REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE

MINISTERE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA RECHERCHE SCIENTIFIQUE

UNIVERSITE YAHIA FARES DE MEDEA



Faculté de **T**echnologie

Mémoire de fin d'études de Master

Filière : Génie Electrique
Spécialité : Automatique et systèmes

Identification du locuteur en utilisant les réseaux de neurones

---

Présenté par

Mlle. Faiza MERAZI et Mlle. Fethia BOUCHERIT

Proposé et Dirigé par

Dr. Ziane KECHIDI et M. Mohamed Nasr eddine TEMMAR

Promotion 2019/2020

Dédicace

Nous dédions ce travail à nos parents le dieu les garde.

À nos chères sœurs et leurs enfants.

À nos frères.

À nos amis et leurs encouragements.

À toute la famille BOUCHERIT, et MERAZI.

À tous ceux qui nous aiment.

Remerciement

A la fin de ce travail, nous aimerions remercier « Allah » d'abord qui nous a donné la volonté et la patience pour aboutir à la réalisation de ce travail.

Elhamdulilah

Nous exprimons notre grande reconnaissance et Remerciements à notre promoteur Mr. KECHIDI Ziane et le Co-promoteur Mr TEMMAR Mohamed Nasr eddine Pour ses conseils et ses savoirs faire.

Nous tenons à remercier les membres du jury M. REGGAI Djahid et M.GUEDMANI Mohamed pour leurs remarques et suggestions qui ont aidé a amélioré la qualité de notre mémoire et d'avoir honoré à nous évaluer et qu'ils trouvent ici toutes les expressions de considérations de gratitude et du respect distingué.

Et nous aimons de remercier aussi madame BOUKIRAT Samia pour son aide et ses encouragements.

Nous présentons enfin, notre profonde gratitude à notre famille, ainsi qu'à tous ceux qui nous ont aidés de loin ou de près à réaliser ce travail.

Faiza et Fethia

Résumé

ملخص

الكلام هو أبسط أشكال التواصل والتفاعل البشري. فبينما كانت معالجته من أكثر المجالات الواعدة في معالجة الإشارات، مكنت تقنية تحديد المتحدث التطبيقات الناشئة المختلفة من التعرف على الصوت البشري. والهدف الرئيسي من تحديد المتحدثين هو تطوير تقنيات وأنظمة لتصنيف المتحدثين باستخدام الشبكات العصبية الاصطناعية، وفي الواقع يتم ذلك من خلال المعالجة المسبقة للإشارة، واستخراج الميزات، وتصنيف الشبكة العصبية الاصطناعية، وتحليل الأداء. أما الميزات المستخرجة فهي علو الصوت والتي تعد أكثر الميزات استخدامًا في أنظمة تحديد (MFCC) ودرجة التردد الميل (MFCC) والمعاملات الصوتية المتحدث وقد صممت صوت المتحدث بشكل أقرب من أي تقنية أخرى. تتناول هذه الرسالة تحديد مدخلات الكلام.

الكلمات المفتاحية: الصوت، الشبكة العصبية الاصطناعية، المعالجة، الميزات.

Résumé

La parole est la forme la plus simple de communication et d'interaction humaine alors que son traitement est l'un des domaines les plus prometteurs du traitement du signal.La technologie d'identification du locuteur a permis à diverses applications émergentes pour identifier la voix humaine.

L'objectif principal de l'identification du locuteur est de développer des techniques et des systèmes de classification du locuteur. Ce mémoire traite l'identification de la parole d'entrée à l'aide de réseaux neuronaux artificiels, en fait cela se fait par le prétraitement du signal, l'extraction des caractéristiques, la classification des réseaux neuronaux artificiels et l'analyse des performances.

Les caractéristiques extraites sont la hauteur de la voix et les Coefficients Cepstraux de Fréquence Mel (CCFM) qui sont les caractéristiques les plus couramment utilisées dans les systèmes d'identification du locuteur et qui ont modélisé la voix du locuteur plus étroitement que toute autre technique.

Mots-clés : la parole ; traitement de signal ; identification du locuteur ; la voix humaine ;réseaux neuronaux artificiels ;l'extraction de caractéristiqueset les Coefficients Cepstraux de Fréquence Mel (CCFM).

Abstract

Speech is the most simple form of human communication and interaction whereas its processing has been one of the most promisingfieldsin signal processing. Speakeridentification technology has made it possible for various emerging applications to identify human voice. The main goal of speakeridentification is to develop techniques and systems for speakerclassification.

This dissertation deals with identification of input speech using artificial neural networks. In fact,this workis performedusing signal preprocessing, feature extraction, artificial neural network classification, and performance analysis. The extracted features are pitch and Mel Frequency Cepstral Coefficients (MFCC), whichare the most commonly utilized features in speaker identification systems and have modeled the speaker voice more accurately than any other technique.

Key words: Speech of human communication; signal processing; Speaker identification technology; human voice; artificial neural networks;feature extraction;Mel Frequency Cepstral Coefficients (MFCC).

Liste des figures

Chapitre I

Figure I.1 : Système phonatoire…….…….. 7

Figure I.2 : Echantillonnage d'un signal…….…….9

Figure I.3 : Méthode d'échantillonnage…….……..... 9

Figure I.4 : Spectre d'un signal échantillonné…….…….……...10

Figure I.5 : Erreur de quantification d'un signal triangulaire…….…….....10

Figure 1.6 : Quantification d'un signal…….……...11

Figure 1.7 : Chaîne de traitement…….……...……12

Figure I.8 : Transformées de Fourier de deux signaux différents…….…..13

Figure I.9 : Hiérarchisation des domaines du traitement de la parole…….…….14

Chapitre II

Figure II.1 : Modèle d'un neurone biologique……..…….…. 20

Figure II.2 : Transmission de l'information entre les neurones biologiques…….…..... 22

Figure II.3 : Schéma d'un neuronebiologique……...……23

Figure II.4 : Neurone Artificiel……24

Figure II.5 : Mise en correspondance neurone biologique / neurone artificiel……25

Figure II.6 : Structure d'un neurone artificiel……...…...25

FigureII.7 : Schéma d'un réseau de neurones monocouche……..…….26

Figure II.8 : Schéma d'un réseau de neurones non bouclé (Perceptron multicouches)...27

Figure II.9 : Schéma d'un réseau de neurones à connexions locales……28

Figure II.10 : Schéma de Réseau de neurone bouclé…….……..……29

Figure II.11 : Forme canonique d'un réseau de neurones bouclé…….…… 30

Figure II.12 : Quelque fonction d'activations……...31

Figure II.13 : i le neurone amont, j le neurone aval et wij le poids de la connexion....31

Figure II.14 : Réseau de 3 neurones (les 2 entrées sont considérées comme deux neurones) pour la résolution du problème exprimé table 2……..……..32

Chapitre III

Figure 1 : Modèle de reconnaissance vocale utilisant un réseau neuronal…….40

Figure 2 : Le signal après le filtrage……...…….41

Figure 3 : Le signal sans bruit à l'aide de la méthode de pitch……..…….……42

Figure 4 : Schéma fonctionnel du fonctionnement MFCC pour les données de fonction

Vocale…….……..…….…….43

Figure 5 : Banc de filtres Mel typique……..…….....…….43

Figure 6 :Matrice de confusion après l'entrainement…..…..……48

Figure 7 : La performance de 10 locuteurs……..……..…….52

Figure 8 : La performance de 20 locuteurs…….53

Figure 9 : La performance de 30 locuteurs…….53

Figure 10 : La performance de 30 locuteurs…….…..…….54

Figure 11 : Performance de 30 locuteurs par la méthode trainscg…….……57

Figure 12 : Performance de 30 locuteurs par la méthode trainlm …….59

Liste des tableaux

Chapitre II

1 : La loi de Hebb…….……..…..…31

2 : Base d'exemples d'apprentissage pour la loi de Hebb…….……..…..……33

3 : Analogie entre le neurone biologique et le neurone formel…...…….……...…….…...37

Chapitre III

1 :Le nombre de parleurs dans la banque de données du train…….…….....…44

2 : Le nombre de parleurs dans la banque de données du test……..…..45

3 : Résultats des nombres de fichier d'audio dans chaque étiquette…….…….46

4 :Résultats des nombres de fichier d'audio dans chaque étiquette après la normalisation……..…….47

5 : Résultats des nombres de fichier d'audio dans chaque étiquette après le teste……49

6 : Résultats final d'apprentissage…….....…….51

Liste des abréviations

ANN : (artificial neural network) réseau de neurone artificielle.

KNN : (K-Nearest Neighbor) k plus proche voisin.

MFCC :Mel frequency coefficient cesptral .

LM : Levenberg-Marquardt

BR : Régularisation bayésienne

SCG : Gradient conjugué mis à l'échelle

Introduction générale

Introduction générale

L'informatique est la science du traitement automatique de l'information. Son développement est souvent confondu avec celui des machines de traitement : les ordinateurs. Dès le début(ENIAC 1945) Même aujourd'hui, les ordinateurs sont beaucoup plus puissants.

Cependant, cette augmentation de puissance ne permet pas de toujours résoudre les problèmes d'une application informatique dans un domaine particulier. L'idée s'est donc installée que ce n'était peut-être pas tant le matériel que le logiciel qui pêchait par manque de puissance. La construction de logiciels s'appuie sur plusieurs approches. Deux parmi les plus utilisées sont l'approche algorithmique et l'approche basée sur la connaissance.

Une approche algorithmique nécessite l'écriture (avant la transcription dans un quelconque langage de programmation) du processus à suivre pour résoudre le problème. Lorsque le problème est complexe, ce peut être une étape coûteuse ou impossible. D'autre part, les ordinateurs sont des machines complètement logiques (et même binaires) qui suivent à la lettre chacune des instructions du programme. C'est un avantage lorsque tous les cas ont été prévus à l'avance par l'algorithmique. Ce n'est hélas pas toujours possible. Dans ce cas, dixit l'informaticien : "c'est une faute de la machine". Rien de plus faux ! Ainsi les systèmes informatiques embarqués (à bord des avions, de la navette spatiale, etc.) tentent de pallier à ce manque (prévisible) de clairvoyance de l'algorithmique en triplant les logiciels, chacun étant développés indépendamment par une équipe différente, dans des langages différents. Les risques de laisser l'ordinateur aux prises avec une situation imprévue, où son comportement ne serait pas adapté, sont ainsi considérablement réduits. Rappelons-nous le haro lancé sur les programmes boursiers lors de la chute de la bourse en 1987.

La seconde approche possible est celle de l'intelligence artificielle (appelée IA par commodité), avec pour applications les plus connues les systèmes experts. Ici, la résolution du problème est confiée à un ensemble de règles données par l'expert humain du domaine. Il n'en demeure pas moins que toutes les règles doivent avoir été exprimées préalablement au traitement, et que le programme demeure binaire dans son exécution. Les cas qui n'ont pas été prévus par l'expert ne seront pas correctement traités. L'introduction de la logique floue ne change pas la nature des limitations d'emploi du programme : l'exécution reste totalement déterministe. En fait, l'approche basée sur la connaissance se limite à des domaines d'application où la modélisation de la connaissance, par exemple sous forme de règles, est possible. Ces domaines sont souvent ceux des sciences dites "exactes" comme l'électronique, la mécanique, la physique, etc., par opposition aux sciences dites "humaines" comme la

médecine, la psychologie, la philosophie, etc., où la connaissance est plus empirique. L'IA se révèle donc être principalement un moyen commode de stocker de la connaissance sous forme explicite.

Ces deux approches ne suffisent pas à répondre à tous les problèmes existants. Citons les domaines de la reconnaissance de formes (images ou signaux), du diagnostic, du contrôle moteur, de la traduction automatique, de la compréhension du langage, depuis longtemps explorés à l'aide des approches algorithmiques et à base de connaissances, qui n'ont pas rencontré le succès escompté. Pourtant, des êtres vivants relativement simples sont capables de réaliser certaines de ces opérations apparemment sans difficulté. Il suffit pour s'en rendre compte de lever les yeux, suivre le vol de la mouche et essayer de la capturer. Que dire alors du déplacement au sonar de la chauvesouris, etc.

Une troisième approche au traitement automatique de l'information semble donc s'offrir à nous, où l'on cherche à s'inspirer du traitement de l'information effectué par le cerveau.

L'hypothèse principale, à la base de l'essor des réseaux de neurones artificiels, est que le comportement intelligent est sous-tendu par un ensemble de mécanismes mentaux. Ces mécanismes étant basés sur des processus neurophysiologiques, nous supposons donc que la structure du système nerveux central est à la base du développement d'un comportement intelligent.

Nous examinons quelques notions biologiques relatives au cerveau, et à ses constituants les neurones et leurs synapses. L'organisation en réseaux des neurones permet d'illustrer les notions d'apprentissage et de mémorisation (modification des connexions). Ces données nous sont nécessaires pour aborder le second chapitre qui montre le passage des modèles de réseaux neuronaux biologiques à des modèles mathématiques : les réseaux de neurones artificiels.

Nous établissons un tableau des correspondances biologique/artificiel, avec notamment des modèles de neurones et de synapses et quelques topologies pour l'organisation en réseaux. Au travers d'un exemple simple, nous décrivons le fonctionnement des réseaux de neurones artificiels et leurs propriétés d'apprentissage à partir d'exemples, de résistance au bruit, d'adaptabilité et de tolérance aux pannes. Il existe de nombreux modèles de réseaux de neurones artificiels, nous en présentons successivement quelques-uns choisis principalement

selon des critères de nature pédagogique. Le Perceptron est historiquement le premier modèle, son fonctionnement est particulièrement intéressant pour la suite de notre étude. De fait, il demeure un modèle de base, de même que les cartes auto-organisatrices plus vraisemblables d'un point de vue biologique. Ces deux modèles nous permettent d'introduire les concepts d'apprentissage supervisé et non supervisé. Des modèles plus élaborés sont étudiés par la suite tant au niveau de leur architectures, des techniques d'apprentissage que des performances. Ce sont les mémoires associatives, le réseau ART et une version plus complexe et surtout plus efficace du perceptron : le Perceptron multicouche.

- **Objectifs de notre travail**
    1. (la reconnaissance du locuteur consiste à identifier le locuteur à partir d'un signal audio
    2. Ce mémoire montre comment nous pouvons utiliser des algorithmes de vision par ordinateur pour collecter un ensemble de données d'identification de locuteur.
    3. la perte de contraste est utilisée pour former le réseau de vérification des locuteurs à l'aide d'une stratégie de préformation

- **Structure de la thèse**

Notre mémoire est structurée en trois chapitres, de la manière suivante :

Dans le premier chapitre on présentera des généralités et des définitions fondamentales en identification de locuteur, des généralités sur le traitement numérique de la parole, et la Reconnaissance Automatique du Locuteur.

Le deuxième chapitre exposera un aperçu général sur les réseaux de neurones artificiels, leurs principes de fonctionnement, l'architecture des réseaux de neurones, une vue d'ensemble des fonctions d'activations, les techniques d'apprentissage et leurs algorithmes, les domaines d'application des RNA et l'analogie entre le neurone biologique et le neurone formel. Les simulations et résultats que nous avons obtenus sont présentés dans le dernier chapitre.

Et comme tout travail notre mémoire se termine par une conclusion générale.

Introduction générale

Introduction générale

Chapitre I

Généralités sur l'identification de locuteurs

Chapitre I : Généralités sur l'identification de locuteurs

I.1 Introduction

La parole est la manière naturelle et, en conséquence, la forme la plus commune de communication humaine. À la différence d'autres moyens électroniques de communication, les systèmes utilisant la parole offrent à l'utilisateur non entraîné un accès simple et naturel. Il semble improbable que, dans l'avenir, même l'impact de la technologie des ordinateurs et de réseaux numériques (fax, e-mail et autres) puisse changer la préférence des gens pour la parole comme moyen fondamental de communication entre les humains.

Les travaux sur le traitement de la parole ont une histoire de plus de cinquante ans : c'est en 1939 que, pour la première fois, un chercheur des laboratoires Bell aux États-Unis, M. Dudley, proposa un appareil nommé Vocodeur (« Voice Coder » ou « codeur de voix ») visant à coder électriquement le signal vocal selon des paramètres limités, puis à le transmettre avec un débit d'informations réduit et à le reproduire enfin dans un système de synthèse effectuant l'opération réciproque du Vocodeur. Plus tard, le même chercheur réalisa un autre système de synthèse électrique de la parole, actionné par un clavier. Enfin apparut, en 1947, le sonagraphe, premier analyseur de la composition du signe vocal en fréquence et en amplitude. Depuis lors, un dénombrement de tous les événements et développements de produits serait fastidieux puisque des avances ont été très importantes depuis cette date, fondamentalement, dans le domaine des télécommunications. Voyons donc, sans plus de préambule, les pas nécessaires et les techniques modernes pour le traitement du signal.

I.2 Production de la parole

La parole est le résultat de deux fonctions mécaniques de bases : la phonation et l'articulation. La phonation est la production du signal acoustique par le mouvement du larynx. L'articulation est la modulation du signal acoustique par les articulateurs (les lèvres et la langue) et la résonance de ce signal dans les cavités (la bouche et le nez). [1]

I.2.1 L'appareil vocal humain

L'appareil vocal humain peut être comparé à la fois à un instrument de musiqueà vent et à corde. Il comprend une source de vent, les poumons ; une structure qui vibre, les cordes vocales dans le larynx ; et une série de caisses de résonance que forment le pharynx, la bouche et les fosses nasales. [2]

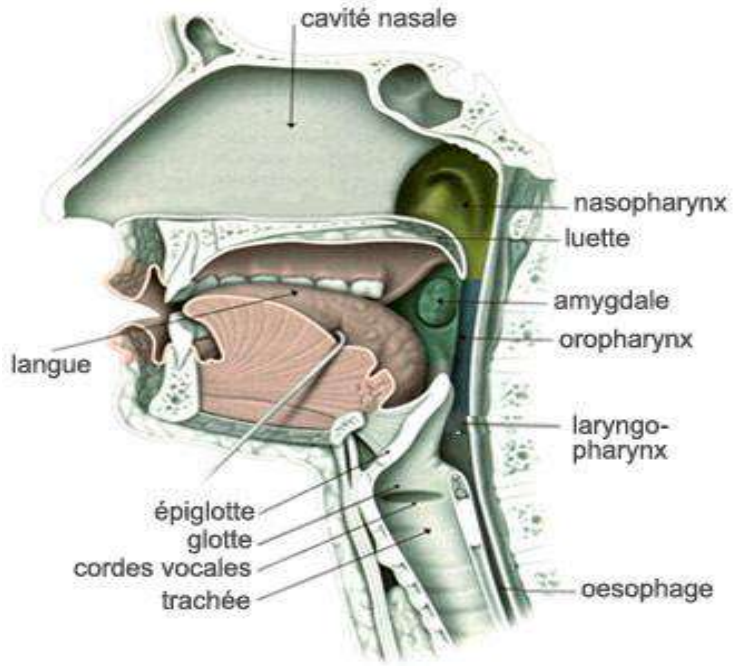


Figure I.1 : Système phonatoire [1]

I.2.2 Caractéristiques de la parole

Comment tout cela fonctionne-t-il ensemble ? Première composant : l'air pulmonaire, que l'on pourrait définir comme le " générateur ". Quand on parle, les phases d'inspiration de notre respiration deviennent plus rapides et plus courtes. On respire également davantage par la bouche, alors que l'inspiration est normalement exclusivement nasale. Du côté de l'expiration, le volume et la pression de l'air expiré sont augmentés pour pouvoir faire vibrer les cordes vocales situées dans le larynx.

Le larynx joue le rôle d'une porte qui préserve les bronches et les poumons des aliments et autres corps étrangers. Le larynx se compose d'une série de muscles et de cartilages plus ou moins mobiles qui peut être relevé ou abaissé. Lors de la déglutition, le larynx s'élève tandis que l'épiglotte, cette lame cartilagineuse située à l'entrée du larynx, se rabat vers l'arrière. Ce faisant, elle bloque l'entrée des voies aérienne supérieure et permet au bol alimentaire d'emprunter l'œsophage pour rejoindre l'estomac.

Lorsqu'on parle, l'air expulsé des poumons emprunte la trachée avant d'arriver dans le larynx où il va rencontrer les cordes vocales. Celles-ci sont en fait une paire de muscle et de ligaments de 20 à 25 millimètres de long et recouverts d'une muqueuse. C'est la seconde composante de l'appareil phonatoire : le " vibrateur ". [2]

I.2.3 Signaux stationnaires

Un processus aléatoire est dit stationnaire si ses propriétés statistiques caractérisées

Par des espérances mathématiques sont indépendantes du temps. Un processus est stationnaire si :

- Sa moyenne ne change pas dans le temps,
- Sa fonction d'autocorrélation ne dépend pas du temps,
- Sa fonction d'auto-covariance ne dépend pas dans le temps.[3]

I.3 TRAITEMENT DE LA PAROLE

Le traitement de la parole est une discipline technologique dont l'objectif est la captation, la transmission, l'identification et la synthèse de la parole.

Dans ce domaine, on peut définir la parole comme un texte oral. On s'intéresse

- à l'intelligibilité, c'est-à-dire à la possibilité, pour la personne qui écoute, de comprendre sans erreur le texte émis ;
- à l'amélioration de l'intelligibilité quand le signal est dégradé ;
- à l'identification de la personne qui parle ;
- à l'établissement automatique d'un texte écrit à partir de la parole ;
- à la synthèse de la parole à partir d'un texte écrit. [1]

I.3.1 Numérisation des signaux

La numérisation est la conversion des informations d'un support (texte, image, audio, vidéo) ou d'un signal électrique en données numériques que des dispositifs informatiques ou d'électronique numérique pourront traiter. [4]

I.3.1.1 Echantillonnage

L'échantillonnage consiste à représenter un signal analogique continus(t) par un ensemble de valeurs s (nTe) avec n entier situées à des instants discrets espacés de Te constante, appelée la période d'échantillonnage cette opération est réalisée par un circuit appelé «échantillonneur».

Le principe de base d'un échantillonneur est réalisé à l'aide d'un interrupteur qui s'ouvre et se ferme périodiquement à la cadence $f_e = \frac{1}{T_e}$ il s'agit d'un échantillonnage naturel. Le signal apparait en sortie quand l'interrupteur est fermé (position1).

Pendant τ secondes et disparait quand l'interrupteur est ouvert (position2) pendant le reste de la période T_e. De ce fait le signal échantillonné est une suite d'impulsions de très faible Durée du signal d'entrée.[5]

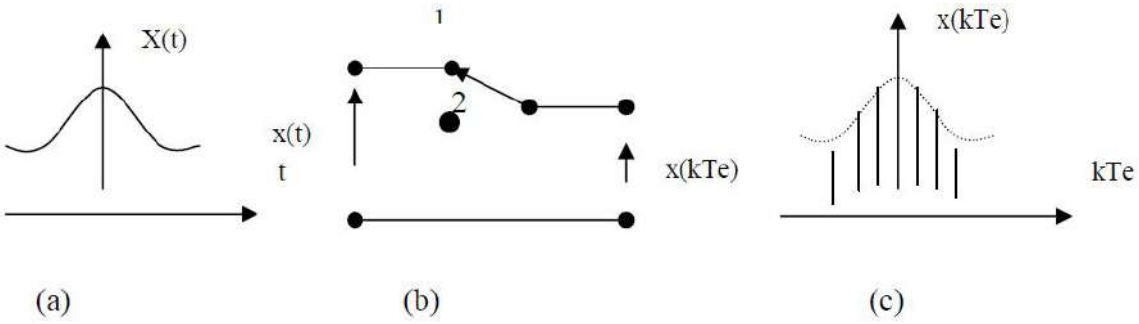


Figure I.2 : Echantillonnage d'un signal [3]

Le signal échantillonné se réduit donc à une simple multiplication du signal analogique à échantillonner s(t), par une fonction δ (t). (Dite : peine de Dirac), qu'est une suite d'impulsions rectangulaires d'amplitude unité, de durée τ et de période Te,

Donc : $S(kT) = S(t).\delta_T(t)$

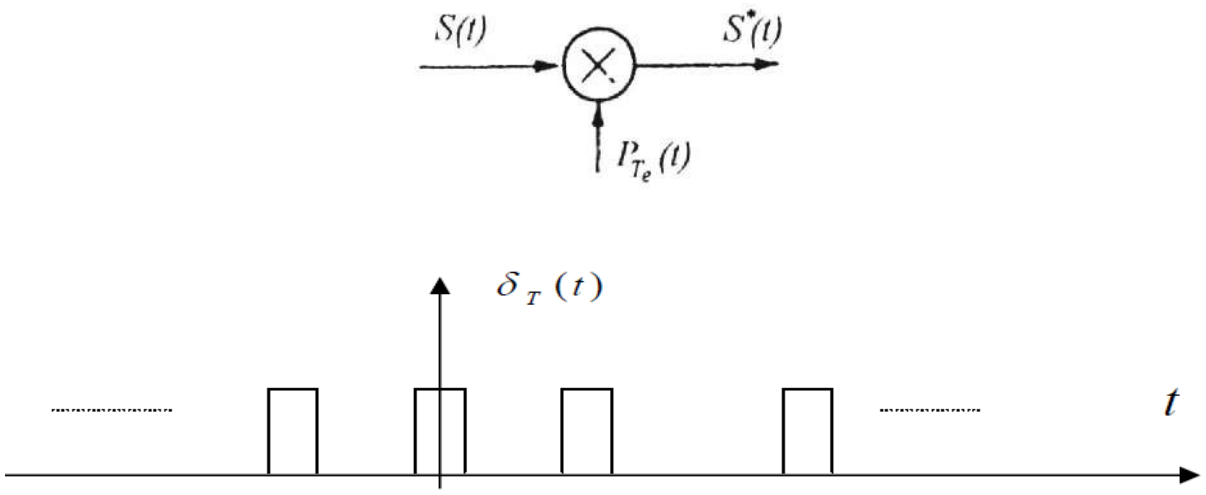


Figure I.3 : Méthode d'échantillonnage [3]

IL existe d'autre types d'échantillonneurs tels que :

- Echantillonneur idéal.
- Echantillonneur à l'aide d'une impulsion de forme quelconque.
- Echantillonneur avec maintien.

L'opération d'échantillonnage et son incidence sur le spectre du signal sont représentées sur la figure (1.4).

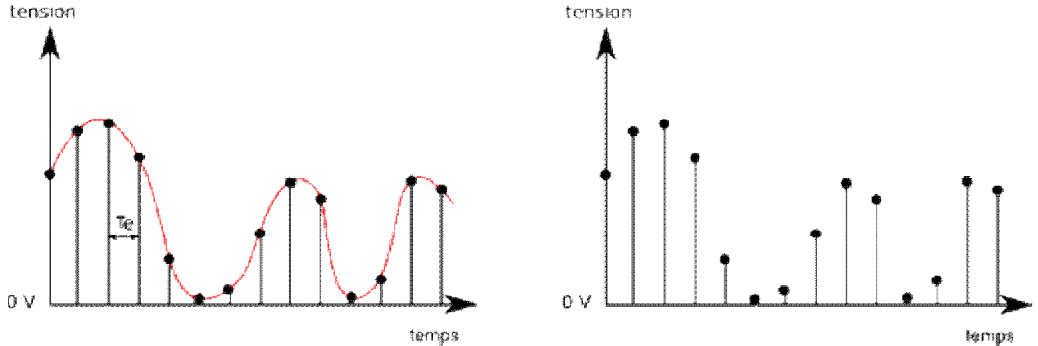


Figure I.4 : Spectre d'un signal échantillonné [3]

I.3.1.2 Théorème de SHANNON [6]

Enoncé du Théorème :

Pour que la répétition périodique du spectre ne déforme pas le motif répété, il faut et il suffit Que la fréquence de répétition (fe) qu'est la fréquence d'échantillonnage, soit égale ou supérieure à$2f_{max}$.

fmax : est la fréquence maximale pouvant être lue sur le spectre, c'est-à-dire la composante Harmonique à la plus grande fréquence contenue dans le signal temporel.

$$f_e \geq 2f_{max} \qquad (1.1)$$

I.3.1.3 Bruit de quantification

Lorsque l'on converti un signal analogique en un signal numérique, on observe obligatoirement une perte de l'information. Cette perte est due à la quantification comme l'illustre la figure ci-dessous:

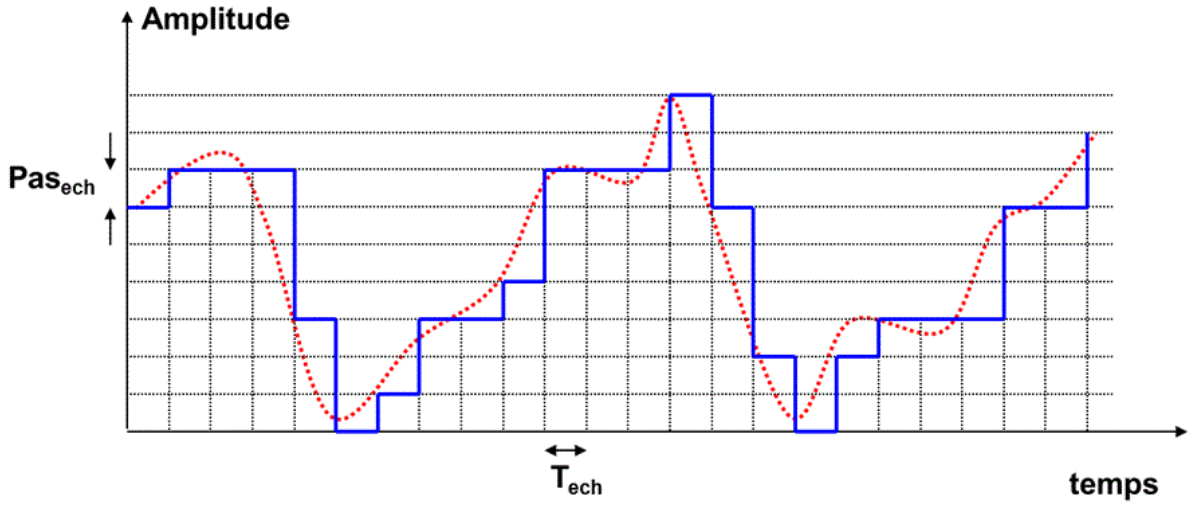


Figure I.5: Erreur de quantification d'un signal triangulaire. [6]

On admettra que ce résultat est validepour un signal pleine échelle triangulaire ou sinusoïdaldès que la résolution est supérieure à 6.

Le rapport signal sur bruit (SNRpour Signal to Noise Ratio) d'un CAN idéal est définitpour une entrée sinusoïdale pleine échelle, c'est le quotient entre la valeur efficace du signal

Veff, sinus et celle du bruit Veff, bruit (s'agissant d'un CAN idéal le bruit se réduit au bruit de Quantification) :

$$SNR = \frac{V_{eff.sinus}}{V_{eff.bruit}}$$

Avec

$$V_{eff.sinus} = \frac{V_{PE}}{2\sqrt{2}} = \frac{2^{N-1}q}{\sqrt{2}}$$

Et

$$V_{eff.bruit} = \sqrt{P_{bruit}} = \frac{q}{\sqrt{12}}$$

D'où $SNR = \sqrt{6.2^{N-1}}$

Soit en dB SNR dB = 20logSNR = 6,02N+1,76

Le SNR d'un CAN augmente avec sa résolution (gain de 6 dB par bit supplémentaire).

Ce résultat représente le SNR maximal atteignable pour un convertisseur (il n'est valable quepour un signal sinusoïdal pleine échelle). [7]

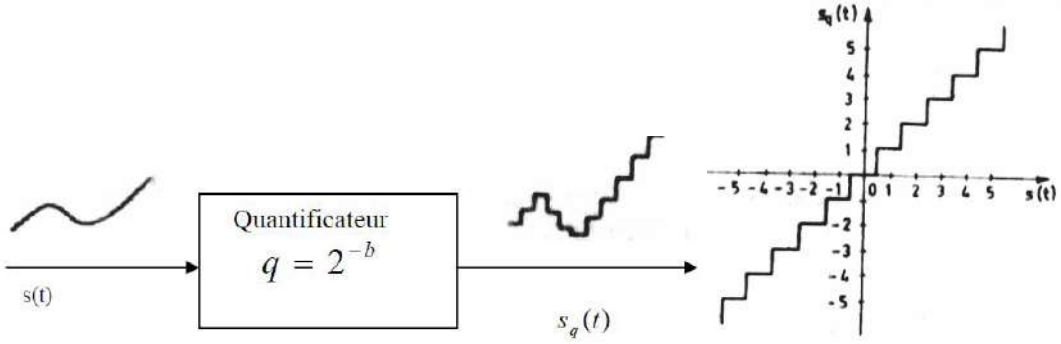


Figure 1.6 : Quantification d'un signal [7]

I.3.2 Outils mathématiques du traitement [1]

I.3.2.1 Analyse temporelle

L'analyse temporelle consiste à analyser directement le signal de parole en tant que vecteur devaleurs prises par ce signal dans le temps. Elle se fait par plusieurs outils mathématiques dont nous utiliserons les suivants :

A) Moyenne

La moyenne d'un signal numérique X notée$\bar{X}$ composé de N échantillons est donnée par la relation :

$$\bar{X} = \frac{1}{N}\sum_{i=1}^{n} x(n) \qquad (1.2)$$

B) Covariance

La covariance nous renseigne sur le sens de variation de deux variables aléatoires. Sa valeur est donnée par la relation :

$$cov(x,y) = \frac{1}{N}\sum_{i=1}^{n}(x_i - \bar{X})(y_{i-}\bar{y})t \qquad (1.3)$$

X et Y représentent 2 signaux numériques.
N représente le nombre d'échantillons des signaux.

I.3.2.2 Analyse fréquentielle

A) Transformation de Fourier

L'analyse de Fourier appliquée à un signal quantifié peut être schématisée par une chaîne de traitement composée de plusieurs fonctions indispensables pour l'analyse spectrale du signal d'entrée. La transformée de Fourier rapide (ou filtre de Fourier), joue un rôle principal dans cette opération de filtrage, sans oublier les autres fonctions telles que l'échantillonnage et la quantification, les opérations de fenêtrage et de lissage nécessaires au procédé de traitement. La figure (1.7) représente la chaîne de traitement comprenant toutes les fonctions nécessaires au calcul du spectre de puissance d'un signal stationnaire. [8]

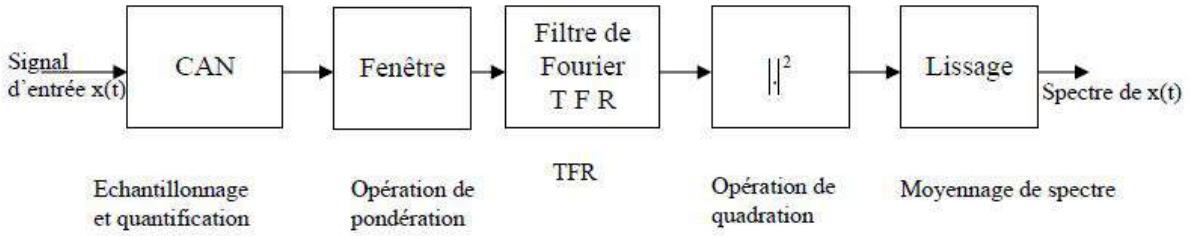


Figure I.7 : chaîne de traitement [8]

On définit la Transformée de Fourier d'un signal par :

$$S(f) = \int_{-\infty}^{\infty} s(t)e^{-j2\pi ft}\,dt \qquad (1.4)$$

La fonction *S*(*f*) est périodique de période 1, et celle est généralement, une fonction complexe de la variable *f*.
La transformation inverse de Fourier est définie par :

$$s(t) = \int_{-\infty}^{\infty} S(f) e^{j2\pi ft}\, df \qquad (1.5)$$

Il y a deux difficultés associées à la relation précédente ; la première est que *f* est une variable continue, on ne peut pas manipuler dans un système de traitement numérique. La deuxième est due au nombre infini d'échantillons du signal *s*(*t*), qu'il est impossible de traiter pratiquement.

Les solutions à ces deux problèmes sont immédiates ; Il faut remplacer la variable continue par variable discrète, et il faut limiter la durée du signal $s_T(k)$. [9]

B)Transformée de Fourier rapide

La transformée de Fourier rapide **TFR** ou **FFT** (Fast Fourier Transform) est simplement un algorithme permettant de réduire le nombre d'opérations, en particulier le nombre de multiplications, pour calculer la TFD. Ce temps de calcul est en effet primordial pour réaliser des systèmes numériques en « temps réel ».[10]
La Transformée de Fourier rapide est définie par :

$$S_m = \sum_{k=0}^{N-1} s_k . W_N^{km} \qquad (1.6)$$

Avec :

$$W_N = e^{-j2\pi/N} \qquad (1.7)$$

C) Spectrogramme

Le signal de parole n'étant pas stationnaire, la transformée de Fourier simple n'est pas adaptée à son analyse.

En effet l'étude d'un signal non stationnaire par la transformée de Fourier ne tient pas compte de l'évolution du signal au cours du temps. Ainsi, si nous avons un signal constitué de deux sinusoïdes successives de fréquences différentes, son analyse par une simple transformée de Fourier ne nous permettra pas de savoir laquelle des deux sinusoïdes est apparue en premier, et le spectre obtenu pourra nous amener à penser que les deux sinusoïdes se chevauchent alors que ce n'est pas le cas.

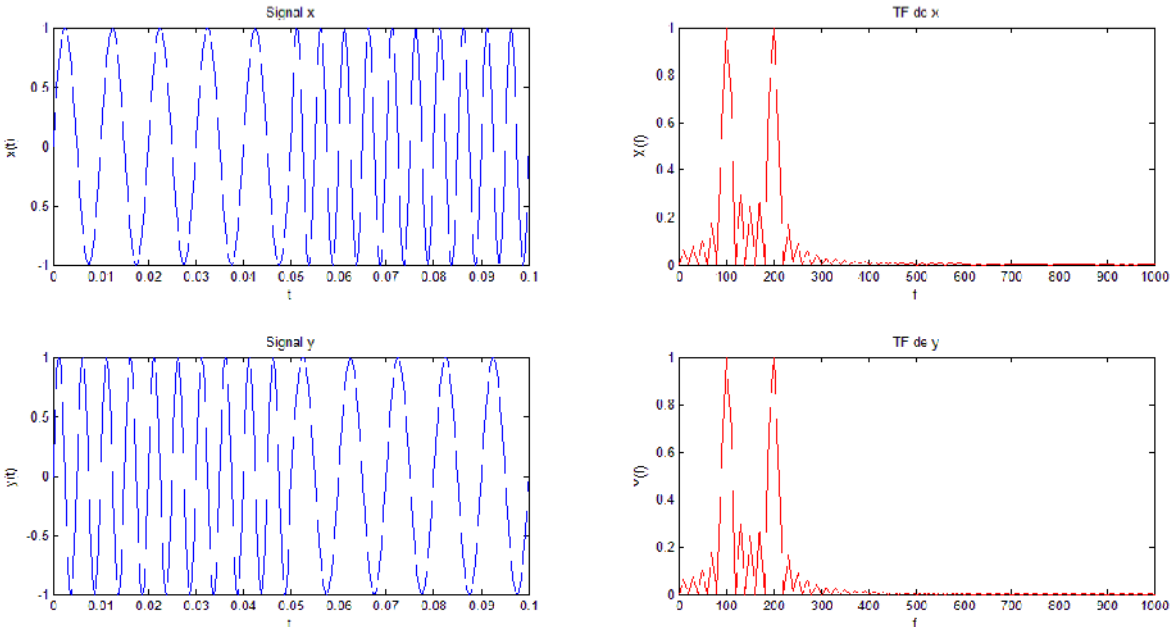


Figure I.8 : Transformées de Fourier de deux signaux différents. [10]

La figure ci-dessus illustre cette situation. Nous avons deux signaux différents dont le premier est constitué d'une sinusoïde de 100 Hz suivie d'une sinusoïde de 200 Hz, tandis que le second présente les mêmes sinusoïdes mais cette fois ci inversées (celle de 200 Hz précède celle de 100 Hz). On constate à droite que les spectres sont identiques, alors que les deux signaux transportent des informations différentes.

Pour remédier à ce problème nous devons utiliser une analyse temps-fréquence. Ce type d'analyse consiste à étudier un signal non stationnaire sur des intervalles de temps où nous sommes sûrs que celui-ci est stationnaire.

Pour cela, on multiplie le signal à analyser par des fenêtres (opération appelée le fenêtrage) qui nous permettent d'analyser seulement une partie du signal complet. La largeur de ces fenêtres est choisie de manière à s'assurer que le signal soit toujours stationnaire sur ces intervalles. Une première méthode, appelée Transformée de Fourier à Court Terme, permet de réaliser cette opération. Comme décrit précédemment, la Transformée de Fourier à Court Terme permet d'étudier un signal sur des fenêtres où celui-ci est toujours stationnaire.[11]

I.4.Extraction des caractéristiques et création du modèle de référence

Suite à la suppression de toutes les informations non pertinentes dans les signaux vocaux, la parole claire est ensuite convertie en flux de coefficients de vecteurs de caractéristiques contenant uniquement les informations sur l'énoncé parlé qui sont nécessaires pour sa reconnaissance. Nous avons cependant utilisé la technique d'analyse spectrale bien établie de l'algorithme du coefficient cepstral de fréquence mel (MFCC) pour l'extraction des vecteurs de caractéristiques en raison de sa robustesse. La figure 4 montre le processus de flux MFCC pour l'extraction des coefficients des vecteurs de caractéristiques du signal de parole.

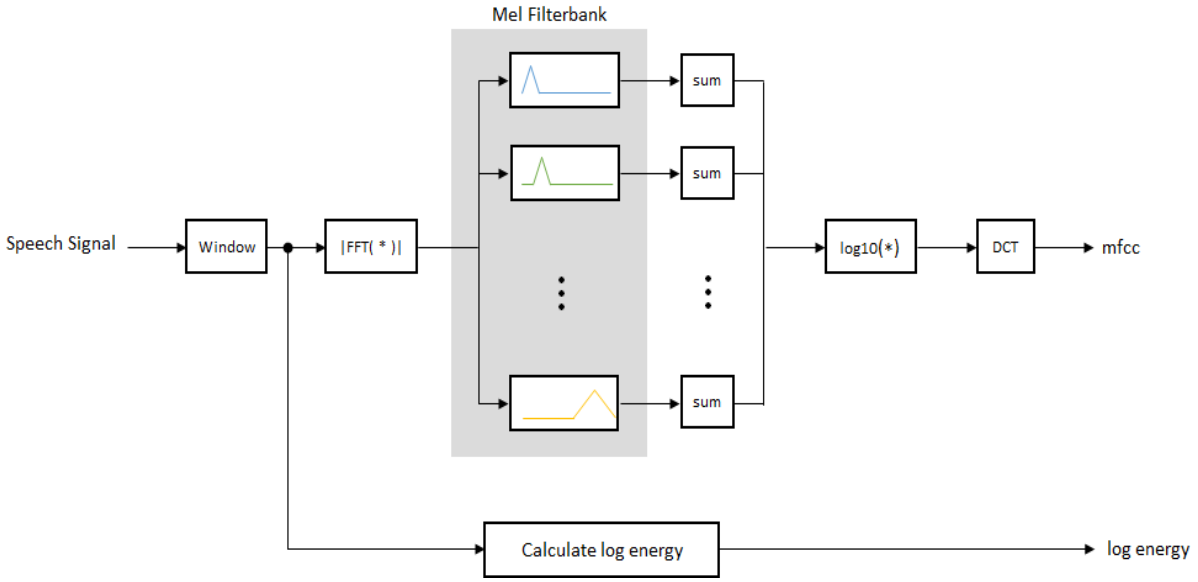


Figure I.9 : Schéma fonctionnel du fonctionnement MFCC pour les données de fonction vocale

Le banc de filtres Mel espace linéairement les 10 premiers filtres triangulaires et logarithmique les filtres restants. Les bandes individuelles sont pondérées pour une énergie uniforme. Vous trouverez ci-dessous une visualisation d'un banc de filtres Mel typique.

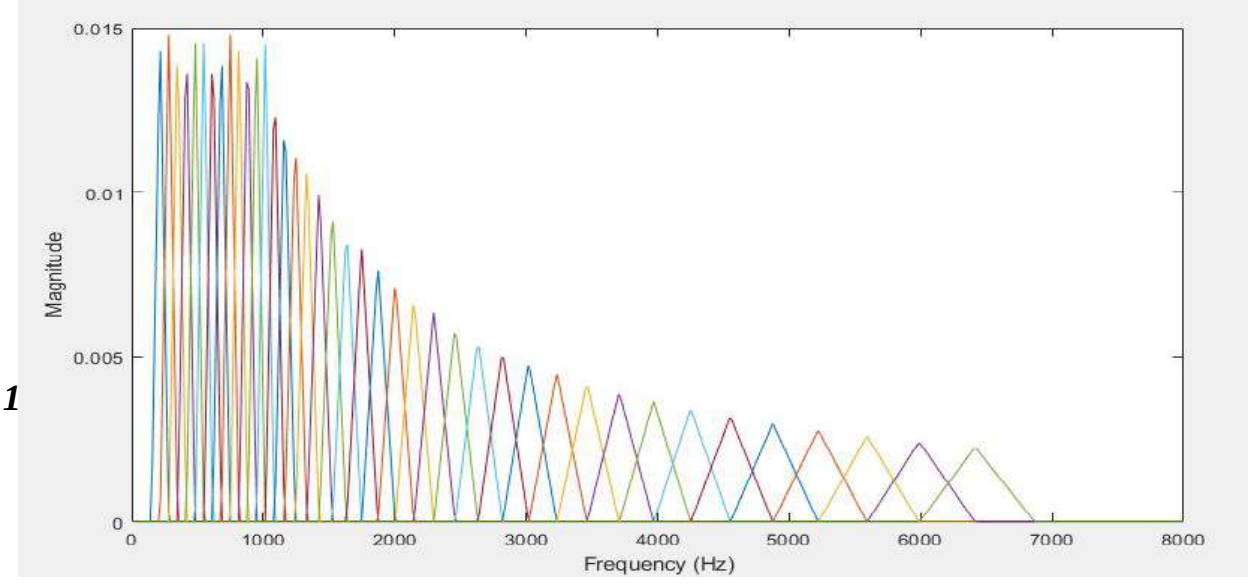


Figure I.10 : Banc de filtres Mel typique

I.5.Reconnaissance automatique du locuteur

On définit le cadre contextuel actuel de la reconnaissance de l'identité de la personne à travers sa voix, on esquisse les concepts de base méthodologiques et de nature informatique. On tente ensuite une investigation des sources de variabilités de la voix et on propose des formalisations abstraites de la vérification et de l'identification. On détaille les diverses méthodologies existantes et on propose leur comparaison, on introduit la reconnaissance par ordinateur. Enfin, on s'intéresse à l'aspect temporel du signal et on fait une sélection fondée sur la relation entre les segments acoustiques mesurés et des considérations physiologiques. [12]

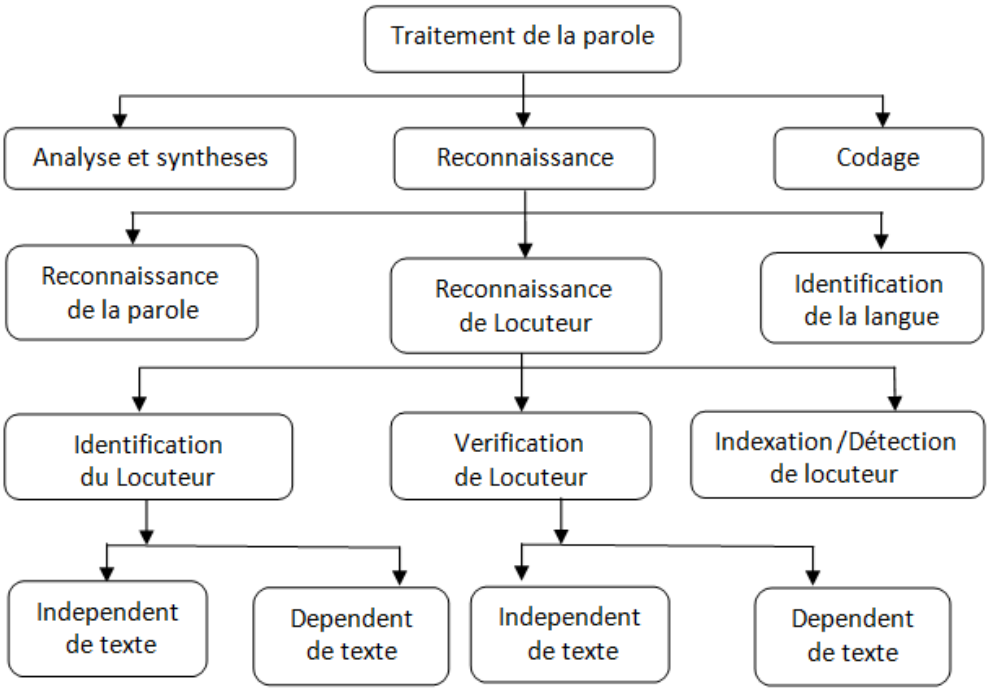


Figure I.9 : Hiérarchisation des domaines du traitement de la parole. [12]

I.6Modélisation du locuteur

Une fois le vecteur extrait à partir de la trace de test le système procède à la création d'unmodèle du locuteur qui sera plus tard comparé à d'autres modèles déjà présents et produits de la même façon lors de l'entraînement.

La technique la plus répandue en RAL pour la modélisation est représentée par le i vecteurs, basé sur le modèle linéaire de Gaussiennes (voir paragraphe suivant, GMM-UBM).

Le locuteur est caractérisé selon une approche descriptive ce qui facilite les mesures statistiques sur ces vecteurs qui obtiennent de meilleurs résultats sur des échantillons plus courts (Li et al, 2016). Un autre moyen de représentation des vecteurs acoustiques est celui nommé d-vecteurs. Cette représentation est moins répandue et plus récente puisqu'elle se relie aux techniques des réseaux neurones (Richardson et al, 2015). Dans cette approche le locuteur est représenté en gardant uniquement les variances les plus pertinentes, la phase d'entrainement devient déterminante pour que le système apprenne quels sont les traits discriminants.

Encore d'autres techniques sont proposées dans les travaux de (Cai et al, 2018), principalement à travers des réseaux neurones, ils représentent peut-être les futurs développements dans le domaine. Une tendance à créer des systèmes hybrides est attestée (Trabelsi et Ayed, 2013) pour maximiser la quantité d'informations contenues dans les vecteurs, ce qui demande un travail plus important par le système mais qui pourra garantir une discrimination plus fine pour certains paramètres, notamment les indices paralinguistiques (Scootz, 2002) comme l'accent ou la prosodie qui pourraient être impliqués dans la modélisation d'un locuteur provenant d'une certaine région ou pays.[13]

I.7Conclusion

Dans ce chapitre, on a évoqué essentiellement des généralités et des définitions fondamentales en identification de locuteur.L'analyse spectrale à base de FFT a été particulièrement exposée. On a présenté aussi, le traitement de la parole et on discute la notion de reconnaissance automatique du locuteur et son modélisation .

Chapitre I : Généralités sur l'identification de locuteurs

Chapitre II

Les réseaux de neurones artificiels

Chapitre II : Les réseaux de neurones artificiels

II.1 Introduction

Le cerveau humain a une fantastique puissance de traitement de l'information si l'on considère ses capacités à prendre en charge certaines taches nécessaires pour obtenir un comportement intelligent. La nature de l'intelligence a longtemps été un sujet difficile et controverse. D'une certaine manière, chacun a une idée assez vague de ce qu'est l'intelligence :

La capacité à observer, à comprendre, à se souvenir, à résoudre des problèmes, à apprendre, à créer,…etc. Deux disciplines sont concernées par la définition et la modélisation du comportement intelligent : ce sont la psychologie cognitive et l'intelligence artificielle (IA en jargon informatique).

La psychologie cognitive est l'étude scientifique des structures de mémorisation et des processus de raisonnement humain. Le but de l'IA est le développement des machines capables d'avoir un comportement intelligent. Le critère pour estimer la qualité d'un modèle en « IA » diffère de ceux de la neurophysiologie ou des sciences cognitives. En « IA », on met l'accent sur le développement des systèmes à hautes performances plutôt que sur la reproduction fidèle ou l'explication du comportement humain et a fortiori des données neuro physiologiques. Au fur et à mesure que progresse l'informatique dans les années cinquante et soixante, il devient clair que les domaines privilégiés à court terme de l'ordinateur sont le calcul scientifique et la gestion. Les performances de celui-ci dans des tâches plus ″intelligentes″ (traduction automatique, vision et reconnaissance des formes sont plus discutables).

Les techniques de l'intelligence artificielle telles que les systèmes experts, la logique floue, les algorithmes génétiques et les réseaux de neurones artificiels (RNA) ont été largement utilisées dans le domaine de l'électronique de puissance et de la commande des machines électriques.

L'objectif recherché dans l'utilisation des techniques de l'intelligence artificielle est d'arriver à l'émulation du raisonnement humain sur un DSP (Digital Signal Processor) (processeur de signal numérique) de telle sorte que le système complet commande - machine puisse penser et réagir intelligemment comme un être humain. Un système commande - machine équipé d'un algorithme développant un calcul par intelligence est appelé système intelligent. En effet, un système intelligent possède la caractéristique d'apprentissage, d'auto-organisation et d'auto-adaptation.

Les techniques de l'intelligence artificielle ont été discutées pendant longtemps et le seront également à l'avenir.

Chapitre II : Les réseaux de neurones artificiels

Actuellement, les techniques de l'intelligence artificielle sont largement utilisées dans de nombreux domaines tels que la régulation de processus industriels, le traitement d'image, le diagnostic, la médecine, la technologie spatiale et les systèmes de gestion de données informatiques. Parmi toutes les techniques intelligentes, les réseaux de neurones artificiels (RNA) semble avoir le maximum d'impact dans le domaine de l'électronique de puissance et dans la commande de machines électriques, ce qui est évident par le nombre important de publications réalisées dans la littérature.

Dans ce chapitre, on vadiscuter surl'historique desréseaux de neurones, Le neurones biologiques ainsi que un aperçu général sur les réseaux de neurones artificiels.

II .2 Historique

Les recherches sur les méthodes neuronales de traitement de l'information en vue de modéliser le comportement du cerveau humain ne sont pas récentes, en :

- 1890 : W. James, célèbre psychologue américain introduit le concept de mémoire associative, et propose ce qui deviendra une loi de fonctionnement pour l'apprentissage sur les réseaux de neurones connue plus tard sous le nom de loi de Hebb.
- 1943 : J. Mc Culloch et W. Pitts, laissent leurs noms à une modélisation du neurone biologique (un neurone au comportement binaire). Ceux sont les premiers à montrer que des réseaux de neurones formels simples peuvent réaliser des fonctions logiques, arithmétiques et symboliques complexes (tout au moins au niveau théorique).
- 1949 : D. Hebb, physiologiste américain explique le conditionnement chez l'animal par les propriétés des neurones eux-mêmes. Ainsi, un conditionnement de type pavlovien tel que, nourrir tous les jours à la même heure un chien, entraine chez cet animal la sécrétion de salive à cette heure précise même en l'absence de nourriture. La loi de modification des propriétés des connexions entre neurones qu'il propose explique en partie ce type de résultats expérimentaux.

Les premiers succès :

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

- 1960 : B. Widrow, un automaticien, développe le modèle ADALINE (ADAptiveLInearNEuron). Dans sa structure, le modèle ressemble au Perceptron, cependant la loi d'apprentissage est différente. Celle-ci est à l'origine de l'algorithme de retro propagation de gradient très utilisé aujourd'hui avec les Perceptrons multicouches. Les réseaux de type ADALINE restent utilisés de nos jours pour certaines applications particulières.
- 1969 : M.L. Minsky et S. Papert publient ensuite un ouvrage qui met en évidence les limitations théoriques du Perceptron. Ces limitations concernent l'impossibilité de traiter des problèmes non linéaires en utilisant ce modèle.
- 1972 : T. Kohonen présente ses travaux sur les mémoires associatives et propose des applications à la reconnaissance de formes.

Le renouveau :

- 1982 : J.J. Hopfield est un physicien reconnu à qui l'on doit le renouveau d'intérêt pour les réseaux de neurones artificiels. Il présente une théorie du fonctionnement et des possibilités des réseaux de neurones.

La levée des limitations :

- 1983 : La machine de Boltzmann est le premier modèle connu apte à traiter de manière satisfaisante les limitations recensées dans le cas du Perceptron. Mais l'utilisation pratiques'avère difficile, la convergence de l'algorithme étant extrêmement longue (les temps de calcul sont considérables).
- 1985 : La rétro propagation de gradient apparaît. C'est un algorithme d'apprentissage adapté aux réseaux de neurones multicouches (aussi appelés Perceptrons multicouches). Sa découverte réalisée par trois groupes de chercheurs indépendants indique que « la chose était dans l'air ». Dès cette découverte, nous avons la possibilité de réaliser une fonction non linéaire d'entrée/sortie sur un réseau en décomposant cette fonction en une suite d'étapes linéairement séparables. De nos jours, les réseaux multicouches et la rétro propagation de gradient reste le modèle le plus productif au niveau des applications.

Depuis ce temps, le domaine des réseaux de neurones fourni constamment de nouvelles théories, de nouvelles structures et de nouveaux algorithmes. Dans ce chapitre, nous allons tenter d'exposer les plus importants. [14]

II .3 Le neurones biologiques

Chapitre II : Les réseaux de neurones artificiels

Le cerveau humain contient environ 100 milliards de neurones. Ces neurones vous permettent, entre autre, de lire ce texte tout en maintenant une respiration régulière permettant d'oxygéner votre sang, en actionnant votre cœur qui assure une circulation efficace de ce sang pour nourrir vos cellules, etc.

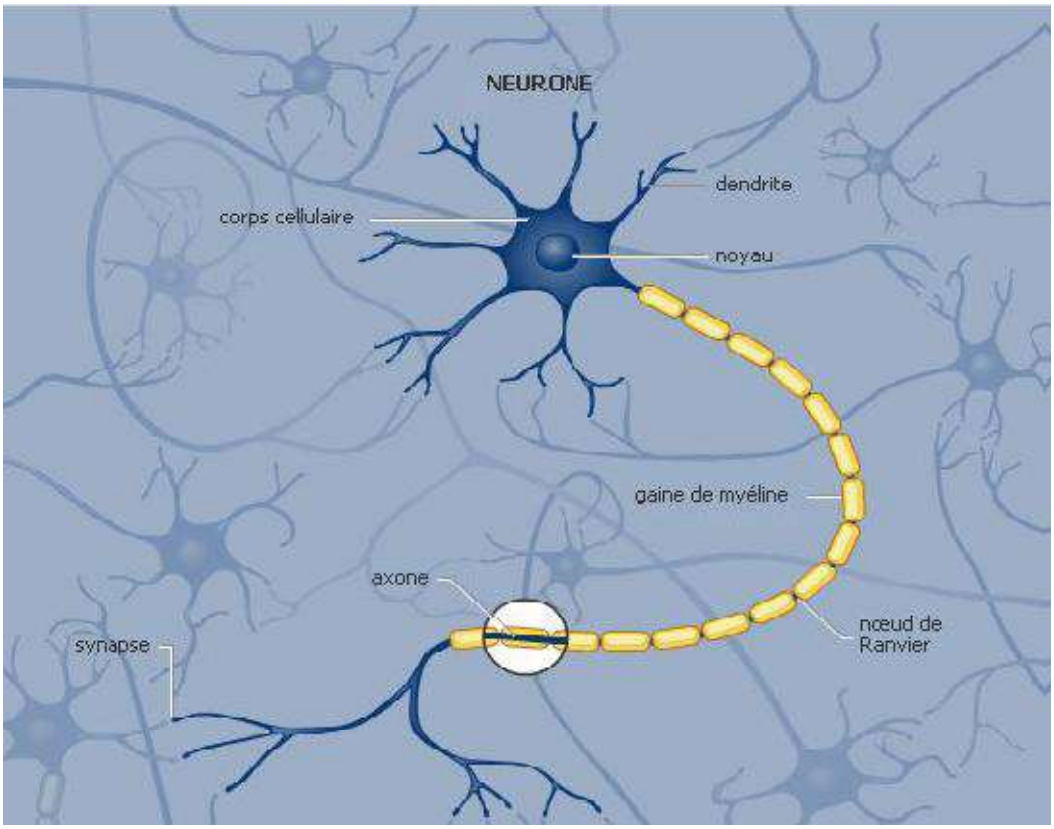


Figure II.1 : Modèle d'un neurone biologique. [15]

Chacun de ces neurones est par ailleurs fort complexe. Essentiellement, il s'agit de tissu vivant et de chimie. Les spécialistes des neurones biologiques (neurophysiologistes) commencent à peine à comprendre quelques-uns de leurs mécanismes internes. On croit en général que leurs différentes fonctions neuronales, y compris celle de la mémoire, sont stockées au niveau des connexions (synapses) entre les neurones. C'est ce genre de théorie qui a inspiré la plupart des architectures de réseaux de neurones artificiels que nous aborderons dans ce chapitre.

L'apprentissage consiste alors soit à établir de nouvelles connexions, soit à en modifier des existantes.

Chapitre II : Les réseaux de neurones artificiels

Un neurone biologique est une cellule particulière comme la montre (FigureII.1). Elle possède des extensions par lesquelles elle peut distribuer des signaux (axones) ou en recevoir (dendrites).

Dans le cerveau, les neurones sont reliés entre eux par l'intermédiaire des axones et des dendrites. En première approche. On peut considérer que ces sortes de filaments sont conductrices d'électricité et peuvent ainsi véhiculer des messages depuis un neurone vers un autre. Les dendrites représentent les entrées du neurone et son axone sa sortie.

Un neurone émet un signal en fonction des signaux qui lui proviennent des autres neurones.

On observe en fait au niveau d'un neurone, une intégration des signaux reçus au cours du temps, c'est à dire une sorte de sommations des signaux. En général, quand la somme dépasse un certain seuil, le neurone émet à son tour un signal électrique. [15]

II.3.1 Le corps cellulaire

Le corps cellulaire contient un noyau entouré des structures et des corpuscules nécessaires à son métabolisme. Le corps cellulaire peut se situer au milieu, sur le trajet principal du neurone comme dans les neurones bipolaires, vers l'extrémité terminale du neurone, comme dans le neurone multipolaire, ou sur une branche dérivé de la cellule nerveuse comme pour le neurone unipolaire représenté.[16]

II.3.2 Le noyau

le noyau est une structure cellulaire présente dans la majorité des cellules eucaryotes et chez tous les organismes eucaryotes, et contenant l'essentiel du matériel génétique de la cellule (ADN). Il s'agit donc d'une importante structure cellulaire. Généralement c'est la plus visible de la cellule à l'échelle microscopique. [17]

II.3.3 Les dendrites

Sont des prolongements du corps cellulaire des neurones dont elles partagent les organites (à l'exception du noyau et des lysosomes). Ces dendrites peuvent se diviser par dichotomies successives, s'affinant ainsi de leur origine trapue vers la périphérie, ce qui leur confère un aspect arborescent. [18]

II.3.4 La synapse

La notion de synapse explique la transmission des signaux entre un axone et une dendrite. Au niveau de la jonction (c'est à dire de la synapse), il existe un espace vide à travers lequel le signal électrique ne peut pas se propager. La transmission se fait alors par l'intermédiaire de substances chimiques, les neuromédiateurs. Quand un signal arrive au niveau de la synapse, il provoque l'émission de neuromédiateurs qui vont se fixer sur des récepteurs de l'autre côté de l'espace inter synaptique (FigureII.2).

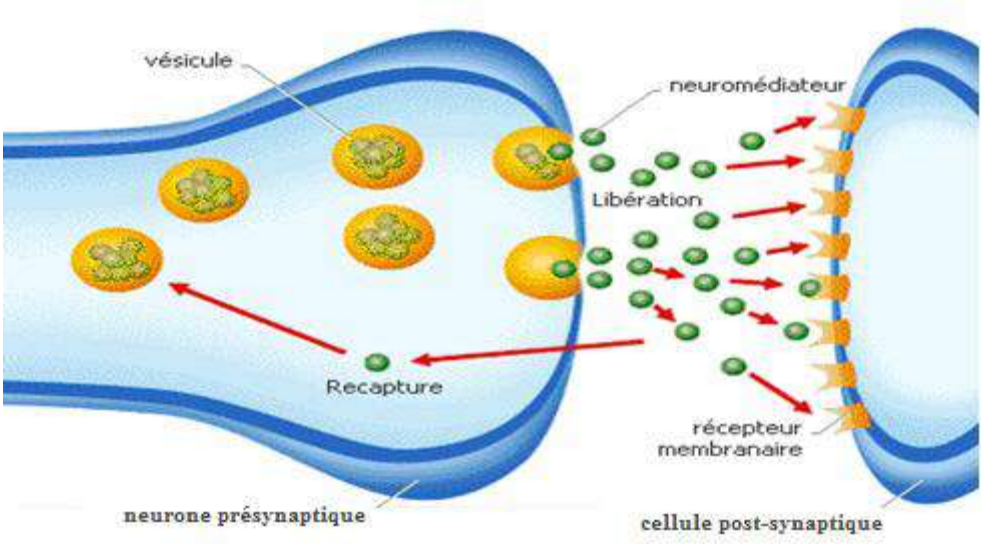


Figure II.2 : Transmission de l'information entre les neurones biologiques.[19]

Quand suffisamment de molécules se sont fixées, un signal électrique est émis de l'autre côté et on a donc une transmission. En fait, suivant le type de la synapse, l'activité d'un neurone peut renforcer ou diminuer l'activité de ces voisins. On parle ainsi de synapse excitatrice ou inhibitrice.[19]

II.3.5 L'axone

L'axone est un long prolongement qui émerge du corps cellulaire du neurone. Sa longueur est variable et peut atteindre plus d'un mètre. Son extrémité se divise en "branches" qui se connectent à d'autres neurones par l'intermédiaire d'une structure particulière : la synapse.

[20]

II.3.6 La gaine de myéline

La gaine de myéline est une enveloppe faite d'un tissu graisseux blanc qui recouvre les neurones du système nerveux périphérique. [21]

II.3.7 Le nœud de Ranvier

Les nœuds de Ranvier correspondent à un amincissement de la gaine isolante, appelée gaine de myéline, qui entoure les neurones.

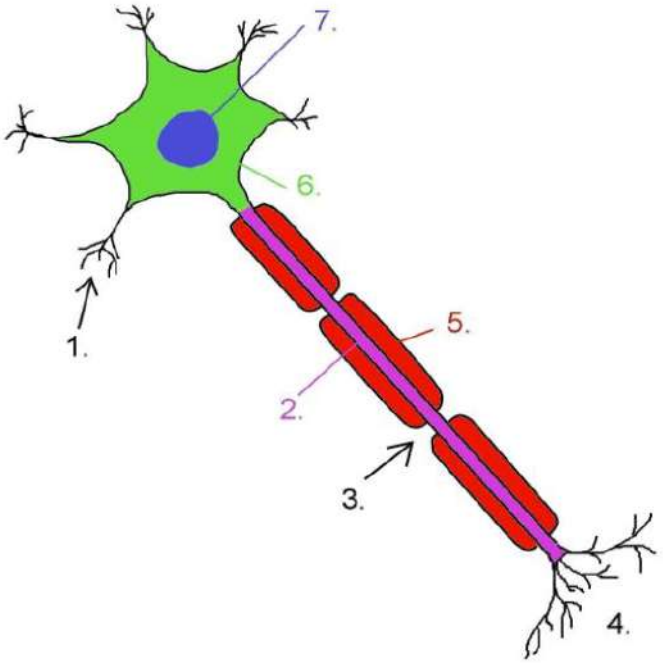


Figure II.3 : Schéma d'un neuronebiologique.[22]

On y trouve une dendrite (1), un axone (2), un nœud de Ranvier (3), l'extrémité de l'axone (4), la myéline (5), le corps cellulaire (6) et le noyau (7). Grâce à la myéline et aux nœuds de Ranvier, la vitesse de progression de l'influx nerveux dans l'axone est maximisée.[22]

II.4 Réseaux de neurones artificiels

II.4.1 Définition d'un RNA :

Un réseau de neurones est un assemblage de constituants élémentaires interconnectés (appelés «neurones» en hommage à leur modèle biologique), qui réalisent chacun un traitement simple mais dont l'ensemble en interaction fait émerger des propriétés globales complexes.

Les RNA constituent une technique de traitement de données bien comprise et bien maitrisée. Ces techniques s'intègrent parfaitement dans les stratégies de commande. En effet, elles réalisent des fonctionnalités d'identification, de contrôle ou de filtrage, et prolonge les

techniques classiques de l'automatique non linéaire pour aboutir à des solutions plus efficaces et robustes. [15]

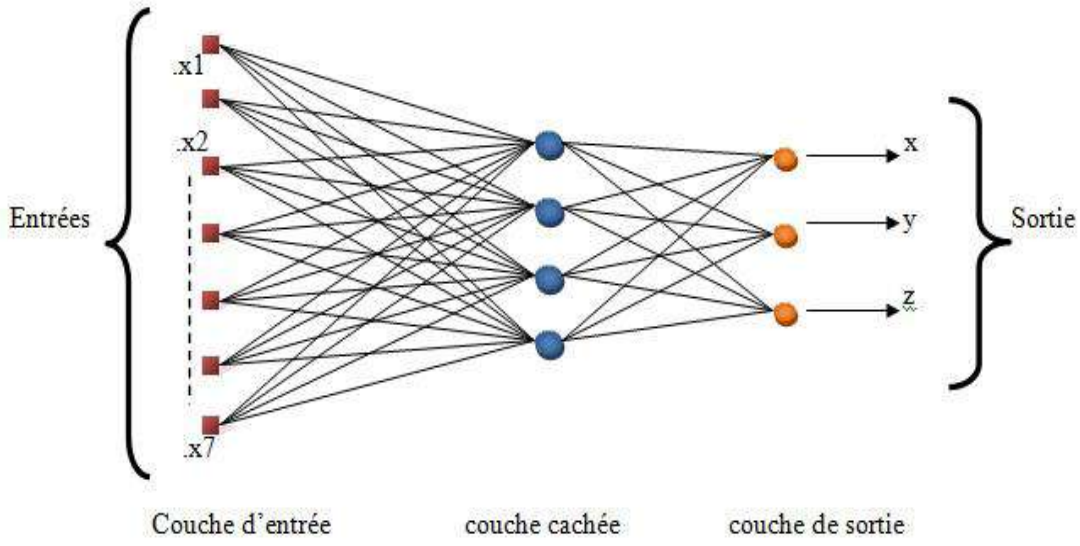


Figure II.4 : Neurone Artificiel. [15]

II.4.2 Principe de fonctionnement d'un RNA :

L'origine des réseaux de neurones vient de l'essai de modélisation mathématique du cerveau humain les premiers travaux datent de 1943 et sont l'œuvre de Warren .Mc. Culloch et Walter. Pitts. Ils supposent que l'impulsion nerveuse est le résultat d'un calcul simple effectué par chaque neurone et que la pensée née grâce à l'effet collectif d'un réseau de neurone interconnecté.

Un réseau de neurone ne se programme pas, il est entraîné grâce à un mécanisme d'apprentissage. Les tâches particulièrement adaptées au traitement par réseau de neurones sont: l'association, la classification, la discrimination, la prévision ou l'estimation, et la commande de processus complexes. Chaque neurone fonctionne indépendamment des autres de telle sorte que l'ensemble forme un système massivement parallèle. L'information est stockée de manière distribuée dans le réseau sous forme de coefficients synaptiques ou de fonctions d'activation, il n'y a donc pas de zone de mémoire et de zone de calcul, l'une et l'autre sont intimement liés.

Les réseaux de neurones artificiels consistent en des modèles plus ou moins inspirés du fonctionnement cérébral de l'être humain en se basant principalement sur le concept de neurone. [15]

II .4.3 Les composants d'un réseau de neurone

La figureci-dessous montre la structure d'un neurone artificiel. Chaque neurone artificiel est un processeur élémentaire. Il reçoit un nombre variable d'entrées en provenance de neurones amont. A chacune de ces entrées est associé un poids w abréviations de weight (poids en anglais) représentatif de la force de la connexion. Chaque processeur élémentaire est doté d'une sortie unique, qui se ramifie ensuite pour alimenter un nombre variable de neurones avals. A chaque connexion est associé un poids.

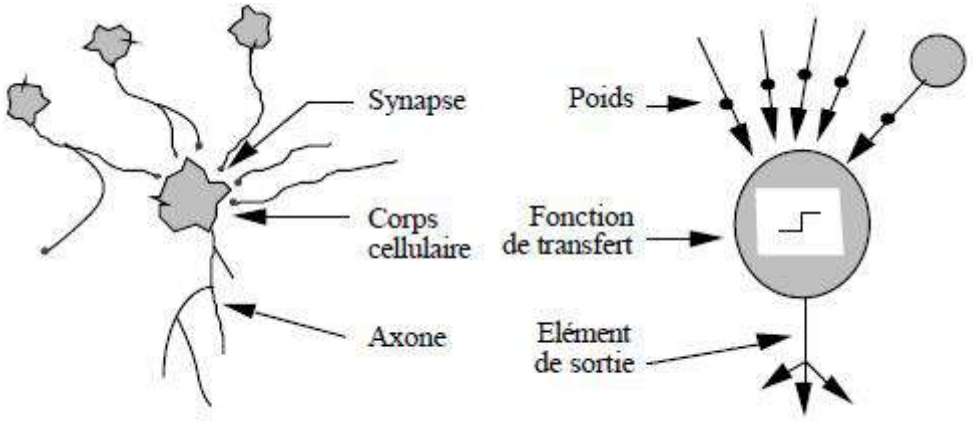


Figure II.5 : Mise en correspondance neurone biologique / neurone artificiel. [23]

La figure suivante donne les notations que nous utilisons dans ce travail.

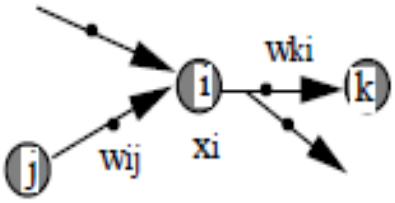


Figure II.6 : Structure d'un neurone artificiel. [23]

Pour le neurone d'indice i, les entrées sur Celui-ci est de poids w_{ij}alors que les connexions avals sont de poids w_{ki}.[23]

II .4.4 Architecture des réseaux de neurones

L'architecture d'un réseau de neurones est l'organisation des neurones entre eux au sein d'un même réseau. Autrement dit, il s'agit de la façon dont ils ordonnés et connectés. La majorité des réseaux de neurones utilise le même type de neurones. Quelques architectures

plus rares se basent sur des neurones dédiés. L'architecture d'un réseau de neurones dépend de la tâche à apprendre.

Un réseau de neurone est en général composé de plusieurs couches de neurones, des entrées jusqu'aux sorties. On distingue deux grands types d'architectures de réseaux de neurones : les réseaux de neurones non bouclés et les réseaux de neurones bouclés. [24]

II.4.4.1 Les réseaux de neurones non bouclés

Un réseau de neurones non bouclé réalise une (ou plusieurs) fonctions algébriques de ses entrées, par composition des fonctions réalisées par chacun de ses neurones. Un réseau de neurones non bouclé est représenté graphiquement par un ensemble de neurones "connectés" entre eux, l'information circulant des entrées vers les sorties sans "retour en arrière"; si l'on représente le réseau comme un graphe dont les nœuds sont les neurones et les arêtes les "connexions" entre ceux-ci, le graphe d'un réseau non bouclé est acyclique. Le terme de "connexions" est une métaphore : dans la très grande majorité des applications, les réseaux de neurones sont des formules algébriques dont les valeurs numériques sont calculées par des programmes d'ordinateurs, non des objets physiques (circuits électroniques spécialisés) ; néanmoins, le terme de connexion, issu des origines biologiques des réseaux de neurones, est passé dans l'usage, car il est commode quoique trompeur. Il a même donné naissance au terme de connexionnisme. [15]

- **Réseaux de neurones monocouches**

La structure d'un réseau monocouche est telle que des neurones organisés en entrée soient entièrement connectés à d'autres neurones organisés en sortie par une couche modifiable de poids (FigureII.7).

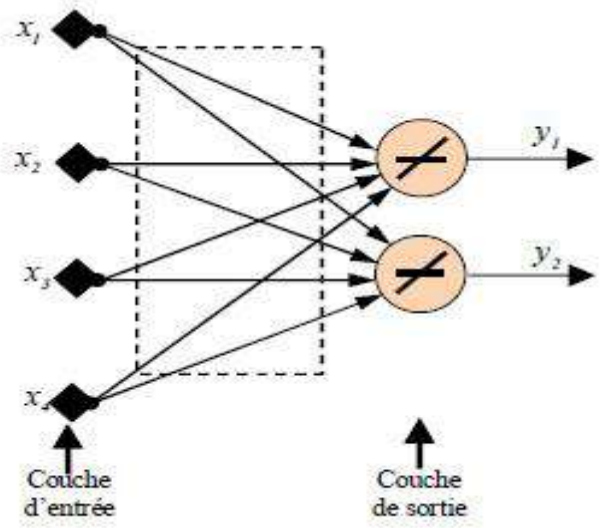


Figure II.7 : Schéma d'un réseau de neurones monocouche. [15]

- **Réseaux de neurones multicouches**

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 de couches avales. 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 appelées couches cachées.

La figure II.8 représente un réseau de neurones non bouclé qui a une structure particulière, très fréquemment utilisée : il comprend des entrées, deux couches de neurones cachés et des neurones de sortie. Les neurones de la couche cachée ne sont pas connectés entre eux. Cette structure est appelée Perceptron multicouches. [15]

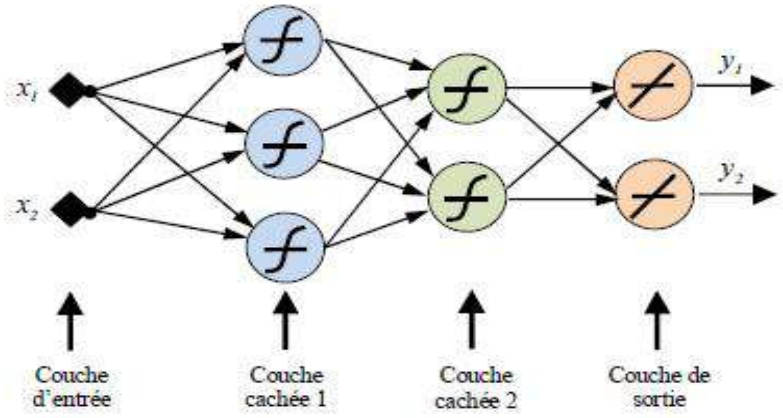


Figure II.8 : Schéma d'un réseau de neurones non bouclé (Perceptron multicouches). [15]

On note aussi que Les réseaux multicouches sont beaucoup plus puissants que les réseaux simples à une seule couche. En utilisant deux couches (une couche cachée et une couche de sortie), à condition d'employer une fonction d'activation sigmoïde sur la couche cachée, on peut entraîner un réseau à produire une approximation de la plupart des fonctions, avec une précision arbitraire (cela peut cependant requérir un grand nombre de neurones sur la couche cachée). Sauf dans des rares cas, les réseaux de neurones artificiels exploitent deux ou trois couches. [15]

- **Réseaux de neurones 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 II.9).

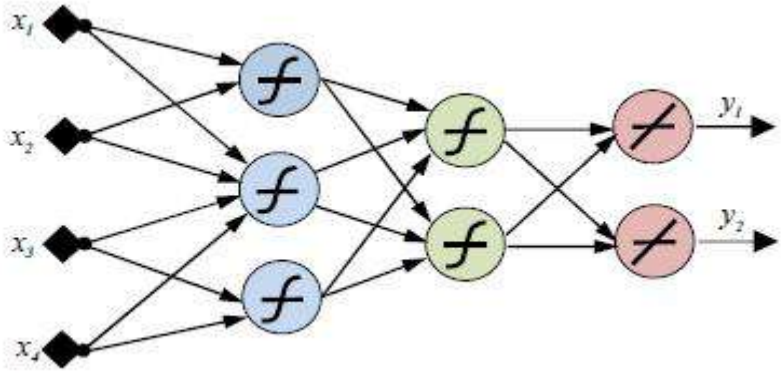


Figure II.9 : Schéma d'un réseau de neurones à connexions locales. [15]

Les réseaux de neurones non bouclés sont des objets statiques : si les entrées sont indépendantes du temps, les sorties le sont également. Ils sont utilisés principalement pour effectuer des tâches d'approximation de fonction non linéaire, de classification ou de modélisation de processus statiques non linéaire.

II .4.4.2 Réseaux de neurones bouclés

Contrairement aux réseaux de neurones non bouclés dont le graphe de connexions est acyclique, les réseaux de neurones bouclés peuvent avoir une topologie de connexions quelconque, comprenant notamment des boucles qui ramènent aux entrées la valeur d'une ou plusieurs sorties. Pour qu'un tel système soit causal, il faut évidemment qu'à toute boucle soit associé un retard : un réseau de neurones 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.

Il s'agit donc de réseaux de neurones avec retour en arrière (feedback network or récurrent network), (Figure II.10).

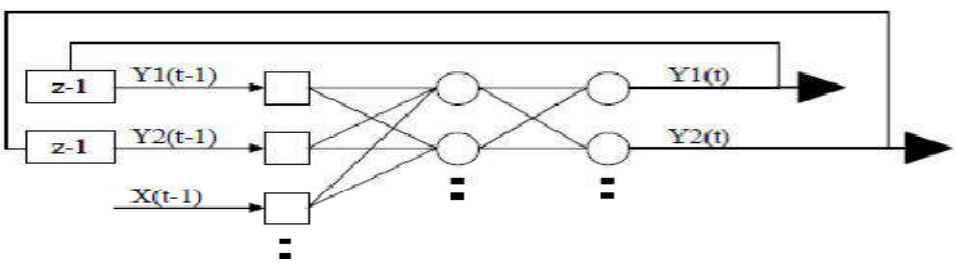


Figure II.10 :schéma de Réseau de neurone bouclé.

Un réseau de neurones bouclé à temps discret est donc régi par une (ou plusieurs) équations aux différences non linéaires, résultant de la composition des fonctions réalisées par chacun des neurones et des retards associés à chacune des connexions.

La forme la plus générale des équations régissant un réseau de neurones bouclé est appelée forme canonique ;

$$x(k+1) = \varphi[x(k), u(k)] \quad \textbf{(II.1)}$$

$$y(k) = \psi[x(k), u(k)] \quad \textbf{(II.2)}$$

Où φ et ψsont des fonctions non linéaires réalisées par un réseau de neurones non bouclé (mais pas obligatoirement, un perceptron multicouche), et où k désigne le temps (discret). La forme canonique est représentée sur la figure 11. Tout réseau de neurones, aussi compliqué soit-il, peut être mis sous cette forme canonique, de manière complètement automatique.

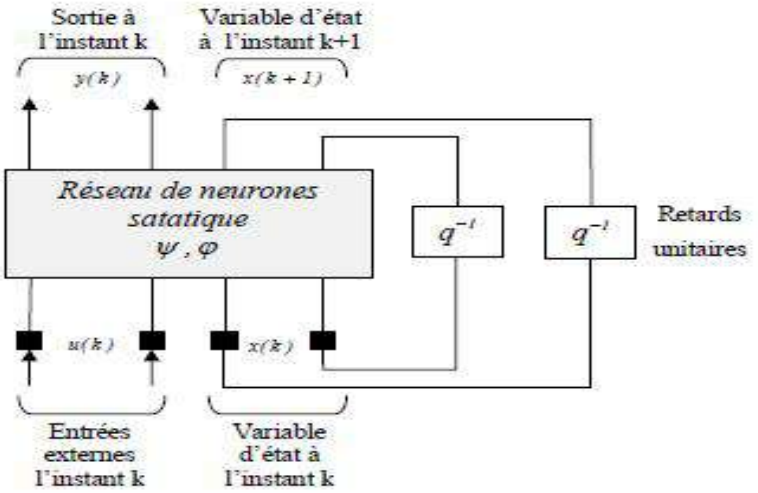


Figure II.11 : Forme canonique d'un réseau de neurones bouclé.

Les réseaux de neurones bouclés sont utilisés pour effectuer des tâches de modélisation de systèmes dynamiques, de commande de processus, ou de filtrage.

II .4.5 Les fonctions d'activations

C'est une fonction non-linéaire qui définit l'état interne du neurone en fonction de son entrée totale elle peut être une fonction binaire à seuil, linéaire à seuil ou multi-seuil sigmoïde stochastique ou une autre fonction généralement choisie croissante et impaire. [15]

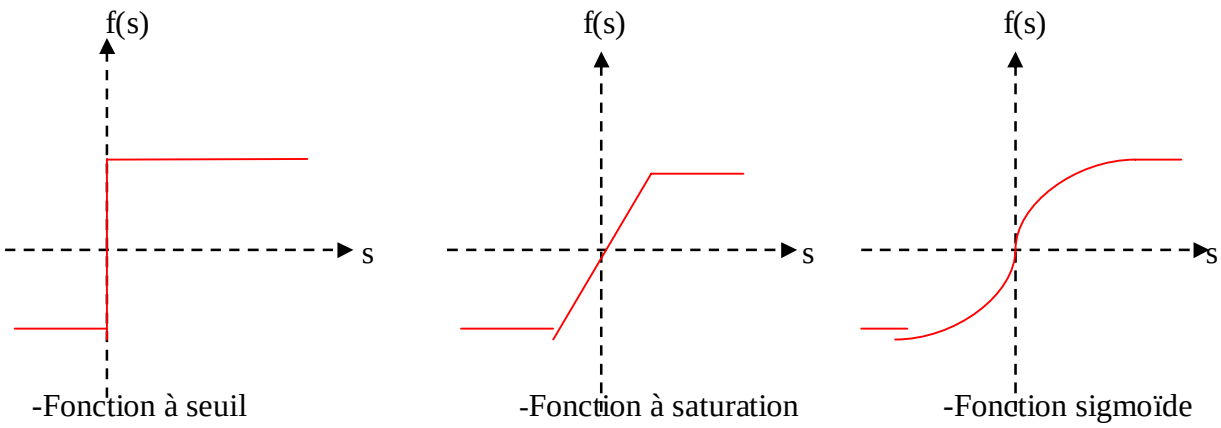


Figure II.12 : Quelque fonction d'activations. [15]

II.4.6 Les techniques d'apprentissage

L'apprentissage est un processus dynamique et itératif permettant de modifier les paramètres d'un réseau en réaction avec les stimuli qu'il reçoit de son environnement. Le type d'apprentissage est déterminé par la manière dont les changements de paramètre surviennent.

Il existe de nombreux types de règles d'apprentissage qui peuvent être regroupées en trois catégories : les règles d'apprentissage non supervisé, renforcé et supervisé. Mais l'objectif fondamental de l'apprentissage reste le même : soit la classification, l'approximation de fonction ou encore la prévision.

II.4.6.1 Apprentissage non-supervisé :

Ce type d'apprentissage est choisi lorsqu'il n'y pas de connaissances à priori des sorties désirés pour des entrées données. En fait, le réseau est laissé libre de converger vers n'importe quel état final lorsqu'on lui présente des entrées.

II.4.6.2 Apprentissage supervisé :

Un apprentissage est dit supervisé lorsque l'on force le réseau à converger vers un état final précis, en même temps qu'on lui présente un motif.

De façon générale, l'apprentissage artificiel supervisé consiste à créer un modèle de prédiction (classification ou prévision) à partir d'une base d'apprentissage comprenant les exemples d'entrée ainsi que les sorties désirés associées. Les paramètres du modèle vont ainsi s'adapter en comparant à chaque fois les sorties obtenues et les sorties désirées, d'où l'appellation supervisé. [25][26]

II.4.6.3 Apprentissage par renforcement :

L'apprentissage renforcé est une technique similaire à l'apprentissage supervisé à la différence qu'au lieu de fournir des résultats désirés au réseau, on lui accorde plutôt un grade (ou score) qui est une mesure du degré de performance du réseau après quelques itérations.

Les algorithmes utilisant la procédure d'apprentissage renforcé sont surtout utilisés dans le domaine des systèmes de contrôle.

L'apprentissage consiste à modifier le poids des connections entre les neurones, alors il existe plusieurs règles de modification :

- Loi de Hebb.
- Règle de Windrow –Hoff(delta-rule).
- Règle de Grossberg.

II.4.6.3.1 Loi de Hebb[21]

La loi de Hebb (1949) s'applique aux connexions entre neurones, comme le représente la figure II.13.

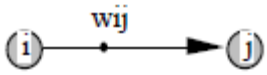


Figure II.13. i le neurone amont, j le neurone aval et wij le poids de la connexion. [21]

Elle s'exprime de la façon suivante

"Si 2 cellules sont activées en même temps alors la force de la connexion augmente".

La modification de poids dépend de la Co activation des neurones présynoptique et post synaptique, ainsi que le montre la table 1. xi et xj sont respectivement les valeurs d'activation des neurones i et j, ∂wij (dérivée partielle du poids) correspond à la modification de poids réalisée.

Table II.1. La loi de Hebb. [21]

Xi	Xj	∂wij
0	0	0
0	1	0

1	0	0
1	1	+

La loi de Hebb peut être modélisée par les équations suivantes (w(t+1) est le nouveau poids,
wij(t) l'ancien) :

$$wij(t+1) = wij(t) + \partial wij(t)$$

$\partial wij(t) = xi.$ xj(la Co activité est modélisée comme le produit des deux valeurs d'activation)

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

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

2/ Présentation d'une entrée El = (e1, ... en) de la base d'apprentissage.

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

$$a = \sum(wi \,.\, ei) - S$$

(la valeur de seuil est introduite ici dans le calcul de la somme pondérée)

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

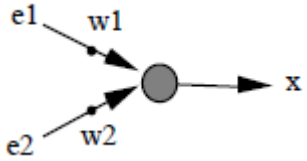
4/ Si la sortie x est différente de la sortie désirée dl pour cet exemple d'entrée El alors modification des poids (μ est une constante positive, qui spécifie le pas de modification des poids) :

$$wij(t+1) = wij(t) + \mu\,(x\,i\,.\,xj)$$

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

Exemple d'application de l'algorithme d'apprentissage de Hebb :

Choisissons pour les neurones un comportement binaire. Les entrées e1 et e2 sont considérées comme des neurones (figure II.14).



FigureII.14. Réseau de 3 neurones (les 2 entrées sont considérées comme deux neurones) pour la résolution du problème exprimé table 2.

e1	e2	x	
1	1	1	(1)
1	-1	1	(2)
-1	1	-1	(3)
-1	-1	-1	(4)

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

Table II.2. Base d'exemples d'apprentissage pour la loi de Hebb.

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

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

3/ $a = w1.e1 + w2.e2 - S = 0.0.1 + 0.0.1 - 0.0 = 0 ; a \geq 0 => x = -1$

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

$w1 = w1 + e1.x = 0.0 + 1.1 = 1$

$w2 = w2 + e2.x = 0.0 + 1.1 = 1$

2/ On passe à l'exemple suivant (2) :

3/ $a = 1.1 + 1.-1 - 0.0 = 0 ; a \leq 0 => x = -1$

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

$w1 = 1 + 1.1 = 2$

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

II.4.7. Algorithme d'apprentissage

A)Algorithme de rétro-propagation[19][20]

La rétro-propagation du gradient BP (Back Propagation) est une généralisation de l'algorithme LMS conçu spécialement pour le MLP .C'est un algorithme supervisé dont l'objectif est d'ajuster les poids du réseau MLP de façon à minimiser une fonction de coût différentiable, telle que l'erreur quadratique entre la sortie du réseau et la sortie désiré:

$$E(n) = \|d(n) - x_L(n)\|^2 \qquad \text{(II.3)}$$

Où $x_L(n)$ est la sortie du réseau au temps n et d(n) la sortie désirée. L'algorithme de BP assure la descente du gradient sur le critère d'erreur pour atteindre le minimum. Le gradient de E(n) est calculé pour tous les poids de la manière suivante :

$$\frac{\partial E(n)}{\partial W_{ijk}(n)} = \frac{\partial E(n)}{\partial y_{ik}} \cdot \frac{\partial_{yik}}{\partial W_{ijk}} = \frac{\partial E(n)}{\partial y_{ik}} \cdot x_{i-1j} \quad \textbf{(II.4)}$$

Dans le cas de la couche de sortie (i=L) on évalue le terme d'erreur de sortie δ_{Lk} comme suit:

$$\delta_{LK} = \frac{\partial E(n)}{\partial y_{lk}} = 2\acute{f}(\partial y_{lk})(d_k - x_{lk}) \quad \textbf{(II.5)}$$

Où $f'(\ )$ est la dérivée de la fonction d'activation:

$$\acute{f}(x) = \frac{df(x)}{dx} \quad \textbf{(II.6)}$$

Pour les couches cachées, le terme d'erreur δ_{ik} du neurone (i,k) est donné par:

$$\delta_{iK} = 2\acute{f}(y_{lk}) \sum_{j=1}^{N(i+1)} \delta_{i+1j} W_{i+1kj} \quad \textbf{(II.7)}$$

La modification des poids et des biais est obtenue selon les équations suivantes :

$$w_{ijk}(n+1) = w_{ijk}(n) + \mu\delta_{ik}x_{i-1j} + \alpha\big(W_{ijk}(n) - W_{ijk}(n-1)\big) \textbf{(II.8)}$$

$$b_{ik}(n+1) = b_{ik} + \mu\delta_{ik} \textbf{(II.9)}$$

μ est le pas d'apprentissage et α le terme d'inertie (moment Um).Le choix du pas d'apprentissage influe beaucoup sur la rapidité de convergence, un pas trop petit ralenti l'apprentissage, un pas trop important provoque un risque d'instabilité. Il est même possible que l'algorithme rencontre un minimum local. Le terme d'inertie (moment Um $0< \alpha <1.0$) permet de sortir du minimum local et de limiter les oscillations lors de l'apprentissage ; en prenant en compte les évolutions des étapes précédentes et donc de converger plus rapidement. La phase d'apprentissage est souvent arrêtée lorsque l'erreur calculée sur l'ensemble de la base d'apprentissage est inférieure à un seuil déterminé par l'utilisateur. Il est possible d'arrêter l'apprentissage en fixant une limite au nombre d'itérations.[27]

B) Algorithme de Levenberg-Marquardt, ou algorithme LM

Permet d'obtenir une solution numérique au problème de minimisation d'une fonction, souvent non linéaire et dépendant de plusieurs variables. L'algorithme interpole l'algorithme de Gauss-Newton et l'algorithme du gradient. Plus stable que celui de Gauss-Newton, il trouve une solution même s'il est démarré très loin d'un minimum. Cependant, pour certaines fonctions très régulières, il peut converger légèrement moins vite. L'algorithme fut découvert par Kenneth Levenberg, puis publié par Donald Marquardt. C'est un problème qui se présente souvent en régression linéaire et non linéaire.[28]

- **Application à la méthode des moindres carrés**

Son application principale est la régression au travers de la méthode des moindres carrés : étant donné un certain nombre de paires de données (ti, yi), on cherche le paramètre a de la fonction f(t|a) de sorte que la somme des carrés des déviations :

$$S(a) = \sum_{i=1}^{m} \left[y_i - f(t_i/\alpha) \right]^2 , \qquad \textbf{(II.10)}$$

soit minimale.La procédure de l'algorithme est itérative. On part d'un paramètre initial, que l'on suppose « assez proche » d'un minimum, et qui constituera le vecteur p de départ. Dans beaucoup de cas, un paramètre de départ « standard », tel que pT=(1,1,…,1) fonctionnera sans problème. Dans certains cas, il n'y a convergence que si le vecteur de départ n'est pas trop éloigné de la solution.

À chaque itération, on remplace p par une nouvelle estimation p + q. Afin de déterminer q, les fonctions fi(p + q) sont approchées en étant linéarisées :

$$f(p + q) \approx f(p) + Jq \qquad \textbf{(II.11)}$$

Où on a noté **J** la jacobéenne de **f** en **p**.

À un minimum de la somme des carrés S, on a $\nabla_q S = 0$. En dérivant le carré de l'expression de droite, qui s'annule donc, on obtient :

$$(J^T J)q = J^T[y - f(p)] \qquad \textbf{(II.12)}$$

D'où l'on tire aisément q on inversant J^TJ.

Dans l'inversion matricielle, tout va dépendre du rapport entre la valeur propre la plus "grande" en norme, et la valeur propre la plus petite ; ce rapport, appelé "conditionnement de matrice", va concrètement refléter la robustesse de l'algorithme face au bruit. Le point essentiel de l'algorithme de Levenberg-Marquardt est d'approcher cette équation, en l'« amortissant »un peu. (On parle alors de "chargement de la diagonale" afin de contourner le mauvais conditionnement le cas échéant, problème que l'on retrouve avec l'algorithme de Capon et que l'on peut résoudre par le procédé "Single Value Décomposition") :

$$(J^TJ + \lambda.\mathrm{diag}(J^TJ))q = J^T[y - f(p)]. \qquad \textbf{(II.13)}$$

Le facteur d'amortissement positif λ est ajusté à chaque nouvelle itération. Si la diminution de S est rapide, on peut utiliser une valeur plus faible, ce qui rapproche l'algorithme de celui de Gauss-Newton. Si en revanche une itération est peu efficace, on peut augmenter λ, ce qui rapproche cette fois l'algorithme de celui du gradient. Un tel facteur d'amortissement est utilisé par exemple dans la régularisation de Tikhonov, utilisée pour résoudre certains problèmes linéaires.

Si on a effectué plus d'un certain nombre d'itérations, ou bien que l'on s'est approché suffisamment d'un minimum, la procédure se termine et renvoie le paramètre **p** comme estimation de la solution. [1]

- **Choix du paramètre d'amortissement**

De nombreux arguments, plus ou moins heuristiques ont été proposés afin de déterminer le meilleur facteur d'amortissement λ. Des démonstrations théoriques montrent que certains choix garantissent une convergence locale - mais peuvent afficher une convergence faible près de l'optimum .La valeur absolue de tout choix dépend de l'échelle du problème. Marquardt recommandait de commencer à partir de λ_0 et avec un facteur $\nu>1$. On pose alors au départ $\lambda=\lambda_0$ et calculons la somme des carrés des déviations S(**p**) après une itération, en utilisant le facteur d'amortissement $\lambda=\lambda_0$, puis en utilisant λ/ν. Si les deux derniers renvoient un point moins bon que le point de départ, on augmente λ en le multipliant par ν, jusqu'à atteindre un point meilleur avec un nouveau facteur $\lambda\nu^k$ pour un certain k. Si l'utilisation du +l'algorithme continue. Si l'utilisation de λ/ν donne une somme plus importante, mais que l'utilisation de λ donne une somme plus faible, alors λ est conservé.[29]

II.4.7 Domaines d'application des réseaux de neurones artificiels

Aujourd'hui, les réseaux de neurones artificiels ont de nombreuses applications dans des secteurs très variés :

- Traitement d'images : reconnaissance de caractères et de signatures, compression d'images, reconnaissance de forme, cryptage, classification, etc.
- Traitement du signal : filtrage, classification, identification de source, traitement de la Parole…etc.
- Contrôle : commande de processus, diagnostic, contrôle qualité, asservissement des robots, systèmes de guidage automatique des automobiles et des avions…etc.
- Défense : guidage des missiles, suivi de cible, reconnaissance du visage, radar, sonar, lidar, compression de données, suppression du bruit…etc.
- Optimisation : planification, allocation de ressource, gestion et finances, etc.
- Simulation : simulation du vol, simulation de boîte noire, prévision météorologique, recopie de modèle…etc.**[30]**

II .4.8 Les avantages et les inconvénients

II .4.8.A Les Avantage

- Aucune connaissance basée sur les règles.
- Le modèle mathématique non requis.
- Plusieurs algorithmes d'apprentissages sont disponibles.

II .4.8.B Inconvénients

- Boite noir (manque de traçabilité).
- L'adaptation aux environnements différents est difficile el le réapprentissage est souvent obligatoire.

II.5 L'analogie entre le neurone biologique et le neurone formel (artificiel)

La modélisation consiste à mettre en œuvre un système de réseau de neurones sous un aspect non pas biologique mais artificiel, cela suppose que d'après le principe biologique on aura une correspondance pour chaque élément composant le neurone biologique, donc une modélisation pour chacun d'entre eux. On pourra résumer cette modélisation par le tableau II.3, qui nous permettra de voir clairement la transition entre le neurone biologique et le neurone formel.

Tableau II.3 : Analogie entre le neurone biologique et le neurone formel.

Neurone biologique	**Neurone formel**
Synapses	Poids des connexions
Axones	Signal de sortie
Dendrites	Signal d'entrée
Noyau ou Somma	Fonction d'activation

II.6Conclusion

Dans ce chapitre, nous avons donné un aperçu général sur les réseaux de neurones artificiels, leurs principes de fonctionnement, l'architecture des réseaux de neurones ainsi que les composants de ces dernierssont présentées. En plus nous avons présenté une vue d'ensemble des fonctions d'activations, les techniques d'apprentissage et leurs algorithmes, les domaines d'application des RNA et l'analogie entre le neurone biologique et le neurone formel.

Chapitre III

SIMULATIONS ET RESULTATS OBTENUS

III.1Introduction

L'objectif de cette étude est un apprentissage automatique pour identifier les personnes en fonction des caractéristiques extraites de la parole enregistrée. Les caractéristiques utilisées pour entraîner le classificateur sont: la hauteur des segments vocaux de la parole et les coefficients de Mel FrequencyCepstrum Coefficients(MFCC). Il s'agit d'une identification d'enceinte fermée - l'audio de l'enceinte testée est comparée à tous les modèles d'enceintes disponibles (un ensemble fini) et la correspondance la plus proche est renvoyée.

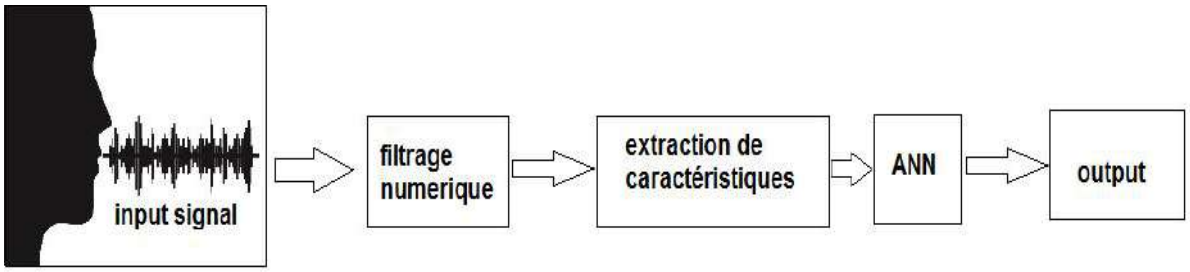


Figure III.1 : Modèle de reconnaissance vocale utilisant un réseau neural.

III.2. Traitementdu signal vocal

Nous avons utilisé le MATLAB pour programmer cette étude, nous avons utilisé un sous-ensemble de la base de données globale (AN4) pour entraîner le réseau neuronal et tester notre système. Ce sous-groupe se compose de 10 locuteurs, chaque locuteur prononce des nombres de un à 10. Les fichiers audio sont enregistrés via un microphone haut sensibilité, échantillonnés à 16 kHz et codés à 16 bits.

III.3.Opérations de préparation des fichiers de parole

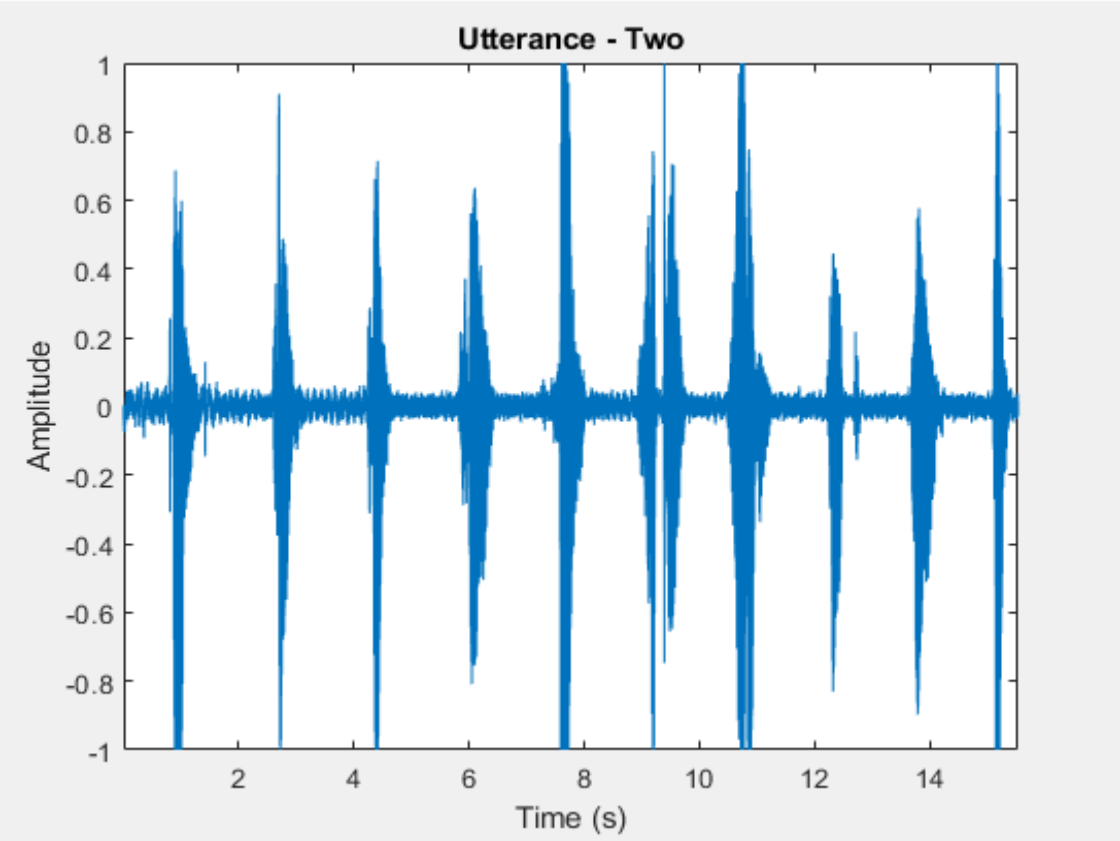
Avant d'utiliser les fichiers de parole de la base de données AN4-Raw nous avonssupprimé le bruit. Peut être utilisé pour enregistrer, écouter, traiter, analyser et convertir un fichier audio.

III.3.1. Opération de supprimer le bruit

Les signaux vocaux capturés du locuteur contiennent des signaux de bruit de fond qui proviennent de différentes sources telles que le bruit ambiant, la borne de microphone et le canal de communication. Pour supprimer les signaux de faible bruit de fond, le signal de parole discrétisé de l'énoncé parlé est passé à travers un filtre numérique passe-bas ButterWorth IIR (réponse impulsionnelle infinie) du second ordre pour le filtrage (g [n] * h [n]). Le choix de ce filtre IIR repose sur sa flexibilité pour gérer facilement des signaux de

nature fortement non linéaire en raison de ses caractéristiques de phase non linéaires et de sa flexibilité des exigences pour répondre à des contraintes telles qu'une réponse arbitraire. Etant donné que la majeure partie du contenu énergétique du signal vocal est concentrée dans la plage des basses fréquences, la fréquence de coupure du filtre est fixée à 0,40 kHz.

Pour supprimer le silence ou les composants non vocaux dans l'énoncé parlé, le signal de parole continu traité est divisé en trames se chevauchant de manière à permettre à la distribution de signal dans chaque trame d'être analysée indépendamment et représentée de manière appropriée par un vecteur de caractéristiques.



Après l'exécution on obtient le signal suivant dans la figureIII.3.

FigureIII.2 :signal d'entrée avant le filtrage

III.3.2.La méthode de Pitch

La parole peut être largement classée comme audible et inaudible. Dans le cas de la parole vocale, l'air des poumons est modifié par les cordes vocales et produit une excitation quasi-périodique. Le son qui en résulte est dominé par une oscillation à fréquence relativement basse, appelée hauteur. Dans un discours silencieux, l'air des poumons passe par une contraction des voies vocales et devient perturbé, comme une excitation bruyante. Dans le mode de réalisation du filtre de parole source, l'excitation est

appelée la source et le tractus vocal est appelé le filtre. La caractérisation de la source est une partie importante de la caractérisation du système vocal.

En tant qu'exemple de parole vocale et non vocale, considérons une représentation dans le domaine temporel du mot "two" (/ T UW /). La consonne / T / (parole non vocale) ressemble à du bruit, tandis que la voyelle / UW / (parole vocale) est caractérisée par une fréquence fondamentale forte.

La méthode la plus simple pour distinguer la parole vocale et non vocale consiste à analyser le taux de passage à zéro. Un grand nombre de passages par zéro implique qu'il n'y a pas d'oscillation basse fréquence dominante.

Une fois que vous avez isolé une région de parole exprimée, vous pouvez la caractériser en estimant la hauteur. Cet exemple utilise la hauteur pour estimer la hauteur. Il utilise l'approche d'autocorrélation normalisée par défaut pour calculer la hauteur.

Appliquez la détection de hauteur au mot «two» pour voir comment la hauteur change au fil du temps. Ceci est connu sous le nom de contour de hauteur et est caractéristique d'un locuteur.

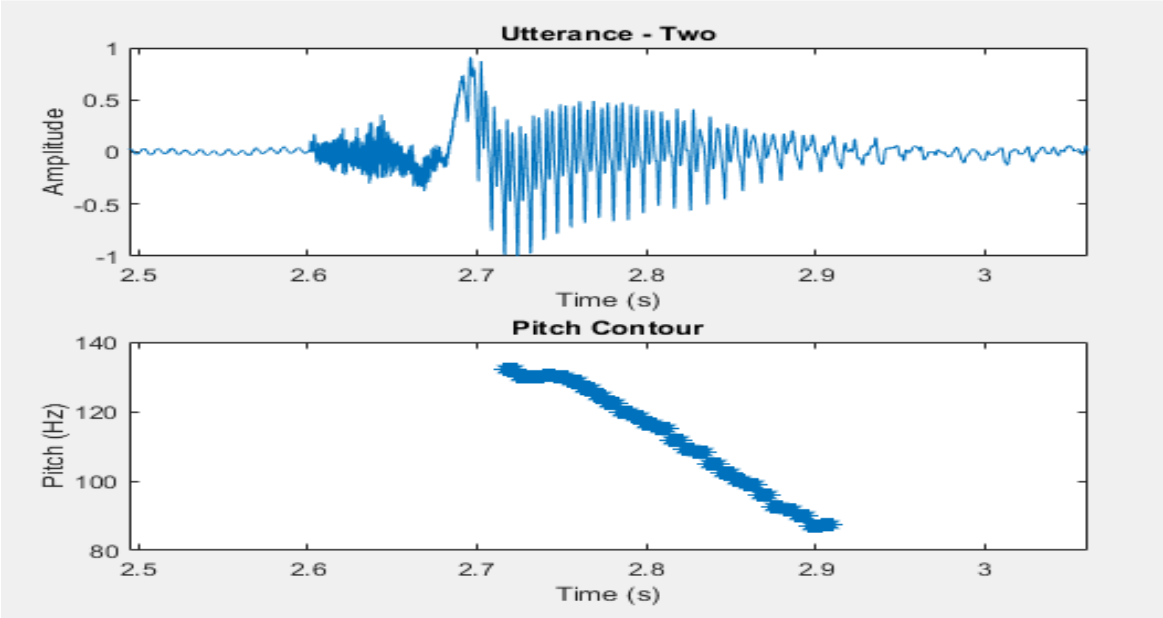


Figure III.3 : Le signal après le filtrage du mot "two"

III.4.Traitements des fichiers de parole

III.4.1.Base de données

Cette étude utilise la base de données de dénombrement robuste du groupe de reconnaissance vocale CMU (également appelée base de données AN4). L'ensemble de données contient des enregistrements de sujets masculins et féminins parlant des mots et des chiffres. La base de

données est disponible sur internet. Les fichiers audio sont divisés en sous-répertoires selon les désignations correspondantes des locuteurs.

Tableau III.1 : Le nombre de parleurs dans la banque de données du train.

Label	Count
fejs	10
fmjd	10
fsrb	10
ftmj	10
fwxs	10
mcen	10
mscb	10
msjm	10
Msjr	10
msmn	7

Affichez la banque de données et le nombre de locuteurdans la banque de données de test.

Tableau III.2 : Le nombre de parleurs dans la banque de données du test.

Label	Count
fejs	3
fmjd	3
fsrb	3
ftmj	3
fwxs	2
mcen	3
mscb	3
msjm	3
Msjr	3
msmn	2

III.4.2. Extraction de caractéristique :

Les fonctionnalités Pitch et MFCC sont extraites de chaque signal à l'aide defonction HelperComputePitchAndMFCC qui effectue les actions suivantes sur les données lues à partir de chaque fichier audio :

1-Collectez les échantillons dans des trames de 30 ms avec un chevauchement de 75%.

2-Pour chaque trame, utilisez audiopluginexample.SpeechPitchDetector.isVoicedSpeech

Pour décider si les échantillons correspondent à un segment de parole vocale.

3-Calculez la hauteur et 13 MFCC (avec le premier coefficient MFCC remplacé par le log-énergie du signal audio) pour l'ensemble du fichier.

4-Conservez les informations de hauteur et MFCC relatives aux trames voisées uniquement.

5-Obtenez le nom du répertoire du fichier. Cela correspond au nom du locuteur et sera utilisé comme étiquette pour l'apprentissage du classificateur.

Les MFCC et l'étiquette (nom du locuteur) sous forme de colonnes pour chaque trame de 30 ms.

Nom de fichier	pitch	MFCC 1	MFCC 2	MFCC 3	MFCC 4	MFCC 5	MFCC 6	MFCC 7	MFCC 8	MFCC 9	MFCC 10	MFCC 11	MFCC 12	MFCC 13	étiquette
'an36-fejs-b.flac'	237.44	-4.4218	3.3816	0.73331	0.98626	0.47093	0.13808	-0.083348	0.069072	0.2345	0.3403	-0.14417	-0.15605	-0.022106	'fejs'

Tableau III.3 : résultats des nombres de fichier d'audio dans chaque étiquette.

‘an36-fejs-b.flac’	242.42	-4.3104	4.7899	0.80432	0.7148	0.46027	0.32963	-0.28647	0.38366	0.01449	0.0093271	-0.2559	-0.17832	-0.11693	‘fejs’
‘an36-fejs-b.flac’	231.88	-3.6432	5.0192	0.74801	0.58299	0.50475	-0.014551	-0.32653	0.39201	0.20982	-0.20282	-0.25637	-0.20576	-0.27675	‘fejs’
‘an36-fejs-b.flac’	230.89	-3.0024	5.2132	0.46784	0.57104	0.64546	-0.085145	-0.23453	0.55408	0.44134	0.17966	0.47125	0.22141	0.22027	‘fejs’
‘an36-fejs-b.flac’	112.45	-2.9718	5.3249	0.48934	0.66976	0.56446	-0.14691	-0.26824	-0.4536	0.31515	0.21356	-0.34067	0.21872	-0.14108	‘fejs’
‘an36-fejs-b.flac’	111.89	-2.6202	5.2746	0.53966	0.55468	0.50989	0.012264	-0.26755	0.3318	0.32108	-0.18096	-0.44212	-0.21208	-0.21385	‘fejs’
‘an36-fejs-b.flac’	111.11	-2.6138	5.0492	0.68513	0.40281	0.36792	0.13352	-0.07321	0.25863	0.25314	-0.1787	-0.51149	-0.14679	-0.077431	‘fejs’
‘an36-fejs-b.flac’	110.1	-2.4483	5.5192	0.64449	0.44857	0.25178	0.25716	0.042426	0.32466	0.17774	-0.194	-0.70127	-0.16868	-0.041083	‘fejs’

Nom de fichier	pitch	MFCC 1	MFCC 2	MFCC 3	MFCC 4	MFCC 5	MFCC 6	MFCC 7	MFCC 8	MFCC 9	MFCC 10	MFCC 11	MFCC 12	MFCC 13	étiquette
‘an36-fejs-b.flac’,	0.905 35	-1.8778	0.11469	0.25866	0.41449	0.97803	0.34062	0.22379	0.031962	0.62995	0.81708	0.29036	-0.47	0.04532	fejs

On Note que le pitch et le MFCC ne sont pas sur la même échelle. Cela biaisera le classificateur. Normaliser les entités en soustrayant la moyenne et en divisant l'écart type de chaque colonne.

Tableau III.4 : résultats des nombres de fichier d’audio dans chaque étiquette après la normalisation.

‘an36-fejs-b.flac’	0.98367	-1.8143	1.2196	0.32634	-0.65383	0.9623	-0.52854	-0.61983	0.70533	0.40654	-0.066892	-0.60354	-0.53907	-0.51924	‘fejs’
‘an36-fejs-b.flac’	0.81806	-1.4342	1.3996	0.27266	-0.77005	1.0279	-0.61349	-0.69793	0.72491	0.56842	-0.6335	-0.60484	-0.62735	-1.0637	‘fejs’
‘an36-fejs-b.flac’	0.8025	-1.1209	1.4881	0.0057061	-0.78058	1.2356	-0.7397	-0.49906	1.1048	0.39759	-0.57164	-0.36655	-0.67672	-0.87127	‘fejs’
‘an36-fejs-b.flac’	-1.0579	-1.0516	1.6394	0.026102	-0.69355	1.116	-0.85012	-0.58429	0.86925	0.83105	-0.66218	-0.84113	-0.66903	-0.60151	‘fejs’
‘an36-fejs-b.flac’	-1.0674	-0.85121	1.5999	0.074074	-0.79501	1.0355	-0.56555	-0.58294	0.5838	0.84584	-0.5711	-1.1255	-0.64767	-0.8494	‘fejs’
‘an36-fejs-b.flac’	-1.0797	-0.8476	1.4231	0.21274	-0.92891	0.82602	-0.34877	-0.20402	0.41321	0.67644	-0.56908	-1.3199	-0.43764	-0.38468	‘fejs’
‘an36-fejs-b.flac’	-1.0955	-0.75325	1.7918	0.174	-0.88856	0.65463	-0.12772	0.021446	0.56706	0.48842	-0.60995	-1.8518	-0.50805	-0.26085	‘fejs’

III.4.3.Former un classificateur

Maintenant que nous avons rassemblé les fonctionnalités des dix locuteurs, nous pouvons former un classificateur basé sur celles-ci. Dans cetteétude, nous utilisons un classificateur K-plus proche voisin. Le K-plus proche voisin est une technique de classification naturellement adaptée à la classification multi-classes. Les hyper paramètres du classificateur du voisin le plus proche incluent le nombre de voisins, la métrique de distance utilisée pour calculer la distance aux voisins et le poids de la métrique de distance. Les hyper paramètres sont sélectionnés pour optimiser la précision de validation et les performances sur l'ensemble de test. Dans cetteétude, le nombre de voisins est défini sur 5 et la métrique de distance choisie est la distance euclidienne pondérée inverse au carré.

On Forme le classificateur et imprime la précision de la validation croisée. crossval et kfoldloss sont utilisés pour calculer la précision de validation croisée pour le classificateur KNN.

Précision de la validation = 93.00%

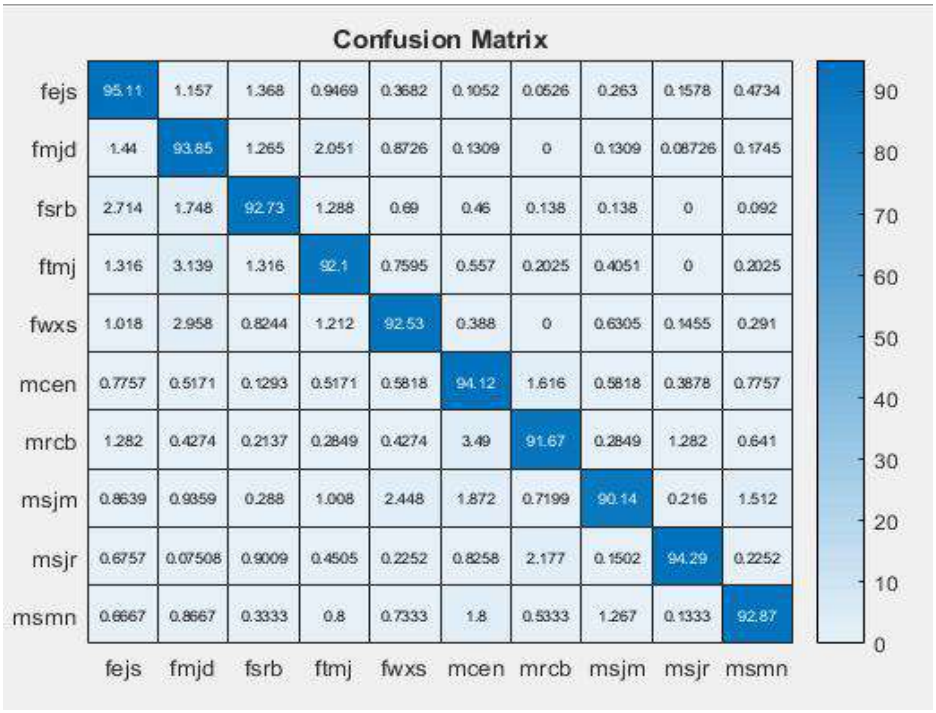


Figure III.6 :matrice de confusion après l'entrainement.

III.4.4.Tester le classificateur

Dans cette section, nous testerons le classificateur KNN formé avec deux signaux vocaux de chacun des dix locuteurs pour voir à quel point il se comporte avec des signaux qui n'ont pas été utilisés pour l'entraîner.

Lise les fichiers, extraie les fonctionnalités de l'ensemble de test et normalise-les.

Tableau III.5 : résultats des nombres de fichier d'audio dans chaque étiquette après le teste.

Nom de fichier	pitch	MFCC 1	MFCC 2	MFCC 3	MFC C4	MFCC 5	MFCC 6	MFCC 7	MFCC 8	MFCC 9	MFCC 10	MFCC 11	MFCC12	MFCC 13	étiquette
'an36-fejs-b.flac'	0.80371	-2.3423	-0.5959	0.03779	-0.544	0.92912	-0.01148	0.19081	0.52413	0.53979	1.3608	1.4529	0.072385	-0.36157	'fejs'
'an36-fejs-b.flac'	0.87164	-1.8071	-0.0160	0.46226	-0.406	0.58718	-0.16943	0.59531	0.74482	1.1505	1.4122	1.0592	0.68586	-0.2466	'fejs'
'an36-fejs-b.flac'	0.71543	-1.5509	0.4363	0.59678	-0.303	0.16468	-0.36499	0.60088	0.86908	1.0306	1.4249	1.2195	1.2085	-0.3206	'fejs'
'an36-fejs-b.flac'	0.60687	-1.3838	0.37673	0.6942	-0.240	0.29205	-0.27508	0.28735	0.49197	0.99254	1.6082	1.1646	0.89275	-0.34946	'fejs'
'an36-fejs-b.flac'	0.51015	-1.2784	0.15729	0.71635	-0.288	0.85877	0.4955	0.35029	0.21002	0.43132	1.444	1.3727	0.75646	-0.3879	'fejs'
'an36-fejs-b.flac'	0.48268	-1.2499	-0.2702	0.351	-0.390	1.0208	0.82701	0.63592	0.84897	0.69597	0.86749	0.65908	0.58337	0.05005	'fejs'
'an36-fejs-b.flac'	0.48248	-1.1657	-0.6657	0.06703	-0.986	0.64687	0.70306	0.32048	0.42064	0.57206	0.66513	0.31176	0.61106	0.19282	'fejs'
'an36-fejs-b.flac'	0.6509	-0.66006	-0.4224	0.23545	-0.910	0.64842	0.61746	0.12066	0.0988	0.33263	0.49967	0.18377	0.54699	0.07464	'fejs'

La fonction HelperTestKNNClassifier exécute les actions suivantes pour chaque fichier dans testDatastore.

1. Lise des échantillons audio et calcule la hauteur et les fonctionnalités MFCC pour chaque trame de 30 ms, comme décrit dans la section extraction de fonctionnalités.
2. Prédise l'étiquette (locuteur) pour chaque trame en appelant la fonction prédire sur le classificateur entraîné.
3. Pour un fichier donné, des prédictions sont faites pour chaque trame de signal. L'étiquette la plus fréquente est déclarée comme locuteur prévu pour le fichier. La confiance de prédiction est calculée comme la fréquence de prédiction de l'étiquette divisée par le nombre total de trames voisées dans le fichier.

Tableau III.6 : résultats final d'apprentissage**.**

Nom de fichier	conférencier actuel	orateur prédit	pourcentage
'cen6-fejs-b.flac'	'fejs'	"fejs"	94.41
'cen6-fmjd-b.flac'	'fmjd'	"fmjd"	75.776
'cen6-fsrb-b.flac'	'fsrb'	"fsrb"	56.219
'cen6-ftmj-b.flac'	'ftmj'	"ftmj"	58.974
'cen6-fwxs-b.flac'	'fwxs'	"fwxs"	71.493
'cen6-mcen-b.flac'	'mcen'	"mcen"	74.359
'cen6-mrcb-b.flac'	'mrcb'	"mrcb"	79.845
'cen6-msjm-b.flac'	'msjm'	"msjm"	60.714
'cen6-msjr-b.flac'	'msjr'	"msjr"	62.759
'cen7-fejs-b.flac'	'fejs'	"fejs"	76.642
'cen7-fmjd-b.flac'	'fmjd'	"fmjd"	74.654
'cen7-fsrb-b.flac'	'fsrb'	"fsrb"	71.279
'cen7-ftmj-b.flac'	'ftmj'	"ftmj"	37.915
'cen7-fwxs-b.flac'	'fwxs'	"fwxs"	71.131
'cen7-mcen-b.flac'	'mcen'	"mcen"	76.25
'cen7-mrcb-b.flac'	'mrcb'	"mrcb"	78.788
'cen7-msjm-b.flac'	'msjm'	"msjm"	59.184
'cen7-msjr-b.flac'	'msjr'	"msjr"	78.295
'cen7-msmn-b.flac'	'msmn'	"msmn"	72.901
'cen8-fejs-b.flac'	'fejs'	"fejs"	85.185
'cen8-fmjd-b.flac'	'fmjd'	"fmjd"	57.317
'cen8-fsrb-b.flac'	'fsrb'	"fsrb"	50
'cen8-ftmj-b.flac'	'ftmj'	"ftmj"	63.547
'cen8-mcen-b.flac'	'mcen'	"mcen"	76.812
'cen8-mrcb-b.flac'	'mrcb'	"mrcb"	72.561
'cen8-msjm-b.flac'	'msjm'	"msjm"	43.165
'cen8-msjr-b.flac'	'msjr'	"msjr"	66.667

'cen8-msmn-b.flac'	'msmn'	"msmn"	87.991

Les locuteurs prévus correspondent aux locuteursattendus pour tous les fichiers testés.

III.5.Réseaux de neurone artificiel

Dans ce travail, nous avons plusieurs variables que nous allons étudier :

-Le nombre de locuteurs utilisés à partir de la base de données.

-Nombre d'échantillons de locuteur pour chaque étape d'apprentissage.

-Nombre de nœuds dans la couche cachée.

-Le nombre des entrées de réseau de neurone.

-Les méthodes d'apprentissage.

- L'algorithme de Gradient conjugué mis à l'échelle (trainscg).
- L'algorithme de Régularisation bayésienne (trainbr).
- L'algorithme de Levenberg-Marquardt (trainlm)

Au départ, nous avons 74 locuteurs dans le flac data, nous changerons le nombre de locuteurs et étudierons l'erreur à chaque fois en gardant les autres variables (l'entrainement 70%, l'essai 15%, la validation 15%, les couches cachées 10).

Pour 10 locuteurs, nous obtenons l'erreurde 39.93%.

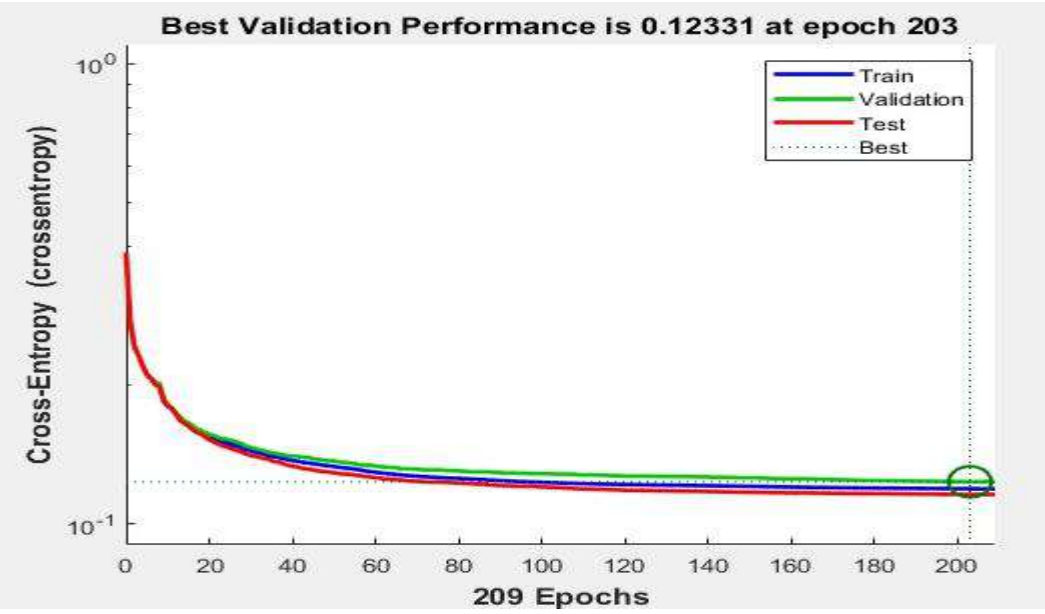


Figure III.7 :la performance de 10 locuteurs

Puis l'erreur de 55.26% pour 20 locuteurs.

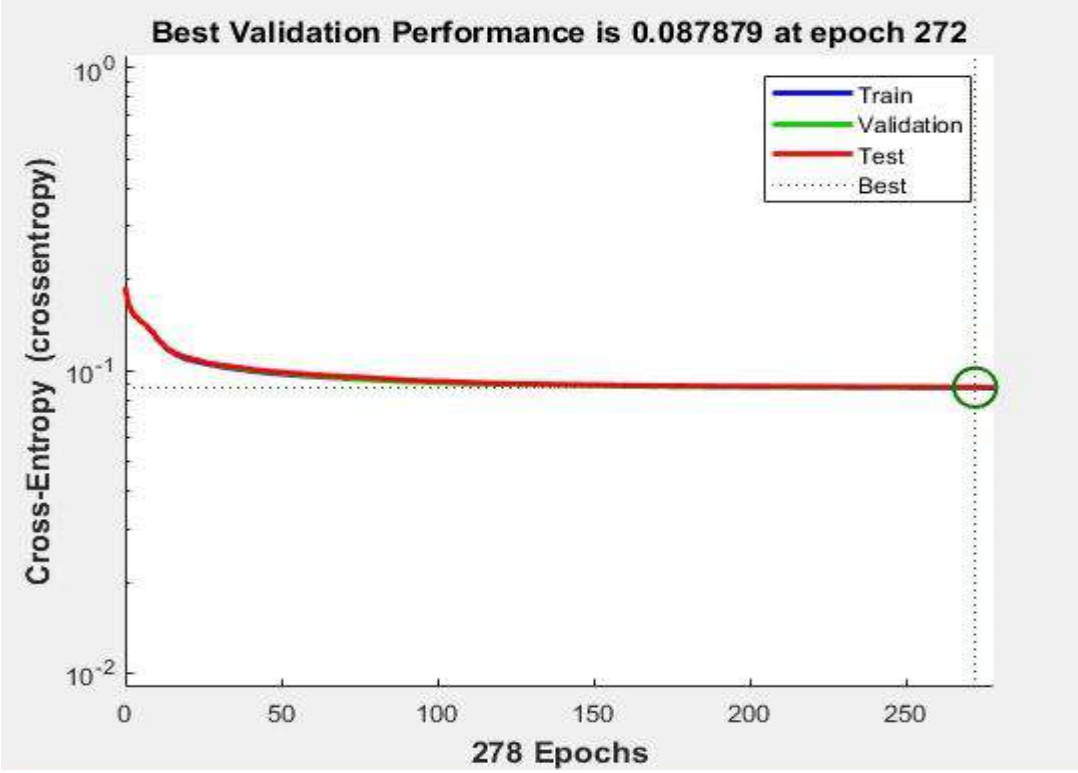
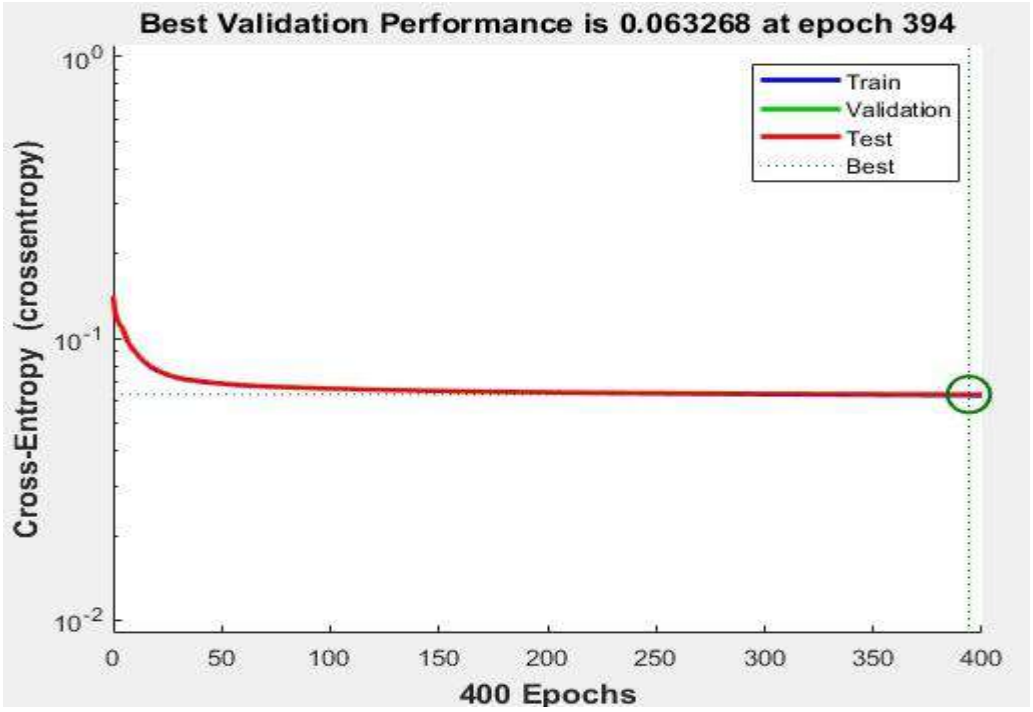


Figure III.8 : la performance de 20 locuteurs



Et pour 30 locuteurs nous avons obtenu l'erreur de 55.62 %.

Figure III.9 :la performance de 30 locuteurs

Enfin pour 50 locuteurs nous avons obtenu llesresultats suivant :

ANN : Pourcentage de confiance moyen des voix testées : 55.5%.

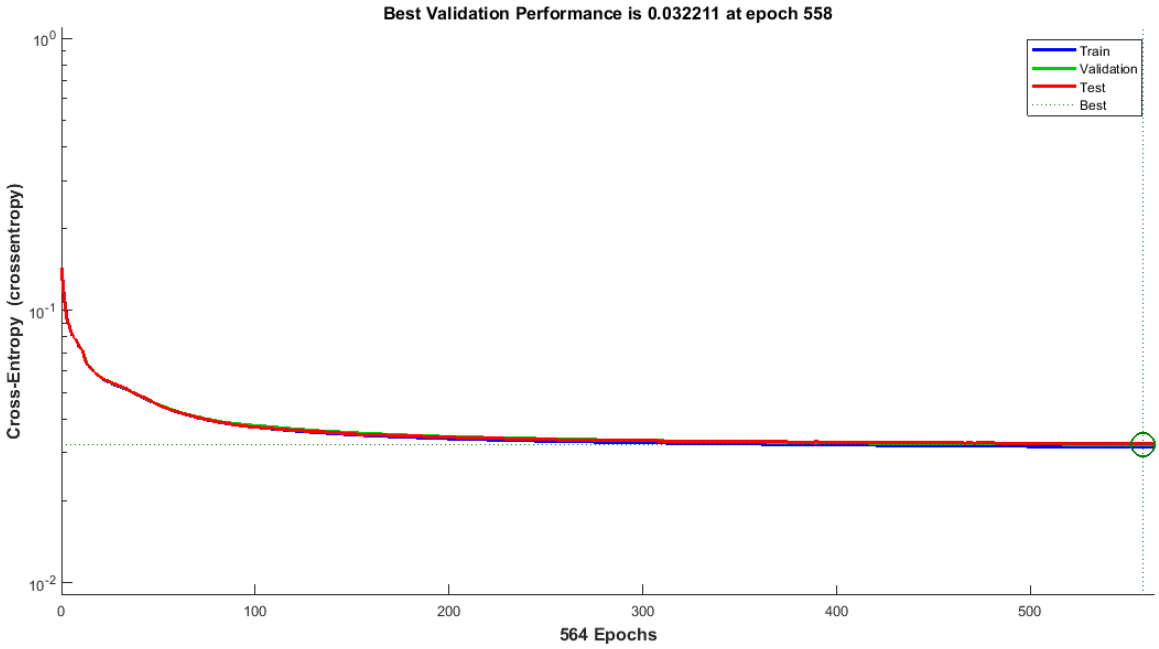


Figure III.10 : la performance de 50 locuteurs

KNN : Pourcentage de confiance moyen des voix testées : 37,73%.

1- k-NN ne nécessite aucun temps d'apprentissage, alors que l'apprentissage des réseaux de neurones est plutôt chronophage. Cependant, k-NN prendra probablement beaucoup plus de temps au moment de l'évaluation, en particulier si vous avez de nombreux points de données et ne recourez pas à la recherche approximative.
2- k-NN est très simple et ne nécessite de régler qu'un seul hyperparamètre (la valeur de k), tandis que l'entraînement au réseau neuronal implique de nombreux hyperparamètres contrôlant la taille et la structure du réseau et la procédure d'optimisation.
3- Les réseaux de neurones ont atteint l'état de l'art dans plus de domaines que k-NN.

4- Il y a plus de garanties théoriques pour k-NN que pour les réseaux de neurones, même si comme nous le savons il y a un grand écart entre les performances théoriques et empiriques des réseaux de neurones.
5- Une fois qu'un réseau de neurones est formé, les données d'entraînement ne sont plus nécessaires pour produire de nouvelles prédictions. Ce n'est évidemment pas le cas avec k-NN.
6- Une fois qu'un réseau neuronal est formé sur une tâche, ses paramètres peuvent être utilisés comme un bon initialiseur pour une autre tâche (similaire). Il s'agit d'une forme d'apprentissage par transfert qui ne peut pas être réalisée avec k-NN.

Après l'étude nous avons remarqué le KNN est simple que le ANN, juste pour un grand nombre de locuteur le ANN est mieux que le KNN, il donne une bonne performance que le Knn.

Donc on conclut que plus il y a de parleurs, plus l'erreur est grande, car le nombre de parleurs augmente la complexité du réseau neuronal.

Après la comparaison entre le ANN et le KNN, nous avons conclu que le KNN est simple par rapport le ANN si nous avons un petit nombre de locuteurs sinon le ANN est mieux quand nous avons un grand nombre de locuteurs (réseau de neurone complexe).

La deuxièmeétape nous avons étudié l'erreur pour 30 locuteurs et nous avons changé l'entrainement ; l'essai et la validation tout en préservant le nombredes nœuds dans la couche masquée.

L'erreur	L'entrainement	l'essai	la validation
55.62%	70%	15%	15%
54.77%	50%	25%	25%
56.88%	90%	5%	5%

On conclut que l'erreur est légèrement augmentée parce que le réseau de neurone a été formé en utilisant un faible pourcentage d'enregistrements, mais l'erreur attendue est comparée car l'ensemble de données se chevauche en utilisant les originaux.

Dans l'étape suivante nous avons étudié l'erreur de 30 locuteurs après de changer le nombre des couches cachéessur les valeurs standards de la commande nprtool (l'entrainement 70% ; l'essai 15% ; la validation 15%).

Nombre de nœuds de couches cachées	L'erreur
2	79%
10	55%
60	35%
90	33%

Nous avons conclu que le pourcentage d'erreur diminueà mesure que le nombre de nœuds augmente.

Dans la quatrième étape nous avons changé le nombre des entrées dans l'extraction de caractéristiques pour 30 locuteurs.

Le nombre d'entrées	L'erreur
14	55%
9	61%
4	76%

Dans cette étape nous avons conclu que le nombre de couches a un grand rôle dans l'obtention de bonnes performances, plus il y a de couchescachées, meilleures sont les performances

Dans la dernière étape où nous parlons sur les méthodes d'apprentissage pour étudier la performance et les comparons, nous avons 3 méthodes

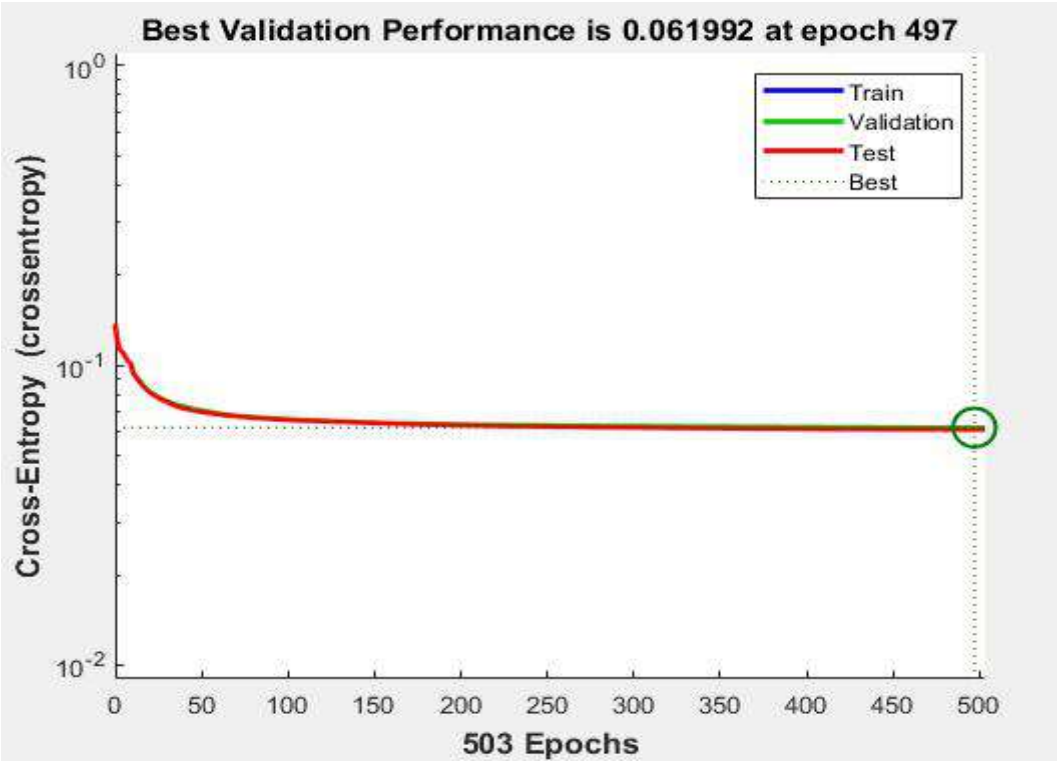
- **L'algorithmede Gradient conjugué mis à l'échelle (trainscg)**

Chacun des algorithmes de gradient conjugué dont nous avons discuté jusqu'à présent nécessite une recherche par ligne à chaque itération. Cette recherche par ligne est coûteuse en calcul, car elle nécessite que la réponse du réseau à toutes les entrées d'apprentissage soit calculée plusieurs fois pour chaque recherche. L'algorithme de gradient conjugué mis à l'échelle (SCG), a été conçu pour éviter la recherche de ligne fastidieuse. Cet algorithme est trop complexe pour être expliqué en quelques lignes, mais l'idée de base est de combiner l'approche modèle-région de confiance (utilisée dans l'algorithme de Levenberg-Marquardt décrit plus tard), avec l'approche du gradient conjugué.

La routine trainscg peut nécessiter plus d'itérations pour converger que les autres algorithmes de gradient conjugué, mais le nombre de calculs à chaque itération est considérablement réduit car aucune recherche de ligne n'est effectuée.[32]

Les résultats de cette méthode sont :

- Le nombre d'itérations = 503itérations.
- Le temps =4mn 12s.



- La performance =0.0619.

Figure III.11 :performance de 30 locuteurs par la méthode trainscg

- **L'algorithme de Régularisation bayésienne (trainbr)**

La procédure de régularisation dans ANN est BR, qui est la combinaison linéaire des méthodes bayésiennes et ANN pour déterminer automatiquement le paramètres de régularisation optimaux.

Dans les modèles ANN régularisés bayésiens (BRANN), les techniques de régularisation impliquent d'imposer certaines distributions a priori aux paramètres du modèle.

Un terme supplémentaire,E_w est ajouté par BRANN à la fonction objective d'arrêt précoce donnée dans l'équation qui pénalise les poids importants en prévision d'une meilleure généralisation et d'une cartographie plus fluide. Comme ce qui se passe dans les pratiques classiques de rétro propagation, une technique d'optimisation basée sur le gradient est ensuite appliquée pour minimiser la fonction donnée.

$$F = \beta E_D(D|w, M) + \alpha E_w(w|M),$$

Où E_w (w | M), est la somme des carrés des poids d'architecture, M est l'architecture ANN (modèle dans le jargon statistique), α et β sont des paramètres de fonction objectifs (également appelés paramètres de régularisation ou hyper-paramètres et prennent les valeurs positives) qui doivent être estimées de manière adaptative. Le deuxième terme du côté droit de l'équation αE_w, est connu sous le nom de décroissance du poids et α, le coefficient de décroissance du poids, favorise les petites valeurs de w et diminue la tendance d'un modèle à sur ajuster.

Pour ajouter une fonction de pénalité quadratique E_w (w | M), α donnant une version de régression de crête non linéaire avec une estimation de w équivalente au maximum bayésien a prior (MAP).

Par conséquent, la forme quadratique (décroissance du poids) favorise les petites valeurs de w et diminue la prédisposition d'un modèle à suradapter les données. Ici, dans l'équation, l'apprentissage implique un compromis entre la complexité du modèle et la qualité de l'ajustement. Si $\alpha >> \beta$, la mise en évidence est sur la réduction de l'étenduedes poids au détriment de la qualité de l'ajustement, tout en produisant une réponse du réseau plus fluide. Si les estimations de Bayes de α sont grandes, les densités postérieures des poids sont fortement concentrées autour de zéro, de sorte que les poids disparaissent effectivement et que le modèle écarte les connexions dans le réseau. Par conséquent, α et β sont prédits de manière adaptative pour gérer la complexité du modèle de compromis et la qualité de l'ajustement.[33]

Les résultats de cette méthode sont :

- Le nombre d'itérations =495 itérations.
- Le temps =20h 13mn 35s.
- La performance =0.0216.

- **L'algorithme de Levenberg-Marquardt**

L'algorithme de Levenberg-Marquardt (LM) est une amélioration de la méthode classique de Gauss-Newton dans la résolution des problèmes de régression non-linéaire des moindres carrés. La méthode est présentée en détail dans Moré (1977). Il s'agit de la méthode recommandée pour les problèmes (de régression) non-linéaires des moindres carrés, car il est plus efficace par rapport aux algorithmes d'optimisation plus généraux (telles que les méthodes Quasi-Newton ou Simplexe .

Considérons l'ajustement de modèle non-linéaire $y = f(\theta,x)$ avec les données X_i et Y_i, i = 1,..., m où X_i est de dimension k et θ est de dimension n. La méthode LM recherche θ la solution de θ (localement) minimisé :

$$g(\theta)=\Sigma_{i=} m_1(Y_i - f(\theta, Xi))^2$$

L'algorithme LM trouve la solution en appliquant la routine :

$$\theta_{J+1}= \theta_j - (J'J + lD)^{-1}J'(Y - f(\theta, Xi))$$

De manière itérative, où :

Y est le vecteur m x 1 contenant $Y_1,..., Y_m$,

X est la matrice m x k contenant $X_1,..., X_m$,

J est la matrice Jacobienne pour $f(\theta, x)$ en respectant θ,

D est une matrice diagonale n x n pour ajuster les facteurs d'échelle. [34]

Nous essayons également l'algorithme de Levenberg-Marquardt (trianlm)

Nous avons obtenu le résultat :

- Le nombre d'itérations =128 itérations.
- Le temps = 2h 56mn 17s.
- La performance = 0.0216.

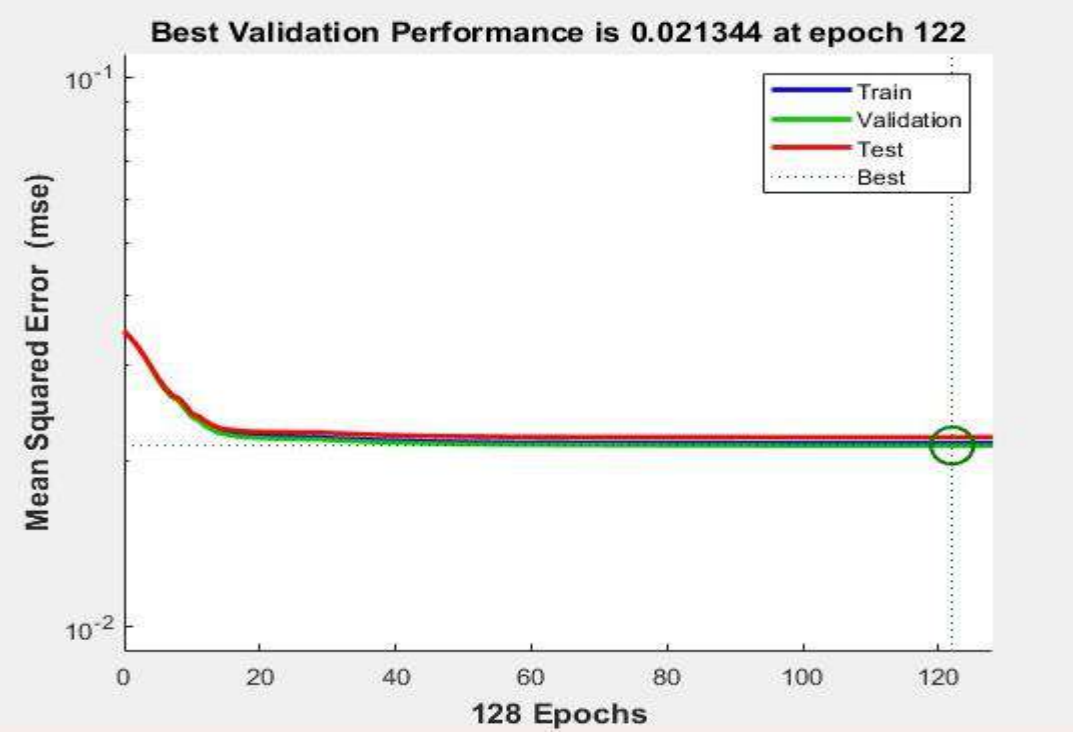


Figure III.12 : performance de 30 locuteurs par la méthode trainlm.

Alors on conclure que l'algorithme de Levenberg-Marquardt est le meilleur algorithme pour obtenir un minimum d'erreur et bonne performance.

III.6.Conclusion

Dans ce chapitre nous avons présenté en détails les méthodes proposées pour le système d'identification automatique de locuteur nous avons commencé par une présentation de la base de données ANN puis les différentes opérations de la phase de prétraitement du signal. Ensuite, nous avons expliqué en détails le principe d'extraction de caractéristique et la création du modèle de référence. En fin sur exposer les séries d'expériences et les résultats obtenus.

Conclusion général

Conclusion général

Dans le cadre de ce mémoire, nous avons abordé une généralité sur l'identification de locuteurs. Nous avons pris un aperçu de production de la parole et le traitement numérique de parole et puis on a parlé sur la reconnaissance automatique du locuteur ainsi que nous avons étudié en détails les réseaux de neurones artificiels suivi par la méthode proposée pour notre travaille. Enfin, on a présenté les simulations et les résultats obtenus.

Au départ, nous avons prisau totale 74 locuteurs dans le flac data. Nous avons changé le nombre du locuteurà chaque fois et on a conservéles autres variables d'entrainement, d'essai et de validation pour étudier la performance et comparer les résultats de classificateur ANN et le classificateur KNN.Nous avons obtenus que le KNN est simple par rapport le ANN dans le cas ou nous avons un petit nombre de locuteurs sinon le ANN est mieux que dans le cas où le nombre de locuteurs est grand(réseau de neurone complexe).

Dans la deuxième étape nous avons étudié la performance des 30 locuteurs avec le changement des autres variables (l'entrainement, l'essai et la validation). Après l'exécution des simulations nous avons conclu que l'erreur augmente au fur et à mesure que le pourcentage d'entrainement augmente. Cela est dû en nombre insuffisant de pourcentage d'essais.

Ainsi dans la troisième étape,nous avons changé le nombre du nœud de couche cachées et nous avons étudié l'erreur à chaque changement. Après exécution et comparaisondes résultats, nous avons trouvé que l'erreur diminuerau fur et à mesure que le nombre des nœuds de couches cachées augmente.

Et dans la quatrième étape,nous avons changé le nombre des entrées dans l'extraction de caractéristiques pour 30 locuteurs. Nous avons montré quele nombre des entrées dans la couche d'entréeà un grand rôle dans l'obtention de bonnes performances.

Enfin,nous avons discuté et comparé les méthodes d'apprentissage pour étudier la performance,le nombre d'itérations et le temps d'apprentissage. Après comparaison, nous avons déterminé que la meilleure méthode est dans le cas d'un compromis entre le temps d'apprentissage et l'erreur d'identification.

Conclusion général

Bibliographie

Bibliographie

[1] :LAIB Abderrzak, Systèmed'authentification de locuteur base sur les réseaux de neurones, université de MSILA, juin 2012.

[2] : Dubuc, Bruno, et al. Le cerveau à tous les niveaux. B. Dubuc, 2002

[3] : Y. K. Lin, Probabilistic Theory of Structural Dynamics, New York, Robert E. Krieger PublishingCompany, juillet 1976, 368 p.

[4] : Anne R. Kenney, Stephen Chapman, Digital Imaging for Libraries and Archives, Ithaca (New-York), CornellUniversity Library, 1996.

[5] : Claude E. Shannon, « Communication in the presence of noise », Proceedings of the Institute of Radio Engineers, vol. 37, no 1, janvier 1949, p. 10–21, consulté le 9 juillet 2013)

[6] : M.Ibnnkahla, "Application of neural networks to digital communication", Signalprocessing 80, pp 1185-1215, 2000.

[7]: Gérald Huguenin , Cours d'Electronique analogiqueBruit ,HE_Arc2014; octobre 2019.

[8] :Solène Kojtych. Introduction au traitement du signal et à l'analyse fréquentielle par transformées de Fourier. [Rapport Technique] École polytechnique de Montréal. 2019.

[9]: H. Mehrez. Étude et classification des algorithmes de TFR application à la conception d'un circuit papillon FFT intégré : unité arithmétique de base d'un processeur.

[10] : P. Duhamel et M. Vetterli, « A Fast Fourier transforms : a tutorial review and a state of the art », Signal Processing, vol. 19, 1990, p. 259-299

[11] : Calliope, "Traitement Automatique de la Parole", édition Masson, Paris, 1989.

[12] : Patrick Corsi. Reconnaissance automatique du locuteur : présentation générale, méthodologies et expérimentation, perspectives d'application. Modélisation et simulation. Institut National Polytechnique de Grenoble - INPG, 1979.

Bibliographie

[13]:Gabriele CHIGNOLI, Reconnaissance Automatique du Locuteur,mécanisme humain et tâche informatique : application de méthodes statistiques, UNIVERSITE PARIS III, 2018.

[14] : Jérôme Damelincourt, Présentation & historique de réseaux neuronaux, magasine de Futura-Sciences, 07/12/2001

[15] : 3-Youcef Djeriri, Comparaison and analysis of direct power control strategiesapplied to doublyfed induction generatorbasedwindenergy conversion system, University of Sidi-Bel-Abbès, 20 septembre 2017.

[16] : Doctissimo, dictionnaire médical.

(https://www.doctissimo.fr/sante/dictionnaire-medical/corps-cellulaire).

[17] : Campbell, Biologie, Pearson, p. 109.2012.

[18] :Tritsch D., Chesnoys-Marchais D., Feltz A., « Physiologie du neurone », 2e édition, 1998, Doin.

[19] : Dale Purves, George J. Augustine, David Fitzpatrick, William C. Hall, Anthony-Samuel LaMantia, James O. McNamara et S. Mark Williams, Neurosciences, Bruxelles, De Boeck Université, coll. « Neurosciences & Cognition », 3e éd., 811 p. chap. 5 (« La transmission synaptique »), p. 93-127, 2005.

[20] :Futura sciences (https://www.futura-sciences.com/sante/definitions/biologie-axone-2535/).

[21]:Doctissimo, dictionnaire médical.(https://www.doctissimo.fr/recherche?search%5Bquery%5D=La+gaine+de+my%C3%A9line+)

[22] : Futurasciences.(https://www.futura-sciences.com/sante/definitions/biologie-noeud-ranvier-14115/).

Bibliographie

[23] : Claude Touzet. Les réseaux de neurones artificiels, introduction au connexionnisme : cours, exercices et travaux. EC2, Collection de l'EERIE, N. Giambiasi ; 1992.

[24] :Sabour, Sara, Nicholas Frosst, and Geoffrey E. Hinton. "Dynamicroutingbetween capsules." Advances in neural information processingsystems. 2017.

[25] : Mitchell, T.M. Machine Learning, volume 4 de McGraw-Hill Series in Computer Science. McGraw-Hill multi-scalar drought index sensitive to global warming: the Standardized Precipitation Evapotranspiration Index – SPEI. Journal of Climate, pp 1696–1718,1997.

[26] : Nilsson, N. J. Introduction to Machine Learning, volume 56. The MIT Press.

[27] : S. Sius, G.J. Gibson,and C.F.N. Cowan, " Decision feedback equalisation using neural network structures and performance comparison with standard architecture,'' IEE

Proceedings, Vol.137, Pt. I, No. 4, pp.221-225, August 1990.

[28]: J. Wolf, Efficient Acoustic Parameters for Speaker Recognition, JASA, Volume 51, pp. 2045-2056, 1972.

[29] : Baoqiao Guo. Identification simultanée des paramètres de rigidité et d'amortissement de plaquesisotropes minces en vibration par la méthode des champs virtuels. Sciences de l'ingénieur [physics]. Arts et Métiers ParisTech, 2007. Français. NNT : 2007ENAM0010.

[30] : Gérard DREYFUS, « RÉSEAUX DE NEURONES », EncyclopædiaUniversalis [en ligne], consulté le 11 octobre 2020.

(https://www.universalis.fr/encyclopedie/reseaux-de-neurones-formels)

[31]:M.Ibnnkahla, "Application of neural networks to digital communication", Signalprocessing 80, pp 1185-1215, 2000.

[32]: The Math Works, Inc. (http://matlab.izmiran.ru/help/toolbox/nnet/backpr59.html)

Bibliographie

[33]: Bayesian regularized neural networks for small n big p Data by Hayrettin Okut,Submitted: October13th2015Reviewed: March 21st 2016Published: October 19th 2016.

(http://dx.doi.org/10.5772/63256)

[34]:Stat soft 1984-2016.

(https://www.statsoft.fr/conceptsstatistiques/glossaire/l/levenberg.html)

TABLES DES MATIERES

Remerciements

Dédicace

Liste des tableaux

Liste des figures

Liste des Abréviations

Introduction générale

CHAPITRE I : Généralités sur l'identification de locuteurs

Introduction

1. Production de la parole …….. 6

1.1 Appareil vocal humain ……..…… 6

1.2 Caractéristiques de la parole …… 7

1.3Signaux stationnaires ……...…… 8

2. Traitement de la parole ……..……... 8

2.1 Numérisation des signaux ……..... 8

2.1.1 Echantillonnage …….…… 8

2.1.2 Théorème de SHANNON …… 10

2.1.3 Bruit de Quantification ……. ….11

2.2 Outils mathématiques du traitement ……...…….11

2.2.1 Analyse temporelle ……..….11

A) Moyenne ……. ….11

B) Covariance …….….. 12

C) Analyse de fourier…….…….. 12

2.2.2 Analyse fréquentielle ……..……. 12

A) Transformation de Fourier ……..……. 12

B) Transformée de Fourier rapide ……...….
12

C) Spectrogramme …….……... 13

3. Reconnaissance automatique du locuteur ……...
14

4. Modélisation du locuteur ……
14

5. Conclusion……..……....15

CHAPITRE II : Réseaux de neurones artificiel

1. Introduction ……..…… 17

2. Historique ……..…….. 18

3. Les neurones biologiques ……..……20

3.1 Le corps cellulaire …….…….. 21

3 .2 . Le noyau …….…… 21

3.3 Les dentrites……...……21

3.4 La synaps…….…. 22

3.5 L'axone…….…… 22

3.6 La gaine de myéline ……..…..…….. 23

3.7 Le nœud de Ranvier ……...…….…….. 23

4. Réseaux de neurones artificiels …….…. 23

4.1 Définition d'un RNA……...…….. 23

4.2 Principe de fonctionnement d'un RNA…….. 24

4.3 Les composants d'un réseau de neurone …… 25

4.4 Architecture des réseaux de neurones …… 25

4.4.1 Les réseaux de neurones non bouclés ……..…..... 26

A) Réseaux de neurones monocouches …… 26

B) Réseaux de neurones multicouches …… 27

C) Réseaux de neurones à connexions locales …… 28

4.4.2 Réseaux de neurones bouclés ……..…… 29

4.5 Les fonctions d'activations ……... 30

4.6 Les techniques d'apprentissage …… 30

2.6.1 Apprentissage non-supervisé……...……. 31

2.6.2 Apprentissage supervisé……..……. 31

2.6.3 Apprentissage par renforcement …….…… 31

2.6.3.1 La loi de Hebb …… 31

4.7 Algorithme d'apprentissage ……..…….……. 33

A. Algorithme de rétro-propagation …….……...……33

B. Algorithme de Levenberg –Marquardt, ou algorithme LM …....…...34

C. Application à la méthode des moindres carrés ……....….... 35

- Choix du paramètre d'amortissement ……...…..... 36

4.8 Domaines d'application des réseaux de neurones artificiels …….….. 36

4.9 Les avantages et les inconvénients …….……..… 37

4.9.1 A Les avantages …….…..….. 37

4.9.2 B Inconvénients…….…… 37

5. L'analogie entre le neurone biologique et le neurone formel (artificiel) ……. 37

6. Conclusion ……. 38

CHAPITRE III : Simulations et résultats obtenus

1. Introduction …… 40

2. Traitement du signal vocal ……..…….. 40

3. Opérations de préparation des fichiers de parole …….…… 40

3.1. Opération de supprimer le bruit …… 40

3.2. La méthode de Pitch ……. 41

4. Traitements des fichiers de parole ……... 42

4.1. Extraction des caractéristiques et création du modèle de référence …….... 42

4.2. Base de données …….. 43

4.3. Extraction de caractéristique ……45

4.4. Former un classificateur …….. 47

4.5. Tester le classificateur ……. 48

5. Réseaux de neurone artificiel …….. 52

L’algorithme de Gradient conjugué mis à l'échelle (trainscg) ……..… 56

L’algorithme de Régularisation bayésienne (trainbr) …….....57

L’algorithme de Levenberg-Marquardt ……58

6. Conclusion ……...60

Conclusion Générale