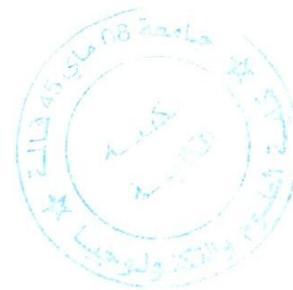


M1621.782

République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche
Scientifique

Université 8 Mai 1945 – Guelma

Faculté de science et science d'ingénierie
Département d'électronique & télécommunication



Mémoire fin d'étude
2^{ème} année Master
Domaine : Sciences et Technique
Spécialité : Systèmes électronique
Filière : électronique

Thème :

Reconnaissance automatique du locuteur

Présenté par :

Griouz Badreddine
Melouki Said

Encadré par :

Nemissi Mohamed

Session juin 2013

Remerciement

Remerciement :



*Mes remerciements vont tout premièrement
à dieu tous puissant, pour la volonté et la
santé et la patience qu'il ma donné.*

*Ainsi, je dois remercier infiniment mon
encadreur Monsieur Nmissi Mohamed pour
les précieux conseils, son soutien son suivi
continu dans la réalisation de ce mémoire
avec la compétence, que chacun de nous lui
reconnaisant.*

*Toute personne qui a été disponible pour
me soutenir matériellement ou normalement
et profondément remerciée.*



Dédicace



Ce mémoire est dédié à mes très chers parents qui ont toujours été là pour moi, et qui ont donné un magnifique modèle de labeur et de persévérance. J'espère qu'ils trouveront dans ce travail toute ma reconnaissance et tout mon amour.

A mes chers frères : Fethi, Abd Erahmen

A ma chère sœur : Madgda

A mes meilleurs amis : Achref, Bilel, Miloud.

Mon binôme dans ce projet et toutes les amies à Dieu à qui j'adresse mon remerciement par Sa grâce infinie pour moi.

dédicaces

AVANTIS TOUTE JE REMERCIE LE BON DIEU DE M'AVOIR MIS
SUR LE BON CHEMIN POUR POUVOIR RÉALISER CE TRAVAIL.

JE DÉDIE CE MODESTE TRAVAIL SPÉCIALEMENT
MES TRÈS CHERS PARENTS QUI ONT TOUJOURS ÉTÉ L'ORIGINE DE
MES SUCCÈS PENDANT TOUT LE CYCLE DE MES ÉTUDES
MA MÈRE LA LUMIÈRE DE MES YEUX, ET SURTOUT POUR SA
SACRIFIÉS ET SA PATIENCE

MON CHER PÈRE POUR SON AMOUR, SON CONFIANCE QU'IL M'A
PERMIS DE SURMONTER LES DIFFICULTÉS E QUI M'A DONNE LE
COURAGE ET L'ESPOIR POUR MES PROJETS D'AVOIR.

À MES FRÈRES ET TOUTE LA FAMILLE

MON BINÔME MELOUKI SAID

À MR NMISSI M, POUR LE SOUTIEN MARAL QU'ILS ONT TÉMOIGNE
TOUT LE LONG DES MES

ETUDES

GRIOZ BADREDDINE

Introduction générale	(1)
Chapitre 1	
1-reconnaissance de forme	(2)
1-1-Introduction	(2)
1-2-Définition de la RDF	(2)
1-3-Objectif de la reconnaissance des formes	(3)
1-4-domaine d'application	(3)
1-5-Domaine En Relation Avec RDF	(3)
1-6-Processus De RDF	(4)
2- Traitement du signal vocal	(4)
2-1- Définition du signal vocal	(4)
2-2-La Numérisation Signal Vocale	(4)
2-2-1- échantillonnage	(5)
2-2-2- Quantification	(5)
2-2-3-codage	(6)
2-3-variabilité du signal vocal	(7)
2-3-1-Variabilités inter locuteurs	(7)
2-3-1-Variabilités intra-locuteurs	(8)
2-3-3-Variabilité due a l'environnement	(10)
2-4- domaine application du traitement de la parole	(10)
3- Reconnaissance Automatique Du Locuteur	(10)

3-1-Définition	(10)
3-2-Principe de La reconnaissance automatique du locuteur:	(10)
4- Extraction Des Caractéristique Par MFCC	(11)
3-3-Identification Du Locuteur	(11)
3-4-Vérification Automatique Du Locuteur	(11)
4-1 Analyses cepstrale	(12)
4-3 Échelle Mel	(13)
4.4 Calcul des coefficients MFCC	(13)
4.5Applications de La reconnaissance automatique du locuteur :	(16)
5-Conclusion	(16)
Chapitre 2	
2.1 Introduction	(17)
2.2 Définition d'un réseau de neurones	(17)
2.3 Domaines d'applications des réseaux de neurones	(18)
2.4 Principes de modélisation des Réseaux de neurones	(19)
2.4.1 Le neurone biologique	(19)
2.4.2 Fonctionnement du neurone biologique	(20)
2.4.3 Le neurone formel	(20)
2.4.4 Les fonctions d'activation	(21)
2.5 Apprentissage des réseaux de neurones	(22)
2.5.1 La règle de Hebb	(23)
2.5.2 La règle de Widrow-Hoff	(24)
2.5.3 Apprentissage compétitif	(25)
2.6 Architecture des réseaux de neurones	(25)

2.7	Modèles neuronaux de base	(27)
2.7.1	Le perceptron	(27)
2.7.2	Classification en utilisant le perceptron	(27)
2.7.3	LVQ	(28)
2.7.4	Classification en utilisant LVQ	(28)
2.8	Conclusion	(29)
	Chapitre 3	
1.	Introduction	(31)
2.	Extraction des caractéristiques	(31)
3.	Exemple d'extraction de caractéristiques	(32)
3.1	Le signal enregistré	(32)
3.3	RFFT du frame	(33)
3.	Phase de classification	(33)
3.1.	Algorithme d'apprentissage LVQ	(33)
3.2.	Exemple de calcul par LVQ	(34)
3.	Base de données	(35)
4.	Résultats	(36)
5.	Conclusion	(37)
	Conclusion générale	(38)

La reconnaissance de la parole est un domaine d'études actif depuis le début des années 50. C'est l'une des composantes fondamentales des sciences moderne qui se situe au croisement du traitement du signal numérique et des techniques de la reconnaissance des formes.

La reconnaissance automatique de la parole peut être basée sur une comparaison de des mots à reconnaître, ou bien sur l'identification d'un ensemble d'unités élémentaires (phonèmes, diphtonges, syllabes). Dans le premier cas, il s'agit d'une reconnaissance dite globale (approche retenue dans ce travail), dans le second cas d'une reconnaissance dite analytique.

Dans ce travail on s'intéresse à la vérification automatique du locuteur. Pour l'extraction des caractéristiques, on se base sur les coefficients cepstreux dans l'échelle de Mels (MFCC) qui sont les plus utilisés. Dans la phase de classification, on utilise le LVQ qui est un réseau de neurones très utilisé dans le domaine de reconnaissance du locuteur.

Le présent mémoire comporte trois chapitres :

Le premier chapitre consiste en une introduction générale à la reconnaissance de la parole et présente les différentes étapes du traitement de la parole.

Le deuxième chapitre consiste en une introduction aux réseaux de neurones.

Le troisième chapitre présente les résultats obtenus sur une petite base de données qu'on réalise.

Finalement, une conclusion générale conclut ce mémoire.

Chapitre 1 : Reconnaissance des formes

1-Reconnaissance des formes

1-1-Introduction

La parole est un signal réel, continu, à énergie finie et non stationnaire. Sa structure est complexe et variable avec le temps.

Le système auditif humain est structure est surtout sensible dans une gamme de fréquence située entre 800 Hz à 8.000 Hz, les limites extrêmes sont respectivement 20 et 20.000 Hz. Par contre, le système vocal est encore plus limité, en résumé, pour des sons vocaliques à des fréquences au-dessus de 4 kHz, les hautes fréquences sont plus de 40 dB en dessous du sommet du spectre. L'information portée par le signal de parole peut être analysée de bien des façons. On en distingue généralement plusieurs niveaux de description non exclusifs : Acoustique, phonétique phonologique, morphologique, syntaxique, sémantique, et Pragmatique.

L'utilisation de la parole comme mode de communication avec une machine présente des avantages certains, notamment dans des situations qui nécessitent souvent la reconnaissance et la synthèse de la parole :

- Utilisateurs ayant déjà les mains ou la vue occupées.
- Utilisateurs occasionnels, non spécialistes d'un système
- Accès à distance : téléphoniques internet télématique vocale
- Utilisateurs handicapés

1-2-Définition de la RDF :

On désigne par reconnaissance de formes un ensemble de techniques et méthodes visant à identifier des objets à partir de données brutes afin de prendre une décision dépendant de la catégorie attribuée à cet objet. On considère que c'est une branche de l'intelligence artificielle qui fait largement appel aux techniques d'apprentissage automatique et aux statistiques.

Les formes à reconnaître peuvent être de nature très variée. Il peut s'agir de contenu visuel (code barre, visage, empreinte digitale...) ou sonore (reconnaissance de parole), d'images médicales (rayon X, EEG, IRM...) ou multi spectrales (images satellitaires) et bien d'autres [2]

1-3-Objectif de la reconnaissance des formes :

- La reconnaissance des formes (RDF) vise la perception et l'interprétation de l'environnement
- Une phase préliminaire consiste à numériser des éléments du monde réel (acquisition des Données)
- Etant donnée une information riche et non structurée (par exemple une image), la RDF a pour but de lui associer une signification symbolique (par exemple une étiquette)
- La RDF peut être perçue comme un procédé méthodique de réduction d'information, pour ne conserver que l'interprétation pertinente [2]

1-4-Domains D'application :

La RDF est utilisé dans de nombreux domaines d'applications :

- L'analyse de signaux sismiques,
- L'analyse d'électrocardiogrammes,
- La reconnaissance de la parole,
- La reconnaissance de l'écriture,
- L'analyse de documents,
- Imagerie médicales (microscope, radiographie, RMN, ...),
- Imagerie satellitaire,
- Applications militaires (observation, guidage, ...)
- La biométrie (reconnaissance d'empreintes, de faces, ...),
- La vision par ordinateur (analyse de scènes 3D) [2]

1-5-Domaine En Relation Avec RDF :

Les méthodes mises en œuvre dans les systèmes de RDF s'appuient sur de nombreux domaines :

- l'analyse numérique,
- les statistiques,
- l'optimisation combinatoire,
- la recherche opérationnelle,
- l'analyse syntaxique,
- la théorie des graphes,
- l'intelligence artificielle,

1-6-Processus De RDF:

Le processus de la RDF comporte généralement quatre étapes :

Objet

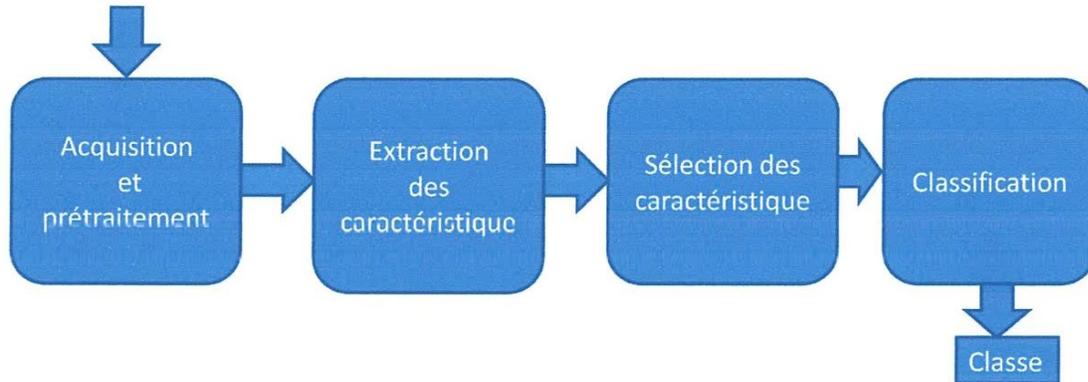


Figure1.1 : les étapes de la RDF

A -Prétraitements

binarisation, localisation, segmentation, élimination du bruit, normalisation, ...

B -Extraction des informations pertinentes

Parole: fréquence fondamentale, harmoniques, énergie, ... quantité

Image: occlusions, concavités, contours, fins de trait, ... d'info.

C- sélection des caractéristiques

Représentation de ces informations en vue de leur classification vecteur , graphe,

Chaîne

D- Classification

Classification de la forme apprentissage/décision, combinaison/fusion [2]

2- Traitement du signal vocal**2-1- Définition du signal vocal**

Le signal de la parole n'est pas un signal ordinaire, il représente un phénomène complexe : la communication parlée. La reconnaissance de la parole pose de nombreux problèmes aux chercheurs depuis 1950. D'un point de vue mathématiques, il est difficile de modéliser le signal de la parole, compte tenu de sa variabilité. [4]

2-2-La Numérisation Signal Vocale

La numérisation du signal vocal se fait en 3 étapes :

- échantillonnage

- quantification
- codification (ou codage)

2-2-1-Echantillonnage

On prélève le signal analogique à intervalle régulier. La norme fixe l'échantillonnage à 8 KHz. Cela correspond à une prise d'échantillon toutes les $1/8000$ s (ou 8 échantillons par Milliseconde soit 1 mesure chaque 125 microseconde).

Cette fréquence est dérivée du théorème de Chanuon, qui dit que la fréquence D'échantillonnage doit au moins être égale au double de la fréquence maximale du signal D'origine.

Un échantillonnage à 8 KHz permet donc la transmission d'un signal de fréquence maximum 4KHz. Le signal vocal analogique 300-3400 Hz passera donc bien. Le signal échantillonné n'est qu'une suite d'impulsions (PAM : Pulse Amplitude Modulation) qui représente l'amplitude du signal analogique lors de chaque échantillonnage.

2-2-2-Quantification

Chaque échantillon est comparé à certains niveaux de quantification à l'aide d'un Convertisseur analogique/digital (CAD), chacun étant représenté par une suite numérique Unique. La suite numérique la plus proche du signal échantillonné est alors utilisée pour Représenter le signal (opération de quantification)

Le nombre de pas détermine par la suite le nombre de bits représentant l'échantillon.

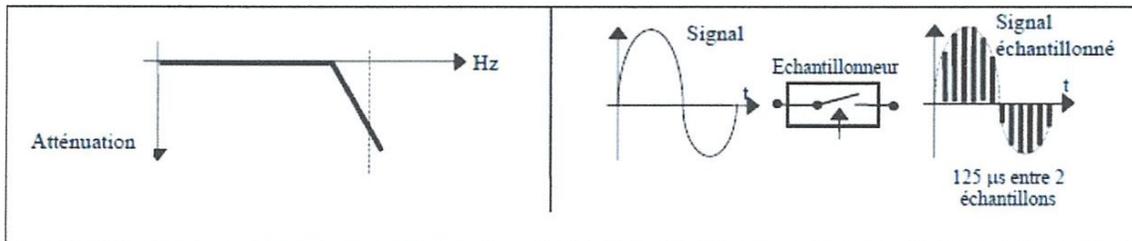


Figure (1.2): Filtrage et Echantillonnage

Pour reproduire fidèlement la valeur analogique, il faudrait un ensemble de valeurs très Important (ou un pas très faible).

En transmission téléphonique, le nombre de niveaux choisi est 256, ce qui correspond à un codage sur un octet. Aux USA, le codage se fait sur 7 bits.

L'approximation de la valeur analogique par une valeur numérique provoque donc une Distorsion du son, dite bruit de quantification. Chaque échantillon mesuré est donc converti En une valeur sur 8 bits.

Le débit de la ligne de transmission pour une communication est donc : 8 bits à transmettre Chaque 125 microseconde ou 8 bits 8000 fois par seconde, ce qui donne un débit de $8 \cdot 8000 = 64000$ bits/s Aux USA où le signal est échantillonné sur 7 bits le débit est de 56 Kbps Par abus de langage, on parle de bande passante de 8KHz. La bande passante ne peut Etre définie que pour un signal analogique. Pour un signal numérique, l'on devrait parler de Débit binaire équivalent à la transmission du même signal en analogique.

La quantification se traduit donc par une approximation du signal réel.

2-2-3-Codification

Le nombre de bits requis pour un échantillon est lié à la valence (ou dynamique) selon la Formule:

$$n = \log_2 D \quad (\log \text{ en base } 2) \quad (1.1)$$

D étant le nombre d'états différents à traduire. Pour la parole la dynamique va de 1 à 4000, Ce qui impose une quantification sur 12 bits. Une opération de codage réduit par Compression ce nombre à 8 bits, dont un bit de signe.

L'erreur relative générée sera d'autant plus importante que le signal est faible. Pour info, L'erreur relative de quantification vaut en moyenne : $(1/2 * \text{amplitude_maxi} / 256) / \text{Valeur_du_signal}$. Pour une amplitude forte, cette erreur est relativement négligeable, mais Pour un signal d'amplitude faible l'erreur relative devient très grande.

Le remède au problème de la trop grande sensibilité aux basses amplitudes consiste à Utiliser une loi de quantification logarithmique. L'effet obtenu est alors équivalent à une Compression.

La règle de compression s'appelle la Loi A en Europe, la loi μ pour les USA et le Japon.

Toutes ces opérations de numérisation de la voix se font dans un composant électronique Appelé CODEC.

A la Réception d'un signal numérisé, l'opération inverse se produit :

- Décodage : Cette opération est effectuée par un convertisseur digital/analogique (CDA).
- Extension : C'est l'opération inverse de la compression.

- Filtrage et Correction : Le signal résultant est filtré pour éliminer les composantes Résiduelles éventuelles. [3]

2-3-Variabilité Du Signal Vocal

Le signal de parole possède une très grande variabilité. Une même personne ne prononce Jamais un mot deux fois de façon identique. La vitesse d'élocution peut varier, la durée du Signal est alors modifiée. Toute altération de l'appareil phonatoire peut modifier la qualité de L'émission (exemple : rhume, fatigue...). De plus, la diction évolue dans le temps. La voix est Modifiée au cours des étapes de la vie d'un être humain (enfance, adolescence, âge Adulte...). [4]

2-3-1-Variabilités Interlocuteurs :

Elles proviennent des différences physiologiques (différences dimensionnelles du Conduit vocal, fréquence d'oscillation des cordes vocales) et de différences de style de Prononciation (accent, niveau social, etc...). Certaines de ces différences, qui influencent le Spectre associé à chaque locuteur, vont nous permettre de les séparer. [1]

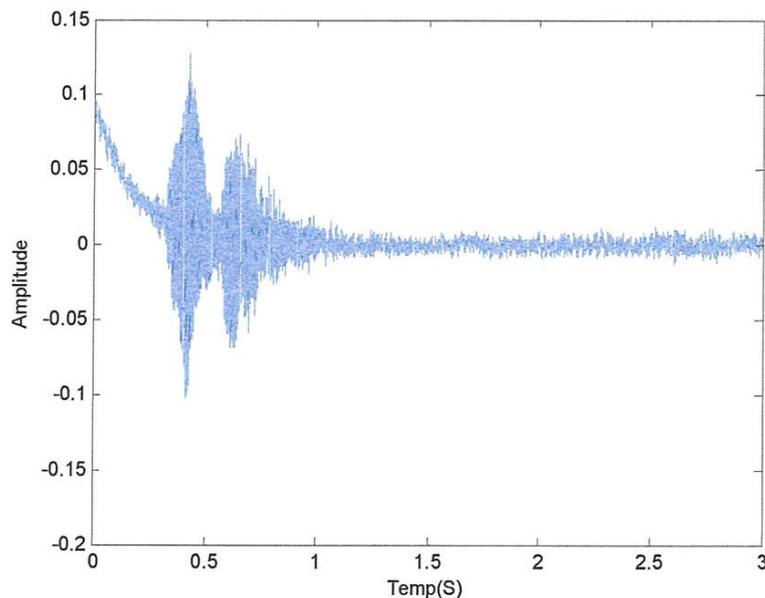


Figure (1.3) : Signal du mot « bonjour » du locuteur1

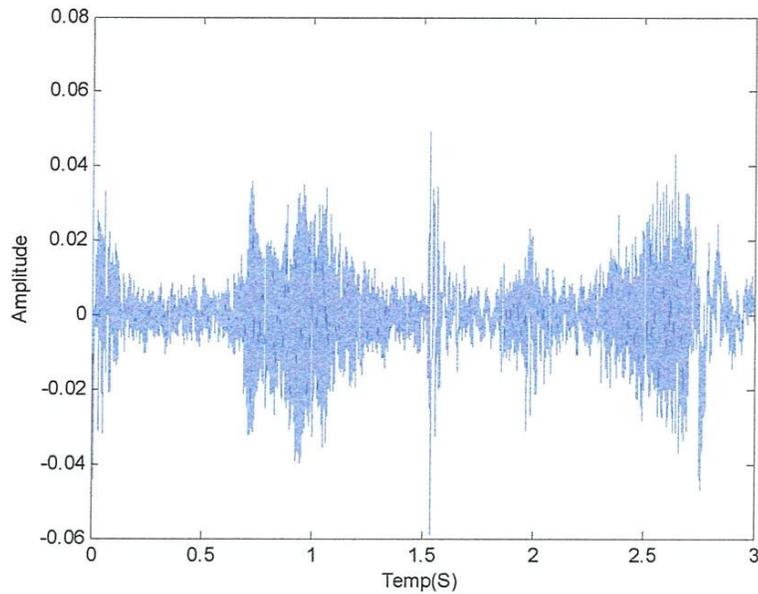


Figure (1.4) : Signal du mot « bonjour » du locuteur2

2-3-2-Variabilité Intra-Locuteur

La voix humaine, à la différence des empreintes digitales, varie avec le temps ou les Conditions physiologiques et psychologiques du locuteur. Cependant, ces variations intra-Locuteur ne sont pas identiques pour tous les humains. En effet, hormis les variations lentes De la voix dues au Vieillessement, certains phénomènes extérieurs tels que la fumée ou l'état de santé d'une Personne ont une influence variable sur sa voix. [1]

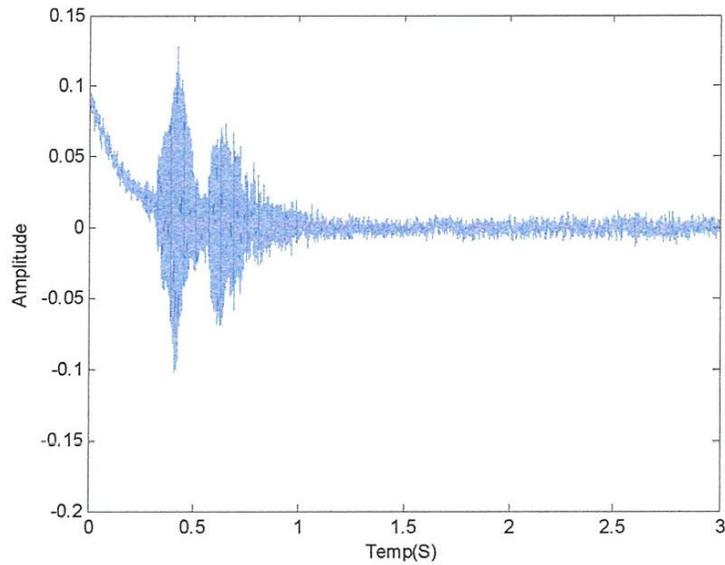


Figure (1.5) : Signal du mot « bonjour » du locuteur 1

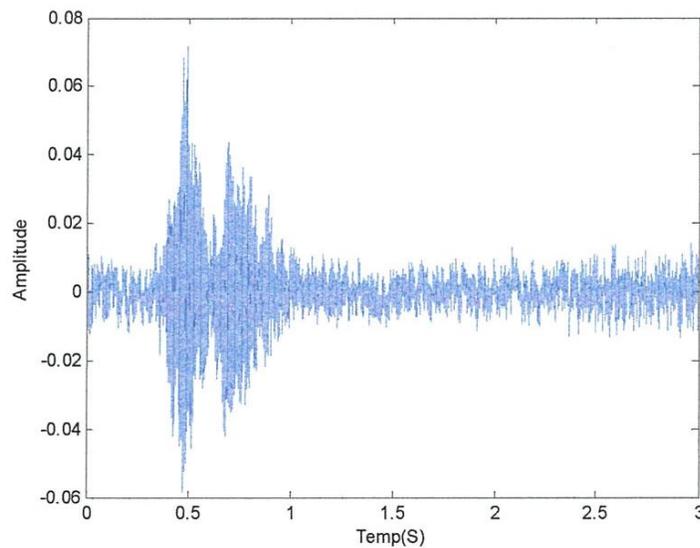


Figure (1.6) : Signal du mot « bonjour » du locuteur 1

2-3-3-Variabilité due a l'environnement

La variabilité liée à l'environnement peut, parfois, être considérée comme une variabilité Intra Locuteur mais les distorsions provoquées dans le signal de parole sont communes à toute personne soumise à des conditions particulières. La variabilité due à l'environnement peut également provoquer une dégradation du signal de parole sans que le

Locuteur ait Modifié son mode d'élocution cette variation, peut être considérée comme du Bruit.

La variabilité environnement due au locuteur peut tout d'abord être de nature physiologique Ainsi, un système mécanique provoquant une déformation du conduit vocal provoquera Immanquablement une variation dans le signal de parole produit.[1].

3- Reconnaissance Automatique Du Locuteur

3-1-Définition

La reconnaissance automatique du locuteur est une technique visant à reconnaître, dans une Suite de signaux sonores, les phonèmes et les phrases prononcées par un locuteur.

La reconnaissance automatique du locuteur se situe à l'intersection de nombreux domaines Tels que l'acoustique, l'électronique, la phonétique, la sémiologie, etc.

3-2-Principe de La reconnaissance automatique du locuteur:

La reconnaissance automatique du locuteur s'insère dans le cadre plus général de la Reconnaissance des formes. Un son correspondant à un mot est prononcé oralement devant Le microphone, qui convertit le signal acoustique en un signal électrique. Celui-ci est soumis à un prétraitement (numérisation, séparation du signal pertinent par rapport au Bruit ambiant), puis il subit la paramétrisation (extraction de paramètres). Ce traitement S'applique tant en phase d'apprentissage qu'en phase de reconnaissance. L'utilisation du Système consiste à comparer un nouveau son à un son de l'ensemble d'apprentissage [3]

3-3-Identification Du Locuteur :

L'identification du locuteur consiste à reconnaître le vrai locuteur parmi une population (Ou Base) composée de N locuteurs connus. L'entrée du système est l'échantillon de parole d'un Locuteur inconnu. La sortie du système correspond à l'identité du locuteur de la base de Référence qui est le plus "proche" du signal de parole inconnu. Dans cette tâche, on fait L'hypothèse que le signal de parole à identifier est prononcé par un des locuteurs de la base De référence (identification en ensemble fermé). [3]

3-4-Vérification Automatique Du Locuteur

3-4-Vérification Automatique Du Locuteur

Un locuteur est connu du système grâce à un ou plusieurs échantillon(s) de sa voix Enregistré(s) au préalable. La vérification automatique du locuteur (RAL) consiste à comparer Ensuite cet enregistrement de référence `a un autre échantillon de parole afin de déterminer S'il s'agit bien du même locuteur dans les deux cas, Ce principe est résumé par La figure 1.3.

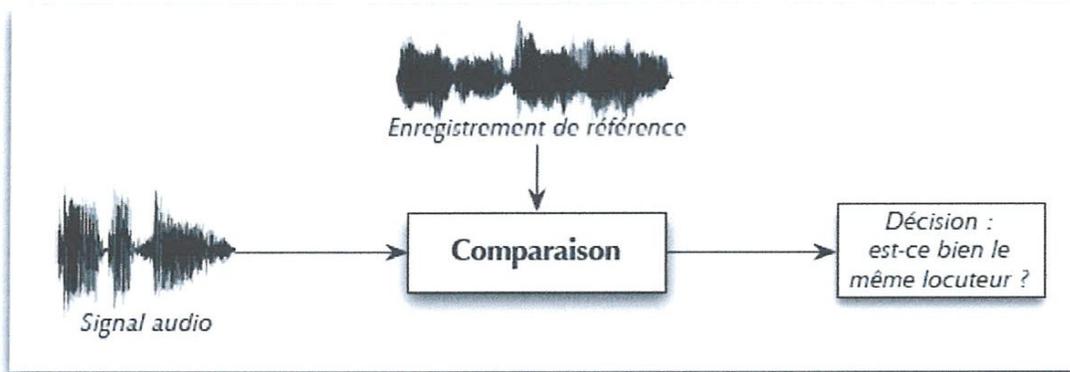


Figure (1.7) – Principe de base de la vérification du locuteur

Ce principe simple trouve son utilité dans un grand nombre d'applications. De plus, les Techniques utilisées dans le cadre de la vérification du locuteur forment la base de la plupart Des autres tâches de la RAL (leur utilisation dans le cadre de l'identification du locuteur en Ensemble ouvert a déjà été évoquée). Leur maîtrise est donc indispensable à la réalisation D'un système de RAL, quelle que soit la tâche visée. Cependant, la diversité des conditions D'application du principe de la RAL fait varier considérablement la difficulté de la tâche, qui Dépend principalement des variations de la voix pouvant intervenir entre l'enregistrement De référence et l'enregistrement de test. [3]

4- Extraction Des Caractéristique Par MFCC:

4-1 Analyses cepstrale :

Le modèle de parole suppose que le signal vocal résulte de la convolution entre la source (Signal excitateur $g(n)$) et le conduit (filtre passif de réponse impulsionnelle $b(n)$) par L'équation :

$$S(z) = G(z) * B(z)$$

Avec $G(z)$ fonction source et $B(z)$ fonction conduit

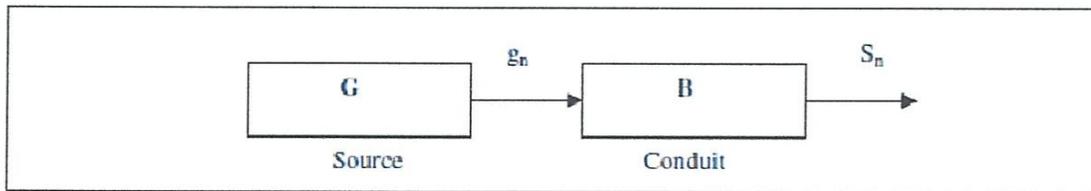


Figure (1.8) signal vocal

En effet le défaut majeur de la FFT pour le calcul du spectre réside dans l'intermodulation Source/conduit qui rend difficile la mesure des formants F_i et la mesure du fondamental (Pitch). [5]

Pour palier à ce manque, le lissage cepstral ou cepstre est utilisé pour séparer leur Contribution respective des deux signaux par déconvolution.

Le cepstre d'un signal est une fonction du temps définie comme la transformée de Fourier Du logarithme de la densité spectrale de puissance. Cette fonction est utilisée pour mettre En évidence les périodicités du spectre. Pour des conditions initiales nulles on peut écrire :

$$\forall n > 0 \quad S_N = G_N * B_N \quad (1.2)$$

Le cepstre est alors obtenue par :

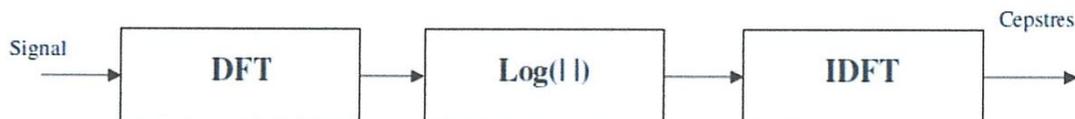


Figure (1.9) calcul des cepstres par analyse cepstrale.

4-3 Échelle Mel :

Il est possible de changer l'échelle de fréquence dans l'espace de Fourier pour obtenir des Coefficients calqués sur une échelle 'physiologique' correspondant mieux à la perception Auditive humaine. Une de ces transformations liée à la perception se fait selon l'échelle de Mel (eq. 1.3) où les coefficients obtenus sont appelés MFCC (Mel-scale Frequency Cepstral Coefficients). Cette transformation compense la sélectivité de l'oreille qui diminue en haute Fréquence. Par analogie, on utilise généralement des bancs de filtres dont la répartition Reproduit cette sélectivité (fig. 1.10). Pour ce type de banc de Filtres, la largeur de chacun Augmente avec la fréquence. Ceci permet, avec peu de filtres et tout en couvrant toute la Bande passante, d'obtenir une résolution fine aux basses fréquences (là où se trouvent les

Premiers formants) et de relever l'information dans les hautes fréquences. Ce type de Paramétrage est actuellement considéré comme des plus performants en reconnaissance de La parole. La relation utilisée pour cela est :[5]

$$\text{Mel}(f) = b \cdot \log_{10}\left(1 + \frac{f}{c}\right) \quad (1.3)$$

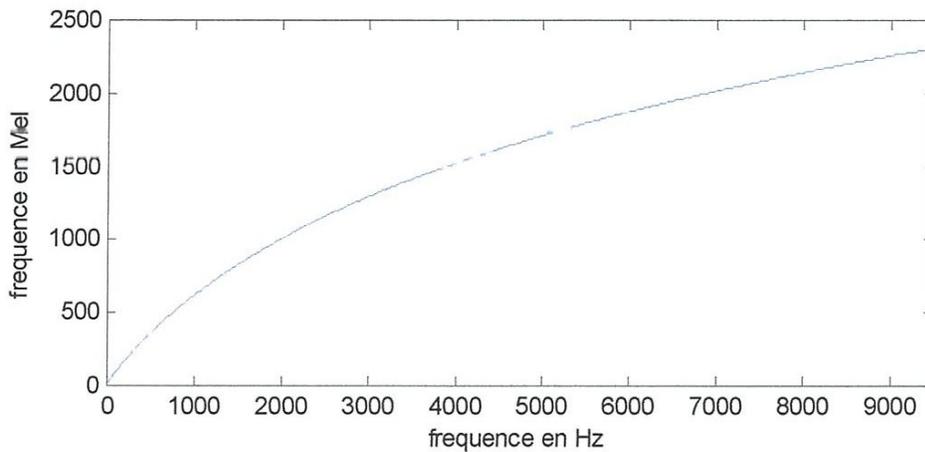


Figure (1.10) : Fréquence en Mel en fonction de la fréquence en Hertz

4.4 Calcul des coefficients MFCC

Les paramètres MFCC sont des coefficients cepstraux obtenus à partir des énergies d'un Banc De filtre en échelle de fréquence Mel .il s'agit en fait d'un calcul classique des Coefficients Cepstraux auquel on a rajoute ,avant le algorithme un filtre de Mel.ces résultats Sont intéressants, car le calcul d'une dizaine de coefficients cepstraux est alors suffisant pour Des expériences de RAP.

$$c_{i,n} = \sum_{k=1}^k \left(\log_{10}(s_{i,k}) \cos\left(n\left(k - \frac{1}{2}\right)\right) \frac{\pi}{k} \right) \quad (1.4)$$

Avec :

$i \in [1, i]$ nombre de frame

$k \in [1, k]$ nombre de filtre

$n \in [1, n]$ nombre de coefficient

$s(i,k)$ la matrice e sortie du $K^{ème}$ filtre pour la $I^{ème}$ frame

La procédure ce calcul des MFCC est la suivante

- Fenêtrage (Hamming) du signal (K fenêtres temporelles de N points chacune).

- Calcul de la DFT dans chaque frame (N points), puis de son module (ou du carré du module).
- Filtrage par les filtres triangulaires en échelle méli (en fait, l'amplitude du spectre est pondérée et sommée dans des fenêtres triangulaire), puis passage au logarithme.
- Calcul de la transformée en cosinus, (ou de la transformée en cosinus inverse, ou de la transformée de Fourier) pour chaque fenêtre.

On se trouve donc une matrice dont les colonnes correspondent aux différents frames, et dont les lignes correspondent aux coefficients cepstraux.

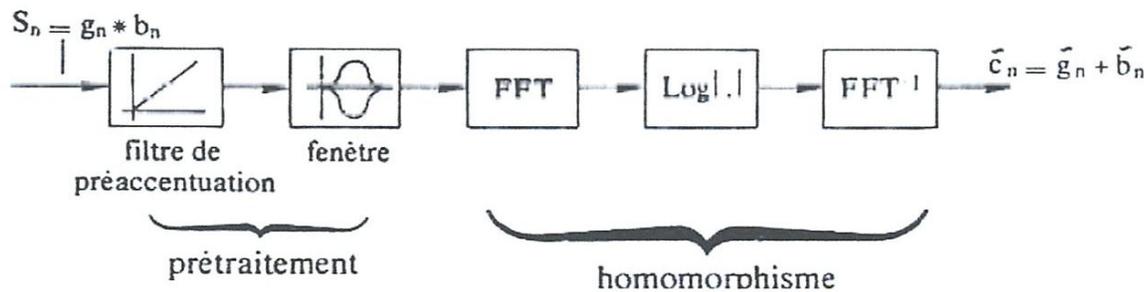


Figure (1.11) : Schéma bloc de la conception du cepstre

Le fenêtrage permet de minimiser la déformation du spectre dans les Hautes fréquences, Due au découpage en trames impose au signal de parole, en employant une Fenêtre dans le Domaine temporel qui réduit progressivement l'amplitude du signal au Commencement et à La fin de chaque trame

Le signal vocal est un signal non stationnaire. Il présente une évolution lente dans le temps Le but du fenêtrage est de découper le signal de parole en petites tranches (chacun de durée 30ms environ) ou il peut être considéré localement comme quasi-stationnaire.[5]

Il existe plusieurs types de fenêtres d'analyse. On en présente quelques-unes :

Fenêtre rectangulaire :

$$\text{Elle est définie par : } f(nT) = \begin{cases} 1 & \text{si } |nT| < t' \\ 0 & \text{ailleurs.} \end{cases} \quad (1.5)$$

$T_e = 1/F_e$ Est le période d'échantillonnage.

T est la moitié de longueur de la fenêtre.

B -La fenêtre de hamming

$$F(nT) = \begin{cases} 0.54 - 0.46 * \cos\left(\pi * n * \frac{T}{T'}\right) & \text{si } |nT| < T' \\ 0 & \text{ailleurs} \end{cases} \quad (1.6)$$

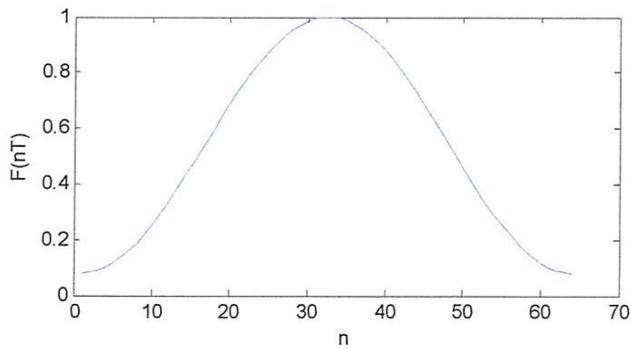


Figure1.12 la fenêtre de hamming

C- Fenêtre de hanning

$$F(nT) = \begin{cases} 0.5 - \left(1 - \cos\left(n * \pi * \frac{T}{T'}\right)\right) & \text{si } |n * T| < T' \\ 0 & \text{ailleurs} \end{cases} \quad (1.7)$$

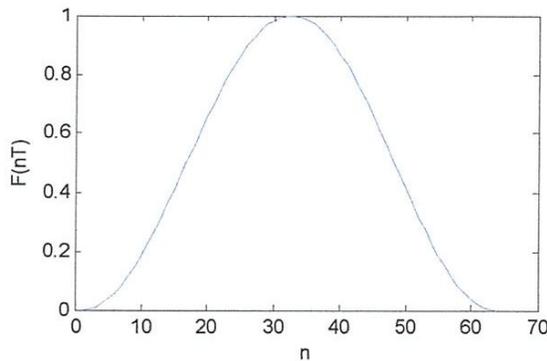


Figure1.13:la fenêtre de hanning

Parmi ces fenêtres, la fenêtre de hamming est la plus convenable à la parole, car elle entraîne un minimum de distorsion spectrale du signal de parole, par rapport aux fenêtres. [6]

4.5 Applications de La reconnaissance automatique du locuteur :

- Sécurité

- * Contrôle d'accès (en complément d'un code, d'un badge)
- * Banques, voitures, entreprises...
- * Consultation de compte bancaire par téléphone...

- Police criminelle (identification de suspects) ?

- * Filtrage de voix suspectes (avec validation humaine)

- * ...Pas assez fiable pour utiliser comme preuve ! (cf. AFCP)

- Indexation multimédia

- * Indexation par locuteur

- Transcription automatique

- *Adaptation des modèles acoustiques à la voix du locuteur

5-Conclusion

Dans ce chapitre nous avons donné un aperçu sur la reconnaissance automatique de Locuteur. Dans ce travail nous réalisons une extraction des caractéristiques donnée par Coefficients MFCC (mel- scaled frequency cepstral coefficients) qui sont les utilisés dans ce Domaine.

Chapitre 2 : Les réseaux de neurones

2.1 Introduction

L'idée d'élaborer un modèle mathématique du cerveau humain, en quête de reproduire ses aptitudes intellectuelles, est à l'origine de la création des réseaux des neurones artificiels. Les premiers travaux sur ces systèmes ont été menés par les neurologues Warren McCulloch et Walter Pitts en 1943. Ces travaux ont été suivis en 1949 par le travail de Donald Hebb qui a proposé une simple règle d'apprentissage dans son ouvrage "The Organization of Behaviour". Le premier processus artificiel capable d'apprendre par expérience a été proposé par Franck Rosenblatt en 1957, c'était le perceptron. Mais, Marvin Lee Minsky et Seymour Papert ont publié un travail énonçant les limitations de ce dernier, et principalement l'impossibilité de traiter des problèmes non linéaires. Cette conclusion pessimiste a étendu tous les modèles des réseaux de neurones et a abaissé leurs intérêts pendant deux décennies. Il a fallu attendre l'apparition du perceptron multicouche, introduit par Rumelhart en 1986, pour donner naissance aux réseaux de neurones. Cette découverte a été une vraie révolution qui a permis aux réseaux de neurones de connaître un essor considérable. Actuellement, la littérature sur les réseaux de neurones est devenue énorme et ne cesse de croître.

Malgré la réussite récente des réseaux de neurones artificiels, la communauté scientifique est toujours loin de la mise en œuvre de machines capables de reproduire les capacités de calcul des systèmes nerveux, même les plus simples. Il s'agit en fait d'une simple modélisation du comportement des neurones biologiques qui ne pourra, en aucune manière, de représenter leurs complexités de fonctionnement. En revanche, cette modélisation permet de mettre en œuvre des processus ayant de grandes capacités de calcul en se basant sur des simples cellules de base : les neurones formels.

2.2 Définition d'un réseau de neurones

Un réseau de neurones est un modèle mathématique qui tente de reproduire quelques fonctions du cerveau humain, telles que : le parallélisme, l'acquisition des connaissances au

Travers d'un processus d'apprentissage, le stockage des connaissances et la possibilité d'utilisation de ces connaissances. [8]

2.3 Domaines d'applications des réseaux de neurones

L'essor des réseaux de neurones dans divers domaines actuels est certainement dû à leurs grandes capacités de calcul et à leurs hautes habiletés d'apprentissage. De plus, l'estimation de leurs paramètres est indépendante de la complexité du problème traité ce qui leur permet d'être bien adaptés aux problèmes actuels qui ne cessent d'être de plus en plus complexes. Les applications des réseaux de neurones peuvent être récapitulées en trois grands domaines :

- La modélisation : La plus part des problèmes industriels ou de recherche, que ce soit Mécanique, physique, chimique ou même économique, nécessitent une représentation à l'aide d'un modèle mathématique permettant de reproduire le comportement du processus mis en œuvre. De telles tâches nécessitent des outils de calcul ayant de grandes capacités de calcul, d'apprentissage et surtout des outils dont leur conception est peu dépendante de la complexité et de la taille du problème traité. Les réseaux de neurones semblent être l'une des solutions les plus adéquates à ce type de problèmes.
- La commande : Commander un processus industriel consiste à concevoir un système permettant le calcul de la commande à appliquer à ce processus de manière à lui assurer un comportement dynamique désiré. Les réseaux de neurones permettent de bonnes performances en tant que partie de commande à cause de leur souplesse d'auto adaptation.
- La classification : Une autre grande catégorie de problèmes industriels consiste à attribuer, de façon automatique, un objet à une classe parmi d'autres classes possibles. La résolution de ce type de problèmes demande de représenter les exemples à classer à l'aide d'un ensemble de caractéristiques. Il s'agit ensuite de concevoir un système capable de classer ces exemples en se basant sur leur représentation et les réseaux de neurones sont particulièrement bien adaptés à ce

Type de problème. La classification est d'ailleurs le domaine privilégié des réseaux de Neurones. [8]

2.4 Principes de modélisation des Réseaux de neurones

2.4.1 Le neurone biologique

Le neurone constitue l'unité de base de l'organisation du système nerveux. Ce sont des Cellules spécialisées dans le traitement des signaux électriques et la transmission du Message nerveux. Le cerveau humain contient plusieurs milliards de neurones Interconnectés, chacun d'eux est connecté à environ dix mille d'autres neurones. La Structure des neurones est parfaitement adaptée à leurs tâches. Un neurone se compose Essentiellement de quatre parties : les dendrites, le corps cellulaire, l'axone et les synapses (fig.2.1).

- Le corps cellulaire : Il contient le noyau du neurone, sa taille est de quelques microns De diamètre. C'est le centre de l'influx nerveux qui représente l'état d'activité du Neurone.
- Les dendrites : Ils constituent les entrées principales du neurone, ils captent les Signaux envoyés vers eux. Leur longueur est de quelques dizaines de microns.

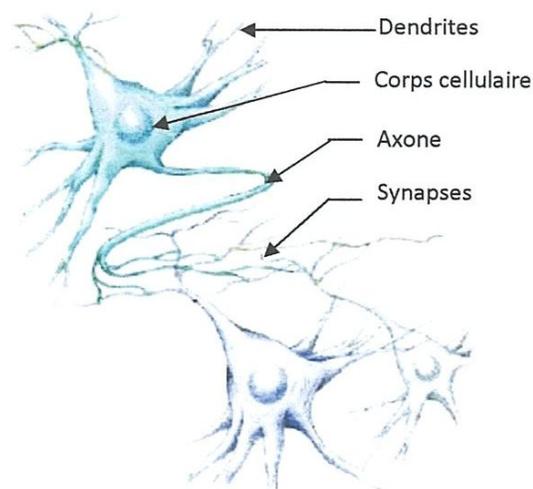


Fig. 2.1. Le neurone biologique (œuvre d'artiste)

- L'axone : C'est la sortie du neurone qui conduit l'influx nerveux sous forme de Potentiel d'action jusqu'aux neurones suivants. Il se termine par une synapse qui Transmet chimiquement le message électrique aux dendrites du neurone prochain. L'axone est plus long que les dendrites, il se ramifie à son extrémité ou il se connecte Aux autres neurones. Sa taille peut varier de quelques millimètres à plusieurs mètres.
- Les synapses : Ce sont des jonctions entre deux neurones et qui sont essentielles Dans le fonctionnement du système nerveux.

2.4.2 Fonctionnement du neurone biologique

La transmission de l'information s'effectue dans un seul sens : des dendrites vers les axones. Chaque neurone reçoit les informations provenant des autres neurones sous forme d'un Signal électrique par ces dendrites. Si la somme de ces signaux est excitatrice, le neurone Emettra à son tour un signal électrique qui se propage par les axones aux connecteurs Terminaux, et ensuite aux dendrites des autres neurones. Selon Hebb , l'apprentissage Des neurones se fait par la modification des résistances électriques des connexions Dendrites.

2.4.3 Le neurone formel

La première modélisation du neurone remonte aux années 40 où McCulloch et Pitts ont Proposé le premier modèle de neurones formels. C'est un modèle mathématique qui réalise Une somme pondérée des signaux qui lui parviennent et déclenche une réponse si cette

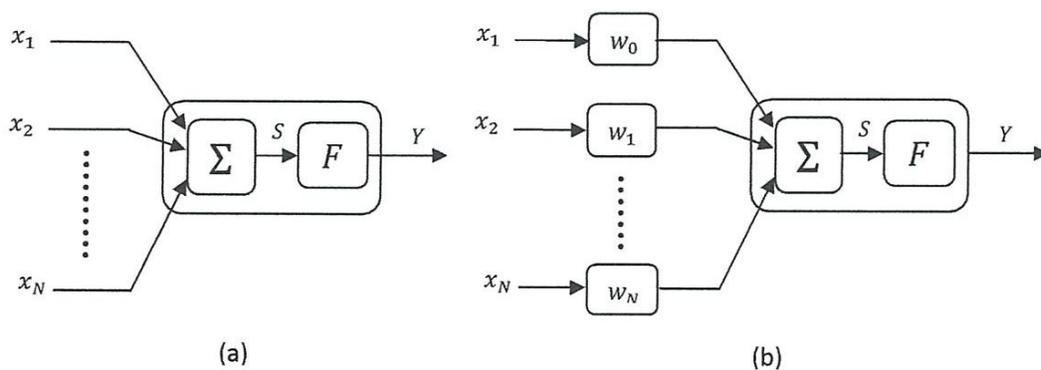


Fig. 2.2. Le neurone formel
 (a) Modèle de McCulloch et Pitts
 (b) Modèle avec capacité d'apprentissage

Somme dépasse un certain seuil. La figure (2.2 a) résume la chaîne de traitement développée par ce neurone. Ce modèle n'a pas possédé une règle d'apprentissage jusqu'à 1949 où Hebb a proposé un principe d'apprentissage sans avoir donné d'équations.

La valeur de la sortie (Y) résulte de la somme des entrées (x_i) pondérées par des coefficients (w_i) et du calcul d'une fonction d'activation (F) de cette somme pondérée. La formalisation Mathématique de son comportement est donnée par :

$$Y = F(S) = F\left(w_0 + \sum_{n=1}^N w_n x_n\right) \quad (2.1)$$

Il est à noter que cette modélisation simplifiée est loin de fournir une explication exacte concernant la complexité de fonctionnement des neurones biologique. Malgré cela, cette formalisation permet d'étudier les connexions entre ces neurones dans d'autres processus plus complexes comportant plusieurs neurones interconnectés.

2.4.4 Les fonctions d'activation

La fonction d'activation peut avoir plusieurs formes différentes. Dans leur modèle d'origine, McCulloch et Pitts ont utilisé des fonctions d'activation à seuil. Les états des neurones avec de telles fonctions sont binaires et les ensembles de valeurs possible les plus couramment utilisées sont $\{-1, 1\}$ ou $\{0, 1\}$. Les deux formes des fonctions à seuil sont :

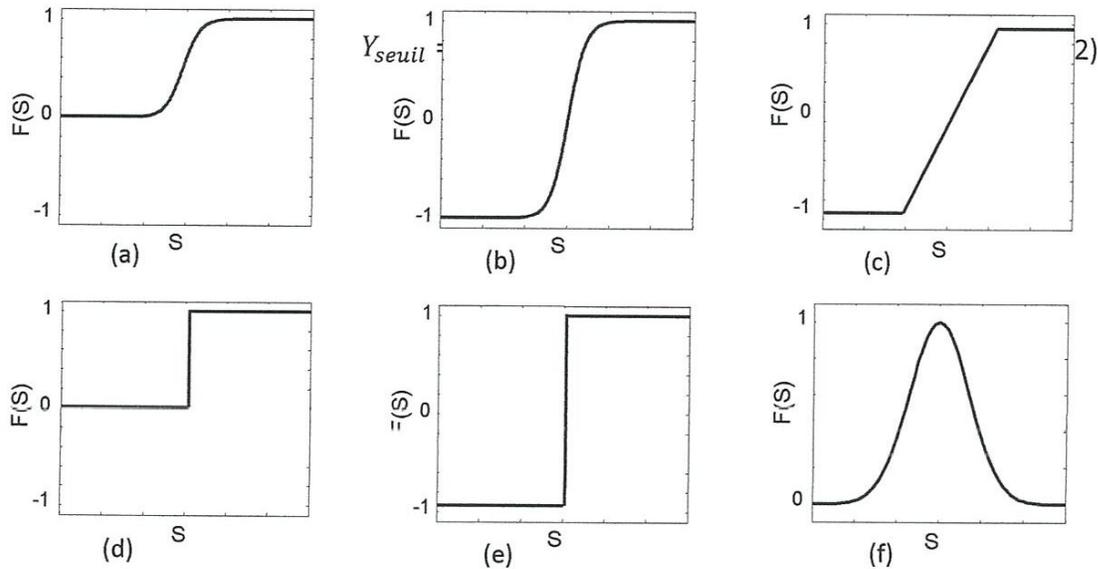


Fig. 2.3. Exemples de fonctions d'activation

- | | |
|---------------------------------|-------------------------------|
| (a) Fonction seuil | (b) Fonction seuil bipolaire |
| (c) Fonction RBF | (d) Fonction sigmoïde |
| (e) Fonction sigmoïde bipolaire | (f) Fonction linéaire saturée |

$$Y_{\text{signe}} = \begin{cases} +1 & \text{si } S > \theta \\ -1 & \text{sinon} \end{cases} \quad (2.3)$$

L'examen du comportement des cellules nerveuses par rapport aux signaux qui leur arrivent a donné lieu à l'utilisation d'un autre type très intéressant comme fonction d'activation : les Fonctions sigmoïdales. Une fonction sigmoïde est donnée par :

$$F_{\text{sigmoïde}}(S) = \frac{1}{1 + \exp(-S)} \quad (2.4)$$

La figure (2.3) donne quelques exemples des fonctions d'activation. Les fonctions RBF forment une catégorie particulière des réseaux de neurones : Les réseaux RBF. [8]

2.5 Apprentissage des réseaux de neurones

L'apprentissage est l'une de des propriétés fondamentales des réseaux de neurones. C'est le processus permettant au réseau de se spécialiser sur un problème spécifique à partir de son expérience. Il consiste généralement à modifier les poids synaptiques jusqu'à ce que le réseau puisse effectuer la tâche désirée. Il s'agit de configurer les valeurs des poids synaptiques censés stocker les informations acquises. D'une façon générale, l'apprentissage

Se traduit par une modification dans la valeur des poids reliant les neurones du réseau. Chaque poids w_{ij} reliant un neurone i au un neurone j à l'itération (t) est modifié selon L'équation générale suivante :

$$w_{ij}(r) = w_{ij}(r - 1) + \Delta w_{ij}(r - 1) \quad (2.5)$$

Où : $w_{ij}(r)$ et $w_{ij}(r - 1)$ sont respectivement les valeurs de ce poids à la $r^{\text{ème}}$ et la $(r - 1)^{\text{ème}}$ itération et $\Delta w_{ij}(r - 1)$ est le changement correspondant.

Selon que l'apprentissage est supervisé ou non, les méthodes d'apprentissage des réseaux De neurones peuvent être catégorisées en deux grandes familles. Dans la première Catégorie, chaque exemple de la base d'apprentissage est couplé à une solution désirée. Cette dernière permet au réseau de connaître ces erreurs et de s'adapter à la présentation De chaque exemple d'apprentissage afin de se rapprocher du résultat souhaité. En revanche, Dans la deuxième catégorie le réseau ne dispose pas de la solution désirée sur les exemples D'apprentissage pour l'aider à ajuster ses paramètres. Il doit chercher à représenter au Mieux l'espace des exemples qui lui sont présentés. Les règles de base d'apprentissage des Réseaux de neurones sont les suivantes :

2.5.1 La règle de Hebb

La règle de Hebb est une règle d'apprentissage supervisé basée sur le renforcement des Connexions entre neurones. Son principe est le suivant : si deux neurones d'une part et D'autre d'une synapse sont activés de façon synchrone et répétée, la connexion synaptique Sera alors renforcée. Bien que Hebb n'a donné aucune équation, le principe d'apprentissage Hebbien a eu une influence philosophique sur les travaux qui lui succédèrent. Il a été en Premier lieu derrière l'idée d'incorporation des poids ajustables dans le modèle de McCulloch et Pitts. Ensuite, Il a été à la base de diverses approches d'apprentissage tel que le Réseau de Hopfield dont la règle d'apprentissage est Donnée par :

$$\Delta w_{ij}(r - 1) = \eta x_i^{(q)} x_j^{(q)} \quad (2.6)$$

2.5.2 La règle de Widrow-Hoff

La règle d'apprentissage de Widrow-Hoff, ou des moindres carrés (LMS, Least Square Sum), est une règle d'apprentissage supervisé basée sur la correction d'erreurs observées en Sortie. Cette règle consiste à minimiser une fonction coût caractérisée par l'erreur Quadratique moyenne. Pour un ensemble d'apprentissage contenant Q paires entrée/sortie Désirée $\{(X^{(q)}/T^{(q)})\}$, $q = 1, \dots, Q$ où $X^{(q)}$ et $T^{(q)}$ représentent respectivement la $q^{\text{ème}}$ Entrée et la $q^{\text{ème}}$ sortie désirée, l'erreur ($e(t)$) à l'itération r est donnée par :

$$e(t) = T(r) - Y(r) \quad (2.7)$$

Où : $Y(r)$ est la sortie calculée du réseau. La fonction coût est :

$$F(X) = e^2(r) \quad (2.8)$$

L'apprentissage selon la règle LMS consiste à calculer le gradient à chaque présentation d'un Exemple d'apprentissage. Le changement de poids est alors :

$$\begin{aligned} \Delta w_{ij}(t) &= -\eta \nabla F(X) \\ &= -\eta \frac{\partial e^2(r)}{\partial w_{ij}} \end{aligned} \quad (2.9)$$

Cette règle de correction permet donc aux neurones d'adapter leurs poids pour se Rapprocher à une valeur désirée correspondante à chaque exemple présenté. Cette règle a Été utilisée pour l'apprentissage de l'ADALINE dans Lequel chaque neurone i corrige ses poids w_{ij} à l'itération r selon l'équation suivante :

$$\Delta w_{ij}(r) = \Delta w_{ij}(r-1) - \eta(t_i - y_i)x \quad (2.10)$$

Où : t_i et y_i sont respectivement la sortie désirée et la sortie calculée correspondantes au Neurone i ; x est l'entrée et η est une constante positive appelée pas d'apprentissage.

2.5.2 Apprentissage compétitif

C'est un apprentissage non supervisé qui consiste à faire une compétition entre les neurones d'un réseau pour déterminer lequel sera actif à un instant donné. Contrairement aux autres types d'apprentissage, où la mise à jour des poids de tous les neurones d'effectue simultanément, ce mode d'apprentissage considère à chaque fois un neurone vainqueur, et parfois un ensemble de voisins du vainqueur, et seuls les poids de ces neurones seront adaptés.

Les poids d'un neurone vainqueur seront modifiés de telle sorte qu'ils se rapprochent de l'exemple X présenté en entrée et pour lequel ce neurone a gagné la compétition contre tous les autres neurones. En revanche, les poids d'un neurone qui ne gagne aucune compétition ne seront pas modifiés. La règle d'apprentissage est donnée par :

$$\Delta w_i = \begin{cases} \eta_1(X - w_i) & \text{si le neurone est vainqueur} \\ \eta_2(X - w_i) & \text{si le neurone est voisin du vainqueur} \\ 0 & \text{Autrement} \end{cases} \quad (2.11)$$

Avec : η_1 et η_2 sont les taux d'apprentissage et $\eta_1 < \eta_2$. Cette incrémentation reflète la distance entre la valeur courante du poids et la valeur d'entrée.

Ce type d'apprentissage caractérise une classe de réseaux de neurones tels que le réseau de Kohonen, où la décision est prise en déterminant quel neurone représente le mieux l'exemple d'entrée.

2.6 Architecture des réseaux de neurones

Bien que la modélisation du neurone formel n'est qu'une simple imitation du neurone biologique, de bonnes organisations de ces simples cellules de base permettent d'avoir des réseaux de neurones ayant de grandes capacités de calcul et d'apprentissage. Néanmoins, l'organisation topologique des réseaux de neurones n'est généralement le résultat d'aucunes imitations des structures neurobiologiques.

Un réseau de neurones artificiels est un ensemble interconnecté de neurones formels fonctionnant en parallèle. On peut distinguer entre les réseaux de neurones selon le fait qu'ils sont à couche ou non. Dans les réseaux à couche, les neurones appartenant à la même

Couche ne sont pas interconnectés et ils ont généralement les mêmes propriétés. Dans ces Modèles, le calcul s'effectue de couche en couche (de l'entrée vers la sortie) et les neurones De la même couche opèrent simultanément.

D'autre part, selon qu'il y a des retours ou non (des neurones de sortie vers ceux d'entrée), On distingue deux grandes familles d'architecture : les réseaux de neurones non bouclés et Les réseaux de neurones bouclés. Dans les réseaux du premier type, l'information se Propage de l'entrée vers la sortie sans retour et les neurones de sortie n'influencent jamais Sur les neurones d'entrée. Cette architecture est la plus utilisée dans la classification, L'approximation des fonctions et la modélisation des procédés. En revanche, les réseaux du Second type (appelés aussi les réseaux récurrents) contiennent des boucles ramenant la Valeur d'une ou de plusieurs sorties vers l'entrée avec un retard. Ils peuvent alors être Considérés comme des systèmes dynamiques. On trouve parmi les réseaux non bouclés: le Perceptron, le Perceptron Multi-Couches, les réseaux RBF...etc. et parmi les réseaux Récurrents : les cartes topographiques de Kohonen et les réseaux de Hopfield.[2]

2.7 Modèles neuronaux de base

2.7.1 Le perceptron

Ayant été introduit par Rosenblatt dans les années 50, le Perceptron est historiquement le Premier réseau de neurones artificiels. Ce modèle comporte une seule couche de neurones Recevant en entrée N valeurs x_1, x_2, \dots, x_N et calcule une sortie Y . L'apprentissage du Perceptron est supervisé, il consiste à ajuster les valeurs des poids associés aux entrées des Neurones qui forment ce perceptron.

2.7.2 Classification en utilisant le perceptron

Le perceptron, ayant été introduit par Resemblat en 1957, est historiquement le Premier réseau de neurones. Comme l'indique son nom, le perceptron a été mis en œuvre Essentiellement pour réaliser des tâches de classification. Ce réseau comporte une seule Couche de neurones permettant de recevoir un vecteur d'entrées et de calculer une sortie (figure 2.4). Un perceptron est défini par un ensemble de poids synaptique et un biais. La Fonction d'activation est une fonction à seuil qui prend ses valeurs dans $\{-1,1\}$.

Pour un perceptron avec N entrées (x_1, x_2, \dots, x_N) , sa sortie Y est donnée par :

$$Y = h(S) = h(w_1x_1 + w_2x_2 + \dots + w_Nx_N + w_0) \quad (2.12)$$

Où : $h(\cdot)$ est la fonction d'activation, w_1, w_2, \dots, w_N sont les poids synaptiques et w_0 est le Biais.

La sortie du perceptron est $Y = 1$ pour les vecteurs d'entrée ayant $S > 0$ et $Y = -1$ pour les Autres vecteurs. Le perceptron divise l'espace caractéristique en deux sous-espaces Délimités par un hyperplan H . Ce dernier est formé par les vecteurs d'entrée dont $w_1x_1 + w_2x_2 + \dots + w_Nx_N + w_0 = 0$. L'espace caractéristique est partitionné donc en demi-Espaces : un demi-espace gauche H^+ et demi-espace droit H^- comme suit :

$$\begin{cases} \text{Si } w_1x_1 + w_2x_2 + \dots + w_Nx_N > w_0 \text{ alors } X \in H^+ \\ \text{Si } w_1x_1 + w_2x_2 + \dots + w_Nx_N < w_0 \text{ alors } X \in H^- \end{cases} \quad (2.13)$$

La figure (2.5) montre un exemple de partitionnement d'un espace caractéristique Bidimensionnel. Il est divisé en deux sous-espaces H^- et H^+ par une droite constituant

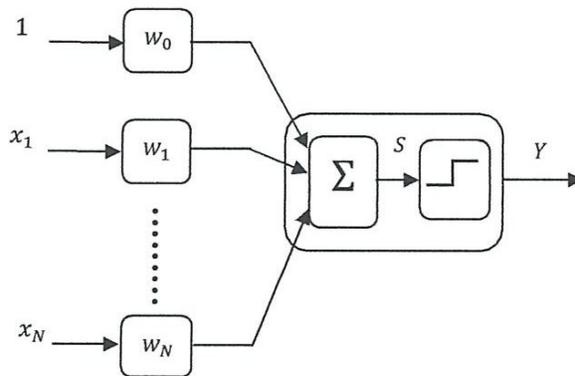
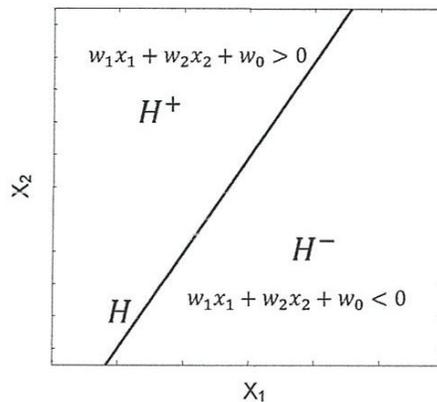


Fig. 2.4. Représentation d'un perceptron

L'hyperplan dans le cas bidimensionnel



La figure (2.4) un exemple de partitionnement d'un espace caractéristique bidimensionnel

2.7.3 LVQ

(LVQ, Learning Vector Quantization) est un réseau de neurones qui combine l'apprentissage Compétitif avec apprentissage supervisé. Le LVQ est constitué d'une couche compétitive et Une couche linéaire (sans compter la Couche d'entrée). Dans la couche d'entrée, chaque Neurone est associé à une variable d'entrée. Différents neurones de la couche compétitive Peut être attribué à la même classe. Chacun de ces les classes sont ensuite affectés à un Neurone dans la couche de sortie.

2.7.4 Classification en utilisant LVQ

LVQ peut être utilisé pour les problèmes de classification. Un ensemble d'apprentissage Constitué de Q-paires vecteur d'apprentissage et sortie cibles sont supposées être donné

$$\{s^{(q)} : t^{(q)}\}, \quad Q=1.2.....Q. \quad (2.14)$$

Où $s^{(q)}$ sont des vecteurs d'apprentissage de dimensions N, et $t^{(q)}$ sont des vecteurs de Sortie cibles de dimensions M. M est le nombre de classes, et il doit être inférieur à Q. Les Vecteurs cibles sont définis par

$$S_i^{(q)} = \begin{cases} 1; & \text{si } s^{(q)} \text{ appartient à la classe } i \\ 0; & \text{autrement} \end{cases}$$

Le nombre des neurones dans la première couche sera égale à N , et le nombre de neurones de la couche de sortie sera égale à M (le nombre de classes)

Dans la couche compétitive, chaque neurone apprend un vecteur prototype qui lui permet de classer une zone de l'espace d'entrée.

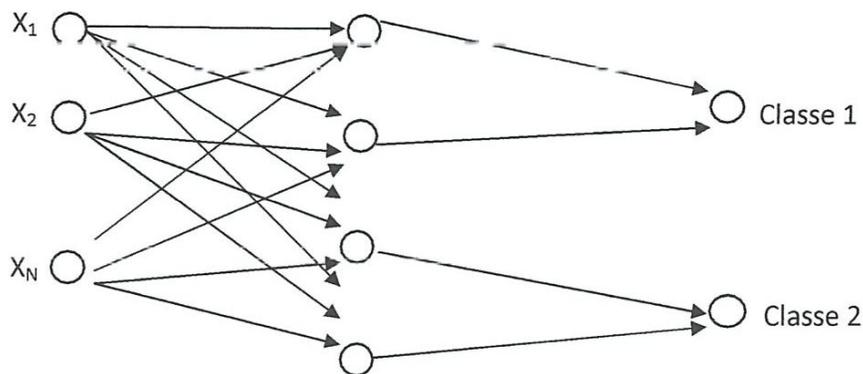


Figure 2.6 : Classification par LVQ

2.8 Conclusion

Les réseaux de neurones ont connu un essor considérable tant qu'en nouvelles architectures qu'en nouveaux algorithmes d'apprentissage et dans ce chapitre nous avons tenté de donner un simple survol sur ces importants outils mathématiques. Cette partie nous a permis de tirer quelques conclusions :

- Les réseaux de neurones semblent bien adaptés aux problèmes actuels qui ne cessent d'être de plus en plus complexes et qui nécessitent des outils mathématiques ayant de grandes capacités de calcul et d'apprentissage.
- Les réseaux de neurones artificiels fonctionnent en se basant sur un calcul parallèle effectué par un ensemble d'unités qui opèrent localement.
- D'un point de vue topologique, les réseaux de neurones offrent une grande gamme d'architectures permettant leur utilisation dans divers problèmes.

- Les réseaux de neurones sont difficiles à paramétrer.
- Les réseaux de neurones sont des boîtes noires qui ne permettent d'interpréter le Modèle obtenu.

Chapitre 3: Application

1. Introduction :

L'objectif de ce travail est la réalisation d'un système de reconnaissance automatique de Vérification du locuteur. Pour l'extraction des caractéristiques, on se base sur les coefficients Cepstraux dans l'échelle de Mels (MFCC) qui sont les plus utilisés dans ce domaine. Dans la Phase de classification, on se base sur les réseaux de neurones et plus précisément sur le Réseau LVQ. Nous réalisons une petite base de données avec des noms et prénoms Prononcés par quatre locuteurs. Cette base de données sera utilisée pour l'apprentissage et Le test du réseau LVQ

2. Extraction des caractéristiques

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

Les différentes étapes pour l'obtention des MFCC sont :

1. Fenêtrage : Découpage du signal en fenêtre et application d'une fenêtre de Hamming sur ces portions.
2. RFFT : Application d'une transformée de Fourier sur chacune des portions, on obtient Le spectre.
3. Filtre de Mel : Création du banc de filtres, il s'agit de plusieurs filtres qui vont chacun Couvrir une fréquence, ils permettent de mieux simuler le fonctionnement de L'oreille humaine.
4. Calcul du Log.
5. DCT : Application d'une DCT (Discrete Cosine Transform) sur les portions et on Obtient alors les coefficients cepstraux :MFCC.

Pour le calcul de ces fonctions nous utilisons les bibliothèques de MATLAB, les fonctions utilisées Sont :

1. Enframe : découpe le signal en utilisant la fonction de pondération.
2. Melbanc : création d'un banc de filtre.
3. Melcepst : création de la matrice des coefficients cepstraux.
4. RTCT : calcul de la transformée en cosinus discrète en réel.

5. RFFT : calcul de la transformées de Fourier discrète réel

3. Exemple d'extraction de caractéristiques

Pour visualiser les différentes étapes de calcul des MFCC, prenons un exemple d'extraction

Des caractéristiques un frame du mot (bonjour) :

3.1 Le signal enregistré

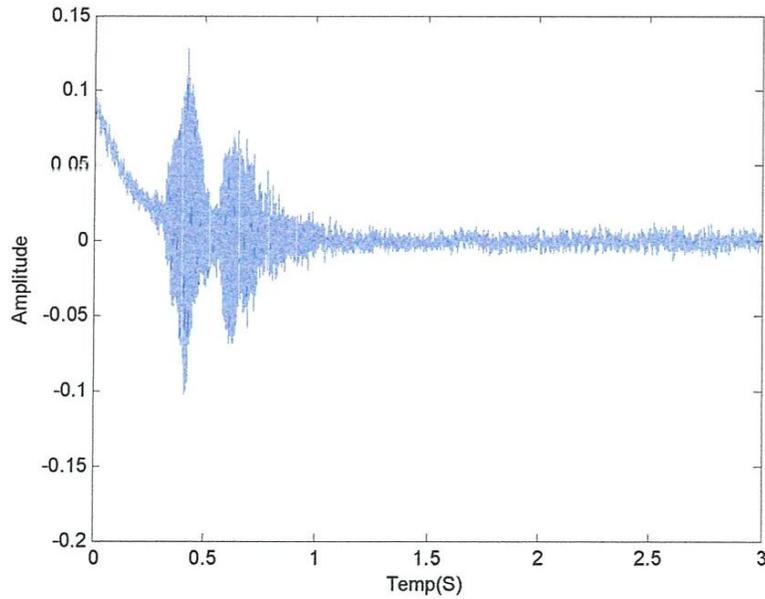


Figure3.1 : Signal enregistré (mot : Bonjour)

3.2 Un frame après fenêtrage :

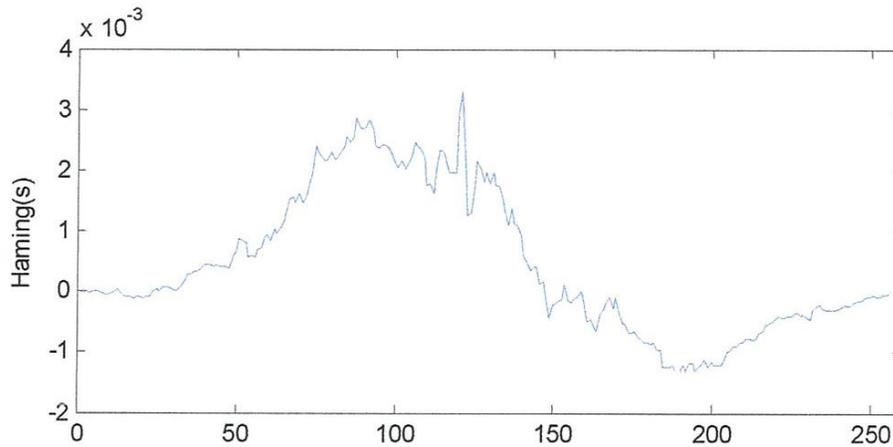


Figure (3.2) : Un frame

3.3 RFFT du frame :

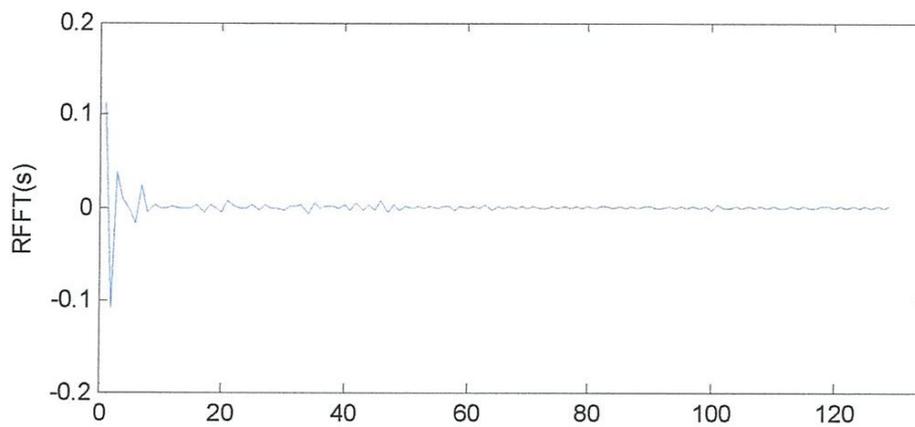


Figure (3.3) : Transformée rapide de fourrier réelle d'un frame

3. Phase de classification

3.1. Algorithme d'apprentissage LVQ

Étape 0

Initialiser les vecteurs de poids pour les M premiers vecteurs d'apprentissage.
(M est nombre de catégories)

Étape 1

Tant que la condition d'arrêt n'est pas atteinte exécuter les étapes 2 à 5

Étape 2

Pour chaque vecteur d'apprentissage X, exécuter les étapes 3 et 4

Étape 3

Déterminer J pour lequel $D(j)$ est minimale

Étape 4

Mise à jour des poids de $J^{\text{ème}}$ neurone comme suit:

Si $T(X) = C_j$

$$W_j(\text{New}) = W_j(\text{old}) + \alpha(X - W_j(\text{old}))$$

Sinon

$$W_j(\text{New}) = W_j(\text{old}) - \alpha(X - W_j(\text{old}))$$

Étape 5

Test de la condition d'arrêt d'arrêt

Cela peut être un nombre d'itérations, l'erreur ou le taux d'apprentissage

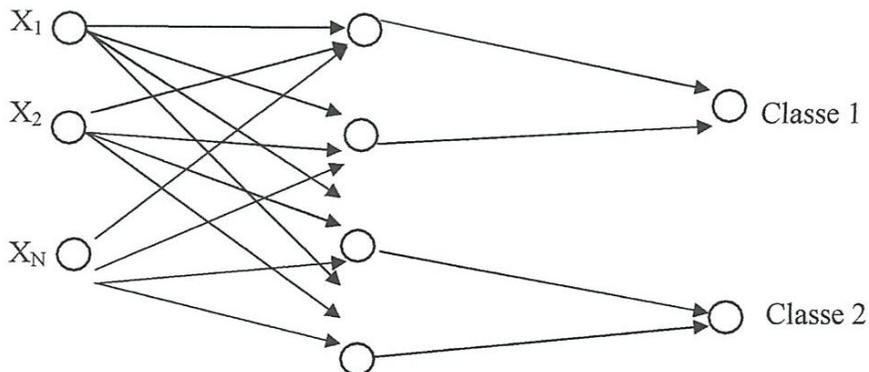


Figure 3.4 : Classification par LVQ

3.2. Exemple de calcul par LVQ

La figure représente un exemple de classification à deux classes. Nous utilisons un LVQ avec deux vecteurs. Les valeurs initiales sont $V1 = [1 \ 3 \ 6 \ 8 \ 9 \ 1]$ et $V2 = [3 \ 4 \ 1 \ 3 \ 1 \ 6]$, le pas

D'apprentissage est $\alpha=0.5$. Le changement des deux vecteurs durant la phase D'apprentissage pour 5 itération est illustré sur la figure.

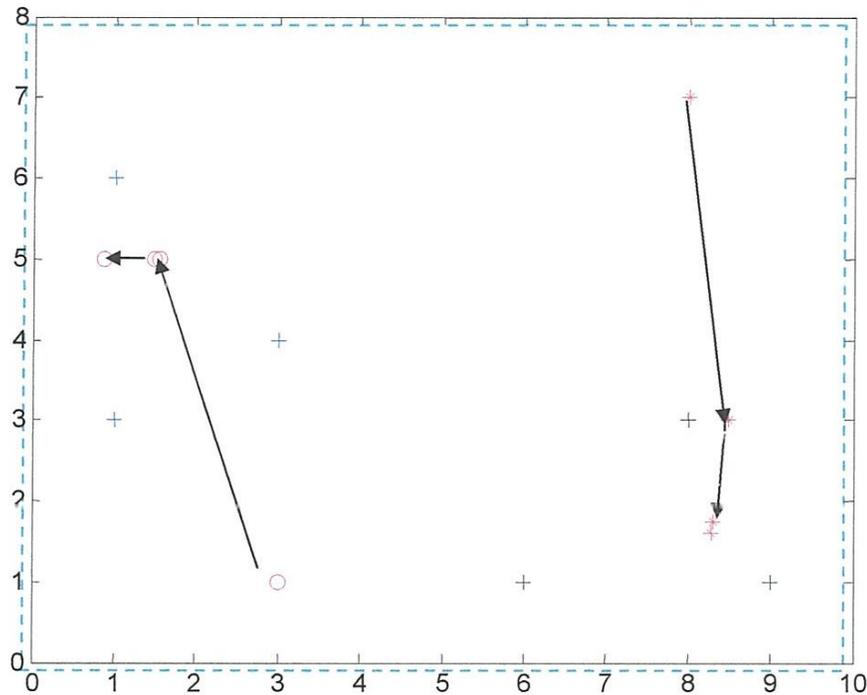


Figure (3.5) : Exemple de classification par LVQ

3. Base de données

Nous réalisons une petite base de données avec des noms et prénoms prononcés par quatre locuteurs :

Locuteur 1 : adulte : masculin

Locuteur 2 : adulte : masculin

Locuteur 3 : enfant : masculin

Locuteur 4 : enfant : féminin

Chaque locuteur prononce son nom et prénom 10 fois et prononce aussi les noms et les prénoms des trois autres locuteurs 10 fois pour chacun. Nous réalisons ensuite une base de données pour chaque locuteur dans laquelle 40 enregistrements : 10 prononcés par le locuteur lui-même et les 30 restants par les autres. Donc nous avons 4 bases de données.

Pour évaluer les performances de généralisation de notre système, nous divisons chaque base en deux parties : une pour l'apprentissage et l'autre pour le test. Nous réalisons cette procédure deux fois (cross validation d'ordre 2)

4. Résultats

Les figures représentent des exemples de l'évolution de l'erreur durant l'apprentissage pour deux locuteurs

4.1 Locuteur 1:

	Itération	Apprentissage	Test		
			Bonne	Mouvais	moyen
Base1	28	100%	100%		100%
Base2	20	100%	100%	95%	99%
Moyenne	24	100%	99.5%		

Tableau (3.1) : Résultat d'apprentissage et de test sur la base locuteur1

4. Locuteur 2 :

	Itération	Apprentissage	Test		
			Bonne	Mouvais	moyen
Base1	24	100%	100%	90%	96%
Base2	20	100%	100%	95%	99%
Moyenne	22	100%	97.5%		

Tableau (3.2) : Résultat d'apprentissage et de test sur la base locuteur2

4.1 Locuteur 3 :

	Itération	Apprentissage	Test		
			Bonne	Mouvais	moyen
Base1	2352	100%	100%		100%
Base2	76	100%	100%	95%	99%
Moyenne	1214	100%	99.5%		

Tableau (3.3) : Résultat d'apprentissage et de test sur la base locuteur3

4.1 Locuteur 4 :

	Itération	Apprentissage	Test		
			Bonne	Mouvais	moyen
Base1	4056	100%	100%	100%	100%

Base2	3452	98%	80%	90%	87%
Moyenne	3754	99%	93.5%		

Tableau (3.4) : Résultat d'apprentissage et de test sur la base locuteur4

4.5 La moyenne

	Itération	apprentissage	test
La moyenne	2192	99,75%	97,5%

Tableau (3.1) : Résultat d'apprentissage et de test sur la base locuteur1

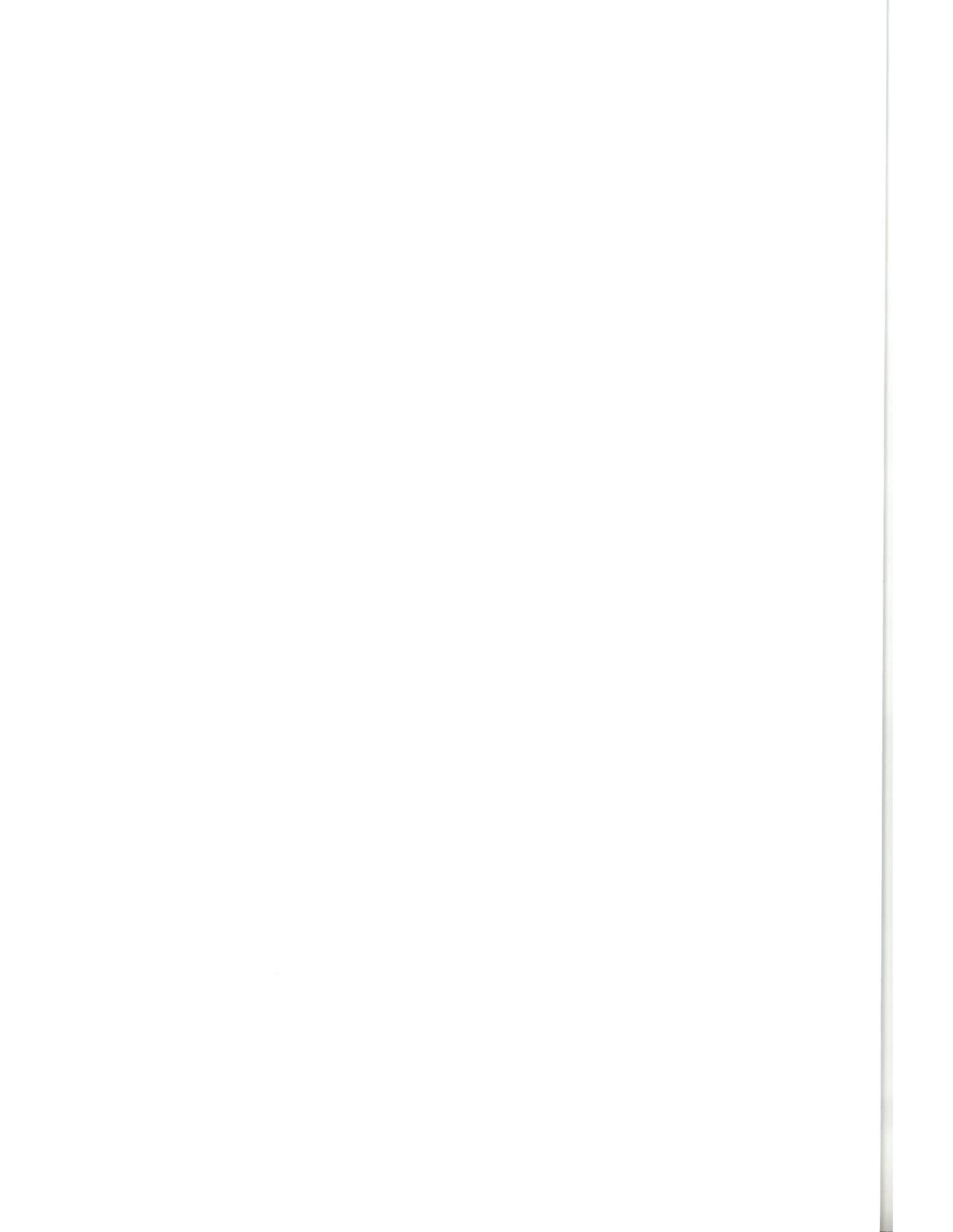
5. Conclusion :

Les testes classification réalisées sur les quatre bases de données nous permettent de constater que les résultats sont satisfaisant ce qui montre que les MFCC permet une bonne Caractérisation des données. D'autre part nous constatons que les réseaux de neurones sont Des bons classificateurs en matière d'apprentissage et de génération .Néanmoins , nous Avons constaté que les réseaux de neurones souffrent du problèmes de manque de Techniques d'initialisation de la difficulté d'interprétation. Les avantages et les inconvénients Des réseaux de neurones peuvent être récapitulés dans le tableau :

Avantage	Inconvénient
-Calcul parallèle	-Manque de technique d'initialisation
-Capacité de généralisation	-Boite noire
-Après apprentissage, classification rapide	-Lenteur d'apprentissage

Dans ce travail on s'est intéressé à la vérification automatique du locuteur. Pour l'extraction des caractéristiques, on s'est basé sur les coefficients cepstreux dans l'échelle de Mels(MFCC) qui sont les plus utilisés. Dans la phase de classification, nous avons utilisé le LVQ qui est un réseau de neurones très utilisé dans le domaine de reconnaissance du locuteur.

On a réalisé une petite base de données avec des noms et prénoms prononcés par quatre locuteurs. Les tests de classification réalisés sur nous permettent de constater que les résultats sont satisfaisants. Ceci montre que les MFCC permettent une bonne caractérisation des données. D'autre part nous constatons que les réseaux de neurones sont de bons classificateurs en matière d'apprentissage et de généralisation. Néanmoins, nous avons constaté que les réseaux de neurones souffrent du problème de manque de techniques d'initialisation et de la difficulté d'interprétation.



- [1] Olivier Morel / Reconnaissance des formes et classification / mémoire de master
- [2] Mémoire master 2 utilisation des réseaux neurones pour la reconnaissance automatique de la parole –gadouche salah eddine +bouzoualegh ahmed.
- [3] Laurent HEUTTE Reconnaissance de Formes Définitions et Applications Université de Rouen
- [4] Teva marlin /AMIRAL, UNE PLATEFORME GÉNÉRIQUE POUR LA RECONNAISSANCE AUTOMATIQUE DU LOCUTEUR — DE L'AUTHENTIFICATION À L'INDEXATION/Thèse de doctorat UNIVERSITÉ D'AVIGNON ET DES PAYS DE VAUCLUSE
- [5] Thèse de doctorat/ la reconnaissance automatique du locuteur avec MFCC /Semet Geatan et Treffot Gregory.
- [6] [Dominique Genoud] Reconnaissance et transformation de locuteurs de / Thèse de doctorat /l'Ecole Polytechnique Fédérale de Lausanne (EPFL)
- [7] Laurent Besacier (un modèle parallèle pour la reconnaissance automatique du locuteur) thèse de doctorat/ présentée à l'Université d'Avignon et des Pays de Vaucluse
- [8] Nmissi Mohamed/thèse de doctorat /Traitement du signal Création de l'empreinte vocale/université de guelma

résumé

Dans ce travail on s'intéresse à la vérification automatique du locuteur. Pour l'extraction des caractéristiques, on se base sur les coefficients cepstreux dans l'échelle de Mels (MFCC) qui sont les plus utilisés. Dans la phase de classification, on utilise le LVQ qui est un réseau de neurones très utilisé dans le domaine de reconnaissance du locuteur.