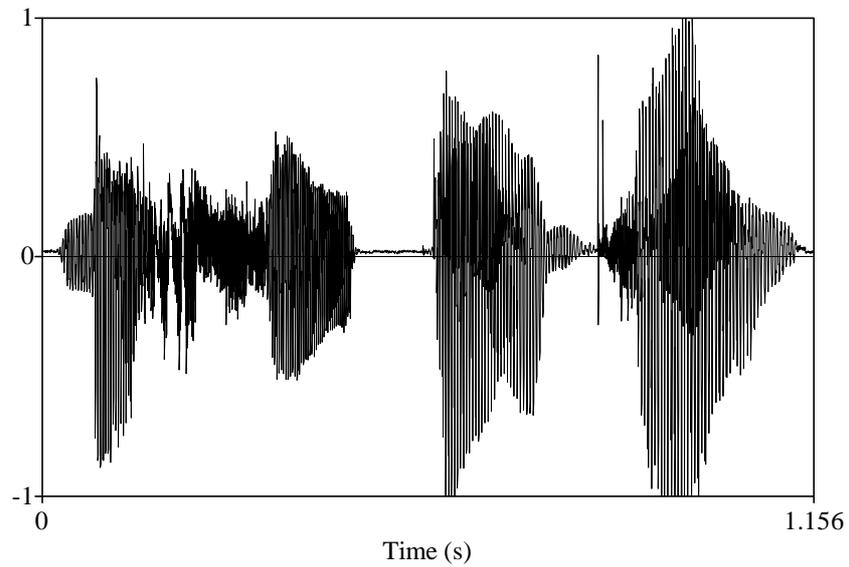


Figure II-3 : La phrase « Merci bon Dieu »

➤ **Fréquence fondamentale :**

La fréquence fondamentale, représente la fréquence de vibration des cordes vocales durant la parole. De nombreuses études ont montré que la fréquence fondamentale était un paramètre utile à la reconnaissance, non seulement parce qu'il est très différent entre les hommes et les femmes, mais aussi entre les personnes du même sexe qui ont des moyennes de pitch différentes.

$$F_0 = \frac{1}{2L_m} \sqrt{\frac{\sigma_c}{\rho}}$$

L_m : représente la longueur des cordes vocales en mètres

σ_c : est la tension longitudinale des cordes quantifiée en Pascal

ρ est la densité des cordes

Elle s'étend approximativement de 70 à 250 Hz chez les hommes, de 150 à 400

Hz chez les femmes, et de 200 à 600 Hz chez les enfants.

II-2-La reconnaissance vocal :

La reconnaissance vocale est une technique informatique qui permet d'analyser un mot ou une phrase captée au moyen d'un microphone pour la transcrire sous la forme d'un texte exploitable par une machine. La reconnaissance vocale, fait partie des techniques de traitement de la parole. Ces techniques permettent notamment de réaliser des interfaces vocales c'est-à-dire des interfaces homme-machine où une partie de l'interaction se fait à la voix. Parmi les nombreuses applications, on peut citer les applications de dictée vocale sur PC où la difficulté tient à la taille du vocabulaire et à la longueur des phrases, mais aussi les applications téléphoniques de type serveur vocal, où la difficulté tient plutôt à la nécessité de reconnaître n'importe quelle voix dans des conditions acoustiques variables et souvent bruyantes (téléphones mobiles dans des lieux publics).

Les techniques modernes de traitement de la parole tendent cependant à produire des systèmes automatiques qui se substituent à l'une ou à l'autre de ces fonctions :

- Les reconnaissances ont pour mission de décoder l'information portée par le signal vocal à partir des données fournies par l'analyse de ce signal. On distingue fondamentalement deux types de reconnaissances en fonction de l'information que l'on cherche à extraire du signal vocal :
 - **la reconnaissance du locuteur**, dont l'objectif est de reconnaître la personne qui parle.
 - **la reconnaissance de la parole**, où l'on s'attache plutôt à reconnaître ce qui est dit.

II-2-1-Modes de fonctionnement des systèmes de reconnaissance :

Un système de reconnaissance peut être utilisé sous plusieurs modes :

II-2-1-1-Dépendant du locuteur (mono locuteur) :

Dans ce cas, le système de reconnaissance est configuré pour un locuteur spécifique. Ce mode de fonctionnement est principalement utilisé pour de petits vocabulaires car il est demandé à l'utilisateur de prononcer tous les mots du vocabulaire une ou plusieurs fois, afin de les apprendre et de s'adapter à sa prononciation. Il faut cependant noter qu'il est souvent demandé de prononcer les mots plusieurs fois car il est impossible de prononcer un mot deux fois exactement de la même façon; il existe toujours des variations dans le débit d'élocution, l'accentuation ou l'intonation du mot prononcé. Cette phase d'apprentissage permet de créer une image du mot qui nous servira de modèle pour la phase de reconnaissance.

II-2-1-2-Pluri locuteur (multi locuteur) :

Le système de reconnaissance est élaboré pour un groupe restreint de personnes.

II-2-1-3-Indépendant du locuteur :

Dans ce cas le système de reconnaissance est capable de reconnaître n'importe qui (tout locuteur peut utiliser le système de reconnaissance).

II-2-2-Le mode d'élocution :

Le mode d'élocution caractérise la façon dont on peut parler au système .il existe trois modes d'élocution (reconnaisseurs) distincts :

➤ Mots isolés :

Chaque mot doit être prononcé isolément, c'est-à-dire précédé et suivi d'une pause (séparer chaque mot par un silence) le débit d'élocution étant de 20 à 100 mots/minutes.

➤ Mots connectés :

Le système reconnaît des séquences de quelques mots sans pause volontaire pour les séparer (exemples : reconnaissance de chiffres connectés ou de nombres quelconques.....).

➤ Parole continue :

Le locuteur prononce n'importe quelle suite de mots de façon continue, son débit entre de 150 à 200mots/minute.

Notre étude sera basée sur la reconnaissance du locuteur sur un ensemble fermé, c'est-à-dire identifier la personne qui a parlé parmi un ensemble de personnes fini et connu à l'avance.

II-3-La reconnaissance du locuteur.

Comme on peut le constater sur la (**Figure II-4**), la reconnaissance du locuteur s'inscrit dans le domaine plus général du traitement de la parole, elle vise à déterminer, si un échantillon de voix a été prononcé par une personne donnée.

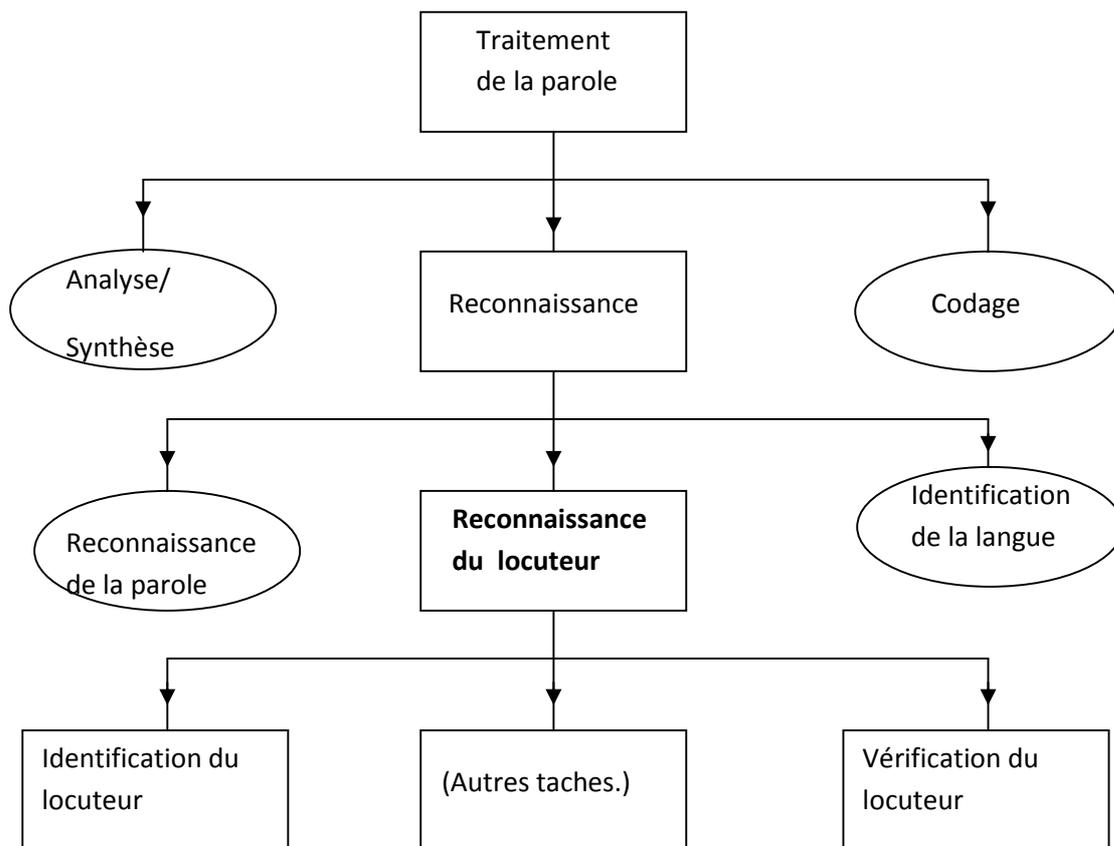


Figure II-4 : Principales tâches traitement de la parole

II-3-1-Identification et vérification du locuteur :

Les tâches courantes en reconnaissance de locuteur peuvent être classées en deux grandes catégories : L'identification et la vérification.

II-3-1-1-La vérification du locuteur (ou l'authentification) :

La vérification du locuteur consiste, après que le locuteur ait décliné son identité, à vérifier l'adéquation de son message vocal avec la référence acoustique du locuteur qu'il prétend être. C'est une décision en tout ou rien.

Dans ce type d'applications, il s'agit donc de trancher entre les deux hypothèses : soit le locuteur est bien le locuteur autorisé, c'est à dire celui dont l'identité est revendiquée, soit nous avons affaire à un imposteur qui cherche à se faire passer pour un locuteur autorisé. Les applications classiquement envisagées pour la vérification du locuteur correspondent donc à l'idée de "serrure vocale" qui peut être utilisée, par exemple, pour valider des transactions bancaires effectuées par téléphone, ou pour compléter un dispositif d'accès (à un bâtiment, un système informatique).

II-3-1-2-L'identification du locuteur :

Les tâches relevant de l'identification du locuteur consistent à rechercher, parmi un ensemble de locuteurs connus, le locuteur possédant la référence la plus proche d'un message vocal donné. Deux modes de vérifications sont possibles :

- L'identification en ensemble fermé : pour lequel le locuteur est identifié parmi un nombre connu de locuteurs.
- L'identification en ensemble ouvert : pour lequel le locuteur à identifier n'appartient pas forcément à cet ensemble (un locuteur inconnu a pu prononcer le message).

Pour une application de vérification ou d'identification, il est nécessaire de disposer d'une base de données contenant des enregistrements de référence correspondant à chacun des locuteurs autorisés. En pratique, on ne conserve, pour chaque locuteur, que les paramètres utiles, extraits des enregistrements de référence. Ces informations constituent les données de référence du locuteur. L'étape préliminaire, qui consiste à bâtir les données ou les modèles de référence propres à chaque locuteur, est appelée phase d'**apprentissage**.

II-4-Système de reconnaissance de locuteur :

La reconnaissance du locuteur peut être interprétée comme une tâche particulière de reconnaissance de formes.

C'est une succession de modules dont l'étape finale est de reconnaître une forme particulière, c'est-à-dire le signal de parole que l'on met à l'entrée de cette chaîne. Un système

automatique de reconnaissance vocale se divise généralement en quatre modules : qui sont présentés par la (Figure II-4).

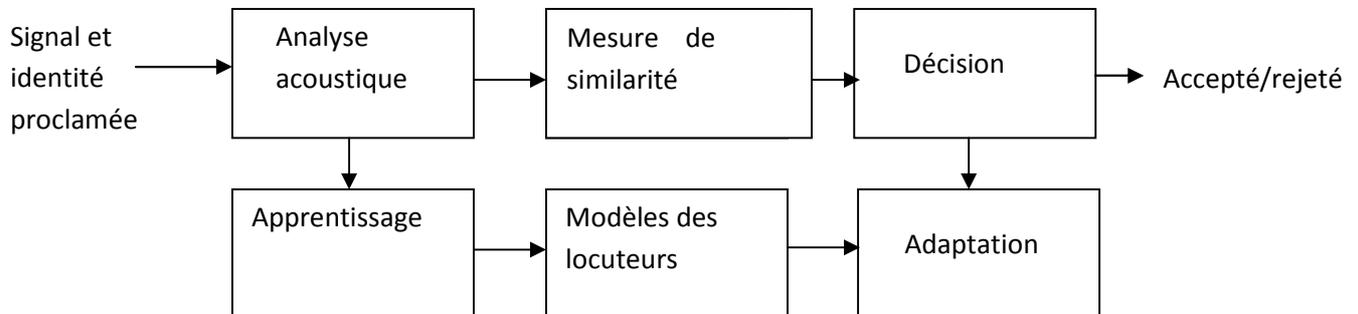


Figure II-4 : Schéma typique d'un système de vérification du locuteur.

Tout d'abord le message vocal doit être soumis à un modèle d'acquisition et de paramétrisation du signal. Celui-ci a en charge de produire une suite de vecteurs de paramètres acoustiques représentant le message vocal sous une forme exploitable par le système.

Dans l'étape d'apprentissage, on crée une référence vocale pour un locuteur à partir d'un échantillon de sa voix (modèle des locuteurs).

Un modèle de calcul de ressemblance (similarité) est utilisé durant la reconnaissance. Il a en charge de mesurer la similarité entre les paramètres acoustiques du signal prononcé et les références correspondant à une personne donnée.

Enfin un module de décision, basé sur une stratégie de décision donnée, fournit la réponse du système.

II-4-1-Paramétrisation :

Un premier module de traitement du signal réalise l'analyse acoustique du signal de parole sur des fenêtres temporelles de courte durée, de 10 à 30 ms. A l'issue de cette étape, le signal est représenté par des vecteurs acoustiques de paramètres qui permettent de discriminer

différents locuteurs, ce qui permet de réduire l'information en quantité et en redondance. Idéalement, les paramètres (ou traits acoustiques) doivent être : fréquents, facilement mesurables, peu sensibles à la variabilité intralocuteur, non affectés par le bruit ambiant ou les variations dues au canal de transmission.

En pratique, il est très difficile de réunir tous ces attributs en même temps. La sélection des traits acoustiques pertinents pour la reconnaissance vocale est donc un sujet largement traité. Il ressort également que les seuls types de paramètres vraiment pertinents et utilisables efficacement, sont les paramètres de l'analyse spectrale.

II-4-2-Modélisation des locuteurs :

Dans le cadre de la reconnaissance vocale, nous modélisons les différentes prononciations qu'un locuteur peut avoir effectuées pour le même motif. En étudiant la parole de locuteurs sur plusieurs prononciations des mêmes motifs, nous pouvons distinguer des variabilités caractéristiques du signal de parole nous permettant de séparer les locuteurs les uns des autres (variabilités inter-locuteurs) et d'autres, intrinsèques au locuteur (variabilités intra-locuteur).

Reconnaître un locuteur revient donc à essayer de le distinguer des autres

Dans ce cadre, différentes approches ont été développées, et peuvent être classées en quatre grandes familles :

II-4-2-1-L'approche vectorielle :

Le locuteur est représenté par un ensemble de vecteur de paramètres dans l'espace acoustique. Les principales techniques utilisées sont la reconnaissance à base de DTW (reconnaissance par déformation temporelle dynamique, en anglais dynamique time warping) et par quantification vectorielle.

II-4-2-2-L'approche statistique :

Consiste à représenter chaque locuteur par une densité de probabilité dans l'espace des paramètres acoustiques et à utiliser les techniques de modélisation par le modèle de Markov caché par les mélanges de gaussiennes et par des mesures statistiques du second ordre.

II-4-2-3-L'approche connexionniste :

Consiste principalement à modéliser les locuteurs par des réseaux de neurones.

II-4-2-4-L'approche relative :

Il s'agit de modéliser un locuteur relativement à d'autres locuteurs de référence dont les modèles sont bien appris.

II-4-3-Décision et mesure des performances :

La stratégie mise en jeu dans cette partie de système de reconnaissance vocale dépend essentiellement du type d'application : identification ou vérification.

Dans le cas vérification du locuteur, l'attitude adoptée en général consiste à fixer un seuil sur la mesure de similarité : en dessous, le locuteur est rejeté, au dessus, le locuteur est accepté (comme étant celui dont l'identité est revendiquée). Toutefois, l'utilisation d'un seuil fixé, identique pour tous les locuteurs, conduit à des taux d'erreur variables en fonction du locuteur.

Dans une application d'identification du locuteur, la stratégie est assez simple puisqu'il s'agit d'évaluer la similarité des caractéristiques mesurées avec toutes les références correspondant à chacun des locuteurs autorisés. Le locuteur identifié est celui pour lequel la similarité est la plus grande. Notons que le coût de calcul de cette opération d'identification, ainsi que le volume des données qu'il est nécessaire de stocker, croissent linéairement avec la taille du groupe de locuteurs autorisés. La situation est plus complexe lorsqu'on a affaire à un

ensemble ouvert car il est, en plus, nécessaire de rejeter les locuteurs n'appartenant pas au groupe de locuteurs autorisés. En général, la démarche adoptée consiste à effectuer d'abord l'identification, puis à utiliser une stratégie de vérification pour rejeter les éventuels imposteurs en considérant que l'identité revendiquée est celle déterminée lors de la phase d'identification.

II-5-Evaluation des performances en reconnaissance du locuteur :

Les techniques de la reconnaissance de locuteur sont basées sur des mesures de ressemblance entre des enregistrements de paroles. Les mesures sont faites sur des paramètres acoustiques extraits par analyse du signal et prennent en compte les informations spécifiques au locuteur, le contenu du message vocal, les informations sur l'environnement et le matériel d'enregistrement.

Pour garantir un niveau de performance acceptable pour les applications de la reconnaissance de locuteur, plusieurs caractéristiques doivent être prises en compte:

- Les locuteurs n'essayent pas de déguiser leur voix.
- Les conditions d'enregistrement et du traitement du signal audio sont connues et/ou contrôlées.
- Des données de paroles, enregistrées dans des mêmes conditions que le signal de test, sont disponibles pour référencer un locuteur dans le système.

II-6-Les limites de la reconnaissance de locuteur (difficultés) :

D'une manière générale, l'amélioration des techniques utilisées permet maintenant d'envisager des conditions de fonctionnement de plus en plus proches des situations rencontrées en pratique (enregistrements de courte durée, variations du canal de transmission, présence de bruit de fond, nombre important de locuteurs autorisés,

La nature très variable du signal de parole est un facteur délicat à gérer.

La voix évolue avec l'âge, l'état physiologique ou pathologique du locuteur (maladie, émotions,...) et les facteurs socioculturels (le locuteur peut changer l'accent).

Le canal de transmission de l'information joue un rôle important en reconnaissance de locuteur; manque de connaissance car la technologie dans ce domaine n'est pas encore bien développée et plusieurs facteurs subjectifs telle que la prononciation, la vitesse de locution et la fréquence de pause, tout cela diffère d'une personne à une autre et encore pour la même personne sous différentes circonstances.

Sans coopération du locuteur, sans contrôle du matériel de prise de son, sans connaissance des conditions d'enregistrement, sans certitude qu'il ne s'agit pas d'une voix falsifiée et dans une moindre mesure sans contrôle du contenu linguistique des messages, il n'est pas possible de déterminer à quel point une ressemblance entre deux messages vocaux est due au locuteur et non à d'autres facteurs.

II-7- Conclusion :

Nous avons brièvement parcouru dans ce chapitre les méthodes utilisées pour extraire les paramètres acoustiques ainsi le système de reconnaissance vocale avec ces diverses constitutions pour cela on utilisant les réseaux de neurones qu'on va traiter précisément dans le chapitre suivant.

Chapitre III

Les réseaux de neurone

Les réseaux de neurones ont été développés pour résoudre des problèmes de contrôle, de reconnaissance de formes ou de mots, de décision, de mémorisation et sont considérés comme une alternative à l'intelligence artificielle.

L'idée fondamentale est consisté certains concepts de cerveau humain. Les réseaux de neurones servent aujourd'hui à toute sorte d'applications dans divers domaines.

III-1-Le neurone biologique :

Le neurone biologique est une cellule vivante spécialisée dans le traitement des signaux électriques et qui assure l'émission et la réception des signaux qu'eux même engendre.

III-1-2-Fonctionnement d'un neurone :

Le neurone possède des caractéristiques qui permettent d'assurer les cinq fonctions suivantes :

1. Recevoir les signaux issus des neurones qui leurs son voisins
2. Traiter ces signaux
3. Engendrer un flux nerveux
4. Assurer son cheminement
5. Le transmettre a un neurone voisin

III-1-3-Anatomie des neurones :

Les neurones sont reliés entre eux par des liaisons appelées axones. Ces axones vont eux même jouer un rôle important dans le comportement logique de l'ensemble. Ces axones conduisent les signaux électriques de la sortie d'un neurone vers l'entrée (synapse) d'un autre neurone.

Les neurones effectuent une sommation des signaux reçus en entrée et en fonction du résultat obtenu vont fournir un courant en sortie. **(Figure III-1)**

La structure d'un neurone se compose de trois parties :

- La somma : ou cellule d'activité nerveuse, au centre du neurone.
- L'axone : attaché au somma qui est électriquement actif, et qui conduit l'impulsion transmise par le neurone.
- Les dendrites : électriquement passives, elles reçoivent les impulsions des autres neurones.

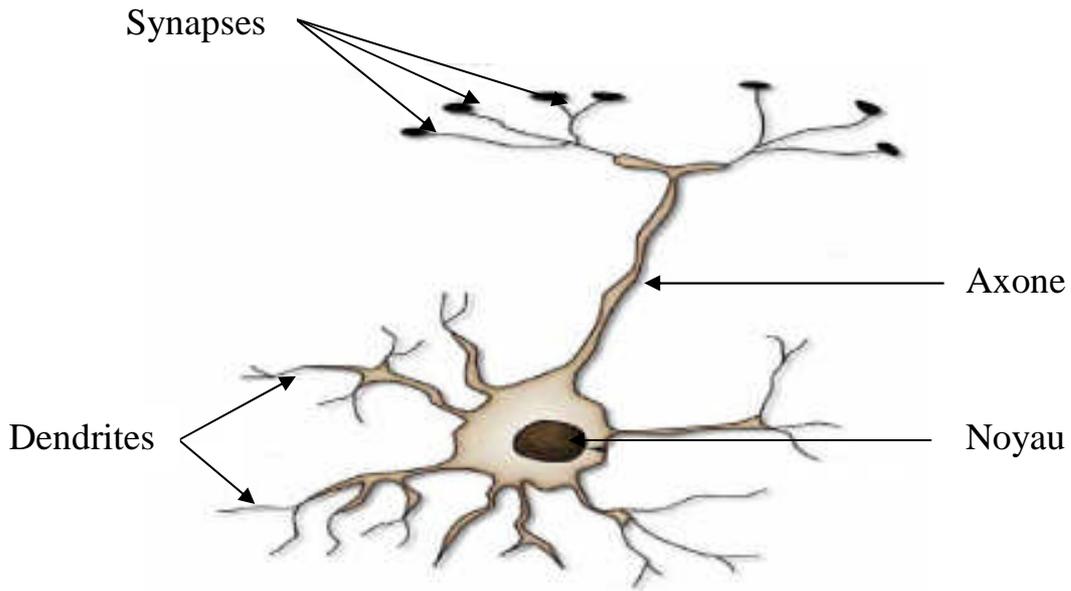


Figure III-1:Le neurone biologique

III-2-Le neurone formel (artificiel) :

Le neurone artificiel (ou cellule) est un processeur élémentaire. Il reçoit un nombre variable d'entrées en provenance de neurones appartenant à un niveau situé en amont. A chacune des entrées est associé un poids w représentatif de la force de la connexion. Chaque processeur élémentaire est doté d'une sortie unique, qui se ramifie ensuite pour alimenter un nombre variable de neurones appartenant à un niveau situé en aval (on parlera de neurones "avals"). A chaque connexion est associé un poids. (**Figure III-2**)

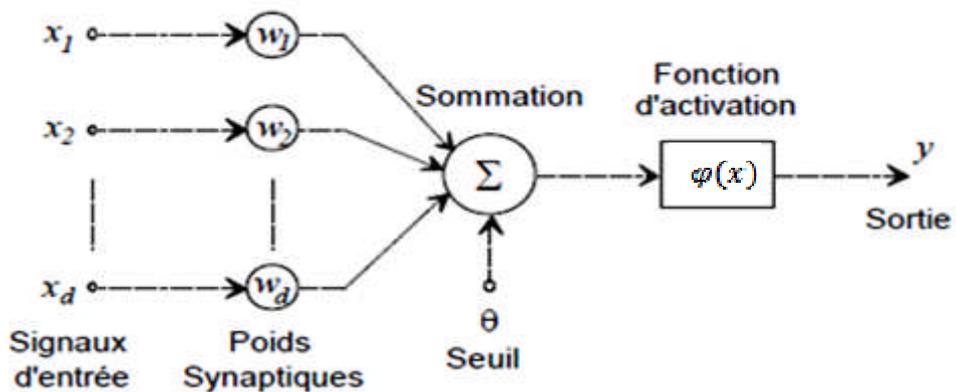


Figure III-2 : Le neurone formel

x_1, x_2, \dots, x_d : représente les vecteurs d'entrée du neurone.

w_1, w_2, \dots, w_d : les poids de ces vecteurs.

θ : le seuil du neurone.

$$y_i = \sum_{i=0}^{i=d} w_i x_i + \theta$$

$\varphi(x)$: la fonction d'activation du neurone (sigmoïde-logarithmique) dans notre exemple

$$\varphi(x) = 1 / (1 + \exp(-x))$$

III-2-1-Modélisation d'un neurone formel :

Les réseaux de neurones formels sont à l'origine d'une tentative de modélisation mathématique du cerveau humain.

La modélisation consiste à mettre en œuvre un système de réseau neuronal sous un aspect non pas biologique mais artificiel ; cela suppose que d'après le principe biologique on aura une correspondance pour chaque élément composant le neurone biologique, donc une modélisation pour chacun d'entre eux.

On pourra résumer cette modélisation par le (**Tableau III-1**), qui nous permettra de voir clairement la transition entre le neurone biologique et le neurone formel.

Neurone biologique	Neurone artificiel
Synapses	Poids de connexions
Axones	Signal de sortie
Dendrite	Signal d'entrée
Somma	Fonction d'activation

Tableau III.1 : La transition entre le neurone biologique et le neurone formel

III-3-Réseaux de neurones :

Concernant notre application, nous nous intéressons aux réseaux de neurones de type perceptron multicouche (multiplier perceptron), appropriés à l'apprentissage supervisé avec la

fonction d'activation de type sigmoïde. (La Figure III-3) présente un schéma synoptique d'un réseau de neurones.

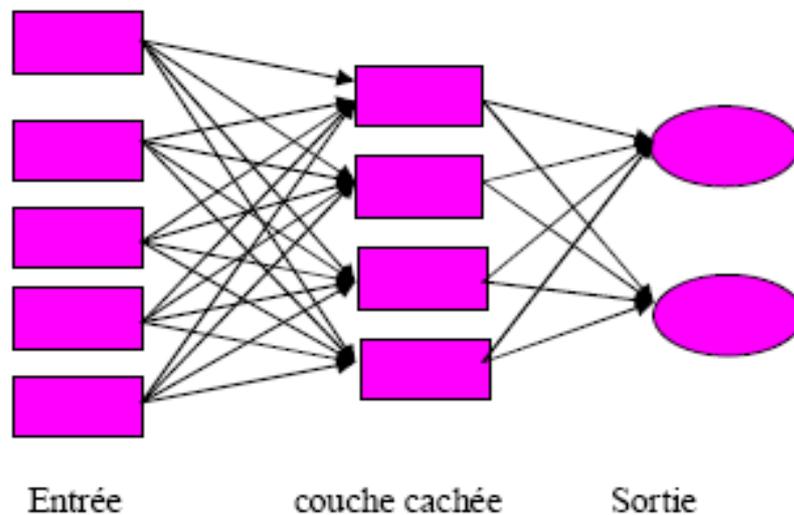


Figure III-3 :Schéma synoptique d'un réseau de neurone

Le réseau de neurone se compose donc de trois couches :

III-3-1-La couche d'entrée :

Elle reçoit les données que l'on veut traiter. Sa taille est donc directement déterminée par le nombre de variables d'entrées.

III-3-2-La couche cachée :

La seconde couche est la couche cachée, dans le sens qu'elle n'a pas de contact direct avec l'extérieur. Les fonctions d'activation sont en général non linéaires. Le choix de sa taille (nombre de neurones) doit être ajusté en fonction de l'application envisagée.

III-3-3-La couche de sortie :

La dernière couche est appelée couche de sortie ; elle donne le résultat obtenu après introduction des données d'entrée dans la première couche cachée.

La taille de cette couche est directement déterminée par le nombre de formes (classe) désirés.

III-4-La fonction d'activation :

La fonction d'activation du neurone définit son état interne en fonction de son entrée totale. Le comportement d'un neurone dépend essentiellement du choix de sa fonction d'activation.

Les fonctions d'activation utilisées généralement sont suivantes :

type	Equation	Allure
Fonctions Heaviside	$h(x) = \begin{cases} 1 & \text{si } x \geq 0 \\ 0 & \text{sinon} \end{cases}$	
Signe	$\text{Sgr}(x) = \begin{cases} +1 & \text{si } x \geq 0 \\ -1 & \text{sinon} \end{cases}$	
Fonction linéaire	$F(x) = x$	
Fonction linéaire a seuil	$F(x) = \begin{cases} x & \text{si } x \in [u, v] \\ v & \text{si } x \geq v \\ u & \text{si } x \leq u \end{cases}$	
Fonction sigmoïde	$f(x) = \frac{1}{1 + e^{-x}}$	

Tableau III-2 : Différents types de fonction d'activation

III-5-Apprentissage :

L'apprentissage est une phase du développement du réseau de neurones durant laquelle on calcule les poids des neurones de telle manière à ce que les sorties du réseau soient aussi proches que possible des sorties désirées. L'apprentissage fait appel à des exemples de comportement du processus de classification.

L'apprentissage est dit supervisé lorsque les exemples sont constitués de couples de valeurs de type :(Valeur d'entrée, valeur de sortie désirée).

Tout le problème d'apprentissage supervisé consiste à faire rapprocher la sortie réelle de la sortie désirée, donc déterminer le vecteur des poids des neurones capables de produire le même vecteur de sortie pour même vecteur d'entrée.

L'apprentissage est qualifié de non supervisé lorsque seules les valeurs d'entrée sont disponibles.

Dans ce cas les exemples présentés à l'entrée provoquent une auto adaptation du réseau afin de produire les valeurs de sortie qui soient proche, en réponse, pour des valeurs d'entrées similaires.

Durant l'apprentissage, les poids sont ajustés itérativement pour minimiser la fonction de performance mesurant l'écart entre les sorties effectives S et les sorties désirées D ; l'erreur la plus couramment employée est l'erreur quadratique.

$$E_p = \frac{1}{2} \sum_{k=1}^m [D_k - Z_k]^2$$

(Z_k, D_k) sont successivement la sortie réelle et sortie désirée de la couche K .

Dans ce qui suit, nous allons décrire l'algorithme d'apprentissage, qui utilise le gradient de la fonction de performance pour minimiser l'erreur.

Le problème pour une ou plusieurs couches cachées est d'évaluer les valeurs optimales des poids de connexion.

La solution est de rétro propagation d'erreur proportionnellement aux poids.

III-6-Le perceptron :

Historiquement le premier réseau de neurone, est le perceptron monocouche de Rosenblatt. C'est un réseau simple, puisque il ne se compose que d'une couche d'entrée et d'une couche de sortie. Il est calqué, à la base, sur le système visuelle et de ce fait a été conçu dans un but premier de reconnaissance des formes. Cependant, il peut aussi être utilisée pour faire de la classification et pour résoudre des opérations logique simple (telle "ET" ou "OU"). Sa principale limite est qu'il ne peut résoudre que des problèmes linéairement séparables. Il suit généralement un apprentissage supervisé selon la règle de correction de l'erreur.

Le perceptron multicouche est une extension du précédent, avec une ou plusieurs couches cachées entre l'entrée et la sortie. Chaque neurone dans une couche est connecté à tous les neurones de la couche précédente et de la couche suivante (excepté pour les couches d'entrée et de sortie) .Il n'y a pas de connexions entre les cellules d'une même couche. Les fonctions d'activation utilisées dans ce type de réseau sont principalement les fonctions à seuil ou sigmoïdes. Cela permet de résoudre des problèmes non linéairement séparables et des problèmes logiques plus compliqués, et, notamment le fameux problème du XOR.il suit aussi un apprentissage supervisé selon la règle de correction de l'erreur.

III-7-Les étapes de fonctionnement d'un réseau de neurone :

Le cycle classique de développement peut être séparé en six étapes :

III-7-1-Collecte des données :

L'objectif de cette étape est de recueillir les données, à la fois pour développer le réseau de neurones et pour le tester. Dans le cas d'application sur des données réelles, l'objectif est de rassembler un nombre de données suffisant pour constituer une base représentative des données susceptibles d'intervenir en phase d'utilisation du système neuronal.

III-7-2-Analyse des données :

Il est souvent préférable d'effectuer une analyse des données de manière à déterminer les caractéristiques discriminantes pour détecter ou différencier ces données.

Ces caractéristiques constituent l'entrée du réseau de neurone.

Il est généralement nécessaire de traiter les données afin d'extraire les caractéristique représentatives. Cette étape de prétraitement permet de réduire la taille du réseau (réduire le

nombre de neurone en entrée), diminuer le temps de calcul et améliorer les performances de classification (augmenter le pouvoir de séparation inter classe)

III-7-3-Séparation des données :

Afin de développer une application à base de réseau de neurone, il est nécessaire de disposer de deux groupes de données un groupe d'apprentissage ou de référence un autre pour le test ou la généralisation.

De façon générale, nous utilisons 75% de l'ensemble des données pour l'apprentissage 25% pour le test.

III-7-4-Choix du réseau de neurones :

Il existe un grand nombre de types de réseaux de neurones, avec pour chacun, des avantages et des inconvénients. Le choix d'un réseau peut dépendre :

- 1) De la tâche à effectuer (classification, association, contrôle de processus, ...)
- 2) De la nature des données.

III-8-Algorithmes de la retro propagation :

L'algorithme de la retro propagation du gradient de l'erreur est certainement à la base du premier succès des réseaux de neurone. Il figure aujourd'hui parmi les algorithmes d'apprentissage les plus utilisés.

Leur principe consiste à ajuster les poids dans le sens inverse du gradient de la fonction de performance comme suite :

$$\Delta w_{j,i} = - \frac{\partial E}{\partial w_{j,i}}$$

E : présente l'erreur entre l'entrée et la sortie.

Pour avoir un bon fonctionnement de l'algorithme de retro propagation, il a lieu d'introduire deux paramètres de plus, qui sont :

Le pas d'apprentissage a :

Qui sert à faire converger les poids vers une solution, en effectuant des sauts discrets pour les poids. Il prend des valeurs comprises dans l'intervalle [0.1, 1]

Le moment (θ) (coefficient d'inertie) :

Il sert à réduire le temps d'apprentissage dans le cas où on utilise un grand pas d'apprentissage. Il prend des valeurs comprises dans l'intervalle [0.6, 0.9].

Les étapes de l'algorithme de rétro-propagation sont les suivantes :

1. Appliquer un vecteur d'entrée aux nœuds d'entrées puis initialiser les poids du réseau.
2. Appliquer à l'entrée du réseau les données permettant l'apprentissage.
3. Calculer les termes d'erreur de signal de la couche de sortie et les couches cachées
4. Mettre à jour les poids de la couche de sortie et couches cachées
5. Répéter ce processus jusqu'à ce que l'erreur E_p devienne acceptable.

III-9-Les applications des réseaux de neurones :

L'intérêt porté aujourd'hui aux réseaux de neurones tient sa justification dans les propriétés intéressantes qu'ils possèdent et qui devraient permettre de dépasser les limites de l'informatique traditionnelle ; pour cette raison, les réseaux de neurones artificiels ont trouvé plusieurs applications et plusieurs domaines, parmi lesquels nous pouvons citer:

III-9-1- L'application à la reconnaissance des formes :

L'application des réseaux neurones pour la reconnaissance des formes a été largement utilisée. Les réseaux bouclés sont intensivement étudiés et utilisés pour la reconnaissance des voix et des images.

III-9-2-L'application à la planification :

L'optimisation est souvent sollicitée pour la planification des actions ou des tâches.

III-9-3-L'application à l'identification :

Les réseaux de neurones, comme nous l'avons vu ci-dessus, offrent des avantages intéressants pour l'approximation de fonctions sur la base d'exemple. Il est, des lors, naturel de penser à appliquer ces techniques à l'identification des processus, surtout pour les phénomènes non linéaires complexes. L'emploi de réseaux de neuronaux à l'identification de processus présente un intérêt évident de commande.

III-9-4-L'application à la commande :

L'application des réseaux de neurones à la commande de processus offre les avantages suivants :

- Le traitement parallèle et distribué des informations servant à la commande,
- L'absence totale des restrictions sur le non linéarité du processus,
- La possibilité de débiter la conduite avec un minimum d'informations à priori sur le processus,
- La robustesse par rapport aux bruit et aux défaillances internes.

III-9-5-L'application au diagnostic :

Les réseaux de neurones sont bien adaptés à la résolution des problèmes de diagnostic et de la classification automatique. Dans ce contexte on distingue plusieurs applications des réseaux de neurones pour le diagnostic des défaillances.

Conclusion :

Ce chapitre a été consacré à la présentation de la théorie de réseaux de neurone en général. Nous allons nous intéresser à l'application de cet opérateur à la reconnaissance du locuteur dans le cas d'un signal de parole.

Chapitre IV

Application à la reconnaissance du locuteur

Nous avons présenté, dans les chapitres précédents, notre approche de classification ainsi que les prétraitements effectués sur notre base de données. Nous allons maintenant aborder l'aspect pratique par une implémentation de la méthode des réseaux de neurones pour la reconnaissance automatique du locuteur en mode dépendant du texte (les locuteurs prononcent le même message). Nous évaluerons la validité de notre système par une analyse des résultats obtenus. Pour cela, il faut décrire d'abord toutes les conditions nécessaires pour effectuer les enregistrements et le matériel utilisé, puis calculer les différents paramètres (le taux de passage par zéro, la fréquence fondamentale et la bande passante) qui constitueront les paramètres d'entrées du réseau.

IV-1-Description et acquisition de la base de données :

Nous disposons d'une base de données composée de cinq (05) locuteurs qui sont : Lynda, Nora, Thanina, Ghani et Samir, chacun prononçant la même phrase «Il n'est jamais trop tard pour bien faire ».

Avant d'expliquer les étapes d'enregistrement et les différentes phases de reconnaissance des voix de ces locuteurs, on décrit d'abord les conditions dans lesquelles ont été effectuées les enregistrements :

- L'identification du locuteur se fait en ensemble fermée (nombre de locuteurs limité).
- La nature du message vocal est en mode dépendant du texte c'est-à-dire que les personnes prononcent la même phrase.
- Les enregistrements des signaux se font dans une salle calme, afin d'éviter le maximum possible le bruit pouvant altérer nos signaux.

Dans notre travail on a utilisé un micro-ordinateur portable et un microphone numérique (casque). Il suffit de raccorder le microphone à l'entrée « Mic » de la carte son situé à côté de l'ordinateur à l'aide des fiches numériques qui permettent la conservation fidèle du son numérique lors de son transfert d'un appareil à l'autre.

Le logiciel utilisé est Matlab (Matrix Laboratory) qui est un système interactif de calcul numérique et de visualisation graphique. Il possède un langage de programmation puissant et simple à utiliser, précis, robuste et rapide. Le logiciel Matlab est basé sur le calcul matriciel numérique.

IV-2-Procédure d'enregistrement :

Après avoir réalisé toutes les conditions nécessaires à l'aide d'un outil de Windows (magnétophone), on commence l'enregistrement : le locuteur prononce la phrase, et dès qu'il fini, on clique sur « stop », puis on choisit le format pour sauvegarder le son sous la fenêtre "File", et en format "Wave" (format sous Windows). La même procédure est effectuée pour l'enregistrement de chaque locuteur.

Une fois, les enregistrements effectués, nous devons les traiter. On transfère tous les signaux enregistrés vers le work de Matlab. On importe le fichier (son.wav) à l'aide de la commande `waveread ('son.wav')`, on obtient par exemple la (**figure IV-1**).

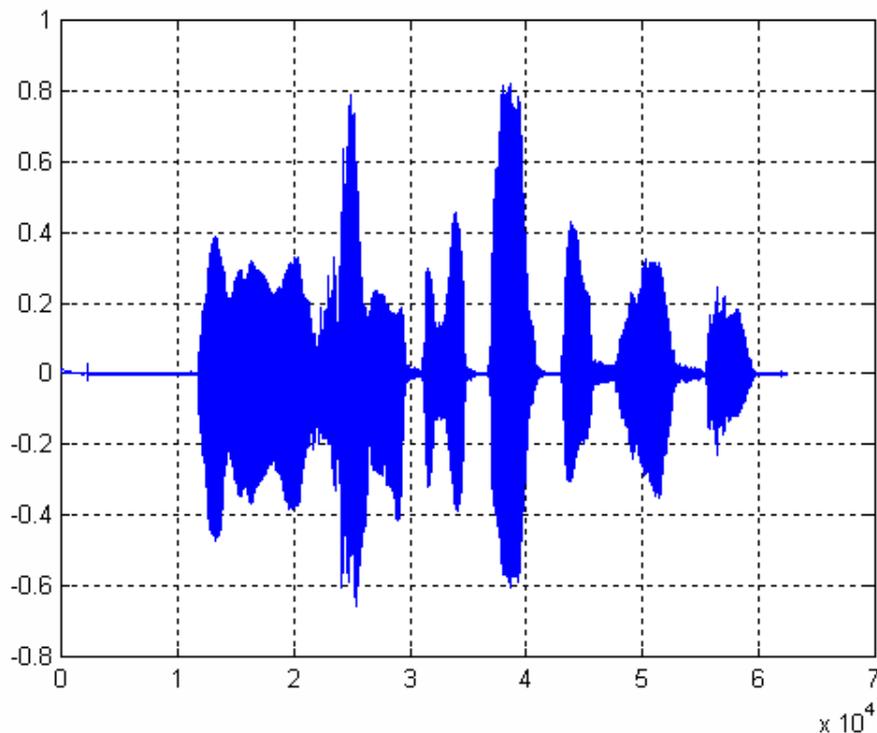


Figure IV-1 : La représentation temporelle du signal de parole «Il n'est jamais trop tard pour bien faire » prononcée par Lynda.

L'étape suivante consiste à extraire les paramètres caractéristiques des locuteurs. Dans notre cas nous optons pour la fréquence fondamentale, la bande passante de chaque signal et le nombre de passage par zéro de chaque son du locuteur 'Nbi' ou bien du taux de passage par zéro, TPZ, N étant le nombre de points du signal.

IV-3-Extraction des paramètres caractéristiques des locuteurs :

Une fois que tous les enregistrements sont présentés sous forme de signaux, on applique un programme qui permet le calcul du nombre de passage par zéro, et la transformée de Fourier pour avoir le spectre d'amplitude de laquelle on va extraire les composantes fréquentielles nécessaires à notre application. Pour avoir la fréquence fondamentale il suffit de lire la valeur de fréquence correspondante au pic qui est caractérisée par l'amplitude maximale ; la bande passante représente la largeur qui contient toutes les informations de la phrase prononcée ; à titre d'exemple, toujours par rapport au locuteur précédent, le spectre de fréquence du signal correspondant et représenté (**Figure IV-II**) et on peut en déduire :

La fréquence fondamentale égale : 740Hz et la bande passante qui de 1999Hz.

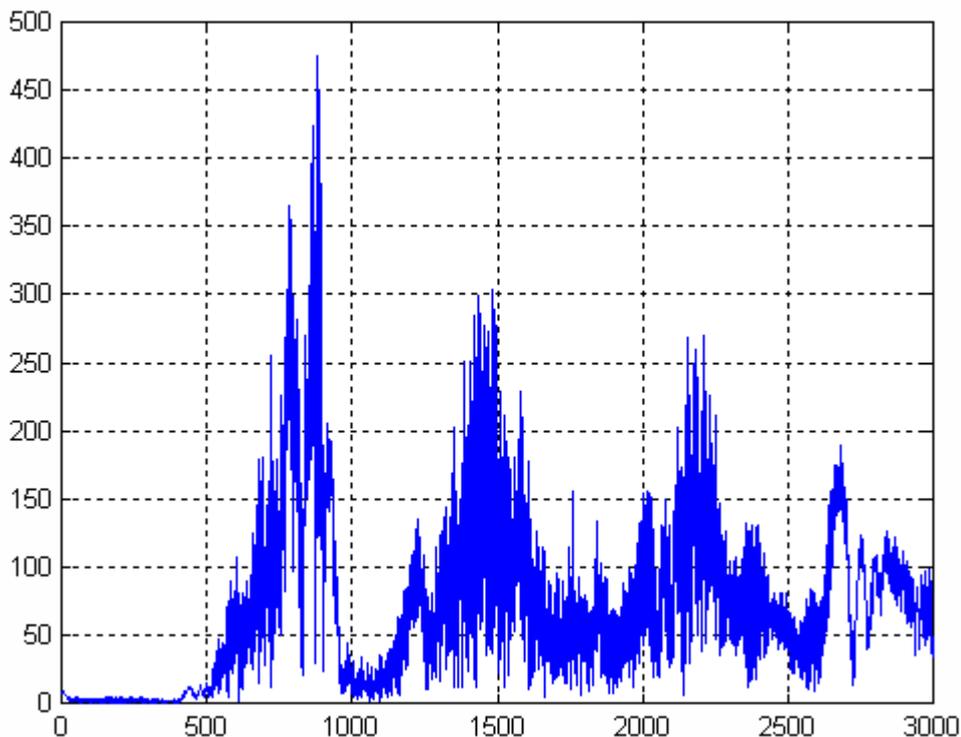
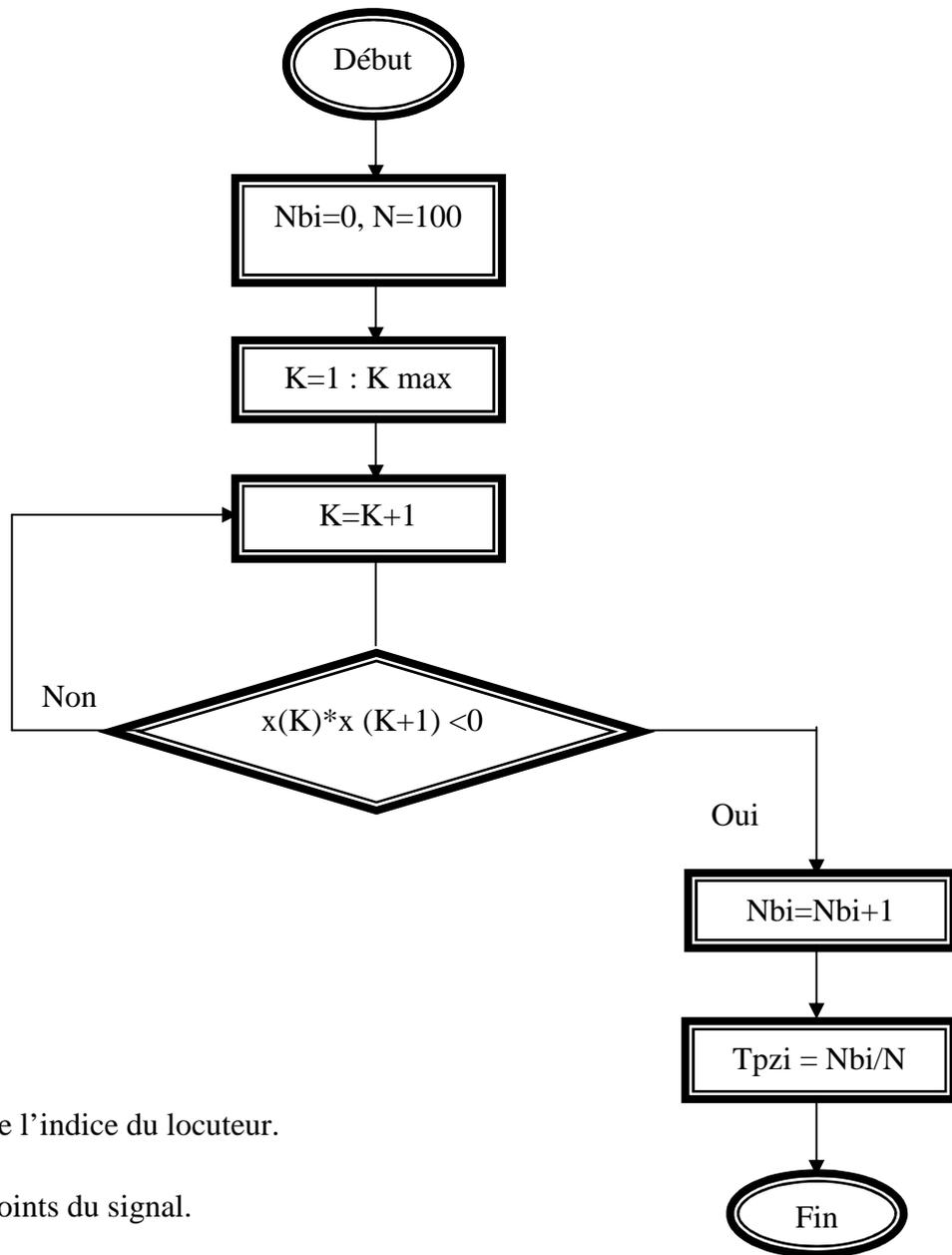


Figure IV-II : Le spectre d'amplitude

Le taux de passage par zéro est de 0.0775

- L'organigramme de calcul du taux de passage par zéro de chaque signal est le suivant :



$i= 1 :5$ représente l'indice du locuteur.

K : nombre de points du signal.

Figure IV-3 : L'organigramme du programme qui calcul le taux de passage par zéro.

De cette façon on détermine tous les paramètres caractérisant les autres locuteurs et on obtient les résultats suivants (**Tableau IV-1**) :

Locuteurs	Fréquence fondamentale Hz	Bande passante Hz	Nombre de points de passage par zéro	Le taux de passage par zéro (TPZ)
Nora	680	2210	112	0.0560
Lynda	740	1999	29	0.0775
Ghani	550	3000	11	0.0055
Thanina	800	3500	7	0.0035
Samir	750	2250	64	0.0325

Tableau IV-1 : Paramètres caractérisant des locuteurs.

IV-4- Choix de l'architecture du réseau de neurones :

Le problème à résoudre maintenant porte sur la détermination de la structure du réseau (nombre de couche, nombre de neurones de la couche cachée et nature des fonctions d'activation).

La structure du réseau dépend de la nature du problème et notamment du nombre des variables d'entrée et de sortie ; il faut d'abord choisir le type de réseau, le nombre de couches et le nombre de neurones dans chaque couche, et tester toutes les possibilités pour choisir celle qui offre la structure la plus efficace.

IV-4-1- Choix des variables d'entrée :

La phase la plus importante de l'étude d'un système est le choix des variables (paramètres) prises en compte pour l'analyse, et qui conditionnent les résultats obtenus. Ces composantes avant tout, de l'expérience(ou du savoir faire) de l'expert. Ce sont des variables correspondant, en général, à l'information délivrée par un capteur mesurant une grandeur physique (ondes acoustiques dans notre cas).

Dans notre application nous avons utilisé des variables indicatrices, qui sont le spectre fréquentielle, la fréquence fondamentale et le taux de passage par zéro du signal de parole.

IV-4-2- Choix de la couche intermédiaires :

En général, l'expérience a montré qu'un réseau à une seule couche cachée doit comporter au moins autant mais, souvent, beaucoup plus de neurones que la couche d'entrée pour être efficace, c'est-à-dire, présenter un caractère de généralisation.

Si le nombre de neurones dans la couche d'entrée est important, il est plus commode et plus efficace de prévoir deux couches cachées avec des nombres plus restreints de neurones.

IV-4-3-Choix des sorties :

Nous avons décidé d'attribuer un vecteur à chaque locuteur, comprenant les différents paramètres acoustiques, c'est-à-dire que chaque code de sortie doit correspondre à un locuteur correspond à ces caractéristiques.

Comme nous avons cinq classes à détecter et classifier, nous avons décidé d'utiliser cinq neurones dans la couche de sortie pour pouvoir coder toutes les entrées.

Lors de la détection les paramètres du locuteur, le réseau doit indiquer un code correspond à ce locuteur.les sorties désirées du réseau sont représentées dans le (**Tableau IV-2**), chaque locuteur à son code associé.

Les sorties du réseau	Le code
Nora	00001
Lynda	00010
Ghani	00100
Thanina	01000
Samir	10000

Tableau IV-2 : Codage de locuteurs

IV-5- La base d'apprentissage :

L'observation des paroles sous forme de spectres permet de construire une base de données (base d'apprentissage) à partir de laquelle le réseau de neurones pourra apprendre.

Celle-ci peut toujours se mettre sous forme de fichier ou de tableaux (matrice)

IV-6- Normalisation des données

Comme dans tout problème de classification, il est nécessaire d'introduire un facteur d'échelle sinon le simple fait de changer d'unité de mesure modifiera les résultats et la classification s'effectuerait alors uniquement sur les paramètres qui auraient la plus grande variabilité.

La normalisation qui ramène toutes les valeurs entre 0 et 1 est :

IV-7-Apprentissage du réseau :

Nous avons développé notre application sous environnement MATLAB version 5.3. L'algorithme utilisé pour l'apprentissage est celui de la retro-propagation du gradient.

La fonction d'activation utilisée dans notre application est celle qui est donnée par

$$a = 1 / (1 + \exp(-x))$$

Dans cette étape d'apprentissage nous avons mené plusieurs campagnes de mesures ayant pour variantes :

- Nombre de neurones de la couche cachée.
- Coefficient d'apprentissage.
- Nombre d'itération.
- Erreur désirée.

IV-8-Evaluation des performances du réseau :

IV-8-1-Diagramme de l'erreur :

C'est une représentation de l'erreur quadratique en fonction du nombre d'itérations, calculée dans la phase d'apprentissage. Le calcul de l'erreur se fait après propagation du vecteur de l'entrée vers la sortie. La sortie obtenue (réelle) est souvent différant de la sortie désirée. Dans ce cas, cette erreur est retro-propagée, c'est-à-dire propager la sortie vers l'entrée en recalculant les vecteurs des poids synoptiques (mise à jour des poids). Cette procédure est itérée plusieurs fois. A chaque introduction d'un vecteur, l'erreur est calculée et comparée à la valeur précédente. Le but est de minimiser cette erreur par conséquent, la convergence de l'algorithme est assurée par un nombre très grand d'itérations afin d'éviter un minimum local. Il est évident que dans le cas idéal, nous obtenons une courbe qui décroît rapidement vers une valeur très proche de zéro.

IV-8-2- L'influence du nombre de neurone de la couche cachée :

Cas 1 :

- Nombre de neurone dans la couche cachée égal à 4
- Coefficient d'apprentissage égal à 0,8
- Le nombre d'itération 80
- L'erreur désirée égale à $1.e^{-3}$

La (**figure IV-5**) montre la variation de l'erreur quadratique moyenne en fonction du nombre d'itérations.

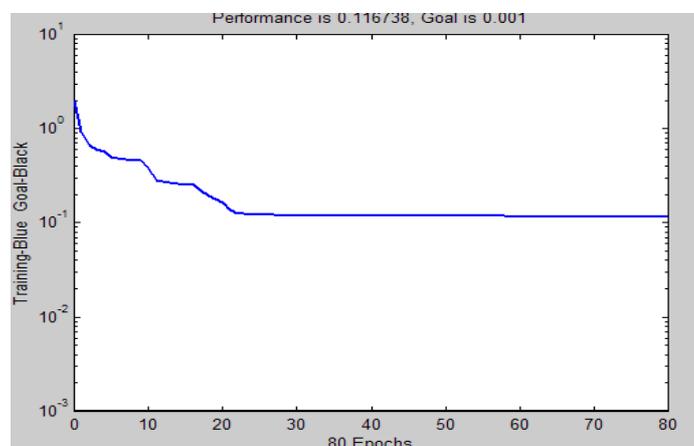


Figure IV-5 :L'erreur quadratique moyenne du réseau.

D'après la figure on remarque que l'erreur quadratique moyenne du réseau diverge de l'objectif prédéfini d'où la performance de l'apprentissage est très faible, cela montre que 4 neurones dans la couche cachée sont insuffisants pour atteindre l'erreur désirée.

Cas 2

- Nombre de neurone dans la couche cachée égal à 15
- Coefficient d'apprentissage égal à 0,8
- Le nombre d'itération 80
- L'erreur désirée égale à $1.e^{-3}$

La (figure IV-6) montre la variation de l'erreur quadratique moyenne en fonction du nombre d'itérations.

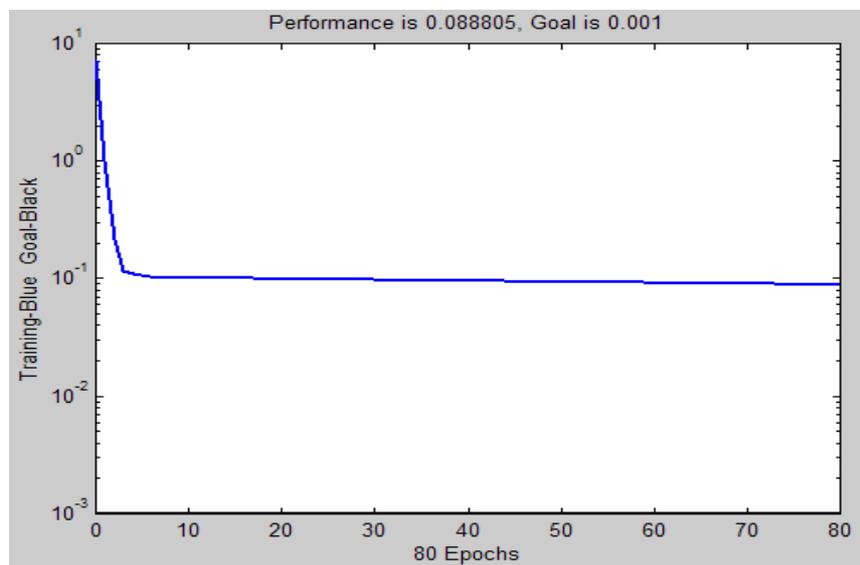


Figure IV-6 :L'erreur quadratique moyenne du réseau.

D'après la figure on remarque que l'erreur quadratique moyenne du réseau n'atteint pas l'objectif prédéfini on va alors agir sur un autre paramètre qu'est le coefficient d'apprentissage.

IV-8-3-Influence du coefficient d'apprentissage :

Cas 1

- Nombre de neurone dans la couche cachée égal à 15
- Coefficient d'apprentissage égal à 0,3
- Le nombre d'itération 80
- L'erreur désirée égale à $1.e^{-3}$

La (figure IV-8) montre la variation de l'erreur quadratique moyenne en fonction du nombre d'itérations.

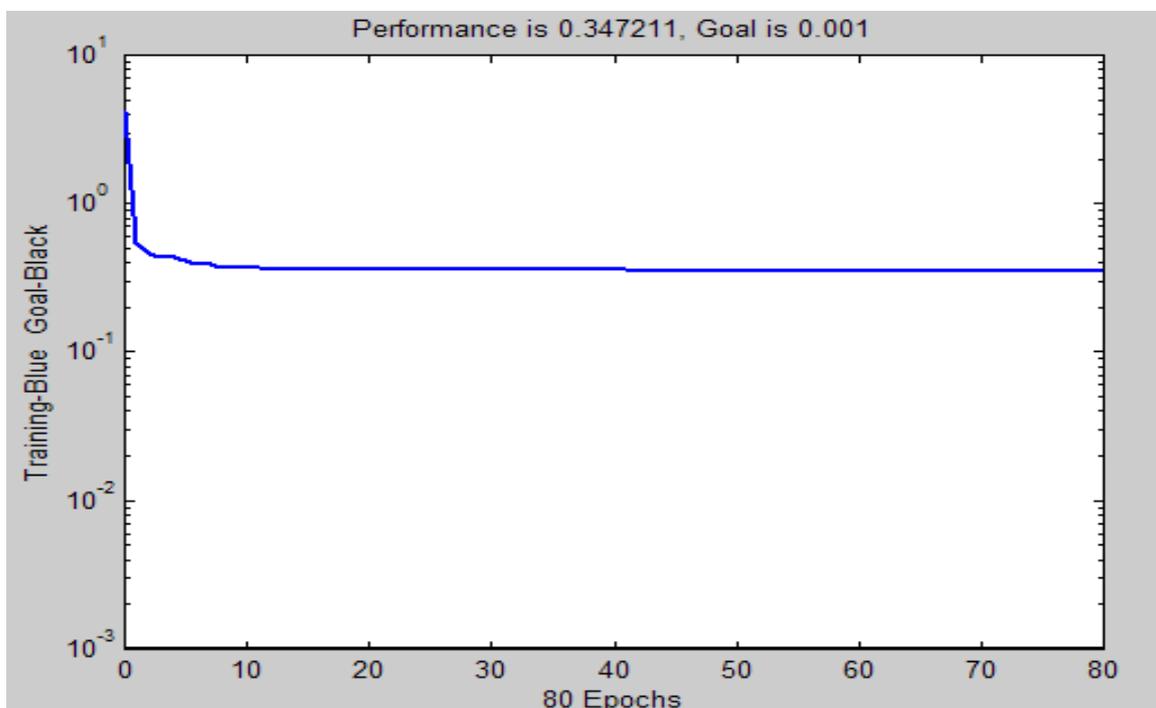


Figure IV-8 :L'erreur quadratique moyenne du réseau.

D'après la figure on remarque que l'erreur quadratique moyenne du réseau diverge de l'objectif prédéfini, cela montre que 0.3 pour coefficient d'apprentissage est une valeur insuffisante pour atteindre l'erreur désirée.

Cas 2

- Nombre de neurone dans la couche cachée égal à 15
- Coefficient d'apprentissage égal à 0,5
- Le nombre d'itération 80
- L'erreur désirée égale à $1.e^{-3}$

La (**figure IV-9**) montre la variation de l'erreur quadratique moyenne en fonction du nombre d'itérations.

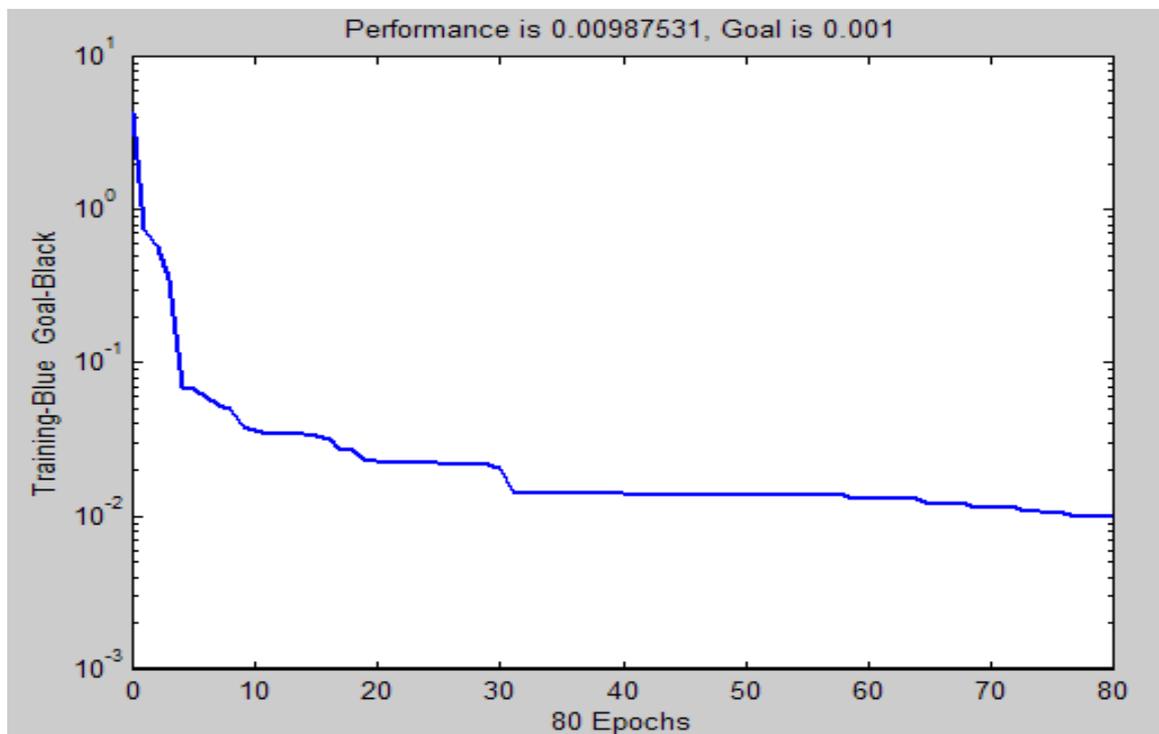


Figure IV-9 :L'erreur quadratique moyenne du réseau.

D'après la figure on remarque que l'erreur quadratique moyenne du réseau n'atteint pas l'objectif prédéfini on va agir alors sur le nombre d'itération.

IV-8-4-Influence de nombre d'itération :

Cas1

- Nombre de neurone dans la couche cachée égal à 15
- Coefficient d'apprentissage égal à 0,5
- Le nombre d'itération 100
- L'erreur désirée égale à $1.e^{-3}$

La (figure IV-10) montre la variation de l'erreur quadratique moyenne en fonction du nombre d'itérations.

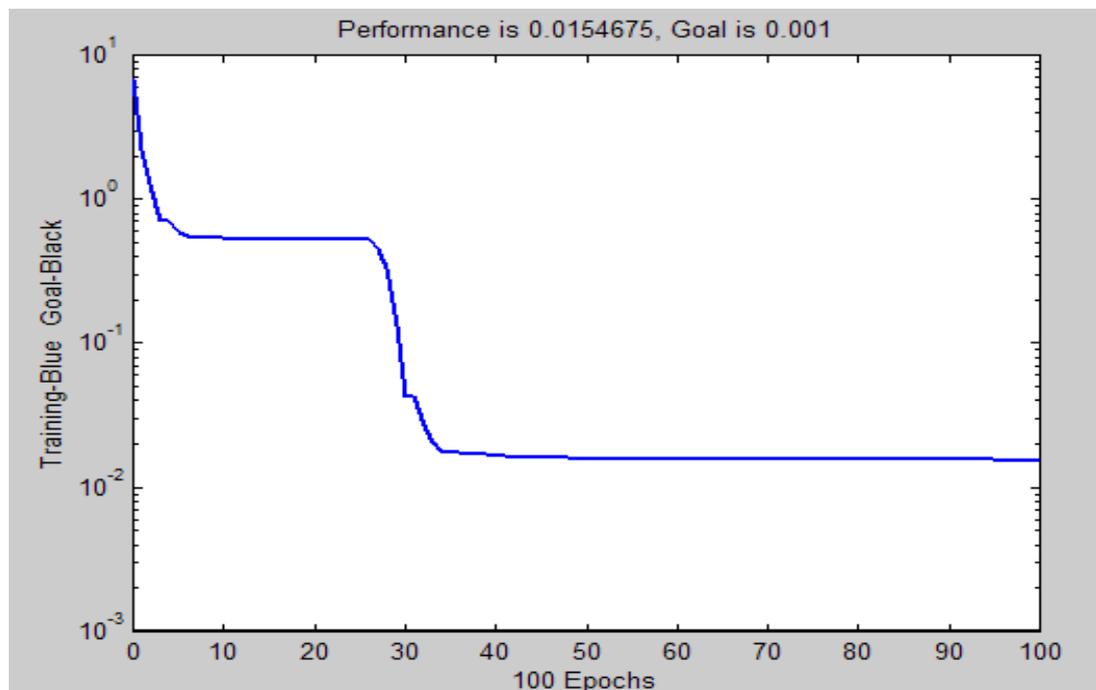


Figure IV-10 :L'erreur quadratique moyenne du réseau.

D'après la figure on remarque que l'erreur quadratique moyenne du réseau diverge de l'objectif prédéfini, cela montre que 100 itérations sont insuffisantes pour atteindre l'erreur désirée.

Cas 2

- Nombre de neurone dans la couche cachée égal à 15
- Coefficient d'apprentissage égal à 0,5
- Le nombre d'itération 200
- L'erreur désirée égale à $1.e^{-3}$

La (**figure IV-11**) montre la variation de l'erreur quadratique moyenne en fonction du nombre d'itérations.

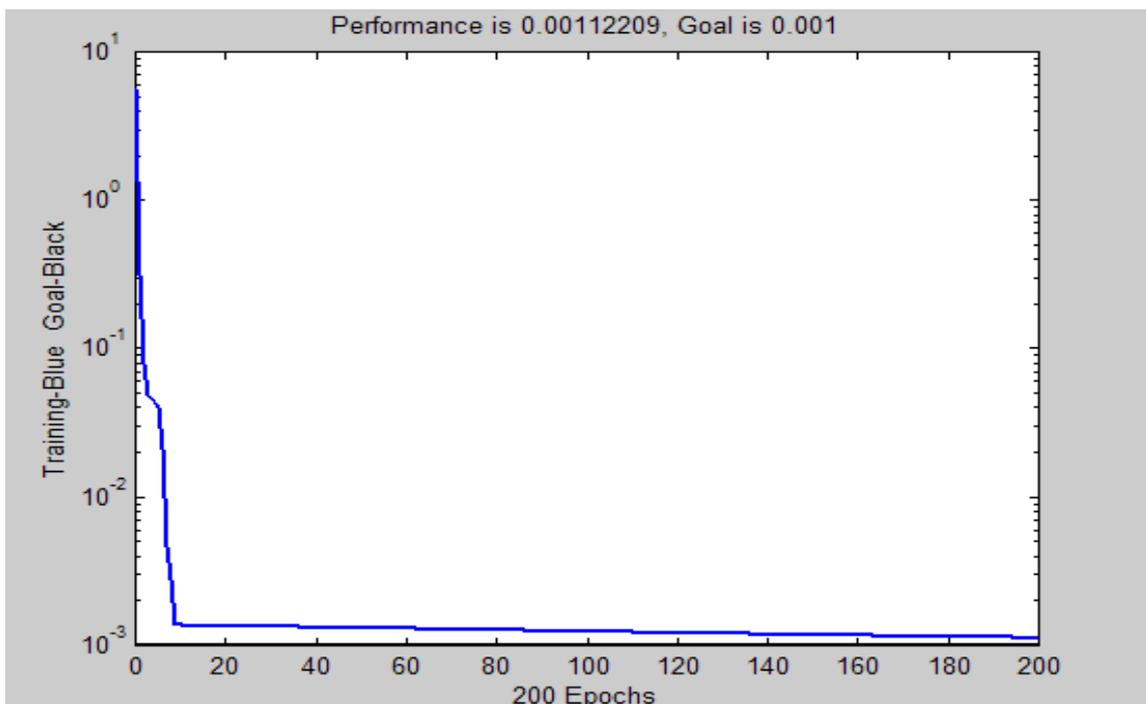


Figure IV-11 :L'erreur quadratique moyenne du réseau.

D'après la figure on remarque que l'apprentissage est achevé après 200 itérations, et il a atteint l'erreur quadratique moyenne désirée qu'on a fixée à 10^{-3} ce qui signifie que l'apprentissage a atteint ces meilleures performances.

IV-8-5-Influence L'erreur désirée :

Cas 1

- Nombre de neurone dans la couche cachée égal à 15
- Coefficient d'apprentissage égal à 0,5
- Le nombre d'itération 200
- L'erreur désirée égale à $1.e^{-5}$

La (figure IV-12) montre la variation de l'erreur quadratique moyenne en fonction du nombre d'itérations.

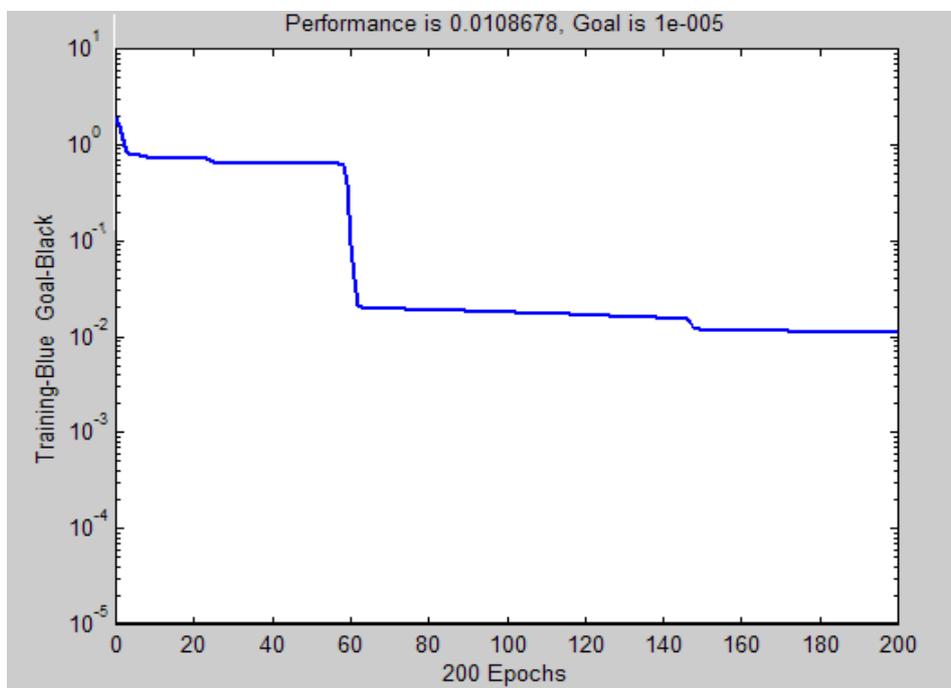


Figure IV-12:-L'erreur quadratique moyenne du réseau.

D'après la figure on remarque que l'erreur quadratique moyenne du réseau n'atteint pas l'objectif prédéfini.

Cas 2 :

- Nombre de neurone dans la couche cachée égal à 25
- Coefficient d'apprentissage égal à 0,5
- Le nombre d'itération 200
- L'erreur désirée égale à $1.e^{-6}$

La (**figure IV-13**) montre la variation de l'erreur quadratique moyenne en fonction du nombre d'itérations.

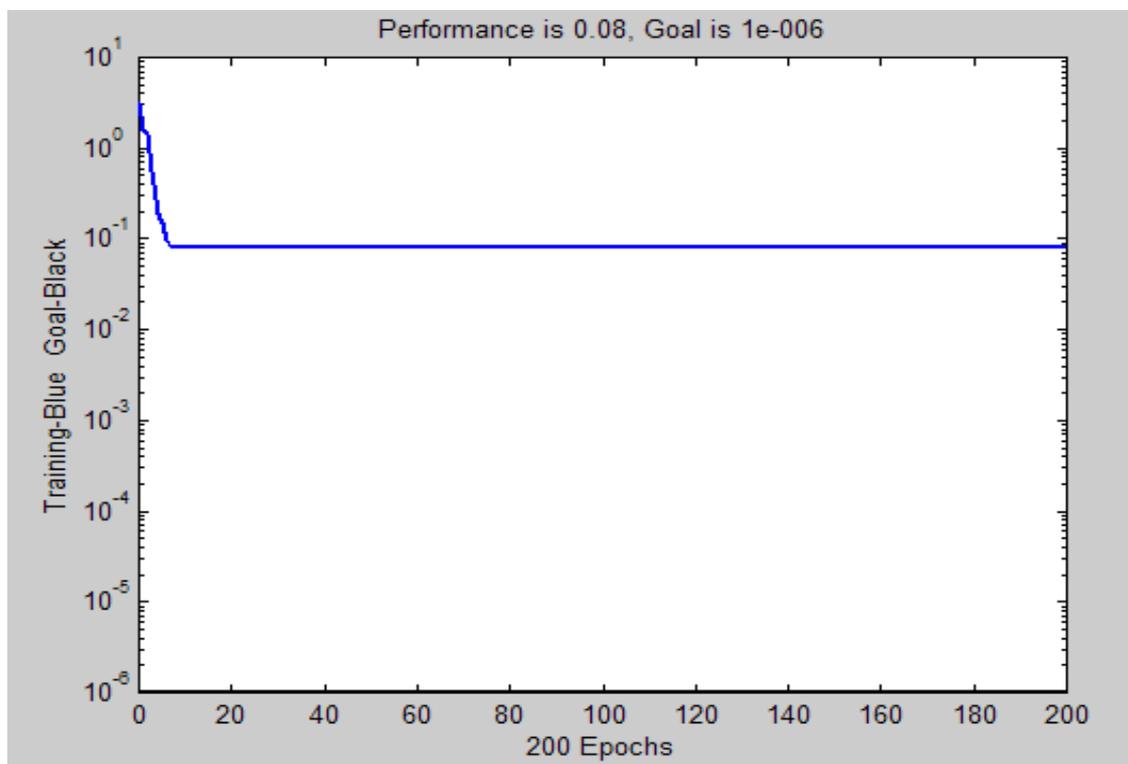


Figure IV-13:-L'erreur quadratique moyenne du réseau.

D'après la figure on remarque que l'erreur quadratique moyenne du réseau n'atteint pas l'objectif prédéfini.

D'après les figures obtenues, nous remarquons qu'avec un pas d'apprentissage égal à 0.5 le réseau atteint l'erreur quadratique moyenne prédéfini après 200 itérations avec des performances importantes (0.0154675).

IV-9-Résultats de l'application du réseau de neurones :

Le choix de réseau étant fait, on exécute l'opération d'apprentissage. Le programme d'apprentissage s'arrête sous l'une des conditions suivantes:

-l'erreur est inférieure à $1.e^{-3}$.

-le nombre d'itération (epochs) égale à 200.

Les résultats de l'apprentissage du réseau sont donnés sous forme de matrice, elle contient 5 lignes suivant le nombre des neurones de la couche de sortie et 5 colonnes suivant le nombre des signaux d'entrée qui sont caractérisés par trois paramètres (la fréquence fondamentale, la bande passante et le taux de passage par zéro).

La lecture des résultats se fait par ligne. Par exemple ; la première colonne correspond à la première personne dont sa sortie désirée est de 00001 (voir le tableau des sorties désirées).

La sortie du réseau est :

a =

-0.0000	-0.0001	-0.0000	-0.0000	1.0000
-0.0000	-0.0000	0.0000	0.9999	0.0000
0.0000	0.0001	1.0000	-0.0000	0.0000
0.0000	1.0000	-0.0001	-0.0000	-0.0000
0.9999	-0.0000	0.0000	0.0000	0.0000

➤ **Remarque : amélioration de la base de données :**

Dans le cas où chaque locuteur de notre base de données (quatre locuteurs) prononce plusieurs fois le même message vocal (par exemple quatre fois), et avec trois paramètres caractéristiques chacun, donc l'entrée de réseau de neurone est une matrice de quatre lignes et de seize colonnes, dans ce cas La sortie du réseau est :

a =

Columns 1 through 8

0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
-0.0000	-0.0000	-0.0000	-0.0000	0.0001	-0.0000	-0.0000	0.0000
0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999

Columns 9 through 16

0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
0.9933	0.9933	0.9933	0.9933	0.9933	0.9933	0.9933	0.9933
-0.0000	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000
0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997

D'après la (**Figure IV-8**) qui montre la variation de l'erreur quadratique moyenne en fonction du nombre d'itérations, on remarque que l'erreur atteint l'erreur optimale 10^{-5} au bout de 59 itérations avec des performances égale à $5.63985 \cdot 10^{-6}$.

D'après ces résultats on constate que, la reconnaissance vocale est meilleure dans le cas où le locuteur répète plusieurs fois le message, ceci peut s'exprimer par l'état et la variabilité interlocuteur et intra-locuteur.

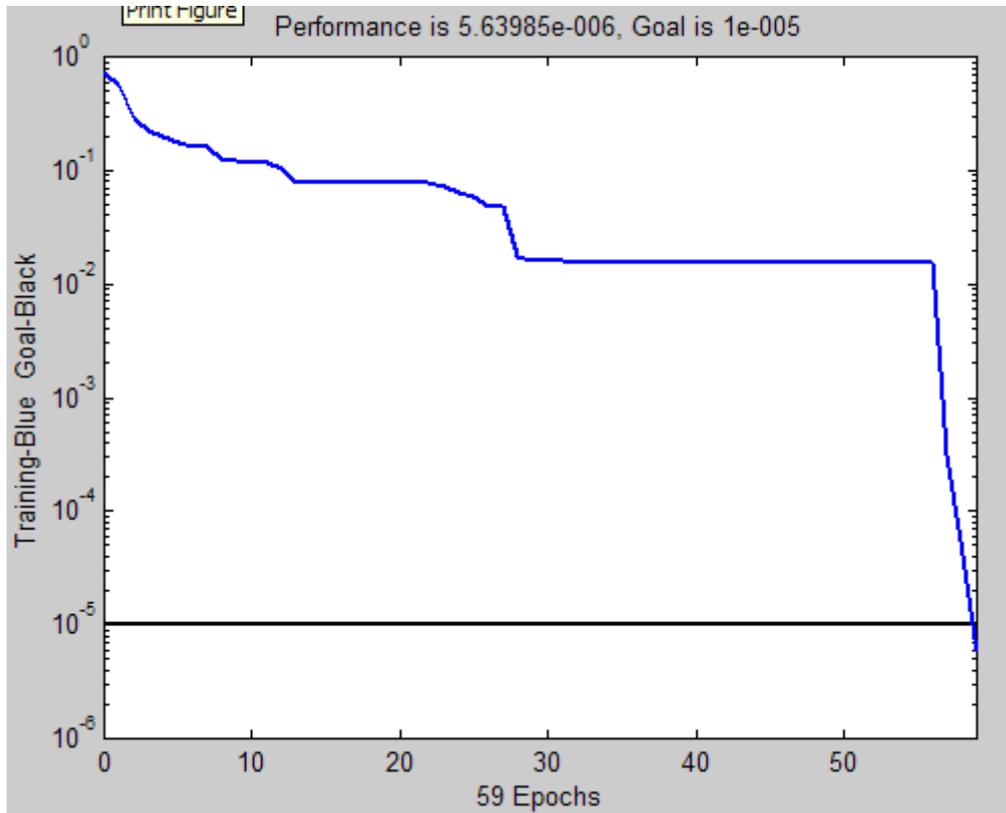


Figure IV-8:-L'erreur quadratique moyenne du réseau

Conclusion :

Les résultats illustrés dans ce chapitre, nous ont permis d'arriver à des remarques importantes qui présentent un intérêt particulier pour notre application. Nous avons déduit que les résultats d'apprentissage dépendent du nombre de neurones de couche cachée, de nombre d'itérations et le du coefficient d'apprentissage. Et nous avons remarqué aussi plus l'erreur est élevé plus il y a une grande différence entre les sorties désirées et les sorties des réseaux donc un apprentissage moins intéressant.

Conclusion générale

Ce travail s'inscrit dans le domaine du traitement du signal et en particulier, du signal de parole. Nous nous sommes intéressés à la reconnaissance du locuteur en mode dépendant du texte et pour cela nous avons donné un aperçu sur le système de production de la parole, ses caractéristiques, son acquisition et sur son analyse par les différentes méthodes temporelle et fréquentielle.

Nous avons opté pour la méthode des réseaux de neurone pour la reconnaissance du locuteur dont on a expliqué le fonctionnement, l'algorithme utilisé ainsi que ses domaines d'applications. L'application de cette méthode sur le signal de parole, nous a permis, d'obtenir de bons résultats c'est-à-dire reconnaître le locuteur qui parle dans le mode dépendant du texte (la phrase est fixée), mais dans des conditions d'enregistrement bien respectées et contrôlées.

Dans le cas de la reconnaissance du locuteur en mode indépendant du texte (la phrase change), l'application de cette méthode ne donne pas des bons résultats, car il est difficile, en particulier, de prévoir les paramètres qui serviront d'entrée du réseau. Dans ce cas l'utilisation d'autres méthodes est nécessaire, parmi lesquelles les modèles de Markov cachés.

A ce jour, il n'y a pas de procédé scientifique qui permette de caractériser de manière unique la voix d'une personne ou d'identifier avec une certitude absolue un individu à partir de sa voix.

Bibliographie

- Bedouheme /Tabani

Les méthodes de débruitage du signal de parole _ mémoire d'INGENIEUR UMMTO 2008

- Guenoun Djamel

Reconnaissance du locuteur _ mémoire d'INGENIEUR UMMTO 2008.

- Khadraoui Mouhamed

Application des réseaux de neurones pour la classification des visages _mémoire d'INGENIEUR UMMTO 2008.

- M^{me} Amrane

Para métrisation et segmentation du signal de parole _thèse magistère UMMTO 2008.

- Chabane Mounia/Bensafia Kahina

Reconnaissance Automatique du locuteur par la méthode du taux de passage par zéro _ mémoire d'INGENIEUR UMMTO 2008.

- Aliouane.A / Belkacem

Utilisation des réseaux de neurones pour la détection et la reconnaissance de défaut sur les machines tournantes_ mémoire d'INGENIEUR UMMTO 2007.

SITE

WWW.youpdf.la reconnaissance vocale.

www.youpd.application application les réseaux de neurones à la reconnaissance vocale.