

11/621.720
République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
Université 8Mai 1945 – Guelma
Faculté des Sciences et de la Technologie
Département d'Electronique et Télécommunications



**Mémoire de Fin d'Etude
pour l'obtention du Diplôme de Master Académique**

**Domaine : Sciences et Techniques
Filière : Electronique
Spécialité : Systèmes Electroniques**

**RECONNAISSANCE VOCAL DES CHIFFRES ARABE ET
FRANÇAIS PAR LES RESEAUX DE NEURONES**

Présenté par :

- **Fernane Mehdi**
- **Brahmia Mohammed el amine**

Sous la direction de :

DR : Nemissi Mohammed



JUIN 2012

REMERCIEMENT



Avant tout, nous tenons a remercier notre dieu qui nous a accordé la Volonté et le courage pour la réalisation de ce projet et a défié les Obstacles.

C'est avec un immense plaisir que nous exprimons nos remerciements les plus chaleureux a notre encadreur « Nemissi Mohammed » qui a Suivi ce travail.

Nous accordons nos remerciement a nos amis qui nos ont aidé par leurs Efforts et a tous. Les personnels et les enseignants d'électronique Et a tous personne ayant contribué de près ou de loin a la réalisation De ce travail.

Nous tenons a remercier aussi les membres de jury qui ont accepté De juger ce travail.

A tout notre enseignant dans notre vie.

À tout on vous dit merci.

MOHAMMED EL ANIME

ET MEHDY



DEDICACE :

Je dédie ce modeste travail à :

La plus chère être que j'ai dans ce monde;

« ma chère mère Atika »

A l'âme de mon père khoudir « rabi yarahmou ».

Mes sœurs Meryem et la petite Yakouta.

Mon petit frère Abd el Basset.

Mon grand père El arbi et mes grandes mères

Yamina Hania et Zineb .

ainsi que mes oncles et tentes.

Mes cousins et cousines.

Mes amis et amies : hamza khamis saleh

À la plus proche sousou

Mes collègues zaki mehdi imen et samira

A toute la promotion 20011 |2012 précisément le groupe

électronique et tous qui me connais

BRANNA MOHAMED EL ANME

DEDICACE :

Je dédie ce modeste travail à :

La plus chère être que j'ai dans ce monde;

« Mes chères parents rabeih et fatima el zahraa »

Mon grand frère farouk.

Ma sœur hiba.

Mon petit frère wassim.

ainsi que mes oncles et tentes chrifa et abd el azize.

Mes cousins et cousines zaki.....

Mes collègues messbah et imen.....

Mes amies khaled et yahya.....

À la plus proche simsim

À toute la promotion 2011 | 2012 précisément le groupe électronique

À tous qui me connais

FERHANE MEHDY

Sommaire

Introduction générale	1
Introduction générale.....	2
CHAPITRE 1 : La reconnaissance automatique de la parole	3
1. Introduction	4
2. Etude Phonétique.....	5
2.1. Production de la parole.....	5
2.2. Audition.....	5
3. Caractéristiques du signal acoustique de la parole.....	7
3.1. Variabilité intra locuteur.....	7
3.2. Variabilité interlocuteur.....	8
3.3. Variabilité due aux conditions d'enregistrement.....	9
4. Le Prétraitement de la parole.....	10
4.1. Acquisition.....	10
4.2. Préaccentuation.....	12
4.3. Fenêtrage.....	12
5. Analyse et traitement d'un signal vocal.....	15
5.1. La fréquence fondamentale	15
5.2. Analyse spectrale.....	15
5.2.1. La transformation de Fourier.....	15
5.2.2. L'échelle des Mels.....	16
5.2.3. Banc de filtres sur l'échelle de Mels	17
5.3. Analyse temporelle	19
5.3.1. Energie totale.....	19
5.3.2. La densité de passage par zéro ZCR (Zero Crossing Rate)	19
5.4. Analyse cepstrale.....	20
5.4.1. Cepstre réel	21
5.4.2. Coefficients MFCC (Mel-scaled Frequency Cepstral Coefficients).....	22
6. Conclusion.....	23

CHAPITRE 2 : Réseaux de neurones	24
1. Introduction	25
2. Définition d'un réseau de neurones	25
3. Principes de modélisation des Réseaux de neurones	25
3.1. Les neurones	25
3.2. Fonctionnement de neurone biologique	27
3.3. Les neurones formels	28
3.4. Fonction d'activation	28
4. Architectures de réseaux de neurones	29
4.1. Les réseaux non bouclés	29
4.2. Les réseaux bouclés	29
4.3. Structure d'interconnexion	29
5. Domaines d'applications des réseaux de neurones	31
6. Apprentissage d'un réseau de neurones	31
6.1. Définition	31
6.2. Protocoles d'apprentissages	32
6.3. Les types d'apprentissage	32
6.4. Règles d'apprentissage	33
7. Réseaux de neurones de base	35
7.1. L'Adaline	35
7.2. Le Perceptron	36
8. Le Perceptron multi couche (MLP)	36
8.1. Architecture du MLP	36
8.2. Apprentissage du MLP par la retro-propagation	38
9. Conclusion	39

CHAPITRE 3 : Applications	40
1. Introduction	41
2. Extraction des caractéristiques	41
3. Exemple d'extraction de caractéristiques	42
3.1. Le signale enregistré	42
3.2. Un frame après fenêtrage	43
3.3. RFFT du farme	43
3.4. Le banc de filtres	44
3.5. Calcul du log	44
3.6. Calcul de la RDCT	45
4. Phase de classification	45
4.1. Algorithme d'apprentissage de la Back-Propagation	45
4.1.1. Etape 1 : introduction des données	45
4.1.2. Etape 2 : détermination des dimensions de données	46
4.1.3. Etape 3 : Apprentissage	46
4.1.4. Etape 4 : Fin	47
4.2. Exemple de classification	47
5. Reconnaissance des chiffres français	48
5.1. Initialisation des nombre des neurones cachés	48
5.2. Tests mono-locuteur	49
5.3. Tests multi-locuteur	50
6. Reconnaissance des chiffres arabe	51
6.1. Tests mono-locuteur	51
6.2. Tests multi-locuteur	52
7. Conclusion	53
Conclusion générale	54
Bibliographie	55
Résumé	58

Liste des figures

Figure 1.1: L'appareil phonatoire.....	5
Figure 1.2: Le système auditif.....	6
Figure 1.3: Le champ auditif humain.....	6
Figure 1.4: Variabilité intra locuteur	7
Figure 1.5: Variabilité interlocuteur.....	8
Figure 1.6: Variabilité due a l'enregistrement.....	9
Figure 1.7: Acquisition.....	11
Figure 1.8: La préaccentuation.....	12
Figure 1.9: La fenêtre rectangulaire	13
Figure 1.10: La fenêtre de Hamming	14
Figure 1.11: La fenêtre de Hanning	14
Figure 1.12: Traitement par transformé de Fourier.....	16
Figure 1.13: Graphe de conversion de la fréquence d'hertz à Mels.....	17
Figure 1.14: Représentation d'un banc de filtres dans l'échelle des Mels.....	18
Figure 1.15: modèle de signal vocal $x(n)$	20
Figure 1.16: Calcul du cepstre complexe.....	21
Figure 1.17: Le calcul des coefficients cepstraux réels.....	22
Figure 1.18: calcul des coefficients MFCC.....	23
Figure 2.1: Le neurone biologique (œuvre d'artiste).....	26
Figure 2.2: Le schéma classique présenté par les biologistes.....	27
Figure 2.3: Neurone formel.....	28
Figure 2.4: Exemples de fonctions d'activation	28
Figure 2.5: Réseau multicouche.....	29
Figure 2.6: Réseau à connexions locales.....	30
Figure 2.7: Réseau à connexions récurrentes.....	30
Figure 2.8: Réseau à connexion complète.....	30
Figure 2.9: MLP avec une seule couche cachée contenant N neurones d'entrée, M neurones cachés et J neurones de sortie.....	37
Figure 3.1: Signale enregistré	42
Figure 3.2: Un frame.....	43

Figure 3.3: Transformée rapide de fourier réelle d'un frame.....	43
Figure 3.4: Banc de filtres.....	44
Figure 3.5: Calcule du log	44
Figure 3.6: Calcul du DRCT.....	45
Figure 3.7: Deux problèmes de classification (le 1 ^{er} linéairement séparable et le 2 ^{ème} non linéairement séparable).....	47
Figure 3.8: Evolution de l'erreur pour les deux problèmes.....	48
Figure 3.9: Evolution de l'erreur de chacun des deux locuteurs.....	49
Figure 3.10: Eévolution de l'erreur des deux locuteurs.....	50
Figure 3.11: Evolution de l'erreur de chacun des deux locuteurs.....	51
Figure 3.12: Evolution de l'erreur des deux locuteurs.....	52

Liste des tableaux

Tableau 3.1: Résultat pour différent nombre de neurones cachés.....	49
Tableau 3.2: Résultat de tests locuteur 1 (Français).....	50
Tableau 3.3: Résultat de tests locuteur 2 (Français).....	50
Tableau 3.4: Résultat de tests multi-locuteur (Français).....	51
Tableau 3.5: Résultat de tests locuteur 1(Arabe).....	52
Tableau 3.6: Résultat de tests locuteur 2(Arabe).....	52
Tableau 3.7: Résultat de tests multi-locuteur (Arabe).....	53
Tableau 3.8: Comparaison entre les résultats de test arabe et français.....	53
Tableau 3.9: Avantages et inconvénients des réseaux de neurones.....	53

Liste des abréviations

FFT	Fast Fourier Transform (transformée de Fourier Rapide)
MLP	Multi Layered Perceptron
ZCR	Zero Crossing Rate (taux de passage par zéro)
GUI	Graphic Utilization Interface
TFD	Transformation de Fourier Discret
MFCC	Mel Frequency Cepstral Coefficients
RAP	Reconnaissance Automatique de la Parole
TZ	Transformation Z

Introduction générale

Introduction générale

La reconnaissance de la parole est un domaine d'études actif depuis le début des années 50. C'est l'une des composantes fondamentale 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, diphtongues, 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 reconnaissance vocale de chiffre arabe et français. Notre objectif est de mettre en œuvre un système de reconnaissance vocal de chiffres. Ce système est à un vocabulaire limité et à 2-locuteurs.

Pour l'extraction des caractéristiques, nous avons basé sur les coefficients cepstraux dans l'échelle de Mels (MFCC) qui sont les plus utilisés dans ce domaine. Dans la phase de classification, nous avons utilisé le PMC qui est le réseau de neurones le plus utilisé dans le domaine de la classification

Le présent mémoire comporte trois chapitres :

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

Le deuxième chapitre constitue une introduction aux réseaux de neurone et leur fonctionnement, ainsi que leurs algorithmes d'apprentissage.

Le troisième chapitre présente les résultats obtenus sur deux petites bases de données que nous avons élaborées. La première base en arabe et la deuxième en français.

Finalement, une conclusion générale conclue ce mémoire.

CHAPITRE 1

La reconnaissance automatique de la parole

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 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 [09] :

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

2. Etude Phonétique

2.1. Production de la parole

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 composé des cavités pharyngienne et buccale pour la plupart des sons. Lorsque la lèvre est en position basse, la cavité nasale vient s'y ajouter en dérivation.

L'air y passe librement pendant la respiration et la voix chuchotée, ainsi que pendant la phonation des sons non-voisés (ou sourds). Les sons voisés (ou sonores) résultent au contraire d'une vibration périodique des cordes vocales. Notons pour terminer le rôle prépondérant de la langue dans le processus phonatoire, elle détermine la hauteur du pharynx, le lieu d'articulation, ainsi que l'aperture, écartement des organes au point d'articulation; elle permet au conduit vocal d'avoir une géométrie et un volume extrêmement variable (figure 1.1) [10].

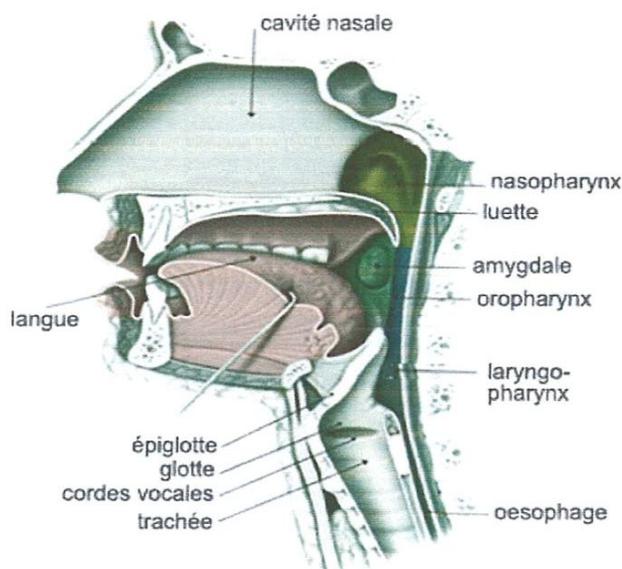


Figure 1.1: L'appareil phonatoire

2.2. Audition

Les ondes sonores sont recueillies par l'appareil auditif, ce qui provoque les sensations auditives. Ces ondes de pression sont analysées dans l'oreille interne qui envoie au cerveau l'influx nerveux qui en résulte.

L'appareil auditif comprend l'oreille externe, l'oreille moyenne, et l'oreille Interne.

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 nerveuses (figure 1.2) [10].

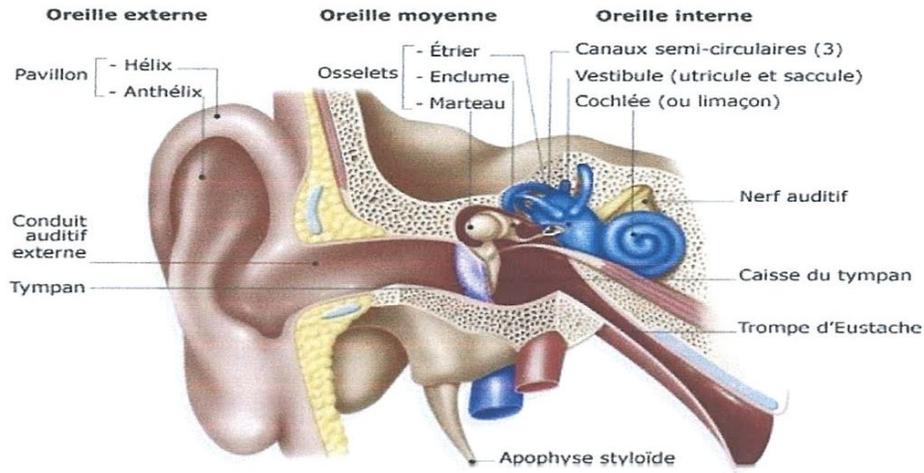


Figure 1.2: Le système auditif

Ainsi, l'oreille ne répond pas également à toutes les fréquences. La figure 1.3 présente le champ auditif humain, délimité par la courbe de seuil de l'audition et celle du seuil de la douleur. Sa limite supérieure en fréquence (16000 Hz, variable selon les individus) fixe la fréquence d'échantillonnage maximale utile pour un signal auditif (32000 Hz).

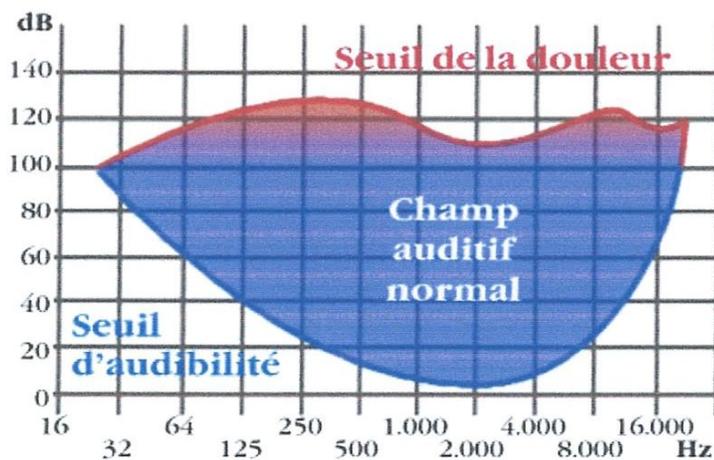


Figure1.3 : Le champ auditif humain

3. Caractéristiques du signal acoustique de la parole

Le signal acoustique de la parole est un peu particulier, il présente des caractéristiques qui rendent l'interprétation très complexe [08]. Ce signal est variable d'un locuteur à un autre (variabilité interlocuteur), et pour le même locuteur (variabilité intra-locuteur). Nous ajoutons les variabilités dues aux conditions d'enregistrement.

3.1. Variabilité intra locuteur

La variabilité intra locuteur exprime les différences dans le signal produit par une même personne. Cette variation peut résulter de l'état physique ou moral du locuteur. Une maladie des voies respiratoires peut ainsi dégrader la qualité du signal de parole de manière à ce que celui-ci devienne totalement incompréhensible, même pour un être humain. L'humeur ou l'émotion du locuteur peut également influencer son rythme d'élocution, son intonation ou sa phraséologie. Il existe un autre type de variabilité intra locuteur liée à la phase de production de la parole ou de préparation à la production de parole, due aux phénomènes de coarticulation (figure 1.4).

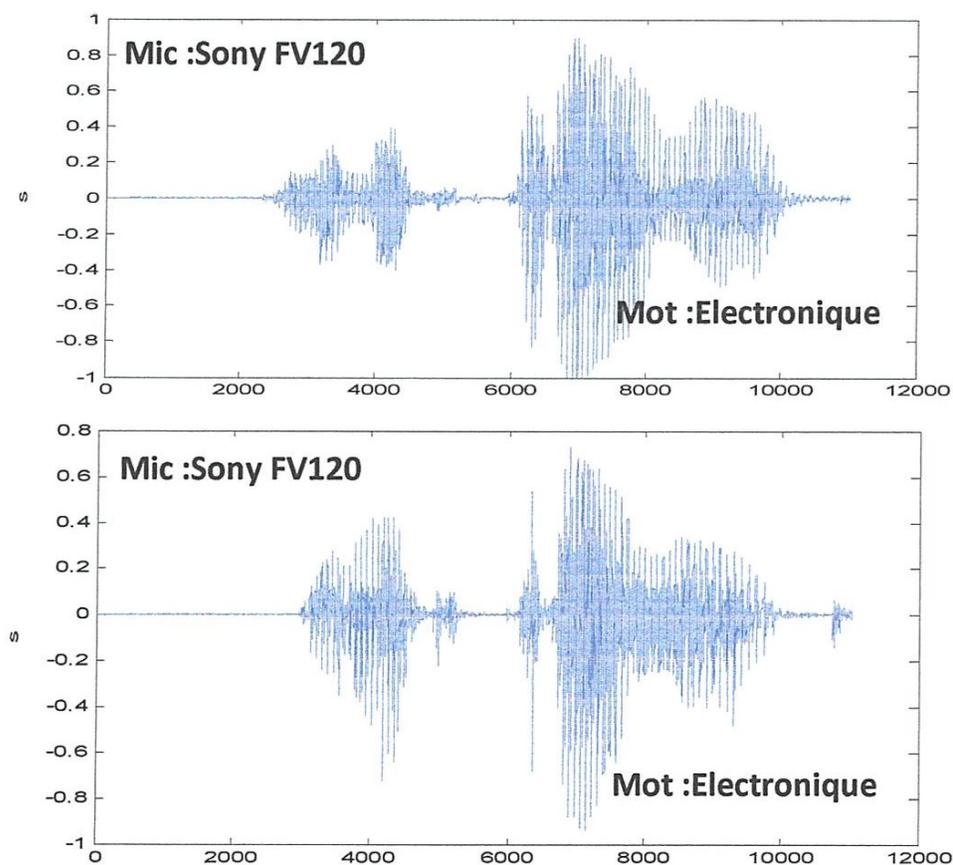


Figure 1.4: Variabilité intra locuteur

3.2. Variabilité interlocuteur

La variabilité interlocuteur est un phénomène majeur en reconnaissance de la parole.

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

La variabilité interlocuteur trouve également son origine dans les différences de prononciation qui existent au sein d'une même langue et qui constituent les accents régionaux (figure 1.5).

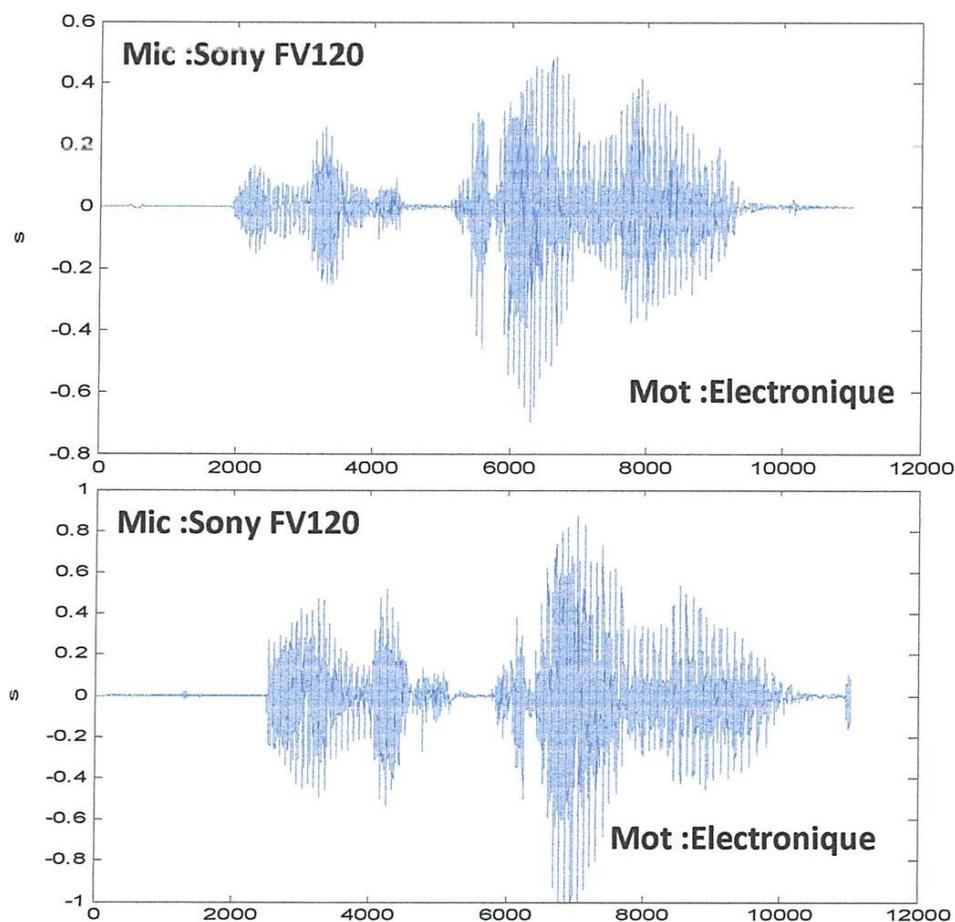


Figure 1.5: Variabilité interlocuteur

3.3. Variabilité due aux conditions d'enregistrement

Pour appliquer dans le commerce un système de reconnaissance vocal, il est important de connaître les effets de la transmission téléphonique sur un signal sonore.

La transmission de la parole par un canal téléphonique entraîne une limitation dans la gamme de fréquence, de 300 Hz à 3400 Hz de la bande passante.

La caractéristique de transfert n'est pas plate mais change de forme selon la ligne sélectionnée. Les spectres fournis par les lignes téléphoniques sont donc limités par la bande passante et également multipliés par une fonction de transfert de forme inconnue. Dans un premier stade, les études ont montré que la limitation des spectres de longue durée à la bande passante caractérisant la qualité du téléphone n'affecte pas sensiblement le taux d'identification.

Cependant, la pondération des spectres par des fonctions arbitraires du transfert, détruit la fiabilité de l'identification parce que, dans certains cas, l'effet de la fonction de transfert sur les spectres est plus important que les caractéristiques des voix (figure 1.6).

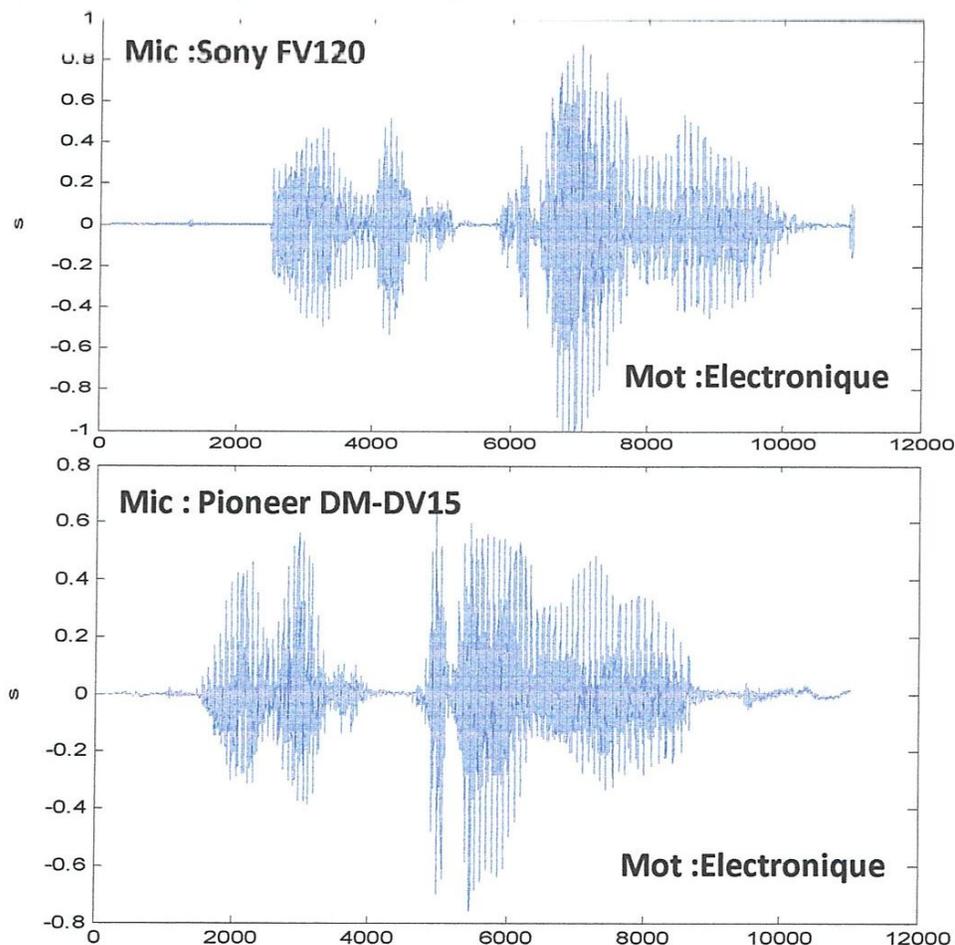


Figure 1.6: Variabilité due à l'enregistrement

4. Le Prétraitement de la parole

La reconnaissance automatique de la parole exige la réduction de la redondance du signal vocal à l'aide de traitements appropriés, afin de diminuer les temps de traitement et l'encombrement en mémoire. Par ailleurs, le traitement du signal vocal permet d'extraire des paramètres pertinents, aussi invariants que possible, pour la reconnaissance [09].

L'objectif des prétraitements est de faciliter la caractérisation de la forme de l'entité à reconnaître en réduisant la quantité d'information à traiter pour ne garder que les informations les plus significatives [03].

Les dispositifs utilisés ont d'abord été analogiques. Avec l'évolution de l'électronique numérique et de l'informatique, les techniques numériques sont désormais généralisées. Après numérisation du signal vocal, les traitements sont alors effectués par logiciel, initialement par des composants spécialisés et désormais par les micro-ordinateurs.

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

- Acquisition.
- Préaccentuation.
- Fenêtrage.

4.1. Acquisition

L'acquisition d'un signal vocal est liée avec la nature de l'application, et puisque on s'intéresse au traitement automatique (ordinateur), le signal de parole doit être numérisé.

De la sorte, on va pouvoir travailler sur une représentation spectrale du signal, décomposant ses différentes fréquences avec leurs amplitudes et leurs harmoniques, aboutissant à des « traits » qu'on appelle formants du signal.

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

La conversion analogique/numérique consiste en l'échantillonnage, la quantification et le codage.

L'échantillonnage est le processus de représentation d'un signal continûment variable comme une séquence de valeurs. La quantification conduit à représenter approximativement chaque échantillon dans un ensemble finit de valeurs.

Le codage consiste à assigner un numéro réel à chaque valeur. Avant l'échantillonnage, un filtre passe-bas de fréquence de coupure égale à la moitié de la fréquence d'échantillonnage est inséré pour éviter l'effet dénommé «repliement» ou « aliasing » postulé par le théorème de Nyquist-Shannon ; Ce filtre est donc appelé filtre « anti-repliement » ou « anti-aliasing ».

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

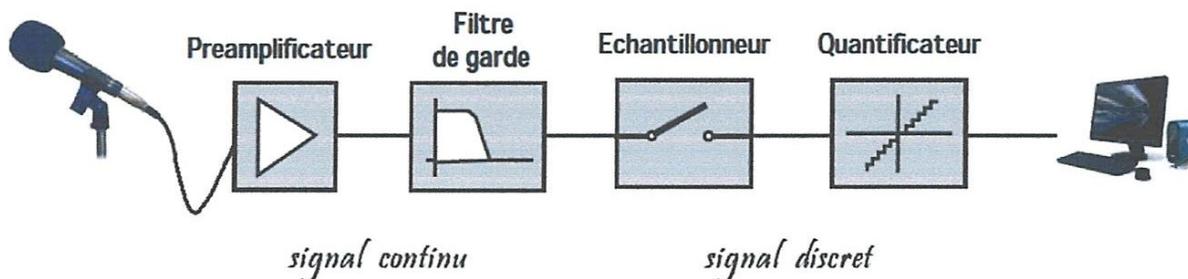


Figure 1.7: Acquisition

4.2. Préaccentuation

En général, le signal vocal se caractérise par une perte de 6 dB/Octave, due à l'influence de la source d'excitation et au rayonnement des lèvres. Une perte de 6 dB/Octave veut dire que les hautes fréquences ont une énergie plus faible que celle des basses fréquences. Pour pallier à cet inconvénient la préaccentuation permet d'égaliser les sons aigus avec les sons graves (figure 1.8).

L'opération consiste à faire passer le signal à travers un filtre de transmittance :

$$H(Z) = 1 - \mu * Z^{-1} \quad \text{où } 0 \leq \mu \leq 1 \quad (1.1)$$

Dans le domaine des signaux discrets (*échantillonnés*) $S(n)$ le problème consiste habituellement à calculer :

$$Y(n) = X(n) - \mu * X(n-1) \quad \text{pour } n \geq 0. \quad (1.2)$$

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

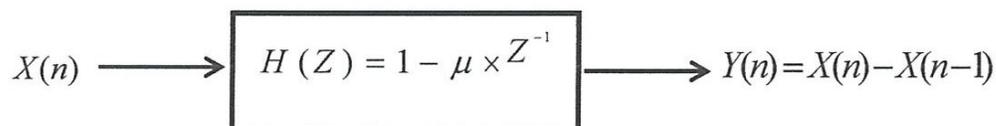


Figure 1.8: La préaccentuation

4.3. Fenêtrage

Le principe de cette phase de traitement est de minimiser la déformation du spectre dans les hautes fréquences, due au découpage en trames imposé 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 [12].

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 (chacune de durée 30ms environ) où il peut être considéré localement comme quasi-stationnaire.

En outre, et pour profiler de l'évolution lente du signal vocal, le fenêtrage permet le traitement en temps réel et il facilite aussi l'analyse des signaux sur la machine. Les ressources d'une machine étant limitées, le signal ne peut pas être traité dans sa globalité.

Du fait que le signal de parole est non-stationnaire il faut mettre en évidence les dépendances qui existent entre les échantillons, pour se faire, deux tranches adjacentes doivent avoir une région commune (de 5 à 12 millisecondes environ).

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

- **Fenêtre rectangulaire :**

Elle est définie par :

$$f(nT) = \begin{cases} 1 & \text{si } |nT| < T'. \\ 0 & \text{ailleurs.} \end{cases}$$

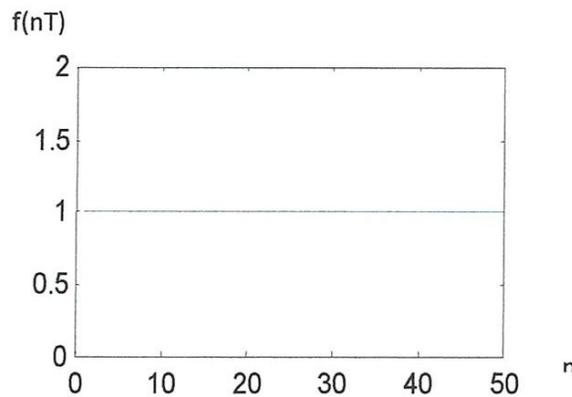


Figure 1.9: La fenêtre rectangulaire et son spectre

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

T' est la moitié de longueur de la fenêtre.

- **La fenêtre de Hamming**

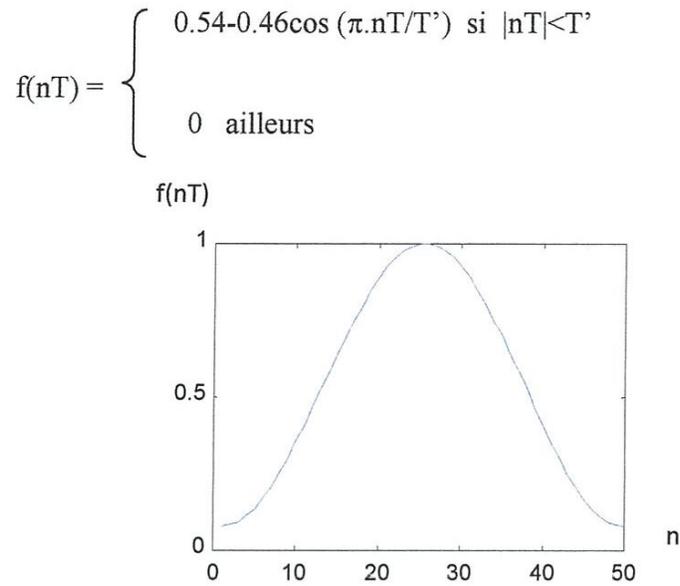


Figure 1.10: La fenêtre de Hamming et son spectre

- **La fenêtre de Hanning**

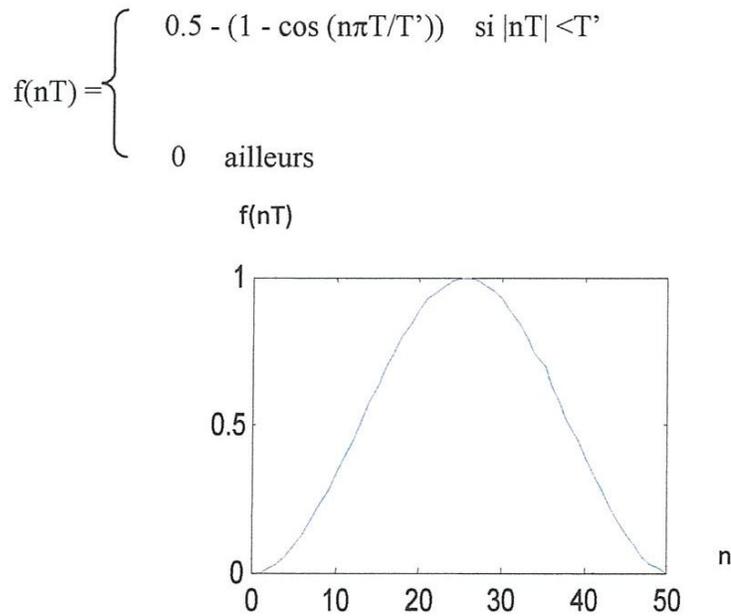


Figure 1.11: La fenêtre de Hanning et son spectre

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

5. Analyse et traitement d'un signal vocal

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

5.1. La fréquence fondamentale

La période du fondamental est par définition la fréquence de vibration des cordes vocales. Elle est aussi appelée (pitch), c'est un paramètre très important dans les différentes applications de la parole [18].

L'opération d'extraction du pitch est une tâche difficile pour trois raisons principales:

- La vibration des cordes vocales n'a pas nécessairement une périodicité complète.
- Il est difficile de séparer le pitch des effets du trait vocal.
- La plage de la fréquence du fondamental est très grande. 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.

5.2. Analyse spectrale

Elle est fondée sur une décomposition du signal sans connaissance à priori de sa structure fine ou de sa source. Dans ce domaine, on trouve principalement la TFD, et le vocodeur à canaux.

5.2.1. La transformation de Fourier

Lorsqu'on veut analyser une entité complexe, une des procédures largement utilisées consiste à la décomposer en une somme d'entités plus simples. L'idée donc est d'exprimer le signal vocal par une combinaison linéaire discrète de fonctions élémentaires de forme simple [09] ; c'est le cas de la transformée de Fourier discrète pour l'estimation du spectre.

La TFD est définie par :

$$s(n) = \sum_{k=0}^{N-1} s(k) \times e^{-j\pi \frac{nk}{N}} \quad (1.3)$$

$S(k)$ est un signal numérique.

N est la longueur du support du signal.

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

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

La séquence des opérations est schématisée sur la figure 1.12.

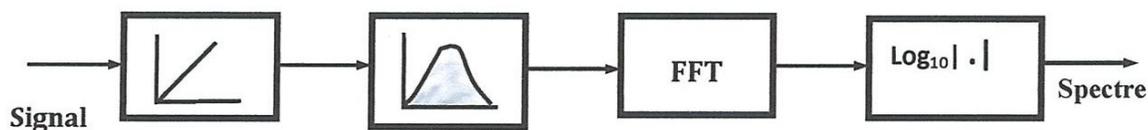


Figure 1.12: Traitement par transformé de Fourier

5.2.2. L'échelle des Mels

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

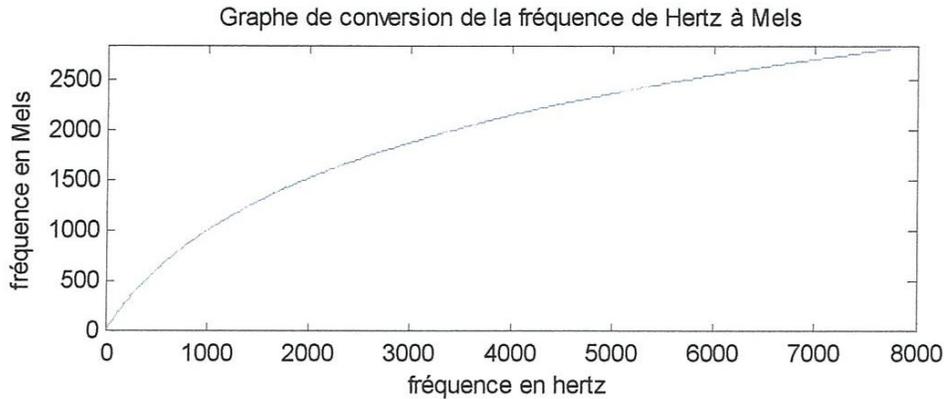


Figure 1.13: Graphes de conversion de la fréquence de Hertz à Mels

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

5.2.3. Banc de filtres sur l'échelle de Mels

Du fait que l'étendue des fréquences présentes dans le spectre est encore très large (par conséquent beaucoup de données à traiter), on a recours au banc de filtres dans l'échelle de Mels. On relie ainsi le système de reconnaissance vocale au fonctionnement de l'oreille humaine. Il s'agit de filtres passes bandes (fonction fenêtre de Hamming) centrés linéairement dans le domaine fréquentiel des Mels et de largeur telle qu'ils divisent l'espace des fréquences de manière égale dans le domaine des Mels et qu'ils se recouvrent chacun par moitié. Les filtres sont donc disposés logarithmiquement dans l'échelle des fréquences usuelles (Hz) : on a ainsi beaucoup de filtres pour les basses fréquences alors que les hautes fréquences sont disposées plus largement.

La formule donnant la fréquence en Mels à partir de la fréquence en Hz est donnée par:

$$m = 2595 \cdot \log_{10} \left(1 + \frac{f}{700} \right) \quad (1.4)$$

Où f est la fréquence en Hertz.

Chaque filtre donne un coefficient cepstrale :

$$s_{i,k} = \sum_{n=0}^{n/2} Y_{i,n} M_{n,k}, \quad k = 0 \dots K \quad (1.5)$$

Avec : $Y_{i,n}$ le $n^{\text{ème}} \in [1, N]$ coefficient de la transformée de la $i^{\text{ème}} \in [1, I]$ frame, et $M_{n,k}$ le $n^{\text{ème}} \in [1, N]$ coefficient du $k^{\text{ème}} \in [1, K]$ filtre. On utilise communément 12 coefficients, on utilise alors $K=13$ filtres (pour obtenir 12 coefficients, il faut un filtre de plus car le $0^{\text{ème}}$ est inutile).

On a donc $S_{i,k}$ la matrice de sortie du $k^{\text{ème}}$ filtre pour la $i^{\text{ème}}$ frame. On a, à cette étape, ce qu'on appelle un Spectre Mel (Spectrum Mel).

Une représentation d'un banc de filtres dans l'échelle des Mels est donnée par la figure 1.14 : c'est une représentation de la matrice $N \times K$ des coefficients du banc de filtres (N : Nombre d'échantillons par frames, K nombres de filtres souhaités).

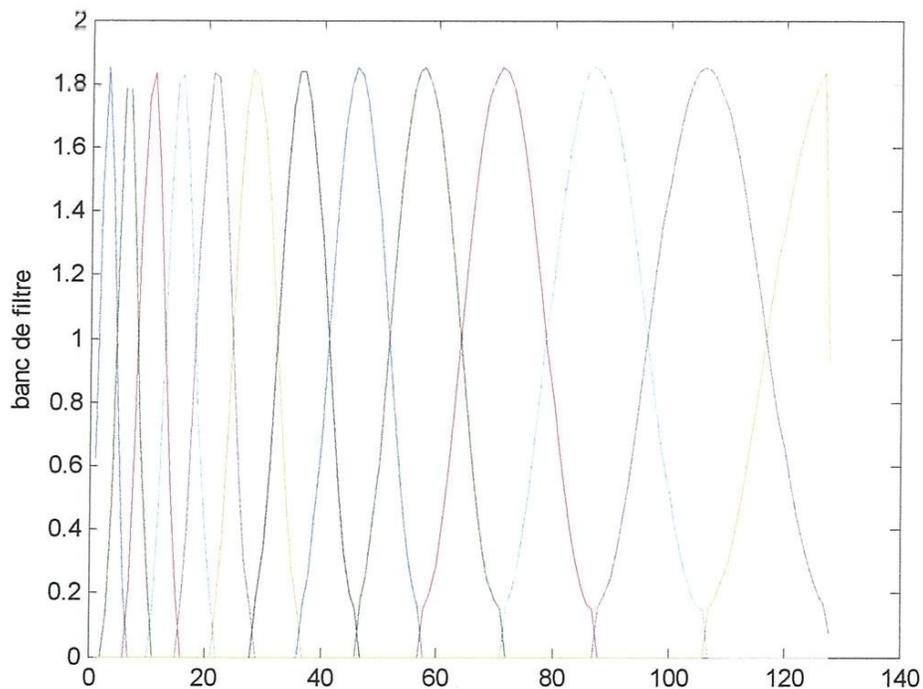


Figure 1.14: Représentation d'un banc de filtres dans l'échelle des Mels

5.3. Analyse temporelle

5.3.1. Energie totale

Elle est évaluée par :

$$E = \frac{1}{N} \sum_{k=0}^{N-1} S^2(k) \quad (1.6)$$

N est la longueur de la fenêtre.

$S(k)$ est le signal à traité.

Elle joue un rôle important pour délimiter les mots, et la localisation des voyelles qui sont caractérisées par une forte énergie par rapport aux autres phonèmes [17].

5.3.2. La densité de passage par zéro ZCR (Zero Crossing Rate)

ZCR correspond au nombre de changements de signe d'un échantillon à son successeur, dans une même trame [17].

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

$$ZCR_x(m) = \frac{1}{L} \sum_{n=m-l-1}^m |\text{sign}(x(n)) - \text{sign}(x(n-1))| \quad (1.7)$$

$$\text{Avec : } \text{sign}(x(n)) = \begin{cases} 1 & \text{si } x \geq 0 \\ 0 & \text{si } x < 0 \end{cases}$$

L : longueur de trame.

m : indice de trame.

n : indice des échantillons.

Elle est utilisée pour distinguer le signal de parole du silence

5.4. Analyse Cepstrale

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

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

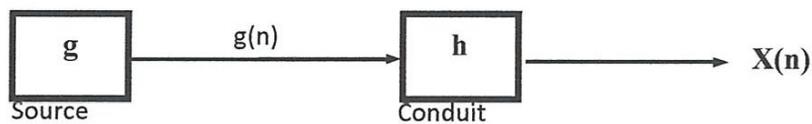


Figure 1.15: Modèle de signal vocal $x(n)$

Donc on peut écrire pour tout $n > 0$:

$$x(n) = g(n) * h(n) \quad (1.8)$$

Pour déconvoluer $x(n)$, c'est à dire pour retrouver les deux composantes $g(n)$ et $h(n)$, il faut se donner une classe de fonctions admissibles pour $g(n)$ (ou pour $x(n)$) : ici on suppose que $g(n)$ est une séquence d'impulsions (périodique pour les sons voisés). Il est évident que l'ensemble de ces hypothèses est très limitatif : en toute rigueur cette analyse ne s'applique théoriquement qu'aux parties stables des sons périodiques (voyelles longues, par exemple), dans la pratique, cependant, cette méthode fournit des résultats acceptables sur l'ensemble du signal [07].

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

D_{*}^{\dagger} est un homomorphisme (application) qui applique l'espace vectoriel des signaux d'entrées muni de la loi « * » (convolution), sur l'espace vectoriel des signaux de sortie muni de la loi « + » (addition), donc on est en face de la situation suivante :

$$x(n) = g(n) * h(n) \xrightarrow{D_{*}^{\dagger}} \hat{x}(n) = \hat{g}(n) + \hat{h}(n) \quad (1.9)$$

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

$$\begin{aligned}\hat{g}(n) &\xrightarrow{D_{*}^{\dagger}} g(n) \\ \hat{h}(n) &\xrightarrow{D_{*}^{\dagger}} h(n)\end{aligned}\quad (1.10)$$

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

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

$$D_{*}^{\dagger} = TZ(\cdot) \circ \log(\cdot) \circ TZ^{-1}(\cdot) \quad (1.11)$$

$$D_{*}^{\dagger} = TZ(\cdot) \circ \exp(\cdot) \circ TZ^{-1}(\cdot) \quad (1.12)$$

Ce qui donne le système schématisé dans la figure 1.16 :

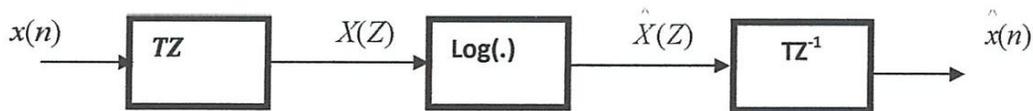


Figure 1.16: Calcul du cepstre complexe

Où :

- TZ est la transformée en Z (TZ^{-1} sa transformée inverse).
- La fonction log est utilisée pour le passage du domaine de la loi « . » (La multiplication) au domaine de la loi « + » (l'addition).

5.4.1. Cepstre réel

La difficulté du logarithme complexe (à cause de la phase) peut être levée dans le cas de la parole (où l'on ne s'intéresse que rarement à l'information de phase) en prenant un log module ($\log|\cdot|$), ce qui garantit dans 3 son inversibilité sans calcul particulier de la phase.

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

$$DM_{*}^{\dagger} = TZ(\cdot) \circ \log|\cdot| \circ TZ^{-1}(\cdot) \quad (1.13)$$

$$DM_{*}^{\dagger} = TZ(\cdot) \circ \exp|\cdot| \circ TZ^{-1}(\cdot) \quad (1.14)$$

Les coefficients du cepstre réel sont définis par :

$$C_n = DM_*^+[x(n)] \quad (1.15)$$

Ces coefficients sont réels, ils conservent le spectre d'amplitude, par contre l'information de phase est perdue.

En pratique, on peut remplacer avantageusement la transformée en Z par une transformation de Fourier rapide (FFT), celle-ci possède les mêmes propriétés de linéarité que la transformée en Z (figure 1.17).

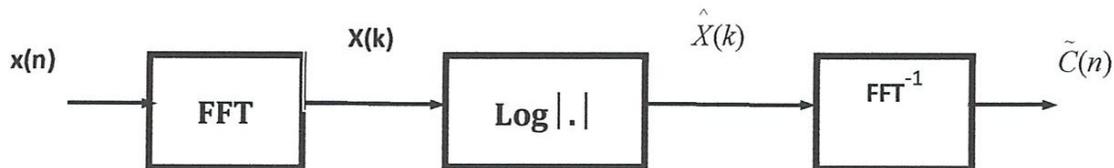


Figure 1.17: Le calcul des coefficients cepstraux réels

Dans ce cas les coefficients \tilde{C}_n sont donnés par :

$$\tilde{C}_n = \frac{1}{N} \sum_{k=0}^{N-1} \log|X(k)| e^{jk12\pi/N} \quad (1.16)$$

Sous certaines conditions, on peut admettre que $\tilde{C}_n \cong C_n$ donc on peut dire que les \tilde{C}_n sont les coefficients cepstraux approchés prenant leurs valeurs dans un domaine pseudo temporel appelé domaine QUEFRENTIEL [11].

5.4.2. Coefficients MFCC (Mel-scaled Frequency Cepstral Coefficients)

Les paramètres MFCC sont des coefficients cepstraux obtenus à partir des énergies d'un banc de filtre en échelle de fréquence Mel. Il s'agit en fait d'un calcul classique des coefficients cepstraux auquel on a rajouté, avant le logarithme un filtre de Mel. Ces résultats sont intéressants, car le calcul d'une dizaine de coefficients cepstraux est alors suffisant pour des expériences de RAP [01][05][11][19].

$$c_{i,n} = \sum_{k=1}^K (\log_{10}(s_{i,k})) \cos\left(n\left(k - \frac{1}{2}\right)\frac{\pi}{K}\right) \quad (1.17)$$

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 de sortie du $K^{\text{ème}}$ filtre pour la $i^{\text{ème}}$ frame

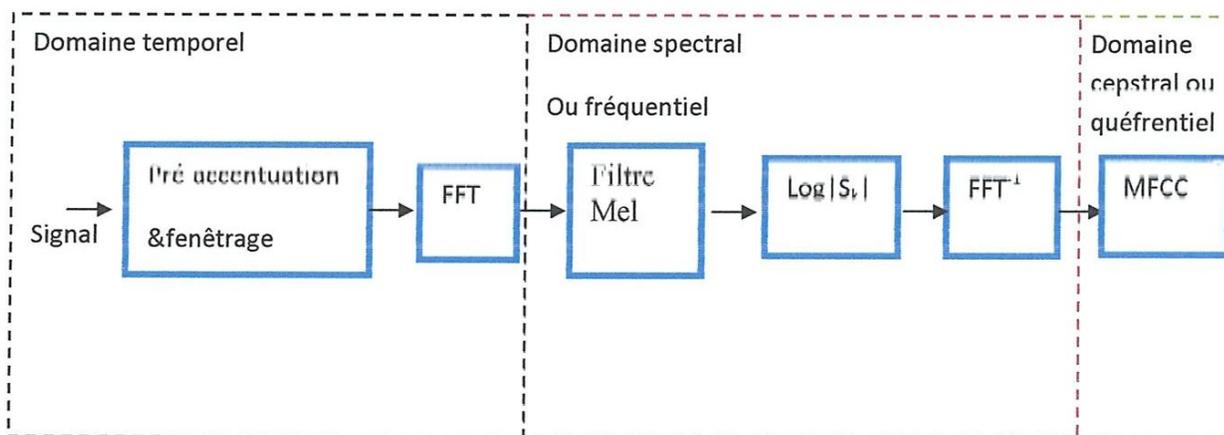


Figure 1.18: Calcul des coefficients MFCC

6. Conclusion

Dans ce chapitre nous avons donné un aperçu sur la reconnaissance automatique de la parole. Dans ce travail nous réalisons une extraction des caractéristiques donnée par coefficients MFCC : (Mel-scaled Frequency Cepstral Coefficients) qui sont les plus utilisés dans ce domaine.

CHAPITRE 2

Réseaux de neurones

1. Introduction

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

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

Inspirés des réseaux neuromémitiques biologiques, ils existent plusieurs modèles de réseaux de neurones artificiels, et chaque modèle se prête bien pour une application particulière (classification, reconnaissance, contrôle ... etc.) ; Mais leurs utilisations restent limitées dans quelques applications, et il reste beaucoup de domaines où les réseaux de neurones n'ont pas trouvé de solutions, telle que la planification par exemple. Les meilleurs systèmes à réseaux de neurones restent assez loin d'imiter des performances telles que celles de l'être humain.

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 [13].

3. Principes de modélisation des Réseaux de neurones

3.1. Les neurones

Les neurones biologiques sont des cellules neurone uses constituant la base du système nerveux central. Ils possèdent trois principales composantes : Les dendrites, le corps cellulaire et l'axiome (figure 2.1).

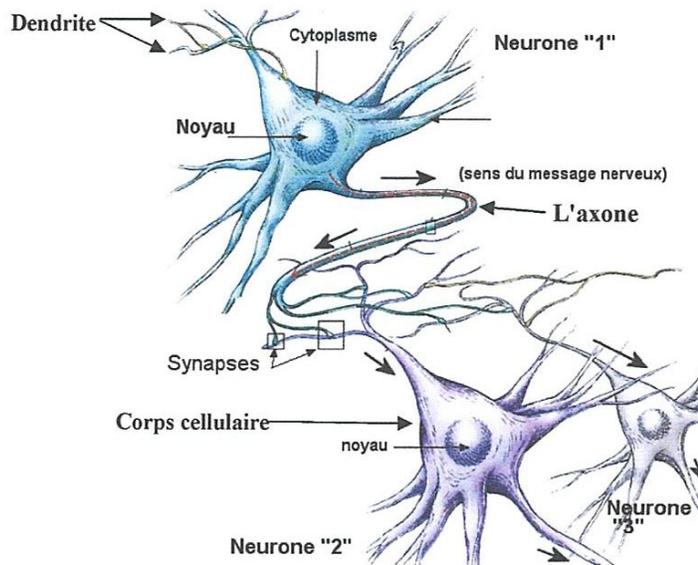


Figure 2.1: Le neurone biologique (œuvre d'artiste)

a) Dendrites

Chaque neurone possède une "chevelure" de dendrites. Celles-ci sont de fines extensions tubulaires, de quelques dizaines de microns. Elles se ramifient, ce qui les amène à former une espèce d'arborescence autour du corps cellulaire. Elles sont les récepteurs principaux du neurone pour capter les signaux qui lui parviennent [06].

b) Corps cellulaire

Il contient le noyau du neurone et effectue les transformations biochimiques nécessaires à la synthèse des enzymes et des autres molécules qui assurent la vie du neurone. Sa forme est pyramidale ou sphérique dans la plupart des cas. Elle dépend souvent de sa position dans le cerveau, ainsi les neurones du néo-cortex ont principalement une forme pyramidale ce corps cellulaire fait quelque microns de diamètre [06].

c) L'axone

L'axone, qui est à proprement parler la fibre nerveuse, sert de moyen de transport pour les signaux émis par le neurone. Il se distingue des dendrites par sa forme et par les propriétés de sa membrane externe. En effet, il est généralement plus long (sa longueur varie d'un millimètre à plus d'un mètre) que les dendrites, et se ramifie à son extrémité, là où il communique avec d'autres neurones, alors que les ramifications des dendrites se produisent plutôt près du corps cellulaire [06].

d) Les synapses

Ce sont des jonctions entre deux neurones et qui sont essentielles dans le fonctionnement du système nerveux.

3.2.Fonctionnement de neurone biologique

Les dendrites forment un maillage de récepteurs nerveux qui permettent d'acheminer vers le corps du neurone, des signaux électriques en provenance d'autres neurones. Celui-ci agit comme une espèce d'intégrateur en accumulant des charges électriques lorsque le neurone devient suffisamment excité, par un processus électrochimique, il engendre un potentiel électrique qui se propage jusqu'à trouver son axiome pour éventuellement venir exciter d'autres neurones. Le point de contact entre l'axiome d'un neurone et la dendrite d'un autre neurone s'appelle la synapse. Il semble que c'est l'arrangement spatial des neurones et de leur axiome, ainsi que la qualité des connexions synaptiques individuelles qui détermine la fonction précise d'un réseau de neurones biologique [08].

- Le schéma classique présenté par les biologistes (figure 2.2) est celui d'un somma qui exécute une sommation de tous les signaux transmis par ses dendrites et envoie un influx nerveux à son axiome ; si la sommation dépasse un certain seuil
- L'évolution des connexions entre les neurones permet la mémorisation et l'apprentissage tel que :

La mémorisation : sert à l'appel des propriétés acquises par l'apprentissage afin d'être utilisé aux modifications.

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

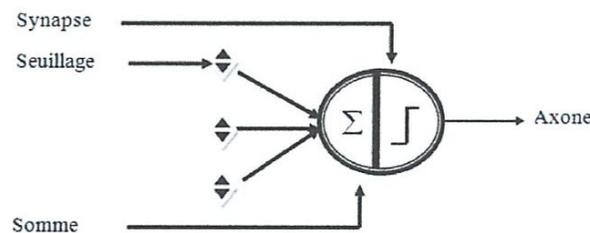


Figure 2.2: Le schéma classique présenté par les biologistes

3.3. Les neurones formels

Un neurone formel est une sorte d'automate, il se compose d'entrées, d'une fonction de sommation, d'une fonction d'activation et d'une ou plusieurs sorties. La figure 2.3 illustre un neurone de ce type.

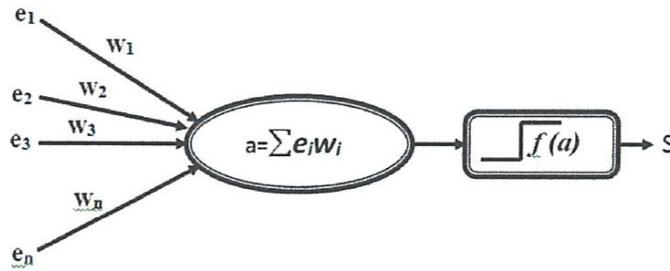


Figure 2.3: Neurone formel

3.4. Fonction d'activation

La fonction d'activation joue un rôle très important dans le comportement du neurone, elle donne une valeur représentative de l'activation du neurone. La fonction d'activation peut avoir plusieurs formes différentes, la figure 2.4 illustre quelques exemples.

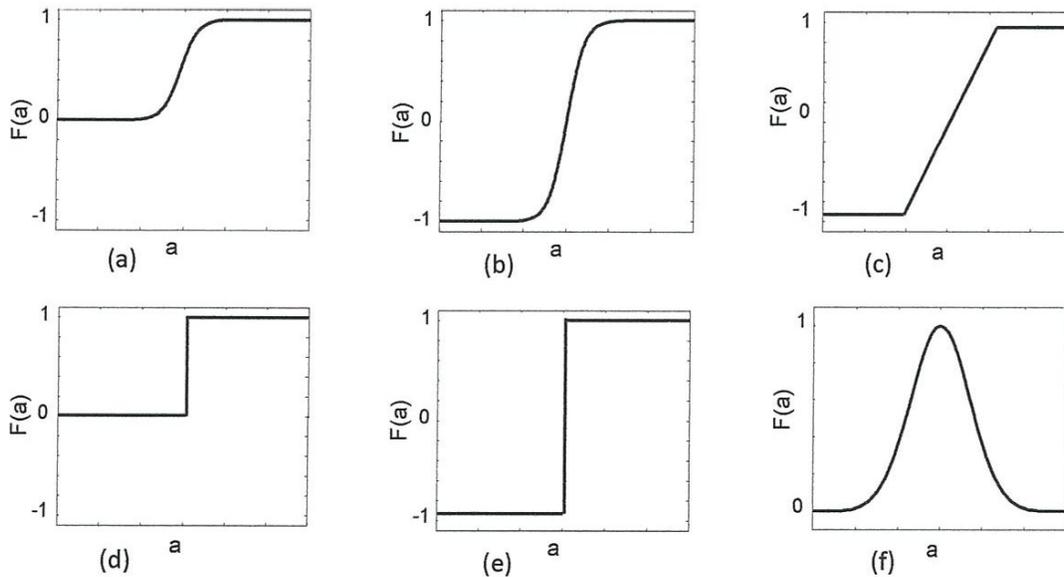


Figure 2.4: Exemples de fonctions d'activation

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

4. Architectures de réseaux de neurones

4.1. Les réseaux non bouclés

Un réseau de neurone est non bouclé, si son graphe ne possède pas de cycle. Dans tel réseau, l'information circule de l'entrée vers la sortie sans aucun retour. Un réseau de neurone non bouclé est dit acyclique. Citons l'exemple des réseaux multicouches.

4.2. Les réseaux bouclés

Un réseau est bouclé, si son graphe possède au moins un cycle. Un réseau bouclé fait ramener une ou plusieurs valeurs à l'entrée. Un réseau de neurone bouclé est donc un système dynamique, régi par des équations différentielles ; comme l'immense majorité des applications sont réalisées par des programmes d'ordinateurs, on se place dans le cadre des systèmes à temps discret, où les équations différentielles sont remplacées par des équations aux différences [07].

4.3. Structure d'interconnexion

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

a) Réseau multicouche

Les neurones sont arrangés par couche. Il n'y a pas de connexion entre neurones d'une même couche et les connexions ne se font qu'avec les neurones des couches avale. Habituellement, chaque neurone d'une couche est connecté à tous les neurones de la couche suivante et celle-ci seulement. Ceci nous permet d'introduire la notion de sens de parcours de l'information (de l'activation) au sein d'un réseau et donc définir les concepts de neurone d'entrée, neurone de sortie. Par extension, on appelle couche d'entrée l'ensemble des neurones d'entrée, couche de sortie l'ensemble des neurones de sortie.

Les couches intermédiaires n'ayant aucun contact avec l'extérieur sont appelés couches cachées [03].

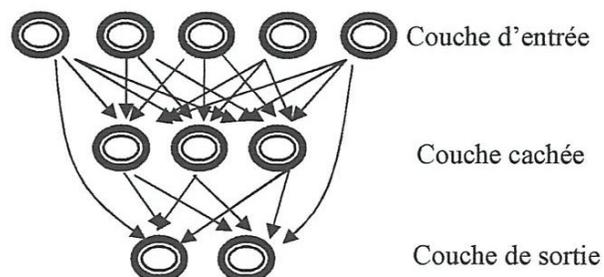


Figure 2.5: Réseau multicouche

b) Réseau à connexions locales

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

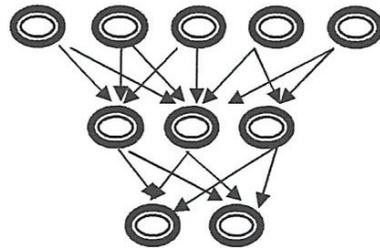


Figure 2.6: Réseau à connexions locales

c) Réseau à connexions récurrentes

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

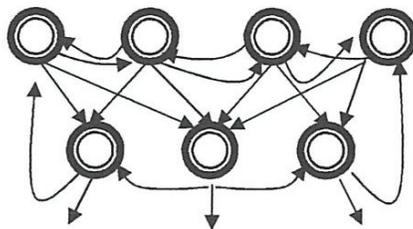


Figure 2.7: Réseau à connexions récurrentes

d) Réseau à connexion complète

C'est la structure d'interconnexion la plus générale. Chaque neurone est connecté à tous les neurones du réseau.

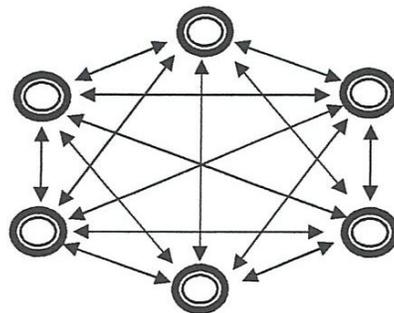


Figure 2.8: Réseau à connexion complète

5. 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 habilités d'apprentissage. De plus, l'estimation de leurs paramètres est indépendante de la complexité du problème traité ce qui leur permet d'être bien adaptés aux problèmes actuels qui ne cessent d'être de plus en plus complexes.

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 [13].
- 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 [13].

6. Apprentissage d'un réseau de neurones

6.1. Définition

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

Dans les algorithmes actuels, les variables modifiées pendant l'apprentissage sont les poids des connexions.

6.2. Protocoles d'apprentissages

L'apprentissage des réseaux de neurones s'effectue généralement en quatre étapes :

Etape 1: Initialisation des poids synaptiques avec des petites valeurs aléatoires

Etape 2: Présentation de l'exemple d'entrée et propagation de l'activation des neurones.

Etape 3: Calcul de l'erreur. Dans le cas d'un apprentissage supervisé cette erreur dépend de la différence entre l'activation des neurones et la sortie désirée.

Etape 4: Calcul du vecteur de correction à partir des valeurs des erreurs, avec lequel on effectue la correction des poids synaptiques.

6.3. Les types d'apprentissage

Les techniques d'apprentissage se subdivisent en trois grandes familles :

a) Apprentissage supervisé

Pour ce type d'apprentissage (perceptron, adaline, etc...), le réseau doit savoir qu'il a commis une erreur et il doit connaître la réponse qu'il doit donner. Pour un stimulus présenté à la couche d'entrée du réseau, la réponse obtenue est comparée avec celle désirée et la modification et les ajustements à apporter aux poids sont déterminés en fonction de l'erreur commise par le réseau. Généralement, les règles d'apprentissage supervisé sont des formes de descente du gradient.

L'apprentissage supervisé nécessite donc la définition d'une base d'exemples d'apprentissage représentative. Chaque exemple présenté au réseau est un couple (entrée, sortie désirée). La minimisation de l'erreur entre la valeur de sortie et la valeur désirée est basée sur le principe de l'erreur quadratique [04].

b) Apprentissage non supervisé

Pour ce type d'apprentissage il s'agit d'atteindre l'ensemble des poids synaptiques pour lesquels le comportement du réseau est optimal. La modification et l'ajustement des poids se font en fonction d'un critère interne, indépendant de la relation entre le comportement du réseau et la tâche qui doit être effectuée.

c) Apprentissage semi-supervisé

L'apprentissage semi-supervisé suppose qu'un comportement de référence précis n'est pas disponible, mais qu'en revanche, il est possible d'obtenir des indications qualitatives (correcte /incorrecte) sur les performances du réseau [04].

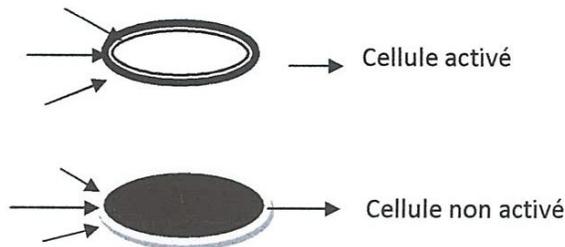
6.4.Règles d'apprentissage

a) La règle de Hebb

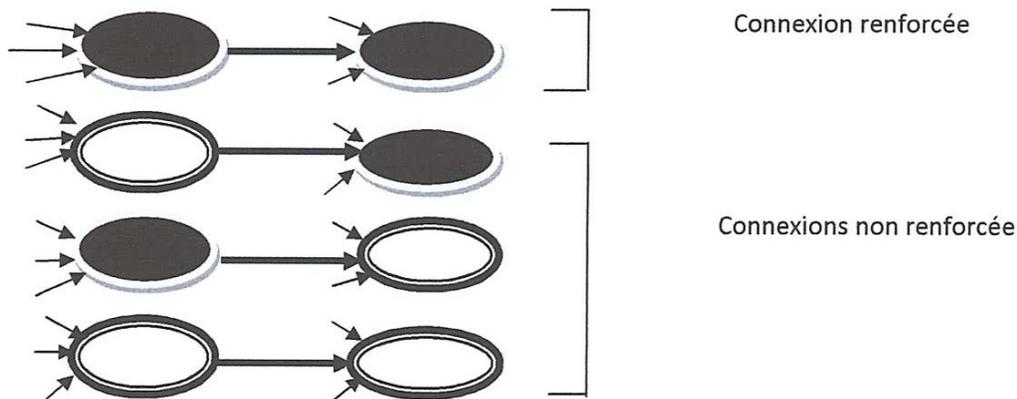
La règle de Hebb est le premier mécanisme d'évolution proposé sur les synapses. Son interprétation pour les réseaux de neurones formels est la suivante :

On considère que deux neurones connectés entre eux sont activés aux mêmes moments, la connexion qui les relie doit être renforcée et elle n'est pas modifiée, dans le cas contraire. C'est-à-dire que le poids w_{ij} d'une connexion entre un neurone i et un neurone j augmente quand les deux neurones sont activés en même temps et il n'est pas modifié, dans le cas contraire [04].

Prenons, à titre d'exemple, les conventions suivantes :



La règle de Hebb donne alors :



Quand la cellule émettrice et la cellule réceptrice s'activent en même temps, il faut augmenter le poids de cette connexion lors de l'apprentissage. La connexion entre ces deux cellules devienne alors très forte. Si la cellule émettrice s'active sans que la cellule réceptrice ne le soit, ou si la cellule réceptrice s'active alors que la cellule émettrice n'était pas activée, cela traduit bien le fait que la connexion entre les deux n'est pas prépondérante dans le comportement de la cellule réceptrice on peut donc dans la phase d'apprentissage laisser un poids faible à cette connexion [04].

En se basant sur ce principe, Hebb a donné la règle d'apprentissage suivante :

$$W_{ij}(t + \partial t) = W_{ij}(t) + \mu A_i A_j \quad (2.1)$$

Avec $W_{ij}(t)$ et $W_{ij}(t + \partial t)$: les poids de la connexion entre le neurone i et le neurone j aux instants t et $(t + \partial t)$.

A_i et A_j : L'activation du neurone i et l'activation du neurone j

μ : C'est le paramètre de l'intensité de l'apprentissage ($\mu > 0$).

b) 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 ($e(r)$) à l'itération r est donnée par :

$$e(r) = T(r) - Y(r) \quad (2.2)$$

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

$$F(X) = e^2(r) \quad (2.3)$$

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 :

$$\Delta w_{ij}(t) = -\eta \nabla F(X) \quad (2.4.a)$$

$$= \eta \frac{\partial e^2(r)}{\partial w_{ij}} \quad (2.4.b)$$

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 :

$$\Delta w_{ij}(r) = \Delta w_{ij}(r-1) - \eta(t_i - y_i)x \quad (2.5)$$

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 [13].

7. Réseaux de neurones de base

Il existe une multitude de réseaux de neurones différents les uns des autres par leurs architectures et leurs méthodes d'apprentissages.

7.1.L'Adaline

Peu de temps après que Rosenblatt proposa son perceptron, Windrow et Hoff proposèrent un autre modèle de neurone qu'ils nommèrent l'Adaline (pour Adaptive LInear Neurone) en 1960.

Ce modèle trouva diverses applications comme composante dans les antennes adaptatives et les modems à haute vitesse. L'Adaline est un seul neurone à valeur d'activation continue et une fonction d'activation linéaire.

$$a_i = \sum W_{ij}a_j - \theta_i \quad (2.6)$$

a_i : Activation du neurone i .

W_{ij} : Poids du neurone j vers le neurone i .

θ_i : Seuil du neurone i .

a_j : Activation du neurone précédant j .

La règle d'apprentissage utilisée est la règle de "Widrow-Hoff" ou règle Delta :

$$\Delta W_{ij} = \eta a_j (d_i - a_i) \quad (2.7)$$

ou :

η : Est le pas d'apprentissage.

a_j : Activation du neurone j.

d_i : Activation attendue du neurone i.

a_i : Activation du neurone i.

7.2. Le Perceptron

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

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

8. Le Perceptron multi couche (MLP)

8.1. Architecture du MLP

Le MLP est une réalisation en série de perceptrons. Ce modèle comprend en plus de la couche d'entrée et la couche de sortie, une ou plusieurs couches cachées. Dans ce réseau, chaque neurone est connecté à tous les neurones de la couche précédente et la couche suivante. Les neurones de la même couche ne sont pas connectés entre eux [04].

L'introduction de couches intermédiaires dans le réseau MLP permet de résoudre des problèmes plus complexes que la simple séparation linéaire. Lorsqu'il existe au moins une couche cachée, les états internes du réseau ne peuvent plus être donnés directement par les exemples et les sorties désirées puisque les sorties des neurones appartenant aux couches intermédiaires sont inconnues [03].

La figure (2.9) illustre un MLP, avec une seule couche cachée, comportant N neurones d'entrée, M neurones cachés et J neurones de sortie. Le $n^{\text{ème}}$ neurone d'entrée est relié avec le $m^{\text{ème}}$ neurone caché par le poids w_{nm} et le $m^{\text{ème}}$ neurone caché est relié avec le $j^{\text{ème}}$ neurone de sortie par le poids u_{mj} . A chaque présentation d'un exemple $X (x_1, x_2, \dots, x_N)$, les composants de son vecteur

caractéristique seront transmis aux neurones de la couche cachée. Les sorties de ces neurones (y_1, y_2, \dots, y_M) seront à leur tour transmises aux neurones de la couche suivante (la couche de sortie). La sortie du $m^{\text{ème}}$ neurone caché est donnée par :

$$y_m = h(r_m) = h\left(\sum_{n=1}^M x_n w_{nm}\right) \quad (2.8)$$

Où : $h(\cdot)$ est la fonction d'activation des neurones de la couche cachée.

Les neurones de la couche de sortie constituent la sortie du réseau. La sortie du $j^{\text{ème}}$ neurone est donnée par:

$$z_j = g(s_j) = g\left(\sum_{m=1}^J y_m u_{mj}\right) \quad (2.9)$$

Où : $g(\cdot)$ est la fonction d'activation des neurones de sortie

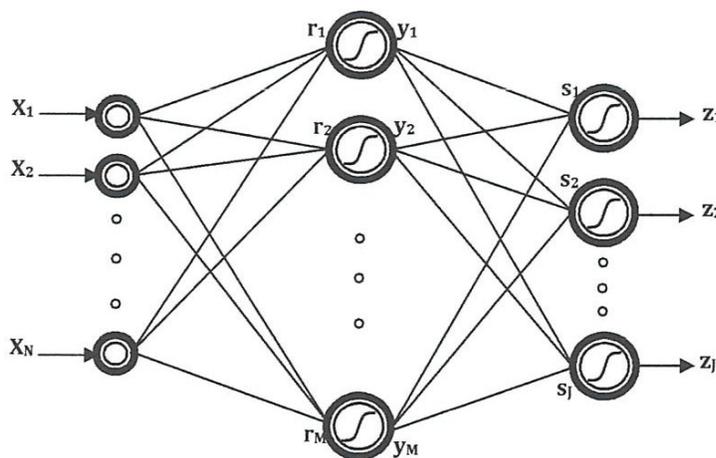


Figure 2.9: MLP avec une seule couche cachée contenant N neurones d'entrée, M neurones cachés et J neurones de sortie.

Dans les problèmes de classification, le nombre de neurones de la couche d'entrée du MLP est égal à la dimension de l'espace caractéristique parce que le rôle de cette couche est de présenter chaque composante du vecteur caractéristique aux neurones de la couche suivante. Le nombre de neurones de la couche de sortie égale le nombre de classes, soit un neurone par classe (la sortie de ce neurone est 1 pour les exemples appartenant à la classe correspondante et les autres sont à 0).

CHAPITRE 3

Applications

1. Introduction

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

Nous réalisons deux petites bases de données avec dix chiffres prononcés par deux locuteurs ; la première en langue arabe et la deuxième en français.

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. Préaccentuation (Préaccentuation du signal, il s'agit de faire ressortir les hautes fréquences avec un filtre passe-haut de la forme $H(z) = 1 - 0.9z^{-1}$)
2. Fenêtrage (Découpage du signal en fenêtre et application d'une fenêtre de Hamming sur ces portions)
3. RFFT (Application d'une transformée de Fourier sur chacune des portions, on obtient Le spectres)
4. Filtre de MEL (Création du banc de filtres, ils'agit de plusieurs filtres qui vont chacun couvrir une fréquence, ils permettent de mieux simuler le fonctionnement de l'oreille humaine)
5. Calcul du Log
6. DCT (Application d'une DCT, Discrete Cosinus 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 frames en utilisant la fonction de pondération.
2. Melbankm : création d'un banc de filtres.

3. Melcepst : création de la matrice de coefficients cepstraux.
4. RDCT : calcul de la transformée en cosinus discrète en réel.
5. RFFT : calcul de la transformée 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 sur une frame du mot (cinq) :

3.1. Le signal enregistré

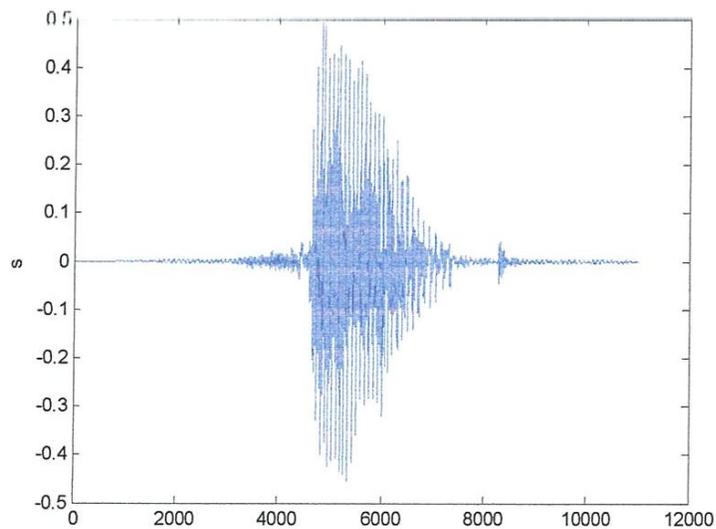


Figure 3.1: Signal enregistré

3.2. Un frame après fenêtrage :

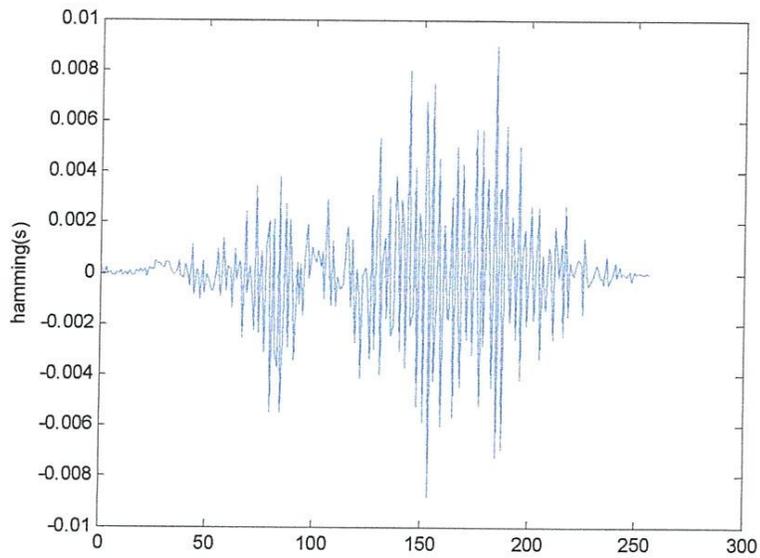


Figure 3.2: Un frame

3.3. RFFT du frame

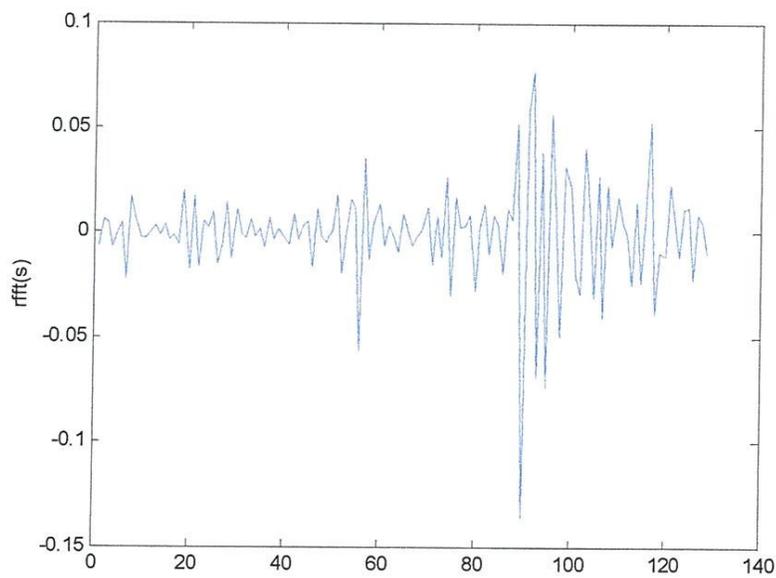


Figure 3.3 : Transformée rapide de Fourier réelle d'un frame

3.4. Le banc de filtres

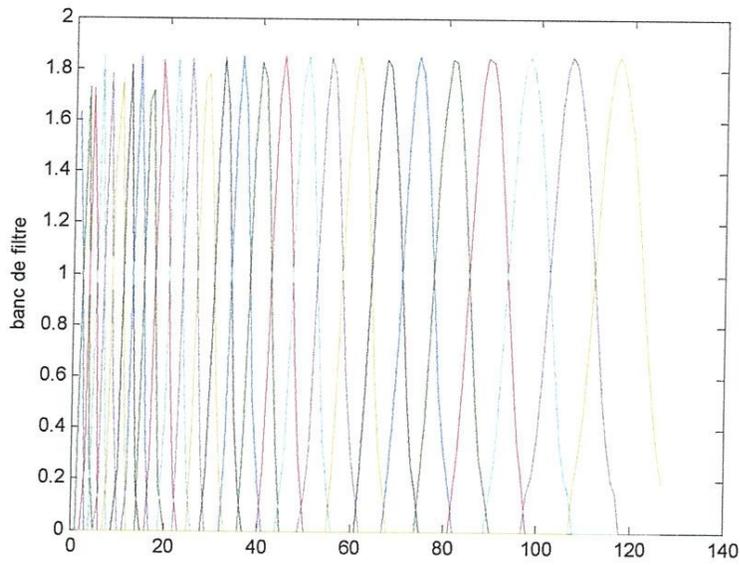


Figure 3.4: banc de filtres

3.5. Calcul du log :

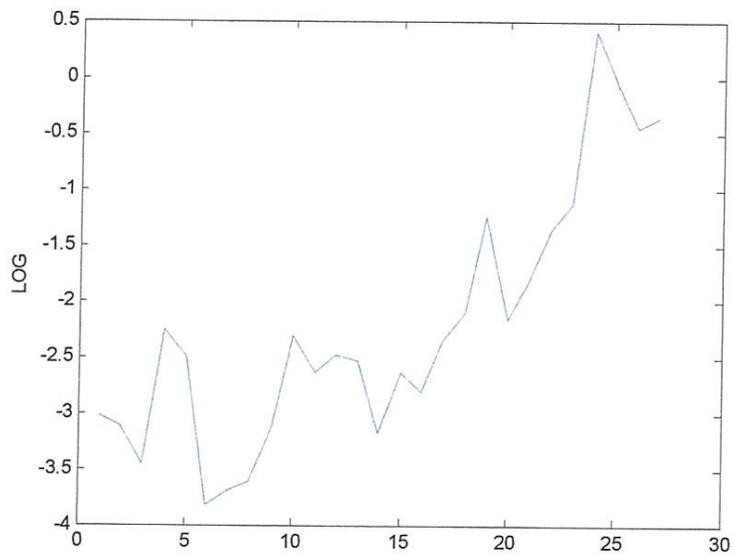


Figure 3.5: Calcule du log

3.6. Calcul de la RDCT :

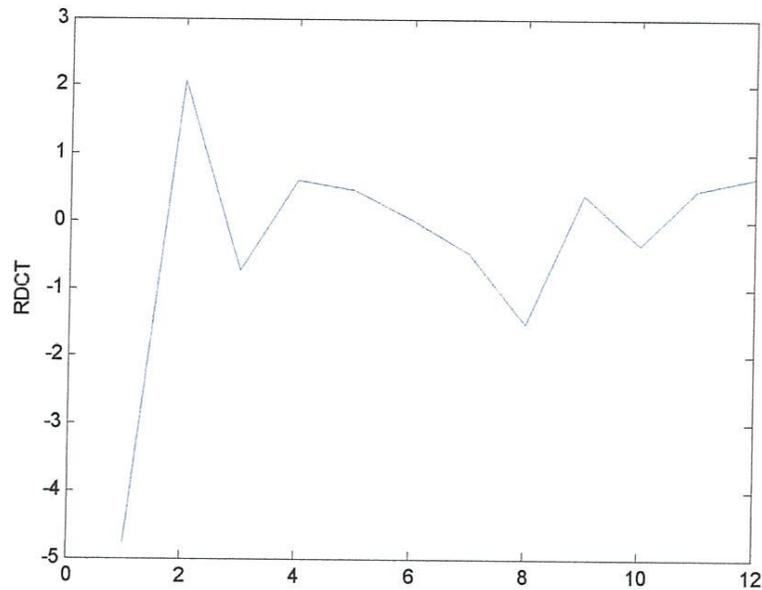


Figure 3.6: Calcul du RDCT

4. Phase de classification

Pour la classification nous utilisons un PMC qui est le réseau de neurones le plus utilisé dans ces problèmes. Le PMC utilisé comporte un nombre de neurones d'entrée égale au nombre de caractéristique (MFCC) donc $N=12$. Ce PMC comporte un nombre de neurones de sortie égale aux nombre de classe donc $J=10$. Pour le nombre de neurones de la couche cachée, nous avons choisie (par expérience) $M=20$.

4.1. Algorithme d'apprentissage de la Back-Propagation

4.1.1. Etape 1 : introduction des données

$X^{(q)}$ (vecteurs caractéristiques)

$T^{(q)}$ (vecteurs de sorties désirées)

4.1.2. Etape 2 : détermination des dimensions de données

N (nombre de caractéristique qui sera le nombre de neurones de la couche d'entrée)

Q (nombre d'exemples d'apprentissage)

J (nombre de classes qui est le nombre de neurones de la couche de sortie)

Initialisation des paramètres

I (nombre d'itération)

η_1 et η_2 (pas d'apprentissage).

M (nombre de neurones de la couche cachée)

U_{mi} et W_{nm} (poids initiaux avec des valeurs aléatoires)

4.1.3. Etape 3 : Apprentissage

Pour $i=1$ à I (pour chaque itération)

Pour $q=1$ à Q (pour chaque exemple)

$X = X^{(q)}$ (introduction du vecteur caractéristique de l'exemple q)

$T = T^{(q)}$ (introduction du vecteur de sortie de l'exemple q)

a) Calcul de la sortie

$$y_m = h(r_m) = h\left(\sum_{n=1}^M x_n w_{nm}\right)$$

$$z_j = g(s_j) = g\left(\sum_{m=1}^J y_m u_{mj}\right)$$

b) La mise à jour des poids

Pour $m=1$ à M

Pour $j=1$ à J

Ajustement de chaque poids u_{mj}

$$u_{mj}^{(i+1)} = u_{mj}^{(i)} + \eta_1 \cdot (t_j - z_j) \cdot g'(s_j) y_m$$

Pour $n=1$ à N

Ajustement de chaque poids w_{nm}

$$w_{nm}^{(i+1)} = w_{nm}^{(i)} + \eta_2 \cdot \left(\sum_{j=1}^J t_j - z_j \cdot g'(s_j) \cdot u_{mj} \right) (h'(r_m)) \cdot (x_n)$$

c) Calcul de l'erreur

$$E = \sum_{q=1}^Q \sum_{j=1}^J (t_j^{(q)} - z_j^{(q)})^2$$

4.1.4. Etape 4 : Fin si le nombre des itérations est achevé ou E est inférieur à une erreur de référence.

4.2. Exemple de classification

Considérons deux problèmes de classification bidimensionnels à deux classes. Les exemples du premier problème sont linéairement séparable tandis que ceux du deuxième problème ne sont pas linéairement séparable (figure 3.7). L'évolution de l'erreur au cours de l'apprentissage d'un PMC pour résoudre ces problèmes sont illustrés sur la figure (3.8). Nous constatons que le PMC converge plus rapidement dans le cas du premier problème.

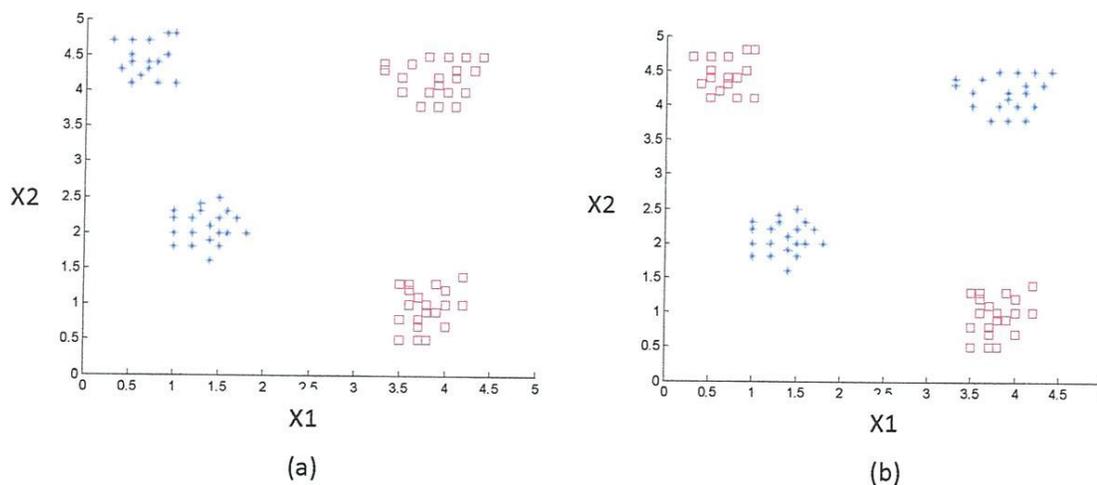


Figure 3.7: Deux problèmes de classification (le 1^{er} linéairement séparable et le 2^{ème} non linéairement séparable)

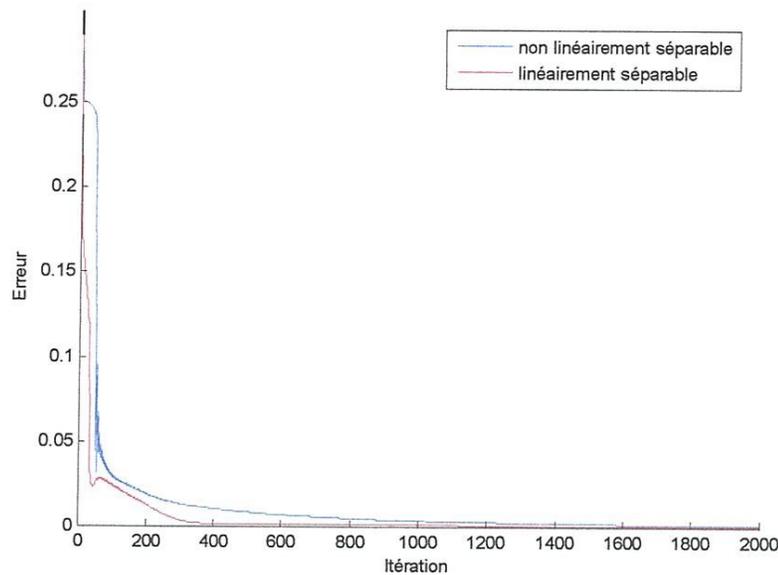


Figure 3.8: Evolution de l'erreur pour les deux problèmes

5. Reconnaissance des chiffres français

Nous avons réalisé une base de données avec 10 chiffres (un, deux, ..., neuf, zéros) prononcés par deux locuteurs. Chacun des deux locuteurs a répété chaque chiffre 15 fois.

Nous avons réalisé deux tests de classification comme suite :

1. Tests mono-locuteur (pour chaque locuteur).
2. Test pour les 2 locuteurs sur tout l'ensemble de données

Dans les deux cas, nous divisons la base de données de chaque locuteur en deux parties : une partie pour l'apprentissage et l'autre pour le test.

Pour évaluer les performances de généralisation de notre système, nous utilisons une validation croisée d'ordre 3. Trois ensembles d'apprentissage, dont chacun se constitue de 10 exemples pour chaque chiffre, sont ainsi obtenus. Les ensembles de test correspondants contiennent 5 exemples pour chaque chiffre.

5.1. Initialisation des nombre des neurones cachés

Pour le nombre des neurones cachés du PMC, il n'existe pas de technique pour le déterminer. Donc nous réalisons des tests pour fixer leur nombre empiriquement. Le tableau (3.1) illustre les résultats obtenus. Nous constatons qu'à partir de 20 neurones cachés, le PMC donne des résultats satisfaisants.

Nombre de neurones cachés	Taux d'apprentissage	Taux de test
5	65%	61%
10	95%	90%
15	100%	96%
20	100%	99%
25	100%	99%
30	100%	99%

Tableau 3.1: Résultat pour différent nombre de neurones cachés

5.2. Tests mono-locuteur

Dans cette partie nous réalisons des tests pour chacun des deux locuteurs. A chaque fois nous exécutons 10 fois le programme d'apprentissage-test. La figure (3.9) représente l'évolution de l'erreur pour chaque locuteur. Les résultats obtenus sont illustrés sur les tableaux (3.2) et (3.3).

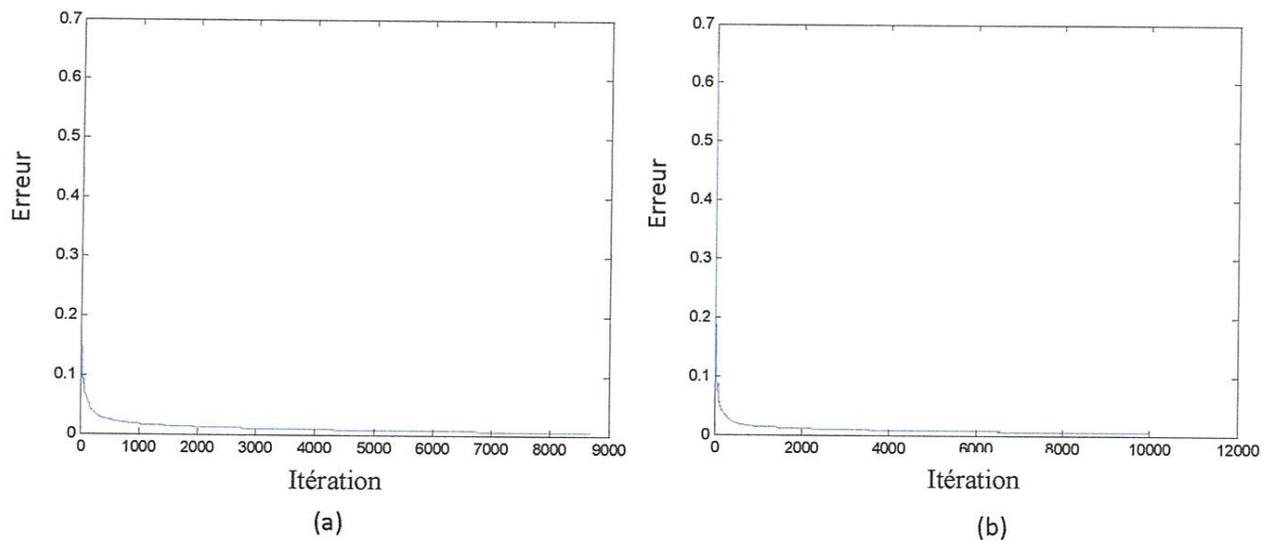


Figure 3.9 : Evolution de l'erreur de chacun des deux locuteurs

(a) Locuteur 1

(b) Locuteur 2

	Taux d'apprentissage	Taux de test			Chiffres mal-reconnus
		Meilleur	mauvais	moyenne	
Base1	100%	100%	96%	96.60%	9-5
Base2	100%	100%	94%	97%	9-5 2-10(Zéro)
Base3	100%	100%	96%	97.60%	9-5
Moyenne	100%	100%	95.33%	97.07%	-

Tableau 3.2: Résultat de tests locuteur 1 (Français)

	Taux d'apprentissage	Taux de test			Chiffres mal-reconnus
		Meilleur	mauvais	moyenne	
Base1	100%	100%	96%	97.80%	5-9 (Zéro)10-6
Base2	100%	100%	94%	97.40%	5-9
Base3	100%	100%	94%	99.20%	5-9
moyenne	100%	100%	94.66%	98.13%	-

Tableau 3.3: Résultat de tests locuteur 2 (Français)

5.3. Tests multi-locuteur

Dans cette partie nous réalisons des tests sur l'ensemble de la base de données (formée des deux locuteurs). A chaque fois nous exécutons 10 fois le programme d'apprentissage-test. La figure (3.10) représente l'évolution des l'erreur. Les résultats obtenus sont illustrés sur le tableau (3.4).

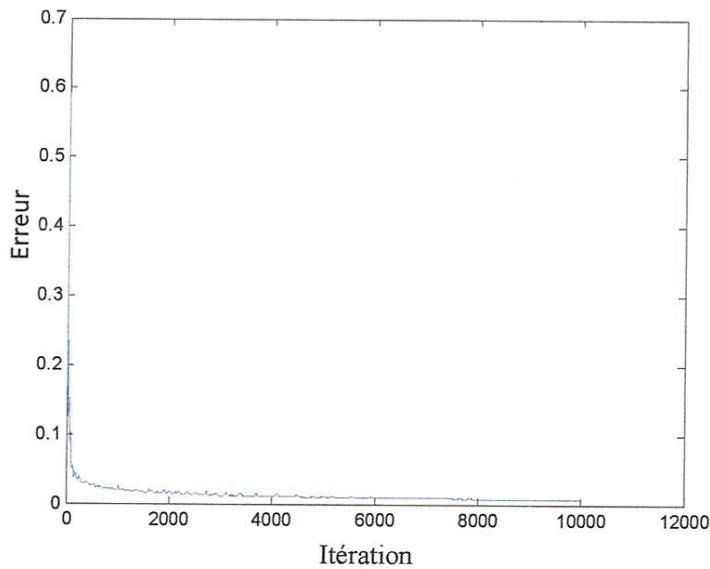


Figure 3.10 : Evolution de l'erreur des deux locuteurs

Taux d'apprentissage	Taux de test			Chiffres mal-reconnus
	Meilleur	Mauvais	Moyenne	
99.75%	98%	95%	96.50%	9-5 10-6 1-9 7-5

Tableau 3.4: Résultat de tests multi-locuteur (Français)

6. Reconnaissance des chiffres Arabe

Nous avons réalisé également une base de données avec 10 chiffres arabe (wahed, Ithnani, ..., tissâa, siffer) prononcés par deux locuteurs. Chacun des deux locuteurs a répété chaque chiffre 15 fois.

Nous avons également réalisé deux tests de classification : mono-locuteur et multi-locuteur et nous utilisons aussi une validation croisée d'ordre 3 pour évaluer les performances de généralisation.

6.1. Tests mono-locuteur

Nous réalisons des tests pour chacun des deux locuteurs. A chaque fois nous exécutons 10 fois le programme d'apprentissage-test. La figure (3.11) représente l'évolution des l'erreur pour chaque locuteur. Les résultats obtenus sont illustrés sur les tableaux (3.5) et (3.6).

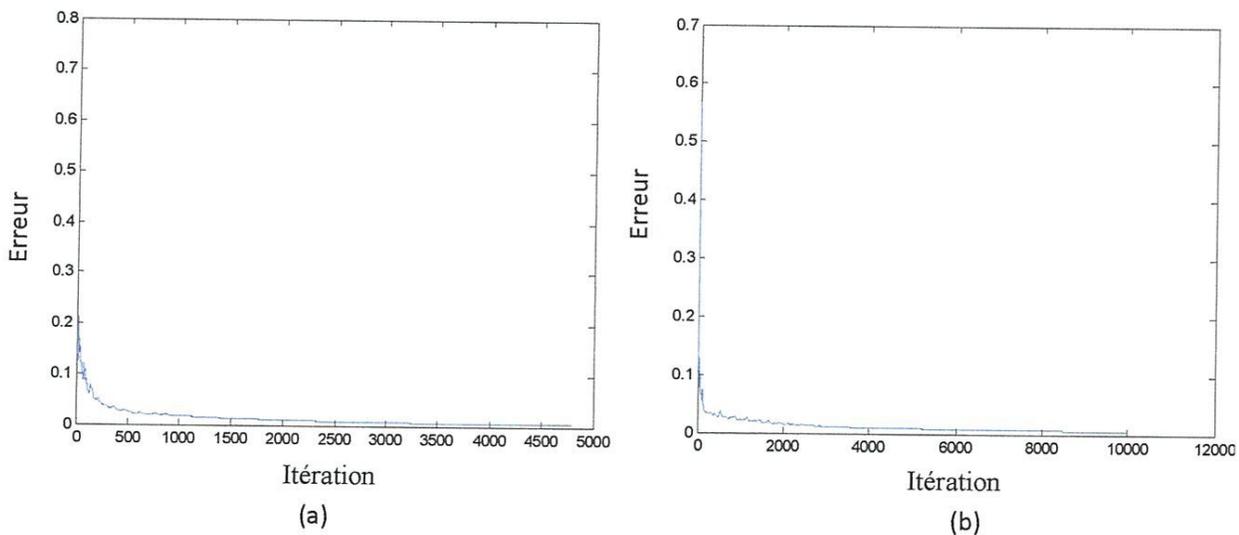


Figure 3.11 : Evolution de l'erreur de chacun des deux locuteurs

(a) Locuteur 1

(b) Locuteur 2

	Taux d'apprentissage	Taux de test			Chiffres mal-reconnus
		Meilleur	mauvais	moyenne	
Base1	100%	100%	98%	99.40%	6-2
Base2	100%	100%	98%	99.80%	7-3
Base3	100%	100%	98%	99.60%	6-2
moyenne	100%	100%	98%	99.60%	-

Tableau 3.5: Résultat de tests locuteur 1(Arabe)

	Taux d'apprentissage	Taux de test			Chiffres mal-reconnus
		Meilleur	mauvais	moyenne	
Base1	100%	100%	98%	99.80%	1-2
Base2	100%	100%	98%	99.40%	1-8
Base3	100%	100%	96%	97.80%	6-5
moyenne	100%	100%	97.33%	99%	-

Tableau 3.6: Résultat de tests locuteur 2(Arabe)

6.2. Tests multi-locuteur

Dans cette partie nous réalisons des tests sur l'ensemble de la base de données (formée des deux locuteurs). A chaque fois nous exécutons 10 fois le programme d'apprentissage-test. La figure (3.12) représente l'évolution des l'erreur. Les résultats obtenus sont illustrés sur le tableau (3.7).

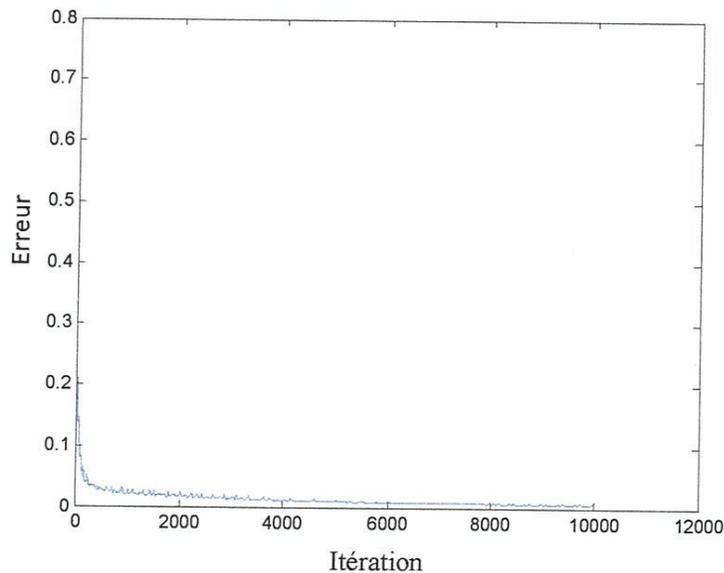


Figure 3.12 : Evolution de l'erreur des deux locuteurs

Taux d'apprentissage	Taux de test			Chiffres mal-reconnus
	Meilleur	mauvais	moyenne	
100%	98%	97%	97.25%	6-2 7-3

Tableau 3.7: Résultat de tests multi-locuteur (Arabe)

7. Conclusion

Les tests de classification réalisés sur les deux bases de données : français et arabe nous ont donnés les résultats représentés sur le tableau (3.8).

Nous constatons que dans les deux cas les résultats sont satisfaisants ce qui montre que les MFCC permet une bonne caractérisation des données. Nous constatons également que les résultats de reconnaissance en arabe sont meilleurs que ceux en français.

	Locuteur 1		Locuteur 2		Multi-locuteur		Chiffres mal-reconnus
	app	test	app	test	app	test	
Français	100%	97.07%	100%	98.13%	99.75%	96.50%	9-5,10-6,1-9,7-5
Arabe	100%	99.60%	100%	99.00%	100%	97.25%	6-2, 7-3

Tableau 3.8: Comparaison entre les résultats de test arabe et français

D'autre part nous constatons que les réseaux de neurones sont des 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 de la difficulté d'interprétation. Les avantages et les inconvénients des réseaux de neurones peuvent être récapitulés dans le tableau (3.9):

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

Tableau 3.9: Avantages et inconvénients des réseaux de neurones

Conclusion général

Dans ce travail nous avons réalisé un système de reconnaissance vocal de chiffres. Ce système est à un vocabulaire limité et à 2-locuteurs. Pour l'extraction des caractéristiques, nous avons basé sur les coefficients cepstraux dans l'échelle de Mels (MFCC) qui sont les plus utilisés dans ce domaine. Dans la phase de classification, nous avons utilisé le PMC qui est le réseau de neurones le plus utilisé dans le domaine de la classification.

Nous avons élaboré deux bases de données avec dix (10) chiffres prononcés par deux locuteurs. La première base en arabe et la deuxième en français.

Nous avons réalisé des tests de classification mono-locuteur et à 2 locuteurs. Les résultats obtenus sont satisfaisants ce qui montre que les MFCC permettent une bonne caractérisation des données et que les réseaux de neurones sont de bons classificateurs en matière d'apprentissage et de généralisation. Néanmoins, nous avons constaté que les réseaux souffrent du problème de manque de techniques d'initialisation de la difficulté d'interprétation.

Nous avons constaté également que les résultats de reconnaissance en arabe sont meilleurs que ceux en français.

Bibliographies

- [1] AMR H Nour-eldin , KABAL Peter, combining frontend-based memory with mfcc features for bandwidth Extension of narrowband speech , ICASSP-IEEE, 2009.

- [2] BAKIR Nadia , Debyeche Mohamed, CHIBANI Youcef, reconnaissance automatique des chiffre arabes en milieu réel par fusion, 10^{ème} congrès français d'acoustique, Lyon, 12-16, avril 2010.

- [3] BENAHMED Nadia , optimisation de réseaux de neurones pour la reconnaissance de chiffres manuscrits isolés , thèse doctorat, université du Québec, 2002.

- [4] Bouyeddou Hocine , reconnaissance de l'écriture arabe imprimée par les réseaux de neurons, mémoire de magister, université de GUELMA, 2007.

- [5] CARLOS S Lima , TAVARES Adriano C , CARLOS A Silva, JORGE F. Oliveira, spectral normalisation mfcc derived features for Robust speech recognition, SPECOM'2004: 9th conference speech and Computer St. Petersburg, Russia September 20-22, 2004.

- [6] DAVALO Eric, NAÏM Patrick , des réseaux de neurones, EYROLLES, 1993.

- [7] DEGHEMICHE Houari, CHIBOUT Menaoeur , utilisation des réseaux neuro-flous pour La Reconnaissance automatique de la Parole, mémoire ingénieur, institut des télécommunications, Oran, 2003.

- [8] HLAOUI Adel , reconnaissance de mots isoles arabes par hybridation de réseaux de, mémoire ingénieur, école nationale d'ingénieurs de Tunis, 2005.
- [9] JEAN Paul , reconnaissance automatique de la parole , dunod, 2006.
- [10] KHEBLIL Abdelmalek, reconnaissance vocal par mélange de Gaussiennes (GMM), thèse magister, école militaire polytechnique 2009.
- [11] LOISELLE Stéphane , exploration de réseaux de neurones à décharges dans un contexte de reconnaissance de parole, thèse doctorat, université du quebec, 2004.
- [12] MEDJDOUB Bensalem , système de vérification des locuteurs basé sur la fusion en scores des paramètres MFCC et LPCC, mémoire ingénieur, école militaire polytechnique ,2008.
- [13] NEMISSI Mohamed , classification et reconnaissance des formes par algorithmes hybrides, thèse de doctorat, université de Guelma, 2009.
- [14] NEMISSI Mohamed , Classification automatique supervisé par les réseaux de neurons, mémoire de magister, université de GUELMA, 2004.
- [15] OHGREN Sara , experiment with adaptation and vocal tract length normalization at automatic speech recognition of children's speech, master thesis in speech technology, Stockholm, 2007.
- [16] RENE Boite Herve , Traitement de la parole, collection électricité, Novembre 1999.

- [17] RENE Boite, MURAT Kunt, Traitement de la parole, presses polytechnique , ROMANDES, 1987.
- [18] TAN Tien Ping , automatic speech recognition for normative speakers, thèse doctorat, université joseph fourier, grenoble, 2008.
- [19] VIBHA Tiwari , mfcc and its applications in speaker recognition , international journal on emerging technologies, 1(1), 19-22, 2010.

Résumé

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

