REPUBLIQUE DU CAMEROUN
*Paix – Travail – Patrie*
********
UNIVERSITE DE YAOUNDE I
ECOLE NORMALE SUPERIEUR D'ENSEIGNEMENT TECHNIQUE D'EBOLOWA
DEPARTEMENT DE DE GENIE INFORMATIQUE
*********

REPUBLIC OF CAMEROUN
Peace – Work – Fatherland
*******
UNIVERSITY OF YAOUNDE I
HIGHER TECHNICAL TEACHER TREAINING COLLEGE OF EBOLOWA
DEPARTMENT OF OF COMPUTER ENGINEERING
*******

**Filière**
**Informatique Industrielle**

# CONCEPTION ET REALISATION D'UN SYSTEME INTELLIGENT D'ASSISTANCE VOCAL POUR LES PERSONNES A MOBILITE REDUITE

Mémoire rédigé et soutenu en vue de l'obtention du Diplôme de Professeur d'Enseignement Technique deuxième grade (DIPET II)

Par : **MATEUZEM NGUEKENG Calixte**
Ingénieur de travaux en informatique

Sous la direction de
**Pr. NDJAKOMO ESSIANE Salomé**
**Maitre de conférences**

**Année Académique : 2019 - 2020**

# DEDICACE

*A ma famille*

# REMERCIEMENTS

Nos sincères remerciements s'adressent :

- A DIEU Tout Puissant plein de miséricorde pour son aide spirituel ;
- A Madame Le Directeur de l'ENSET d'Ebolowa, **Pr SALOME NDJAKOMO ESSIANE** superviseur et encadreur de notre mémoire, elle qui n'a ménagé aucun effort pour nous offrir une formation de qualité;
- A notre Chef de Département Génie Informatique, **Dr OLLE OLLE DANIEL** pour son encadrement, ses conseils, son dévouement, sa présence et son soutien ;
- A **M. NYATTE** membre important de notre équipe d'encadrement, enseignant à L'ENSET d'Ebolowa, pour son encadrement, ses remarques, ses conseils et sa disponibilité ;
- A tout le Staff Administratif et Professoral de l'ENSET d'Ebolowa pour la formation enrichissante et nécessaire pour le suivi de nos jeunes frères et sœurs des lycées et collèges de notre pays ;
- A mes parents **M. NGUEKENG JEAN CLAUDE et Mme TSAFACK CHANTAL** pour leur soutien, leur présence, leurs conseils et leurs multiples encouragements pour faire de nous des personnes responsables et autonomes vis-à-vis de la société;
- A tous mes frères et sœurs pour leur esprit de solidarité, pour leur aide financière sans quoi nous ne serons arrivés au bout de ce travail ;
- A tous nos camarades de l'ENSET d'Ebolowa;
- A toutes les personnes de près ou de loin qui m'ont toujours soutenu et encouragé.

# TABLE DES MATIERES

# LISTE DES FIGURES

# LISTE DES TABLEAUX

# LISTE DES ABREVIATIONS

**IoT** : Internet of Things

**DARPA** : Defense Advanced Research Projects Agency

**IPA** : intelligent personal assistant

**IA** : Intelligence Artificielle

**MFCC** : Mel Frequency Cepstral Coefficients

**PIR** : Passive Infra-Red

**ML** : Machine Learning

**RNA** : Réseau de Neurone Artificielle

**DL :** Deep Learning

**IBM :** International Business Machines

**PLP** : Perceptual Linear Prediction

**SMC** : Short-time Modified Coberence

**CC** : Cepstral Compensation

**DCT** : Discrete Cosine Transform

**DFT :** Transformée de Fourier Discrète

**FFT :** Fourier Fast Transform

# RESUME

Des millions de personnes dans le monde sont atteintes de déficience visuelle, cécité, de vieillesse et toute autre forme de handicap. Ces individus sont soumis à de nombreuses difficultés. Un réel besoin est présent afin d'améliorer leur confort de vie et minimiser au maximum leur handicap, de nombreuses solutions répondant à cette problématique sont actuellement sur le marché.

Notre projet vise à apporter des solutions techniques en utilisant de l'intelligence artificielle pour répondre aux besoins de confort de ces personnes à mobilité réduite de bien se mouvoir dans leur domicile.

Ce projet s'inscrit dans le cadre de la conception et réalisation d'un assistant vocal intelligent basé sur les réseaux de neurones convolutifs (le Deep Learning), pour tenter de rendre automatiques les tâches quotidiennes effectuées par ces personnes et exécuter par de simples commandes vocales tel qu'ouvrir la porte' ou 'allumer les lampes', …etc.

Pour ce faire, nous avons élaboré un système d'aide intelligent basé sur la reconnaissance vocale grâce aux réseaux de neurones, essentiellement qui interagit avec les équipements électroniques afin de leur rendre la vie simple et conviviale.

Notre système est capable de s'adapter aux besoins de chaque personne par : le grand choix de commandes vocales laissé à l'utilisateur pour piloter ses équipements, afin qu'il soit très à l'aise lors des échanges avec le système ; l'interaction du système avec l'usager qui est informé en permanence par des messages vocaux annoncés par le système d'aide.

Le nombre d'entrée/sortie prévu pour le raccordement (équipements et capteurs) est important ce qui offre une multitude de solutions pour s'adapter à l'environnement de l'usager.

**Mots-clés :**

Personnes à mobilité réduite, Deep learning, réseaux de neurones, assistance vocale, interaction, intelligent.

# ABSTRACT

Millions of people around the world suffer from visual impairment, blindness, old age and other forms of disability. These individuals are subject to many difficulties in moving around. There is a real need to improve their living comfort and minimize their disability, and many solutions are currently on the market to address this issue. Our project aims to provide technical solutions to meet the comfort needs of these people with reduced mobility to move well in their homes.

This project is part of the design and development of an intelligent voice assistant based on convolutional neural networks (Deep Learning), to try to make the daily tasks performed by these people automatic and executed by simple voice commands such as 'open the door' or 'turn on the lights', ...etc... To do this, we have developed an intelligent help system based on voice recognition using neural networks, which essentially interacts with electrical equipment to make life simple and user-friendly.

Our system is capable of adapting to the needs of each person by : The large choice of voice commands left to the user to control his equipment, so that he is very comfortable when interacting with the system; The interaction of the system with the user who is permanently informed by voice messages announced by the help system.

The number of inputs/outputs provided for the connection (equipment and sensors) is important, which offers a multitude of solutions to adapt to the user's environment.

**Keywords :**

People with reduced mobility, Deep learning, neural networks, voice assistance, interaction, smart.

# INTRODUCTION GENERALE

Alors que les pays sous-développés sont engagés dans un processus de transition démographique marqué par un accroissement de la population vivant avec des maladies physiques qui diminuent leur espérance de vie et amènent à une augmentation du nombre de personnes dépendantes, la notion de dépendance étant basée sur une déficience des fonctions physiques, sensorielles et cognitives, sur une incapacité d'effectuer certaines activités de la vie quotidienne.

Les personnes en situation de handicap (malvoyant, muet, sourd, personnes ne possédant plus leurs mobilités …) ont parfois des difficultés, voire même une incapacité totale à effectuer les activités quotidiennes de manière autonome et ne possèdent plus les mobilités de contrôler efficacement la multitude d'équipements présents dans leur environnement. Il est donc possible d'améliorer la situation de handicap d'une personne en adaptant l'environnement à ses besoins, et augmenter ainsi son degré d'autonomie car dans la vie de tous les jours, nous avions des grands-parents ou des personnes handicapées que l'on laisse seules à la maison pour la plupart du temps, qui s'ennuient et n'ont personne avec qui parler.

Il en résulte donc ainsi un besoin en aide et assistance par un tiers pour les soins et les tâches de la vie courante. Cette assistance vise à améliorer le confort et le bien-être de l'habitant grâce à des interfaces naturelles permettant de contrôler les différents éléments de la maison. Pour ces personnes, ces dispositifs doivent être prévus et personnalisés pour compenser au mieux leurs handicaps. La parole étant le moyen de communication le plus naturel, un dispositif intelligent de reconnaissance automatique de la parole dans la maison peut donc être pertinent pour la commande des appareils permettant de leur faciliter la vie et maintenir le lien social.

Ainsi donc le but de notre projet sera non seulement d'aider ces personnes à avoir une vie plus confortable mais aussi à commander les appareils électroniques sans avoir besoin de se déplacer comme par exemple : allumer les ampoules, l'ouverture des portes, allumer la télé ou radio, demander l'heure qu'il est au système communiquant ; allumer le ventilateur ; signaler la présence de quelqu'un devant la porte en cas de détection de mouvement.

Les questions auxquelles ce travail voudrait répondre se présentent comme suit :

Comment mettre en place une solution technologique capable d'améliorer les conditions de vies de ces personnes? Quels sont les outils adéquats que nous pourrions utiliser pour répondre aux exigences de notre problème et avoir les résultats souhaités ?

Pour répondre à ces questions nous nous sommes fixés comme principal objectif de la présente étude de développer un système de conception et réalisation d'un système intelligent d'assistance vocale pour les personnes à mobilité réduites.

Et comme objectifs spécifiques :

- S'inspirer de certains résultats de l'état de l'art actuel en utilisant un certain type d'algorithmes et intelligence artificielle appliqués au problème d'assistance vocale pour l'autonomie et l'indépendance des personnes à mobilité réduites ;
- Construire à partir d'un réseau de neurones à convolution avec une précision acceptable, un système d'assistance vocale

Ainsi suivant ces enjeux sociaux, humanitaires et certaines avancées technologiques, notre étude intitulée « **CONCEPTION ET REALISATION D'UN SYSTEME INTELLIGENT DE COMMANDE VOCALE POUR LES PERSONNES A MOBILITE REDUITE** » serait une solution sur laquelle les personnes à mobilités réduites pourraient s'appuyer pour pallier les problèmes sus évoqués dans notre analyse.

La contribution de ce mémoire est principalement l'intégration dans un système embarqué du développement d'un système à commande vocale, utilisant un microcontrôleur Arduino UNO. Le système proposé sera conçu sur la base d'un microcontrôleur basé sur les réseaux de neurones pour la reconnaissance de la parole. Un module de carte SD associé à une carte SD contenant une voix humaine préenregistrée sous forme de fichier audio sera utilisé par le système pour le développement de son système de conversation et doit être suffisamment intelligent pour comprendre leurs langages et savoir avec exactitude ce qu'elles veulent et par la suite exécuter la tâche pour laquelle la personne a sollicité son aide.

Afin d'atteindre nos objectifs, le travail a été divisé en trois chapitres repartis sur deux parties la revue de la littérature et la conception du système proposé. Néanmoins, nous commençons d'abord par situer le problème par une **introduction générale** qui définit la problématique, détermine les hypothèses et présente l'objectif principal et la contribution de cette recherche.

- Introduction générale : Contexte de l'étude, problématique, questions de recherche, les objectifs de l'étude, la méthodologie à suivre, les résultats attendus ;
- Partie 1 : La revue de la littérature ;
    - Chapitre 1 : Généralités sur les assistances vocales : conçue pour donner une vision plus ou moins claire sur les concepts clés de notre sujet d'études, de la présentation des différents travaux ou méthodes ;
- Partie 2 : Conception de la solution ;
    - Chapitre 2 : Méthode de conception du système embarqué et matériels utilisés; il est question de présenter tour à tour : le cahier des charges de notre solution, les méthodes de conception ; le matériel et logiciels utilisés pour concevoir le système ;
    - Chapitre3 : Résultats et Interprétation : qui nous permet de présenter nos résultats obtenus et de les interpréter.
- Conclusion et perspectives.

# PREMIERE PARTIE : REVUE DE LITTERATURE

## ETAT DE L'ART SUR LES SYSTEMES D'ASSISTANCES VOCALES

**<u>Objectif</u> : Mettre en évidence les différents travaux relatifs aux Systèmes d'assistances vocales, les travaux ayant intégrés les réseaux de neurones (intelligence artificielle) dans un système de reconnaissance vocal.**

# CHAPITRE I : GENERALITES SUR LES ASSISTANCES VOCALES

L'assistant vocal est devenu une réalité sous l'impulsion des smartphones et des objets connectés. L'interaction à la voix se pose comme un complément sinon une alternative aux interfaces tactiles. Elle permet à l'utilisateur d'obtenir des réponses à des requêtes énoncées oralement ou par écrit et de commander certaines fonctions du terminal sur lequel il est installé ainsi que la possibilité de transférer des données sur un réseau sans nécessiter une intervention humaine ou l'interaction homme-ordinateur. Ce terme grandit et évolue de plus en plus rapidement, cela implique de créer un réseau d'objets physiques autour de nous et de les connecter au monde numérique [1].

Nous souhaitons concevoir comme solution un système à assistance vocale intelligent, à faible coût, installable, basé sur les réseaux de neurones convolutifs, qui seront principalement bénéfiques pour les personnes à mobilité réduite (personnes malvoyantes, handicapés, âgées). La solution vise à remplir toutes les fonctions qui peuvent être contrôlées par un module de reconnaissance vocale comme éteindre / allumer les lumières ou les appareils, contrôler les chaînes de télévision, ouvrir les portes, maintenir la température, veiller sur l'heure de prise des médicaments de ces derniers, la surveillance, grâce à des commandes vocales.

Ce chapitre est conçu pour donner une vision plus ou moins claire sur les concepts clés de notre sujet d'études en mettant en exergue l'ensemble des méthodes faisant suite à l'État de l'art.

## I. SYSTEMES D'ASSISTANCE VOCALE

### I.1/ Evolution et historique

Le premier outil permettant la reconnaissance vocale numérique a été IBM Shoebox, présenté au grand public lors de l'exposition universelle de Seattle en 1962, après son lancement commercial en 1961 [5]. Ce premier ordinateur, développé près de 20 ans avant l'introduction

du premier ordinateur personnel IBM en 1981, était capable de reconnaître 16 mots parlés et les chiffres de 0 à 9.

La seconde étape dans le développement de la technologie de reconnaissance vocale a été réalisée dans les années 1970 par l'Université Carnegie Mellon de Pittsburgh, en Pennsylvanie, avec un important soutien du Département de la Défense des États-Unis et de son agence, la Défense Advanced Research Projects Agency (DARPA, « Agence pour les projets avancés de défense ») [1]. Leur outil « Harpy » a maîtrisé avec environ 1000 mots le vocabulaire d'un enfant de trois ans. Une dizaine d'années plus tard, le même groupe de scientifiques a développé un système qui pouvait non seulement analyser des mots individuels, mais aussi des séquences entières de mots activées par le modèle de Markov caché (Hidden Markov Model). Ainsi, les premiers assistants virtuels, qui utilisaient un logiciel de reconnaissance vocale, étaient des logiciels automatisés de dictée et de dictée médicale.

Dans les années 1990, la technologie de reconnaissance vocale numérique est devenue une caractéristique de l'ordinateur personnel avec Microsoft, IBM, Philips et Lernout & Hauspie luttant pour l'intérêt des clients et des utilisateurs. Ce n'est que beaucoup plus tard, en 1994, que le lancement du premier smartphone IBM Simon a introduit les bases des assistants virtuels intelligents tels que nous les connaissons aujourd'hui [5].

Le premier assistant virtuel numérique installé sur un smartphone était SIRI [2], qui a été présenté comme une caractéristique de l'iPhone 4S le 4 octobre 2011. En intégrant Siri en 2011 dans l'iPhone, la société Apple est devenu le pionnier en matière d'assistants personnels intelligents (ou IPA pour intelligent personal assistant). Ces applications se basent sur deux technologies complémentaires : la reconnaissance vocale et l'intelligence artificielle. Elles permettent ainsi à un utilisateur de réaliser une recherche web sous forme conversationnelle.

En 2012, Google présente son assistant vocal pour Android baptisé **Google Now** qui s'appuient notamment sur le Knowledge Graph [2]. Google Now est un assistant personnel intelligent qui prend la forme d'une application Android et iOS basée sur la reconnaissance vocale, le traitement du langage naturel par oral ainsi que sur la synthèse vocale pour apporter des réponses aux requêtes des utilisateurs à l'oral et à l'écrit, faire des recommandations et effectuer des actions en déléguant certaines requêtes à des services en ligne. Google Now est inclus par défaut à partir de la version 4.1 d'Android (« Jelly Bean »), et a fonctionné pour la

première fois sur le Galaxy Nexus mais depuis mai 2018, Google Now est remplacé par Google Assistant.

En novembre 2014, Amazon a annoncé **Alexa** en même temps qu'Echo5 . Alexa et ses objectifs ont été inspirés par la voix de l'ordinateur et du système de conversation de Star Trek TNG. Alexa est le nom qui désigne et sert à interpeller l'assistant personnel virtuel développé par le Lab126 d'Amazon.com, rendu populaire par les appareils Echo [2]. Il est capable d'interaction vocale, de lire de la musique, faire des listes de tâches, régler des alarmes, lire des podcasts et des livres audio, et donner la météo, le trafic et d'autres informations en temps réel.

En avril 2014, Microsoft dévoile Cortana, son assistant vocal pour Windows Phone qui est également présent dans Windows 10. **Cortana** est le nom de l'assistant personnel intelligent développé par Microsoft pour sa plateforme Windows Phone à partir de la version 8.1 et désormais sur Windows 10. Cortana existe également sur Android et iOS (en bêta et uniquement dans certaines langues) sous la forme d'une application et est intégré avec le système Cyanogen Mod. La particularité de Cortana réside dans l'utilisation d'un « bloc-notes » dans lequel l'utilisateur peut renseigner et consulter diverses informations telles que ses centres d'intérêt, ses tâches, ses « heures tranquilles », ses lieux favoris, ses rappels, synchronisés entre appareils Windows et Android mais aussi le nom par lequel il souhaite être appelé et en vérifier la bonne prononciation [2].

À partir de 2017, les capacités et l'utilisation des assistants virtuels se développent rapidement, avec l'arrivée de nouveaux produits sur le marché et un fort accent sur les interfaces utilisateur de messageries et vocales [2].

## I.2/ Définition et concept

Un assistant vocal, aussi appelé « assistant personnel intelligent », est une application logicielle basée sur la reconnaissance vocale du langage naturel et la restitution d'informations par synthèse vocale. Il est un agent logiciel qui peut effectuer des tâches ou des services pour un individu [32]. Ces tâches et les services effectuées par l'assistant sont basés sur les entrées fournies par l'utilisateur, la connaissance de l'emplacement de l'utilisateur, les données historiques conservées par l'assistant et la possibilité d'accéder à des informations à partir d'une variété de sources.

Représentant un outil fondamental dans la communication humaine, la parole demeure une forme d'interaction intuitive, invisible et rapide pour réaliser une diversité de tâches à partir des commandes vocales [26]. Ces technologies ont permis de combiner les systèmes de reconnaissance vocale à des méthodes de synthèse et de commande vocale [24]. Par ailleurs, ces dernières années ont été marquées par d'autres progrès dans le domaine de l'intelligence artificielle, et cela dans plusieurs secteurs : robotique, véhicules autonomes, maisons intelligentes, traitement de la parole et compréhension du langage naturel, [3], et cela, dans l'optique, notamment, d'augmenter les capacités humaines en s'appuyant sur une coopération entre l'homme et la machine [12].

## I.3/ Principe de l'assistance vocale

Les progrès de l'IA appliquée à la compréhension de la parole ont favorisé le développement de robots conversationnels (chatbots) et d'assistants vocaux répondant aux requêtes variées des usagers (Siri, Google assistant, Cortana, etc.). D'abord limités à l'utilisation du smartphone, les assistants vocaux dits « intelligents » intègrent dorénavant les habitats avec l'arrivée des enceintes intelligentes telles qu'Alexa d'Amazon ou Google Home [2].

Un assistant vocal intelligent représente un logiciel intégré dans un dispositif tangible, disposant de la faculté de dialoguer avec un humain pour réaliser des tâches sur ordre vocal des utilisateurs (exemples : répondre à des questions en sélectionnant des informations sur le web, rappeler des actions importantes aux utilisateurs, effectuer certaines tâches sur des objets connectés, etc.). Disposant très souvent d'une capacité d'apprentissage automatique, ces intelligences artificielles se développent continuellement sans l'intervention des développeurs, apprennent de leurs erreurs, et cela, tout en intégrant des informations sur les utilisateurs afin de proposer des services contextualisés et personnalisés [4].

Nous considérons que pour développer ces nouveaux systèmes vocaux ou les intégrer à des produits ou services, il importe d'étudier et de comprendre les situations d'utilisation dans lesquelles ils s'insèrent, ainsi que les leviers et les freins à leur appropriation. Proposer des systèmes adaptés à l'activité des utilisateurs en prenant en compte le point de vue de l'humain et de ses activités constitue un des principes fondateurs de la démarche en ergonomie.

### I.4/ Exemple de synoptique

Le schéma simplifié, ci-dessous, permet de mieux comprendre la circulation des informations dans un système d'assistance vocale.

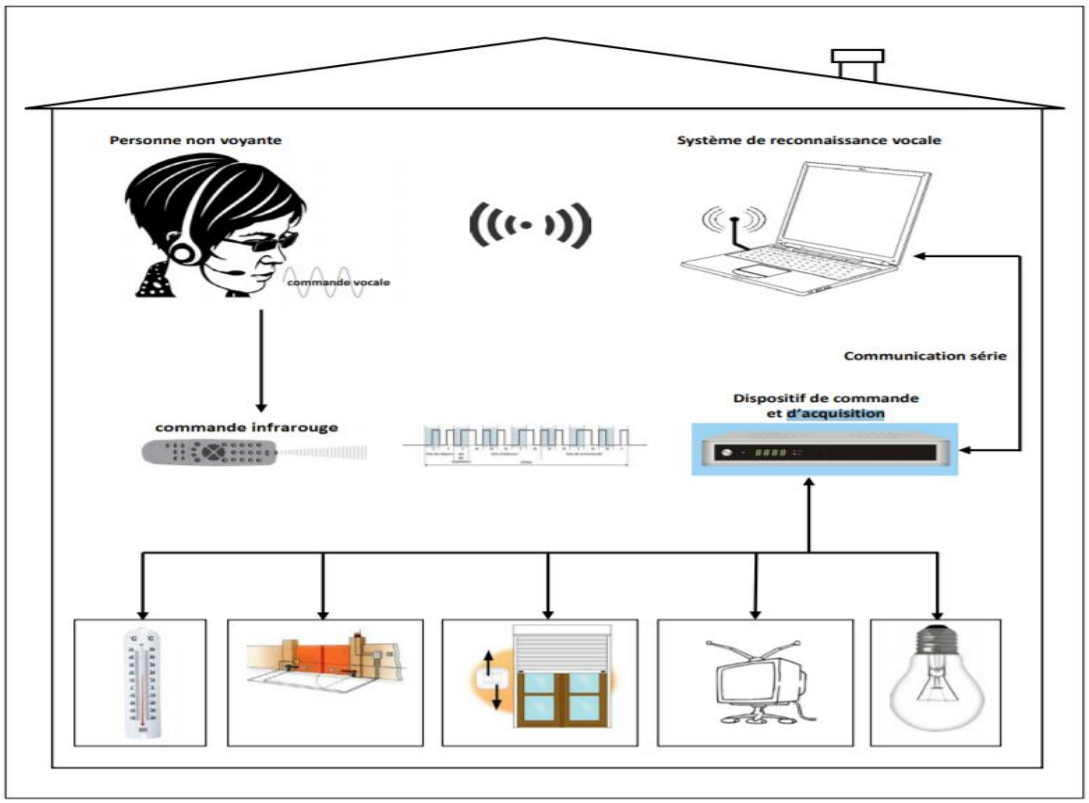


*Figure 1: Exemple synoptique de l'assistance vocale [17]*

## I.5/ Les avantages des technologies vocales

Parmi les utilisations identifiées, nous avons retrouvé des expériences appréciées par les utilisateurs que nous avons catégorisées en trois dimensions : les qualités des formes d'interaction, les qualités du système et les effets des interactions vocales sur l'activité des utilisateurs. Cette section nous permettra d'aborder les problématiques d'utilité et d'utilisabilité afin de saisir de quelle manière les dispositifs répondent aux besoins, aux attentes et aux buts des utilisateurs.

#### a) Les qualités de l'interaction

Parmi les qualités retrouvées dans les interactions vocales, nous retrouvons tout d'abord **l'instantanéité, la facilité et le confort d'utilisation** de mobiliser la parole comme un moyen

d'interaction. Ces caractéristiques permettent une rapidité de commande et de contrôle dépassant largement les actions consistant à presser un bouton ou cliquer à l'aide d'une souris. Les retours vocaux sont notamment rapidement compréhensibles, ils évitent toute ambiguïté que nous pourrions rencontrer via les alarmes sonores et les voyants lumineux [10]. De plus, la commande vocale semble avantageuse pour réaliser des actions complexes nécessitant plusieurs touches et de multiples interfaces [8]. Presque un tiers des répondants se dit même excité par un futur où les assistants vocaux pourront anticiper leurs besoins, agir par eux-mêmes ou faire des suggestions [16].

La **concision et l'adaptation des feedbacks** aux requêtes facilitent l'acceptabilité des technologies vocales. En effet, la pertinence et la justesse des réponses apportées par ces interfaces contribue grandement à la réussite de l'expérience utilisateur [9].

Cette expérience positive est renforcée si l'utilisateur utilise son propre vocabulaire, il s'agit d'une commande vocale libre [11]. Il a d'ailleurs été observé que certains assistants vocaux, tels qu'Alexa, disposent d'un traitement sémantique plus performant que Siri [2] : « *Hey Siri, combien de temps il reste ? ==> J'ai trouvé un article sur le Times. Dois-je te le lire ? (...) Alexa, combien de temps reste-t-il ? ==> Environ 6min et 10 sec* ». Par ailleurs, nous pouvons remarquer que les technologies vocales intègrent naturellement une forme d'interaction pouvant s'adapter à une grande variété d'utilisateurs, dont les personnes en situation de handicap ou « déficientes », les personnes âgées, ainsi que les enfants [5]. Ils créent une relation forte : 36% de ceux qui les utilisent régulièrement avouent aimer leur assistant vocal à tel point qu'ils souhaiteraient qu'il soit réel. Enfin, presque la moitié, (46%) d'utilisateurs potentiels pourrait avoir recourt à la reconnaissance vocale si on leur donnait davantage de garanties quant à la protection de leurs données personnelles ; [12]

### b) <u>Les effets favorables sur les activités</u>

La **capacité à cumuler des tâches** est l'une des conséquences et des opportunités permises par l'interaction vocale qui est grandement plébiscitée par les utilisateurs. Lors de situations contraintes où les mains et le regard sont affairés, les utilisateurs sont ainsi en capacité de réaliser des doubles tâches durant les routines matinales par exemple : « *D'habitude le matin, je mets de la musique avant de sortir du lit. Ou bien, lorsque je sors de la douche avec les mains mouillées, plutôt que d'aller vers le téléphone avec les mains humides, je peux juste lancer la musique pendant que je me prépare pour aller travailler* » [17]. Ce cumul de tâche révèle

notamment des utilisations qui s'ancrent profondément dans les routines en faisant de la commande vocale un élément incontournable du quotidien : « *Allume la radio. Quel temps fait-il aujourd'hui ?* » [17. Ainsi, la possibilité de gérer des commandes vocales depuis n'importe quelle pièce du foyer sans les mains, ni du regard, favorise non seulement une grande mobilité dans l'espace domestique, mais elle permet notamment de soutenir la dynamique et le passage d'une activité à l'autre dans le cours d'action en situation domestique ou à l'extérieur [15].

## I.6/ Les inconvénients des technologies vocales

Nous abordons dans cette section l'ensemble des empêchements et des épisodes d'échec vécus lors des parcours d'usage de ces dispositifs. Les contraintes rencontrées constituent des résultats pertinents à étudier car elles éclairent l'ensemble des freins à l'acceptabilité, c'est-à-dire les raisons pour lesquelles ces technologies ne sont pas « en adéquation avec les pratiques, les ressources, les objectifs, le système de valeurs et les normes éthiques des utilisateurs pour être acceptable » [7]. Nous détaillerons ainsi des formes d'interactions imposées, des défaillances du système évoqué, ainsi que des questions éthiques soulevées.

### a) Des formes d'interaction imposées

- **Les empêchements du système**

L'absence d'écran visuel pour soutenir le vocal afin de proposer une plus grande diversité et richesse des données est souvent considéré comme une contrainte [13]. Portant sur une variété de dispositifs domestiques (un haut-parleur à commande vocale ; un écran tactile mural ; une application mobile ; un robot social) a démontré l'importance du canal visuel auprès des utilisateurs pour disposer d'informations sur la réussite ou non des commandes réalisées et pour obtenir des informations sur la situation en cours et le contexte de la machine.

### b) Défaillances du système

L'échec d'interprétation des commandes et du contexte constitue un autre frein à l'acceptabilité de ces technologies. Parmi cette catégorie nous retrouvons deux principales contraintes. On identifie un problème de fiabilité causé par la mauvaise interprétation des commandes provenant des biais sémantiques évoqués précédemment ou d'une mauvaise reconnaissance pour certaines voix comme celles des personnes âgées, des enfants ou des personnes atteintes d'une maladie [25]. Cet échec de reconnaissance vocale ou de détection des

phrases d'activation peut aussi intervenir en cas de perturbation selon l'ambiance sonore tels que les fonds musicaux ou les voix multiples [19]. Enfin, nous identifions une dernière limite liée à l'utilisation collective. Si l'assistant vocal ne dispose pas de la capacité à distinguer les différentes voix du foyer, son partage auprès de plusieurs membres peut empêcher la personnalisation du système pour l'utilisateur cible et ainsi modifier la compréhension du contexte de l'usager [4].

### c) Questions éthiques soulevées : confidentialité, discrétion et privauté

L'utilisation des assistants vocaux invite à questionner la dimension éthique des données. En effet, cette problématique met en exergue la confidentialité et la sécurisation des données, mises à mal par ces technologies qui doivent cumuler une quantité massive de données personnelles pour fonctionner efficacement [4]. Dans un souci d'efficacité du dispositif et de contextualisation des services, les utilisateurs sont encouragés à **renoncer à la confidentialité** totale de leurs données dont on recommande un stockage permanent sur les serveurs afin d'assurer une mémoire des actions et des requêtes [6].

# II. APPLICATION DE L'INTELLIGENCE ARTIFICIELLE SUR L'ASSISTANCE VOCALE

L'intelligence artificielle (IA) est l'intelligence présentée par les machines. En informatique, une machine « intelligente » idéale est un agent rationnel flexible qui perçoit son environnement et prend des mesures qui maximisent ses chances de réussite à un certain objectif. De manière familière, le terme « intelligence artificielle » est appliqué lorsqu'une machine imite des fonctions « cognitives » que les humains associent à d'autres esprits humains, comme « apprendre » et « résoudre des problèmes ». L'intelligence artificielle est généralement basée sur l'apprentissage automatique.

Les tâches d'apprentissage automatique sont généralement classées en trois grandes catégories, en fonction de la nature du signal d'apprentissage ou de la rétroaction dont dispose un système d'apprentissage :

- **Apprentissage supervisé** : L'apprentissage supervisé (supervised learning en anglais) est une technique d'**apprentissage** automatique où l'on cherche à produire automatiquement des règles à partir d'une base de données d'**apprentissage** contenant des exemples (en général des cas déjà traités et validés).
- **Apprentissage non supervisé** : consiste à tirer de la valeur de données dans lesquelles l'attribut à prédire n'apparaît pas. Apprentissage non supervisé peut être un but en soi (découvrir des tendances cachées dans les données) ou un moyen vers une fin (fonction d'apprentissage).
- **Apprentissage par renforcement** : Un programme informatique interagit avec un environnement dynamique dans lequel il doit effectuer un certain objectif (comme la conduite d'un véhicule), sans professeur dire explicitement si elle est venue près de son objectif.
- **Apprentissage semi-supervisé** : C' est une approche de l'apprentissage automatique qui combine une petite quantité de données étiquetées avec une grande quantité de données non étiquetées pendant la formation. L'apprentissage semi-supervisé se situe entre l'apprentissage non supervisé (sans données de formation étiquetées) et l' apprentissage supervisé (avec uniquement des données de formation étiquetées).

## II.1/ Les Algorithmes utilisés dans l'intelligence artificielle

Il existe plusieurs outils permettant d'implémenter l'intelligence artificielle. On peut citer entre autres :

- **Arbre de décision**

L'apprentissage par arbre de décision utilise un arbre de décision comme un modèle prédictif, qui mappe observations sur un élément à des conclusions au sujet de la valeur cible de l'élément.

- **Les réseaux neuronaux artificiels**

Un réseau de neurones artificiels (RNA) généralement appelé « réseau neuronal » (RN), est un algorithme d'apprentissage qui est inspiré par la structure et les aspects fonctionnels des réseaux de neurones biologiques. Les calculs sont structurés en termes d'un groupe

interconnecté de neurones artificiels, traitement de l'information en utilisant une approche connexionniste de calcul.

Un réseau de neurones décompose les données en couches d'abstraction [36]. Ils sont notamment appliqués pour résoudre des problèmes de classification, de prédiction, de catégorisation, d'optimisation, de reconnaissance des formes et de mémoire associative (Drew et Monson, 2000). Dans le cadre du traitement des données, les RNA constituent une méthode d'approximation de systèmes complexes, particulièrement utile lorsque ces systèmes sont difficiles à modéliser à l'aide des méthodes statistiques classiques. Les RNA sont également applicables dans toutes les situations où il existe une relation non linéaire entre une variable prédictive et une variable prédite. Par leur nature et leur fonctionnement, les RNA peuvent détecter les interactions multiples non linéaires parmi une série de variables d'entrée, ils peuvent donc gérer des relations complexes entre les variables indépendantes et les variables dépendantes [37]. L'objectif général d'un RNA est de trouver la configuration des poids de connexion entre neurones pour qu'il associe à chaque configuration d'entrée, une réponse adéquate. L'utilisation d'un RNA se fait en deux temps. Tout d'abord une phase d'apprentissage qui est chargée d'établir des valeurs pour chacune des connexions du réseau, puis une phase d'utilisation proprement dite, où l'on présente au réseau une entrée et où il nous indique en retour sa sortie calculée.

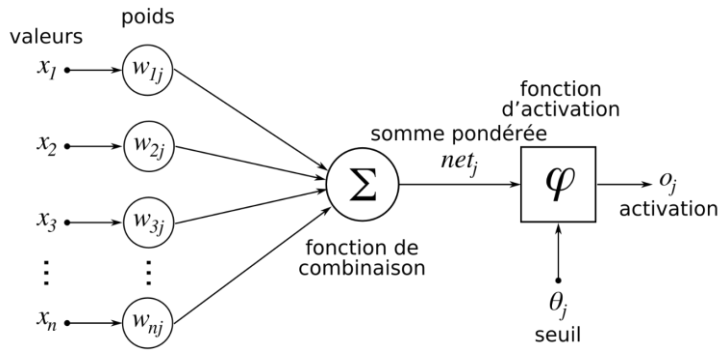


*Figure 2: structure d'un réseau de neurones artificiel[38]*

- **Réseaux Bayésiens**

Un réseau bayésien : modèle graphique acyclique orienté est un modèle graphique probabiliste qui représente un ensemble de variables aléatoires et leurs indépendances conditionnelles par un graphe orienté acyclique. Par exemple, un réseau bayésien pourrait représenter les relations probabilistes entre les maladies et les symptômes. Compte tenu des

symptômes, le réseau peut être utilisé pour calculer les probabilités de la présence de diverses maladies. Des algorithmes efficaces existent qui effectuent l'inférence et l'apprentissage.

- **Le Machine Learning (apprentissage automatique)**

Le Machine Learning (ML) ou apprentissage automatique ou apprentissage artificiel, est un champ d'étude de l'intelligence artificielle qui se fonde sur des approches mathématiques et statistiques pour donner aux ordinateurs la capacité d' « apprendre » à partir de données, c'est-à-dire d'améliorer leurs performances à résoudre des tâches sans être explicitement programmés pour chacune. Plus largement, il concerne la conception, l'analyse, l'optimisation, le développement et l'implémentation de telles méthodes [33].

Les deux objectifs principaux du Machine Learning sont la classification et la prédiction. L'algorithme doit pouvoir distinguer les données et les classer, c'est le cas lors du filtrage de spam sur une boîte mail par exemple. Une fois classées, les algorithmes observent les comportements du passé à partir des données classées, essayent de repérer les schémas récurrents et prédisent avec une certaine probabilité l'issue.

## II-2/ Choix de la méthode d'apprentissage pour le système de reconnaissance vocal

Tous les algorithmes d'apprentissage citées ci-dessus permettant de faire de la reconnaissance vocale mais notre choix s'est penché sur celle du Deep Learning.

Les réseaux Deep Learning communément appelé apprentissage profond peuvent comporter de nombreuses couches (trois à quatre couches). Ce sont des techniques d'apprentissage automatique qui tirent directement des enseignements des données d'entrée. Ces derniers temps, le Deep Learning attire beaucoup l'attention, et pour cause : Il obtient des résultats auparavant irréalisables.

Le Deep Learning est particulièrement bien adapté aux applications d'identification complexe telles que la reconnaissance faciale, la traduction de texte et la reconnaissance vocale.

En fait, l'algorithme va adapter les liaisons entre ses neurones (il va les renforcer ou les détruire), pour qu'en sortie on ait une bonne approximation des données d'entrée. Par exemple,

dans un réseau qui aura appris à prédire le mot qu'un individu va prononcer, s'il dit en entrée "bon", alors en sortie on aura un mot très proche de "bonjour".

Parmi quelques exemples d'algorithmes de Deep Learning nous pouvons citer :

- Les réseaux de neurones artificiels (NN) : ce sont les plus simples et sont souvent utilisés en complément car ils trient bien les informations ;
- Les réseaux de neurones convolutifs (CNN) : spécialisés dans le traitement de l'image, ils appliquent des filtres à des données pour en faire ressortir de nouvelles informations (par exemple, faire ressortir les contours dans une image peut aider à trouver où est le visage), il est le plus adapté pour la reconnaissance vocale et dont le réseau de neurones sur lequel notre système sera basé ;
- Les réseaux de neurones récurrents (RNN) : les plus connus sont les LSTM, qui ont pour faculté de retenir de l'information en mémoire utilisé par les séquences et de la réutiliser peu après. Ils servent pour l'analyse de texte (NLP), puisque chaque mot dépend des quelques mots précédents (pour que la grammaire soit correcte).

## II.3/ Reconnaissance vocale basée sur MFCC adaptatif et apprentissage en profondeur (Deep Learning)

Pour mener à bien notre étude, nous proposons une méthode de reconnaissance vocale améliorée utilisant MFCC (Mel Frequency Cepstral Coefficients) adaptatif et Deep Learning. Pour améliorer le taux de reconnaissance vocale, il est important d'extraire les données audio du signal d'origine. Cependant, les algorithmes existants qui sont utilisés pour supprimer le bruit d'une bande particulière détériorent le signal audio. À la différence du MFCC existant, le filtre proposé est construit de manière compacte dans la zone de densité de données pour réduire la perte de données et imposer la valeur pondérée à la zone de données. En conséquence, il empêche la perte de données qui se traduit par une amélioration du taux de reconnaissance.

Les méthodes de reconnaissance vocale sont généralement basées sur une comparaison entre « Le signal audio à reconnaître » et « le signal audio à régler ». Dans le cas idéal, un système embarqué peut être utilisé à cet effet. Cependant, le signal vocal bruité entraîne une erreur fatale dans la reconnaissance vocale. La méthode d'amélioration de la parole est proposée pour éliminer le bruit du signal audio lui-même, [38]. Pour compenser le modèle endommagé pour

les effets du bruit, la méthode intrinsèquement robuste des fonctionnalités vocales comprend MFCC (Mel-Frequency Cepstral Coefficient), PLP (Perceptual Linear Prediction) [7], SMC (Short-time Modified Coberence) [8] et CC (Cepstral Compensation) [9]. Récemment, la méthode MFCC est relativement plus populaire que d'autres méthodes. Fondamentalement, le MFCC calcule l'énergie du logarithme du banc de filtres configuré en tenant compte des caractéristiques auditives humaines à l'aide de DCT (Discrete Cosine Transform). Même si les performances de reconnaissance diminuent dans les signaux avec un faible rapport signal / bruit, cette méthode montre de bonnes performances

## II.4/ Description du système de reconnaissance vocal

Tous les systèmes de reconnaissance vocale sont divisés en deux parties, une première partie qui représente la phase d'extraction des paramètres, et une deuxième partie qui est le moteur de reconnaissance. Les performances des systèmes de reconnaissance vocale dépendent de façon considérable des paramètres acoustiques utilisés.

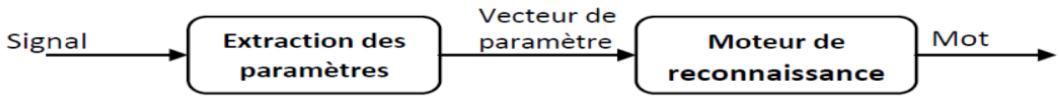


*Figure 3: Schéma bloc d'un système de reconnaissance de la parole*

- **Extraction des paramètres**

Pour résoudre les problèmes liés à la complexité de la parole, il est possible de calculer des coefficients représentatifs du signal traité appelé MFCC (Mel Frequency Cepstral Coefficients). Ces coefficients sont calculés à l'intervalle temporel régulier. En simplifiant les choses, le signal de parole est transformé en une série de vecteurs de coefficients.

Ces coefficients doivent représenter au mieux le signal qu'ils sont censés modéliser, et extraire le maximum d'informations utiles pour la reconnaissance.

Un système de paramétrisation du signal, se décompose en deux blocs, le premier de mise en forme et l'autre de calcul de coefficients. Le signal analogique est fourni en entrée et une suite discrète de vecteurs, appelée trame acoustique est obtenue en sortie.

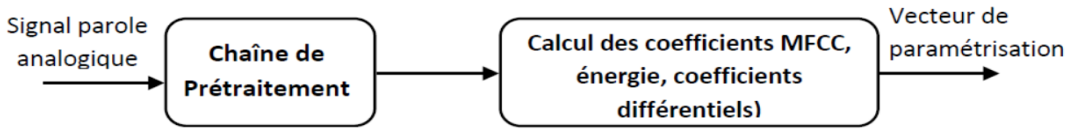


*Figure 4: Phase de paramétrisation acoustique*

- **Chaîne de prétraitement**

Il est nécessaire de mettre en forme le signal de parole. Pour cela, quelques opérations sont effectuées avant tout traitement. Le signal est tous d'abord filtré puis échantillonné à une fréquence donnée. [27]

Une préaccentuation est effectuée afin de relever les hautes fréquences. Qui sont moins énergétiques que les basses fréquences; la préaccentuation $Sn^{'}$ de l'échantillon Sn *à* l'instant *n* est calculée pour une valeur α comprise entre 0,9 et 1 comme :

$$S_n^{'} = S_n - \alpha S_{n-1} \qquad (1)$$

Puis le signal est segmenté en trames. Chaque trame est constituée d'un nombre N fixe correspondant à environs 25 ms de parole (durée pendant laquelle la parole peut être considérée comme stationnaire). Enfin une multiplication par une fenêtre de pondération Wn est effectuée, afin de réduire les effets de bords. Le choix se porte sur la fenêtre de **Hamming:**

$$Ham\min g(n) = 0{,}54 - 0{,}64\cos\left(2\pi\frac{n}{N-1}\right), \quad 0 \le n \le N\text{-}1 \qquad (2)$$

avec

$$S_n^{''} = W_n S_n^{'} \qquad (3)$$

Après cette mise en forme du signal (commune à la plupart des méthodes d'analyse de la parole), une transformée de Fourier discrète DFT en particulier FFT (Transformé de Fourier Rapide) est appliquée pour passer dans le domaine fréquentiel. [26]

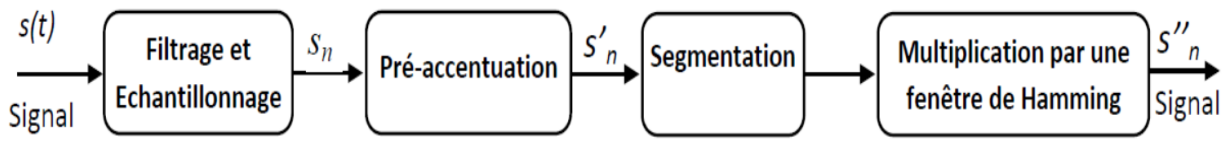

*Figure 5: Chaîne de prétraitement du signal parole*

- **Les coefficients de MFCC**

L'analyse acoustique MFCC est l'une des techniques les plus utilisées pour la paramétrisation du signal en segmentation markovienne de parole.

Cette technique est basée sur deux idées clés [28]. La première consiste à exploiter les propriétés du système auditif humain par la transformation de l'échelle linéaire des fréquences en échelle de Mel. Et la deuxième consiste à effectuer une transformation cepstrale qui permet la décorrélation des composantes spectrales du signal de parole.

Pour transformer une fréquence linéaire en une fréquence Mel, on utilise la formule de transformation suivante:

$$B(f) = 2595 \log_{10}\left(1 + \frac{f}{700}\right) \qquad (4)$$

Où *f* est la fréquence en Hz, *B(f)* est la fréquence mel-échelle de *f*. Les bandes-passantes sont de même taille dans l'échelle Mel.

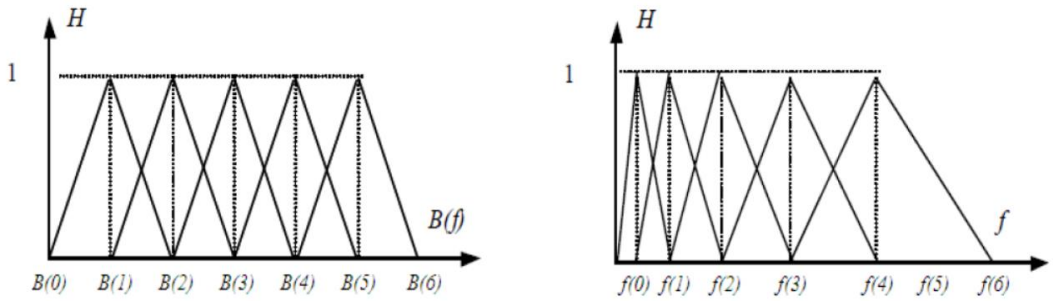


*Figure 6: Les filtres triangulaires passe-bande en Mel-Fréq (B(f)) et en fréquence (f)*

On peut calculer les points frontières *B*(j) des filtres en Mel-fréquence ainsi :

$$B(j) = B(f_l) + j\frac{B(f_h) - B(f_l)}{J+1} \qquad 0 \le j \le N+1 \qquad (5)$$

*N* est le nombre de filtres (N = 22).

On doit calculer les points *f* (j) correspondants dans le domaine de fréquence réelle :

$$f(j) = \frac{N}{F_s} B^{-1} B(j) \qquad (6)$$

Puis on détermine tous les coefficients de chaque filtre :

$$H_j(k) = \begin{cases} 0 & k \le f(j-1) \\ \dfrac{k - f(j-1)}{f(j) - f(j-1)} & f(j-1) \le k \le f(j) \\ \dfrac{f(j+1) - k}{f(j+1) - f(j)} & f(j) \le k \le f(j+1) \\ 0 & k \ge f(j+1) \end{cases} \qquad (7)$$

L'analyse MFCC comporte plusieurs étapes. Le prétraitement consiste à effectuer sur le signal de parole, échantillonné à 16000 Hz et quantifié sur 16 bits, les opérations suivantes :

• Toutes les 10ms (160 échantillons), une trame acoustique de 25ms (400 échantillons) est extraite du signal.

• La composante continue des échantillons constituant cette trame est enlevée.

• Afin de compenser l'atténuation naturelle du spectre du signal de parole, la séquence des échantillons constituant la trame subit une pré-accentuation avec le filtre du premier ordre

$$H(Z) = 1 - 0{,}97 Z^{-1} \qquad (8)$$

L'analyse MFCC proprement dite consiste à effectuer sur chacune des trames résultantes du prétraitement les opérations suivantes :

- La transformation de Fourier permet de calculer le spectre d'amplitude de la trame.
- Pour chacun des 22 filtres triangulaires répartis sur l'échelle des fréquences de Mel, l'énergie du spectre d'amplitude en sortie de ce filtre est calculée. Cette opération donne un vecteur de 22 valeurs énergétiques *E j*. [28]

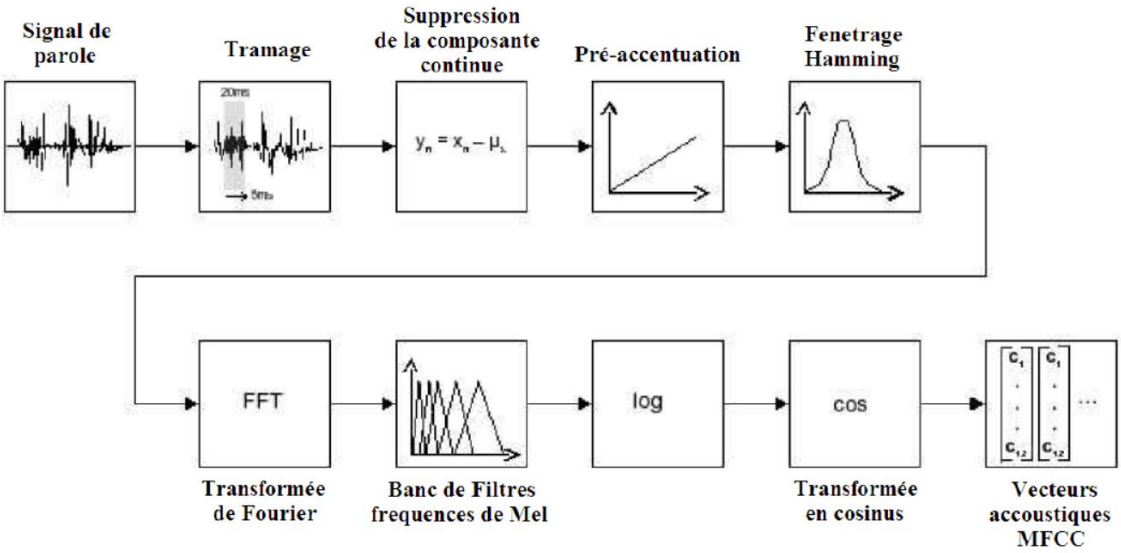


*Figure 7: Schéma en blocs de l'analyse acoustique permettant le calcul des vecteurs MFCC [39].*

- Les logarithmes de ces 22 valeurs sont alors transformés en 12 coefficients MFCC par l'inverse de la transformée en cosinus discrète :

$$c_i = \sqrt{\frac{2}{N}} \sum_{j=1}^{N} \log_{10}(E_j) \cos\left(\frac{\pi i}{N}(j+0{,}5)\right) \qquad (10)$$

Où *ci* est le *i-ème* coefficient Mel-cepstral, *E j* est l'énergie du spectre calculée sur la bande passante du *j-ème* filtre, et *N* est le nombre de filtres ( $N = 22$ ).

- Afin d'augmenter la robustesse de ces coefficients pour le calcul des distances cepstrales, une pondération en sinus (liftering) est appliquée sur les coefficients MFCC *ci* [46] :

$$\hat{c}_i = \left(1 + \frac{L}{2}\sin\frac{i\pi}{L}\right) c_i \qquad 1 \le i \le 12 \qquad (11)$$

Où *cˆi* est le *i*ème coefficient mel-cepstral liftré et *L* est le coefficient du liftering ($L = 22$). Ces pondérations corrigent la décroissance rapide des coefficients MFCC d'indice élevé et permet l'utilisation d'une distance euclidienne.

# III. QUELQUES TRAVAUX QUI ONT DEJA ETE EFFECTUES PAR RAPPORT A CE THEME

Nous avons souligné quatre thèmes de mémoires qui traitaient dans cette technologie dont les thèmes portaient sur :

- **« Extension des commandes vocales pour les malvoyants Interfaçage de Google Home avec Arduino à l'aide Webhooks»** Ce projet est un article réalisé par **Sameer Tuteja**, étudiant du département du génie mécanique, campus technique de l'ACROLE publié dans le journal : **International Journal of Research in Advent Technology, Vol.7, No.4, April 2019** [44].

  Dans ce projet, il était question d'aider les étudiants et les passionnés à créer leur propre IoT appareils pouvant être contrôlé à l'aide de Google Assistant sur leurs smartphones. Ce projet montre que Google Home / Assistant peut-être utilisé pour contrôler les appareils intelligents de bricolage en utilisant des phrases ou commandes. L'inconvénient de ce projet est qu'il utilise les applications sur internet pour pouvoir fonctionner, par conséquent nécessite une très grande connexion internet. L'assistance aux personnes handicapés n'est pas également gérée.

- **« Automatisation IoT basée sur l'intelligence artificielle: contrôle des appareils avec Google et Facebook »** C'est un article réalisé par **ANJAN CHARTTERJEE** Bachelor of Technology in Electrical Engineering JIS College of Engineering publié dans le journal : **International Research Journal of Engineering and Technology (IRJET)** [43].

  Dans cet article, le système proposé est une manière intelligente d'intégrer les appareils électroniques à notre quotidien. L'article implique l'utilisation de l'intelligence artificielle avec l'Internet des objets, mais il montre également une approche du nouveau monde de l'Internet social des objets, où les potentialités des concepts de réseaux sociaux peuvent être fusionnées avec l'IoT afin qu'un appareil physique puisse être intégré directement dans une plateforme sociale. La technologie est bénéfique pour l'établissement et la gestion des relations sociales entre les objets de telle sorte que le réseau social résultant est navigable tout comme les humains sont connectés sur les réseaux sociaux. [1].

La limite de ce projet est qu'il est focalisé sur l'internet des objets connectés via les réseaux sociaux mais ne tient pas en compte de l'interaction homme-machine. Une réelle communication entre une personne et son assistance vocale.

- **« Système interactif d'information auditive pour la mobilité des personnes aveugles dans les transports publics »**, C'est un article réalisé par : G. Baudoin, O. Venard, G. Uzan et A. Paumier, J. Cesbron publié dans le journal **RIGHTS LINK** .

  Ce projet a pour objectif de concevoir, développer et expérimenter un système actif et interactif d'assistance et d'information aux personnes aveugles pour favoriser leur autonomie et leur mobilité dans les transports publics. Ce système est destiné à équiper les points d'arrêt des transports collectifs (bus, tramway) ou à être installé dans un pôle d'échange. Il s'appuie sur des assistants personnels numériques utilisant une synthèse vocale et communiquant par liaison sans fil WIFI avec les bornes fixes installés aux arrêts.

  La limite majeure de ce projet est qu'il est restreint dans le domaine des transports et en cas de vol de son portable ou égarement il sera difficile pour lui de se retrouver.

- **« Acquisition et reconnaissance automatique d'expressions et d'appels vocaux dans un habitat » projet** réalisé par Michel Vacher, Benjamin Lecouteux, Frédéric Aman, François Portet, Solange Rossato publié dans L'archive ouverte pluridisciplinaire **HAL.**

  Ce projet présente un système capable de reconnaître les appels à l'aide de personnes âgées vivant à domicile afin de leur fournir une assistance. Le système utilise une technologie de Reconnaissance Automatique de la Parole (RAP) qui doit fonctionner en conditions de parole distante et avec de la parole expressive. Pour garantir l'intimité, le système s'exécute localement et ne reconnaît que des phrases prédéfinies.

  L'inconvénient de ce projet est que si la personne est malade et a de la grippe sa voix sera transformée d'où le système ne saura plus correctement répondre aux exigences de la personne âgée et lui venir en aide au cas où elle est en difficulté.

## IV. NOTRE APPORT

Partant de ces différents projets des articles suscités, nous est venu l'idée de combler cette insatisfaction chez le public en quête d'automatisation et de confort à petit prix. Nous avons ainsi décidé de réaliser une solution d'assistance vocale pour le particulier qui soit abordable financièrement, tout en alliant robustesse, flexibilité, la sécurité et stabilité, c'est avec ces objectifs en tête, que nous avons décidé de nous lancer dans: « La conception et la réalisation d'un système intelligent d'assistance vocale pour les personnes à mobilité réduite» qui permet la gestion d'une multitude d'équipements dans la maison en développant les pilotes adéquats des cartes, ainsi que les applications permettant le contrôle des équipements qui fonctionnent à base de l'électricité à l'aide des méthodes d'apprentissage profond pour la reconnaissance vocale. Nous nous sommes servi des réseaux de neurones convolutifs pour l'entrainement des données pré-enregistré enfin de reconnaitre les voix et d'exécuter une tâche précise

# CONCLUSION

Malgré les progrès techniques des dispositifs de la maison, ce genre d'installation ne sera à la portée que de quelques privilégiés à cause du prix tout simplement. Il semble donc qu'il faille attendre encore un certain temps avant que les " systèmes d'assistance vocale " soient à la portée de tous. Actuellement, ce sont ces mêmes privilégiés qui profitent des maisons intelligentes et qui peuvent dès à présent « automatiser » l'ensemble de leur maison.

# PARTIE 2: CONCEPTION ET REALISATION DE NOTRE SOLUTION

**<u>Objectif</u>** : **Dans cette partie il sera question de présenter la méthodologie de conception du système embarqué en spécifiant tout d'abord les différentes fonctions et spécifications de notre système et les outils utilisés.**

# CHAPITRE 2 : METHODOLOGIE DE CONCEPTION DU SYSTEME EMBARQUE ET MATERIELS UTILISES

## INTRODUCTION

La mobilité réduite est une situation de handicap due à une diminution des capacités de déplacement dans l'espace public d'une personne, de manière temporaire ou définitive. Cela peut être lié notamment à des déficiences prénatales, des maladies invalidantes, des accidents, ou plus généralement au vieillissement mais aussi à des situations ponctuelles comme c'est le cas pour les femmes enceintes, les personnes sur chaises roulantes. Cette mobilité réduite amène la personne à avoir une autonomie de déplacement limitée ou nulle dans un environnement « ordinaire ». L'amélioration de leurs conditions de vie fait l'objet de notre étude.

Comment les rendre beaucoup plus autonomes et leur rendre la vie facile et confortable ? nous avons donc pour objectif de concevoir un système embarqué dédié à cette cause qui devient primordial car la majorité n'ont souvent personnes au quotidien pour leur rendre service et leur tenir compagnie. Ainsi nous allons à partir d'un système embarqué intelligent :

- **Moderniser leur environnement de vie**, en tenant compte des paramètres humains et sociaux. En associant l'intelligence artificielle à l'électronique moderne.
- **Concevoir un système de reconnaissance vocal**, pour leur venir en aide, ils pourront à l'aide de la voix commander les équipements autour d'eux et à leurs tour ces équipements comprendront leurs langages et exécuteront les tâches demandées.

## I. CAHIER DES CHARGE FONCTIONNEL

Le **cahier des charges fonctionnel** est un document formulant le besoin, au moyen de fonctions détaillant les services rendus par un produit et les contraintes auxquelles il est soumis. Le cahier des charges vise à définir et à faire valider les spécifications d'un produit ou d'un service à réaliser

## I.1 Le concept général et les principaux services attendus

### a) Formulation du besoin

Assurer la sécurité et le confort de vie des personnes à travers un système autonome pour garder ou préserver des vies humaines grâce à un dispositif intelligent, autonome, écologique, simple et design et à moindre coût.

### b) Les clients, utilisateurs et usagers potentiels

- **Les clients**

Les clients peuvent être toutes personnes femmes et hommes. Le client étant celui qui réalise l'acte d'achat, il faut compter parmi les clients les personnes qui offrent le produit.

- **Les utilisateurs et les usagers**

Les **utilisateurs** sont principalement les personnes souffrantes de déficience physique mais ils peuvent aussi être toute personne ayant pour but de rendre son habitat confortable.

## I.2 Contexte du projet et analyse des besoins

Etudes déjà effectuées en amont de l'élaboration du Cahier des Charges Fonctionnel

Avant de constituer le Cahier des Charges Fonctionnel, deux études sont effectuées pour mieux saisir le besoin du client potentiel.

- **Une étude de marché : dégager les fonctions de service**

L'étude de marché permet d'analyser la demande et de répondre au mieux aux attentes du client. Pour satisfaire le client, il faut dégager les fonctions principales répondant au besoin.

Dans notre cas, pour répondre au besoin, il faut travailler sur les fonctions suivantes :

- Permettre la communication l'utilisateur et le système ;
- Exécuter correctement les tâches demandées au système ;
- Prendre en compte l'état psychologique des personnes ;
- Etre simple d'utilisation :
- Etre léger ;
- Etre peu encombrant ;
- Etre esthétique.

- **Etude succincte du marché actuel des systèmes de reconnaissance vocale**

Au Cameroun, il n'existe pas encore un modèle implémenté et fonctionnel d'assistance vocale pour les personnes à mobilité réduite. Nous sommes encore sous l'emprise des anciennes technologies informatique, numérique et électronique. D'où la nécessité d'aider notre pays à s'arrimer à la pointe des nouvelles technologies pour faciliter la vie de ses citoyens.

- **Limite de l'étude**

L'étude effectuée sera limitée à la zone géographique du Cameroun mais le produit sera conforme aux normes françaises et européennes en vigueur.

- **Enoncer le besoin :**

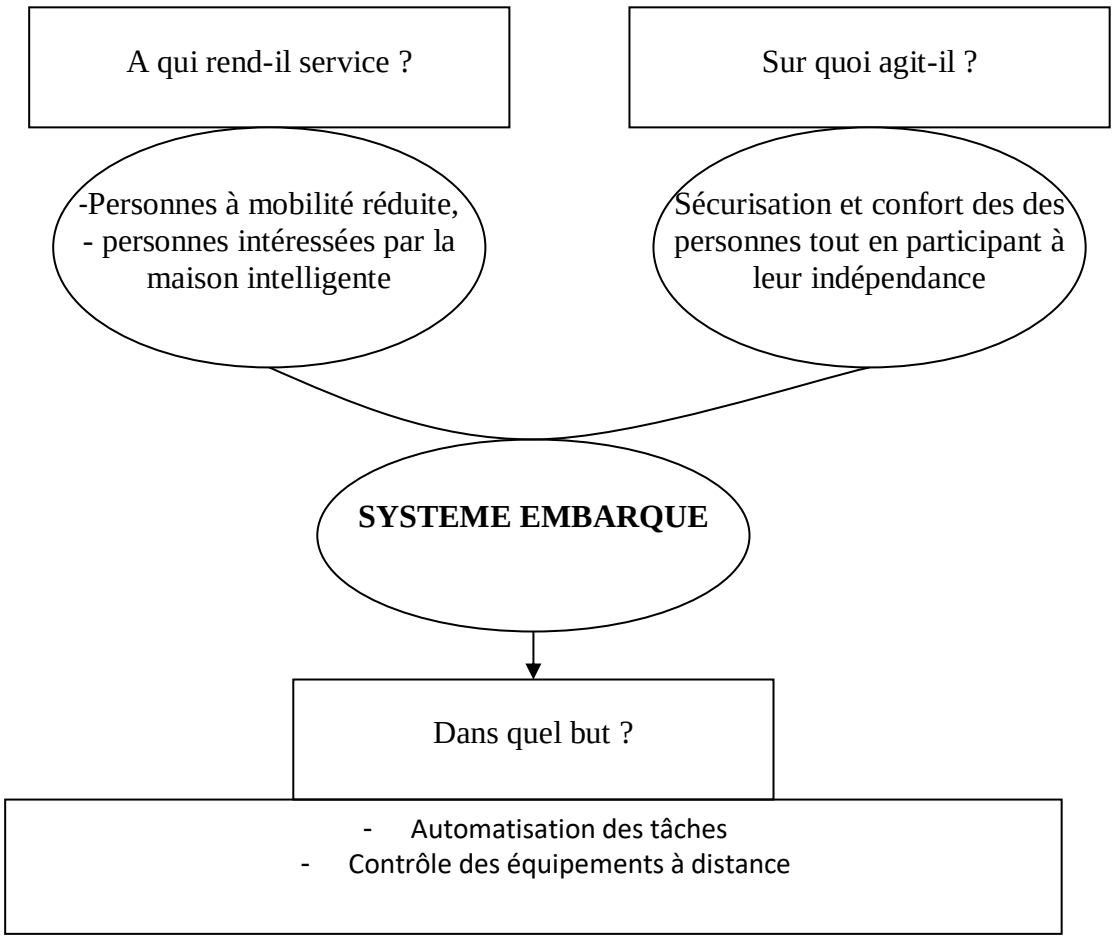


*Figure 8:Diagramme bête à corne*

- **Le besoin peut-il évoluer ?**

Il est impossible d'imaginer qu'un besoin n'évolue pas. En effet, les évolutions technologiques pourraient entraîner des modifications du besoin.

- Par exemple un nouveau système embarqué prenant en compte d'autres paramètres comme la reconnaissance faciale, communication réelle et en plein temps avec son système d'assistance vocale sous forme de robot pouvant se déplacer etc.

- **Le besoin peut-il disparaître ?**

Le besoin peut disparaître si :

- Le système de sécurité sociale des personnes handicapés et âgées soit pris totalement en charge dans notre pays ;
- Les maladies disparaissent (peu probable) ;
- Les accidents de routes disparaissent (peu probable).

Il est donc peu probable que le besoin disparaisse totalement.

# II. METHODES DE CONCEPTION

Comme énoncé à l'introduction générale, l'objectif principal est de concevoir un système embarqué intelligent de reconnaissance vocal multicritères qui a pour objectif spécifique :

- Etablir des communications entre homme-machine ;
- Automatiser les tâches ;
- Contrôler les équipements électriques à l'aide de la voix.

Ce système consiste donc à prédire de façon intelligente ce qu'une personne va dire et reconnaitre la tâche qui lui est allouée. Ceci dit, notre système devra donc être constitué comme suit :

- Une interface homme-machine qui utilisera des réseaux de neurones convolutifs (Deep Learning) pour reconnaitre les mots prononcés et faire l'apprentissage ;

- Un module électronique avec un système à base de réseau du Deep Learning adaptatif aux algorithmes du MFCC : il s'agit d'intégrer du raisonnement humain dans un système électronique en établissant à la sortie des sons vocaux dépendants de l'état psychologique de l'utilisateur. Et aussi de traduire les paroles en texte.

## II.1/ Description de la méthode

Le deep learning est un élément important de la science des données, qui comprend les statistiques et la modélisation prédictive. Il est extrêmement bénéfique pour les scientifiques chargés de collecter, d'analyser et d'interpréter de grandes quantités de données, alors que l'apprentissage en profondeur rend ce processus plus rapide et plus facile.

Nous allons deviner que nous pouvions simplement alimenter des enregistrements sonores dans un réseau neuronal et l'entraîner afin qu'il reconnaisse la tâche qui lui a été confié et l'exécute promptement:

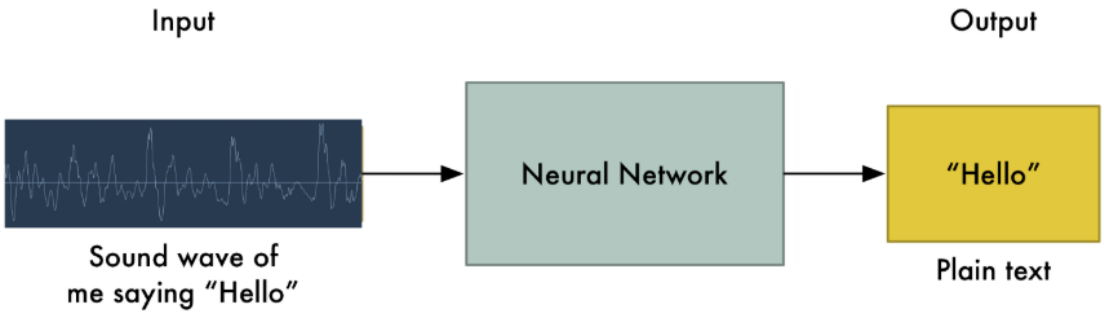


*Figure 9: son appliqué au Deep Learning*

Le gros problème que nous voulons résoudre est la vitesse du discours qui varie. Une personne pourrait dire « Hello » très rapidement et une autre personne pourrait dire « heeeellllllllllllooooo » très lentement, produisant un fichier son beaucoup plus long avec beaucoup plus de données. Les deux fichiers audio doivent être reconnus comme étant exactement le même texte - « hello! ». L'alignement automatique de fichiers audio de différentes longueurs sur un morceau de texte de longueur fixe s'avère assez difficile.

Pour contourner ce problème, nous devons utiliser des astuces spéciales en plus d'un réseau neuronal profond avec plus de précisions. Pour ce fait, nous allons suivre les étapes suivantes :

### a) Transformation des sons en bits

La première étape de la reconnaissance vocale est évidente: nous devons introduire des ondes sonores dans un ordinateur. Mais le son est transmis sous forme d'ondes. Nous allons transformer les ondes sonores en nombres. Utilisons ce clip audio de moi disant "hello":

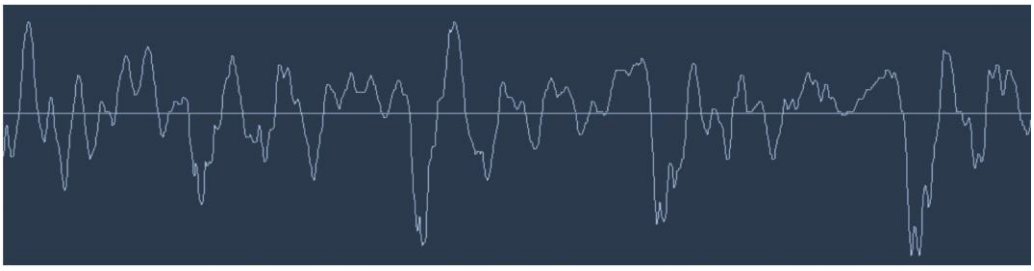

*Figure 10: clip audio sous forme d'onde*

Les ondes sonores sont unidimensionnelles. À chaque instant, ils ont une valeur unique basée sur la hauteur de la vague. Zoomons sur une toute petite partie de l'onde sonore et jetons un œil

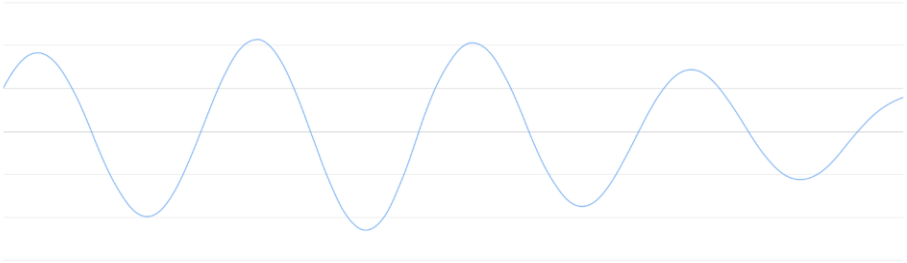

*Figure 11: onde sonore unidirectionnelle*

Pour transformer cette onde sonore en nombres, nous enregistrons simplement la hauteur de l'onde en des points également espacés par échantillonnage:

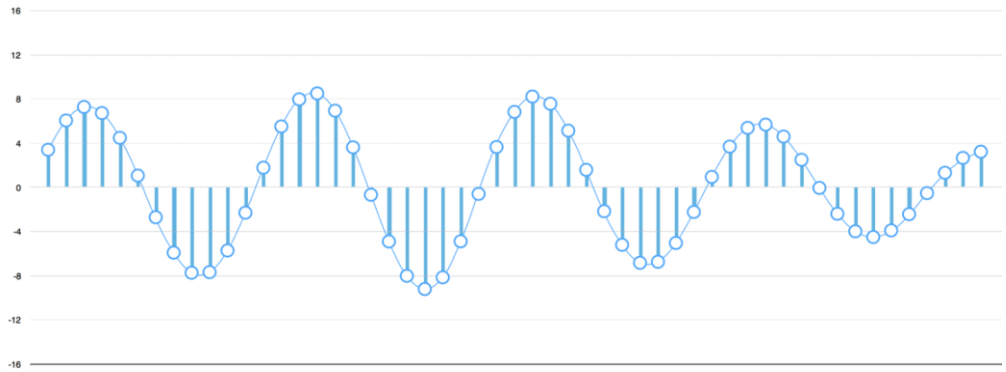


*Figure 12: signal échantillonné*

Nous prenons une lecture des milliers de fois par seconde et enregistrons un nombre représentant la hauteur de l'onde sonore à ce moment-là. C'est essentiellement tout un fichier audio .wav non compressé.

Le son est échantillonné à 44,1 kHz (44 100 lectures par seconde). Mais pour la reconnaissance vocale, un taux d'échantillonnage de 16 kHz (16 000 échantillons par seconde) est suffisant pour couvrir la plage de fréquences de la parole humaine.

Il Permettra d'échantillonner notre onde sonore « hello» 16 000 fois par seconde. Voici les 100 premiers échantillons:

```
[-1274, -1252, -1160, -986, -792, -692, -614, -429, -286, -134, -57, -41, -169, -456, -450, -541, -761, -1067, -1231, -1047, -952, -645, -489, -448
, -397, -212, 193, 114, -17, -110, 128, 261, 198, 390, 461, 772, 948, 1451, 1974, 2624, 3793, 4968, 5939, 6057, 6581, 7302, 7640, 7223, 6119, 5461,
 4820, 4353, 3611, 2740, 2004, 1349, 1178, 1085, 901, 301, -262, -499, -488, -707, -1406, -1997, -2377, -2494, -2605, -2675, -2627, -2500, -2148, -
1648, -970, -364, 13, 260, 494, 788, 1011, 938, 717, 507, 323, 324, 325, 350, 103, -113, 64, 176, 93, -249, -461, -606, -909, -1159, -1307, -1544]
```

*Figure 13: les 100 premiers échantillons*

### b) Analyse de l'échantillonnage numérique

Le processus d'échantillonnage ne crée qu'une approximation approximative de l'onde sonore d'origine, car il ne prend que des lectures occasionnelles. Il y a des écarts entre nos lectures, nous devons donc perdre des données.

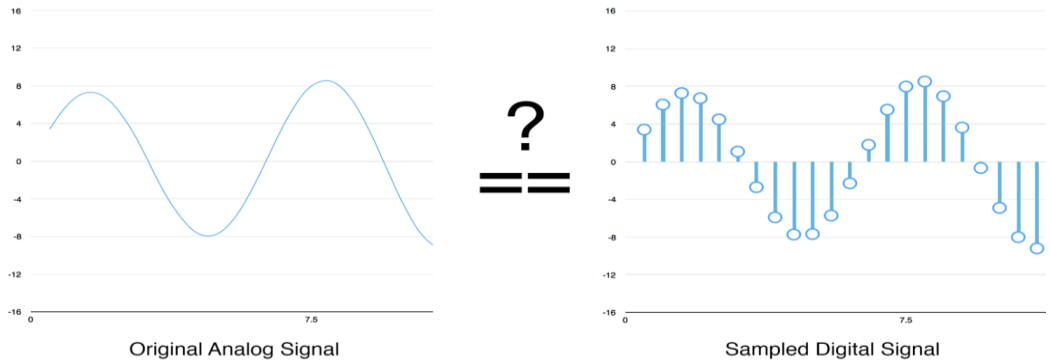


*Figure 14: comparaison du signal original et le signal échantillonné*

Nous nous posons la question de savoir : Les échantillons numériques peuvent-ils parfaitement recréer l'onde sonore analogique d'origine et ses lacunes?

Mais grâce au théorème de Nyquist , nous savons que nous pouvons utiliser les mathématiques pour reconstruire parfaitement l'onde sonore d'origine à partir des échantillons espacés tant que nous échantillonnons au moins deux fois plus vite que la fréquence la plus élevée que nous voulons enregistrer.

### c) Pé-traitement des données sonores échantillonnées

Nous avons maintenant un tableau de nombres, chaque nombre représentant l'amplitude de l'onde sonore à des intervalles de 1/16 000e de seconde.

Nous pourrions introduire ces nombres directement dans un réseau de neurones. Mais il est difficile d'essayer de reconnaître les modèles vocaux en traitant directement ces échantillons. Au lieu de cela, nous pouvons faciliter le problème en effectuant un prétraitement sur les données audio.

Commençons par regrouper notre audio échantillonné en morceaux de 20 millisecondes. Voici nos 20 premières millisecondes d'audio (c'est-à-dire nos 320 premiers échantillons):

```
[-1274, -1252, -1160, -986, -792, -692, -614, -429, -286, -134, -57, -41, -169, -456, -450, -541, -761, -1067, -1231, -1047, -952, -645, -489, -448
, -397, -212, 193, 114, -17, -110, 128, 261, 198, 390, 461, 772, 948, 1451, 1974, 2624, 3793, 4968, 5939, 6057, 6581, 7302, 7640, 7223, 6119, 5461,
 4820, 4353, 3611, 2740, 2004, 1349, 1178, 1085, 901, 301, -262, -499, -488, -707, -1406, -1997, -2377, -2494, -2605, -2675, -2627, -2500, -2148, -
1648, -970, -364, 13, 260, 494, 788, 1011, 938, 717, 507, 323, 324, 325, 350, 103, -113, 64, 176, 93, -249, -461, -606, -909, -1159, -1307, -1544,
-1815, -1725, -1341, -971, -959, -723, -261, 51, 210, 142, 152, -92, -345, -439, -529, -710, -907, -887, -693, -403, -180, -14, -12, 29, 89, -47, -
398, -896, -1262, -1610, -1862, -2021, -2077, -2105, -2023, -1697, -1360, -1150, -1148, -1091, -1013, -1018, -1126, -1255, -1270, -1266, -1174, -10
03, -707, -468, -300, -116, 92, 224, 72, -150, -336, -541, -820, -1178, -1289, -1345, -1385, -1365, -1223, -1004, -839, -734, -481, -396, -580, -52
7, -531, -376, -458, -581, -254, -277, 50, 331, 531, 641, 416, 697, 810, 812, 759, 739, 888, 1008, 1977, 3145, 4219, 4454, 4521, 5691, 6563, 6909,
6117, 5244, 4951, 4462, 4124, 3435, 2671, 1847, 1370, 1591, 1900, 1586, 713, 341, 462, 673, 60, -938, -1664, -2185, -2527, -2967, -3253, -3636, -38
59, -3723, -3134, -2380, -2032, -1831, -1457, -804, -241, -51, -113, -136, -122, -158, -147, -114, -181, -338, -266, 131, 418, 471, 651, 994, 1295,
 1267, 1197, 1291, 1110, 793, 514, 370, 174, -90, -139, 104, 334, 407, 524, 771, 1106, 1087, 878, 703, 591, 471, 91, -199, -357, -454, -561, -605,
-552, -512, -575, -669, -672, -763, -1022, -1435, -1791, -1999, -2242, -2563, -2853, -2893, -2740, -2625, -2556, -2385, -2138, -1936, -1803, -1649,
 -1495, -1460, -1446, -1345, -1177, -1088, -1072, -1003, -856, -719, -621, -585, -613, -634, -638, -636, -683, -819, -946, -1012, -964, -836, -762,
 -788]
```

*Figure 15: audio recueilli après 20ms*

Le fait de tracer ces nombres sous forme de graphique linéaire simple nous donne une approximation approximative de l'onde sonore d'origine pour cette période de 20 millisecondes:

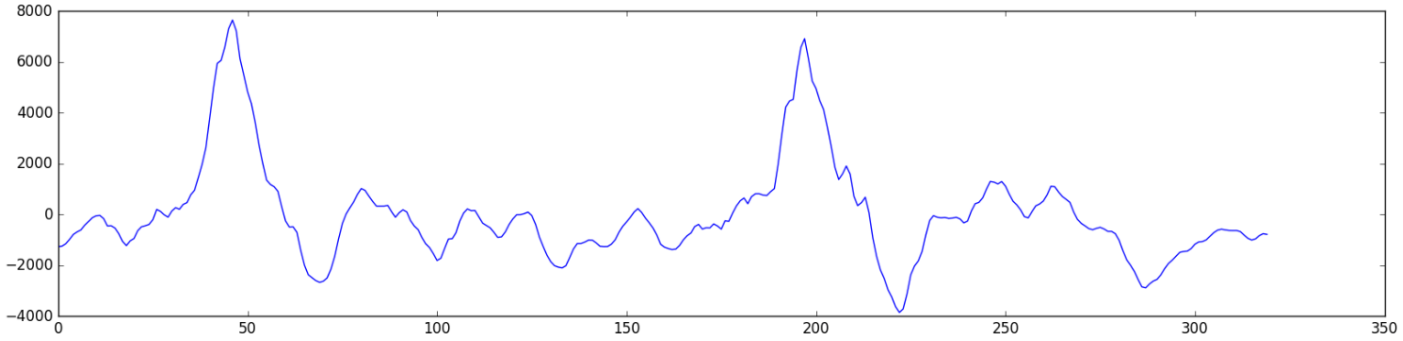


*Figure 16:signal approximatif de l'audio après 20ms*

Cet enregistrement est seulement 1/50e de seconde longue. Mais même ce court enregistrement est un méli-mélo complexe de différentes fréquences sonores. Il y a des sons

graves, des sons de milieu de gamme et même des sons aigus arrosés. Mais prit ensemble, ces différentes fréquences se mélangent pour former le son complexe de la parole humaine.

Pour faciliter le traitement de ces données par un réseau de neurones, nous allons séparer cette onde sonore complexe en ses composants. Nous décomposerons les parties graves, les prochaines parties les plus basses, etc. Ensuite, en additionnant la quantité d'énergie dans chacune de ces bandes de fréquences (de faible à élever), nous créons une sorte d'empreinte *digitale* pour cet extrait audio.

Pour ce faire, nous utilisons une opération mathématique appelée *transformée de Fourier* . Il brise l'onde sonore complexe en ondes sonores simples qui la composent. Une fois que nous avons ces ondes sonores individuelles, nous additionnons la quantité d'énergie contenue dans chacune.

Le résultat final est un score de l'importance de chaque plage de fréquences, depuis la hauteur basse. Chaque nombre ci-dessous représente la quantité d'énergie dans chaque bande de 50 Hz de notre clip audio de 20 millisecondes:

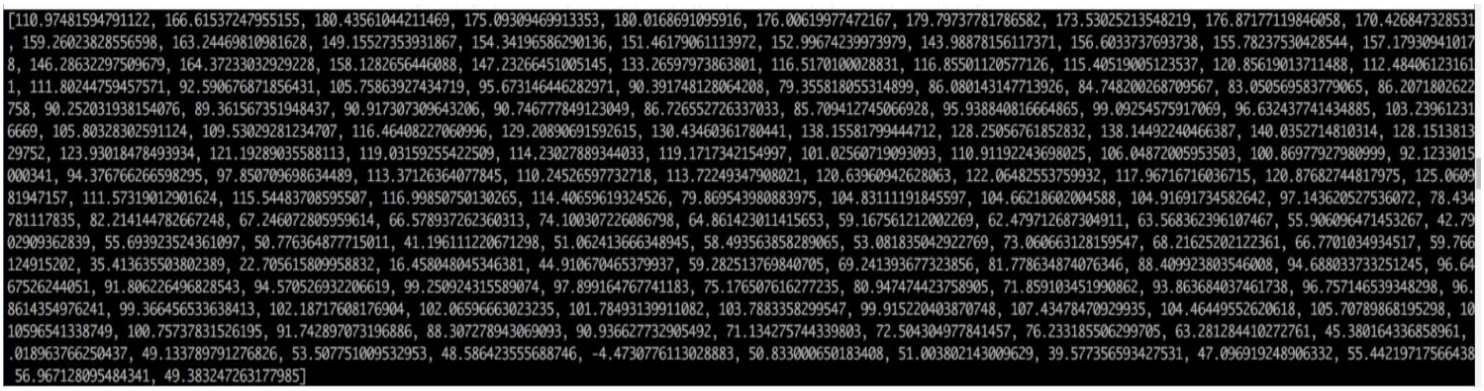


*Figure 17: quantité d'énergie dans une bande de fréquence de 50HZ*

Mais c'est beaucoup plus facile à voir lorsqu'on dessine cela sous forme de graphique:

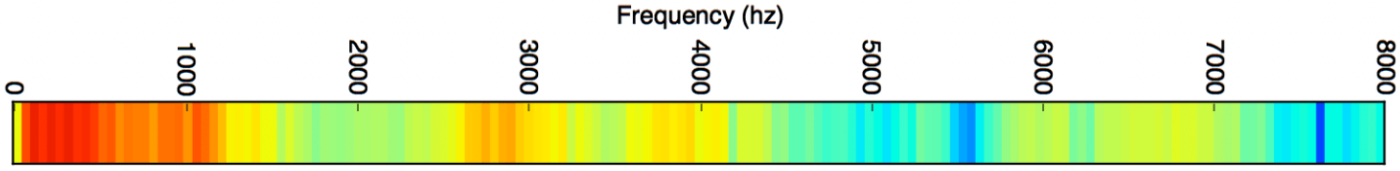


*Figure 18: forme graphique de la quantité d'énergie*

Nous pouvons voir que notre extrait de son de 20 millisecondes a beaucoup d'énergie à basse fréquence et pas beaucoup d'énergie dans les fréquences plus élevées. C'est typique des voix « masculines ».

Si nous répétons ce processus sur chaque bloc audio de 20 millisecondes, nous nous retrouvons avec un spectrogramme (chaque colonne de gauche à droite est un bloc de 20 ms):

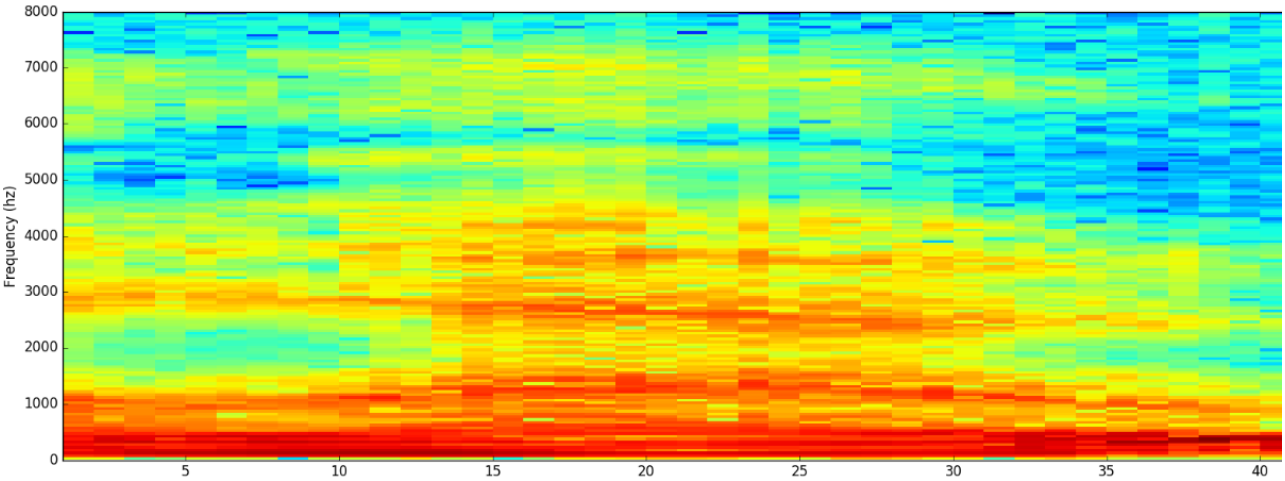


*Figure 19: Le spectrogramme complet du clip audio «hello»*

### d) Reconnaissance des caractères des sons courts

Maintenant que nous avons notre audio dans un format facile à traiter, nous allons l'intégrer dans un réseau neuronal profond. L'entrée du réseau neuronal sera constituée de morceaux audio de 20 millisecondes. Pour chaque petite tranche audio, il essaiera de comprendre la lettre qui correspond au son en cours de prononciation.

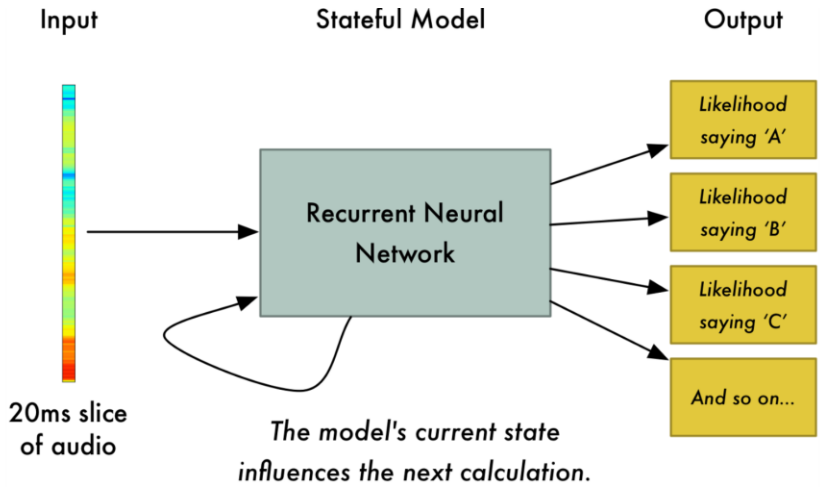


*Figure 20: reconnaissance des caractères des sons*

Nous utiliserons un réseau neuronal convolutif c'est-à-dire un réseau neuronal qui a une mémoire qui influence les prédictions futures. C'est parce que chaque lettre qu'il prédit devrait également affecter la probabilité de la prochaine lettre qu'il prédira. Par exemple, si nous avons

dit « HEL » jusqu'à présent, il est très probable que nous dirons « LO ». Donc, avoir cette mémoire des prédictions précédentes aide le réseau neuronal à faire des prédictions plus précises à l'avenir.

Après avoir exécuté l'intégralité de notre clip audio à travers le réseau de neurones (un bloc à la fois), nous nous retrouverons avec un mappage de chaque bloc audio avec les lettres les plus probablement prononcées au cours de ce bloc. Voici à quoi ressemble ce mappage pour moi en disant « HELLO » :

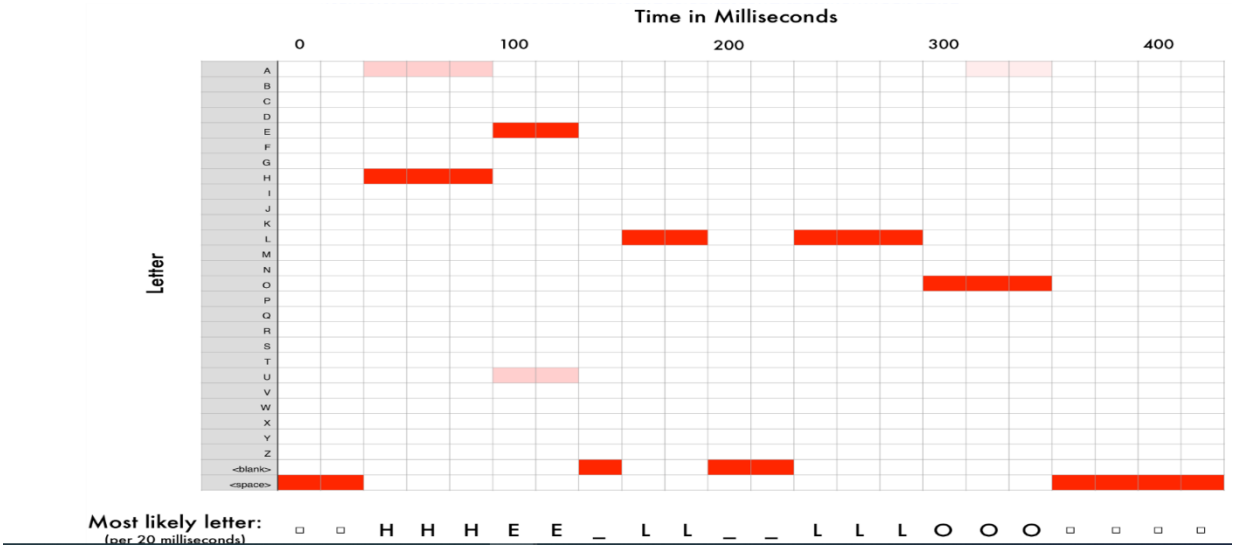


*Figure 21: mappage des bocs audio avec des lettres*

Notre réseau neuronal prédit qu'une chose probable que l'on ait dite était « HHHEE_LL_LLLOOO ». Mais il pense aussi qu'il était possible qu'on dise « HHHUU_LL_LLLOOO » ou même « AAAUU_LL_LLLOOO ». Nous avons quelques étapes à suivre pour nettoyer cette sortie. Tout d'abord, nous remplacerons tous les caractères répétés par un seul caractère:

- HHHEE_LL_LLLOOO devient HE_L_LO
- HHHUU_LL_LLLOOO devient HU_L_LO
- AAAUU_LL_LLLOOO devient AU_L_LO

Ensuite, nous supprimerons tous les blancs:

- HE_L_LO devient HELLO
- HU_L_LO devient HULLO
- AU_L_LO devient AULLO

Cela nous laisse avec trois transcriptions possibles – « HELLO », « Hullo » et « Aullo ». Si nous les disons à voix haute, tous ces sons ressemblent à « Hello ». Parce qu'il prédit un personnage à la fois, le réseau de neurones proposera ces transcriptions très saines.

L'astuce consiste à combiner ces prédictions basées sur la prononciation avec des scores de vraisemblance basés sur une grande base de données de textes écrits (livres, articles de presse, etc.). Nous jetons les transcriptions qui semblent les moins susceptibles d'être réelles et conservons la transcription qui semble la plus réaliste.

De nos transcriptions possibles « HELLO », « Hullo » et « Aullo », évidemment « Hello » apparaîtra plus fréquemment dans une base de données de texte (sans parler de nos données de formation audio originales) et est donc probablement correcte. Nous choisirons donc «Hello» comme transcription finale au lieu des autres.

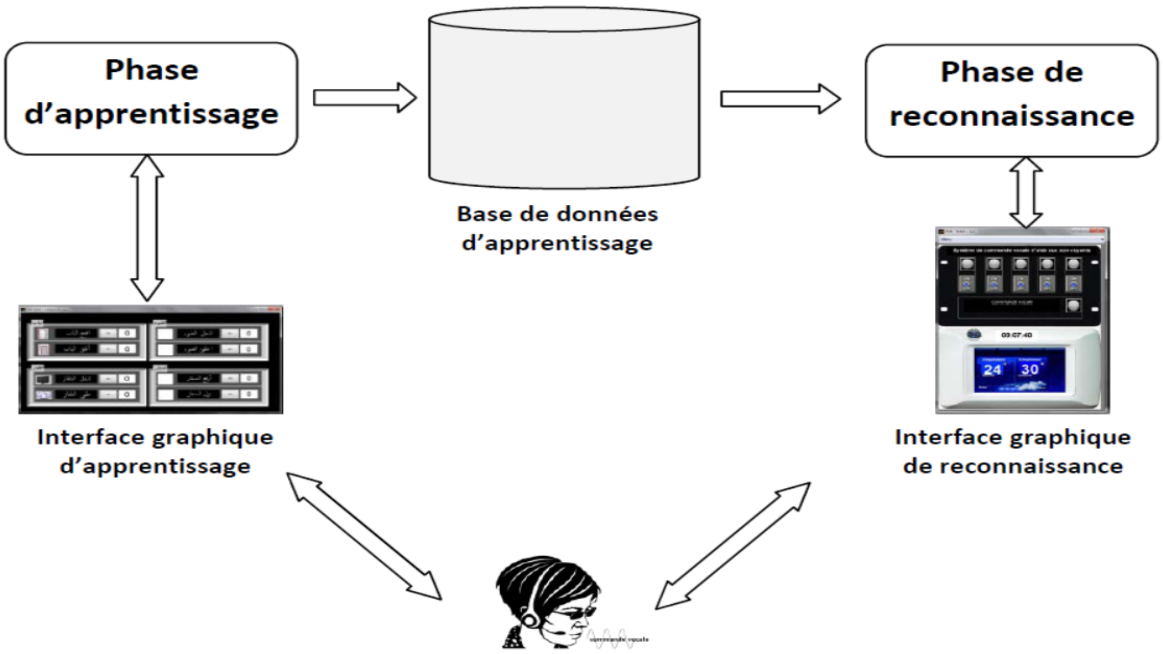


*Figure 22: synoptique du système d'apprentissage*

## II.2/ Synoptique général du système

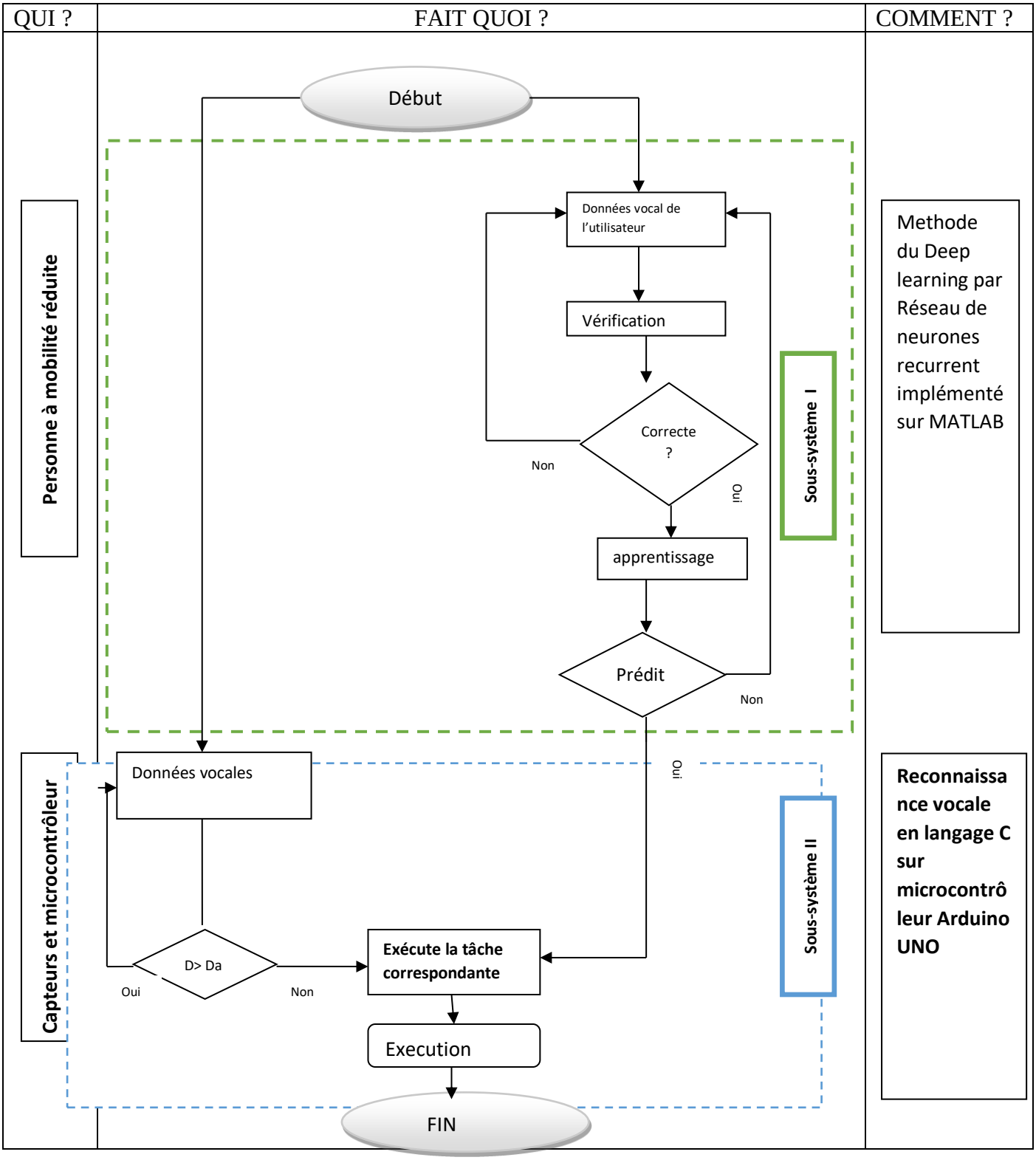


*Figure 23: Synoptique général*

### II.3/ Organigrammes associes des différents capteurs fars du système embarqué

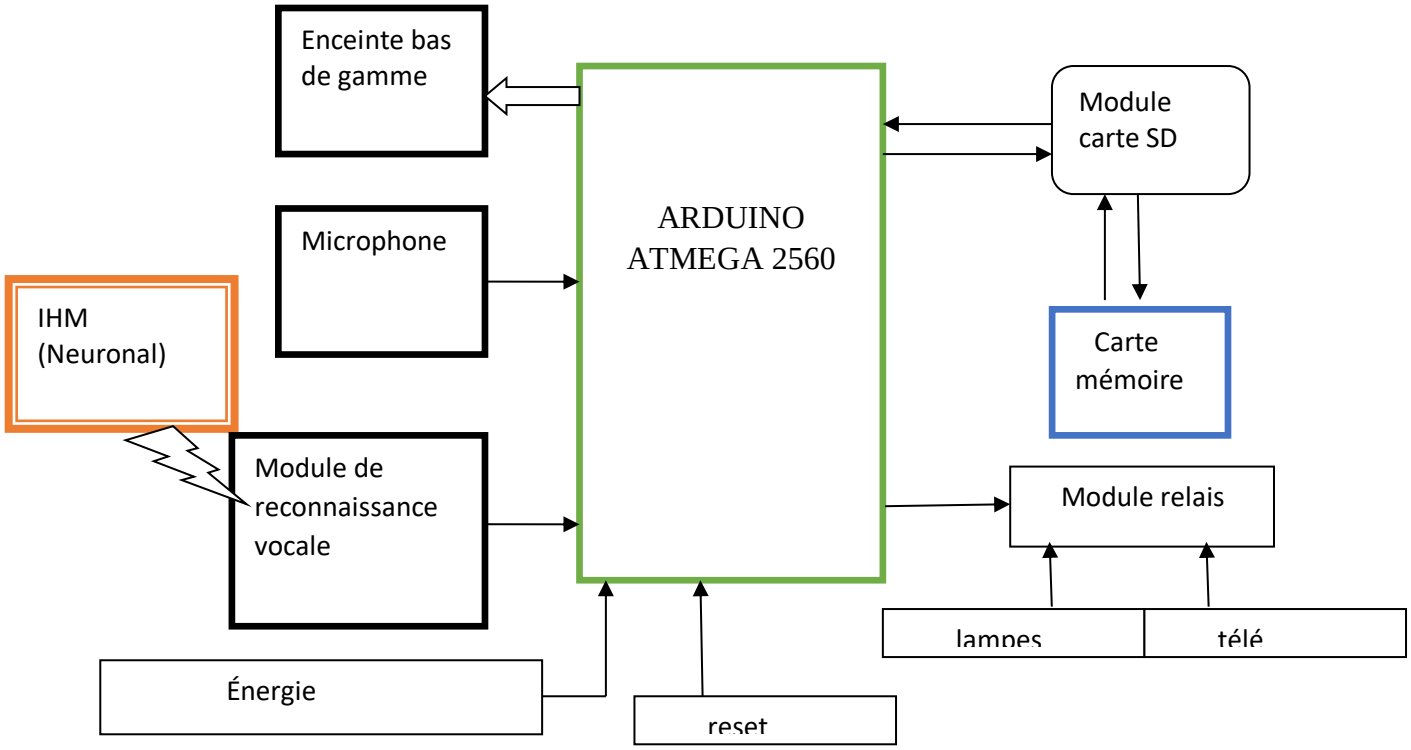


*Figure 24: Diagramme bloc du système embarqué*

En conclusion, il était question de donner la méthodologie de conception de notre système embarqué en passant par la conception du système neuronal après avoir défini les paramètres d'entrée et de sortie du système. Pour mettre en œuvre ce système, il nous faut du matériel adéquat pouvant répondre aux spécifications du cahier de charge.

## III. MATERIELS UTILISES

Pour mener à bien cette étude, il nous a fallu faire une recherche poussée des différents composants électroniques circonstanciels à notre système qui peut être entre autre:

- Le cœur du projet qui est l'Arduino ATMEGA 2560 ;
- Le module de reconnaissance vocal ;
- Le microphone ;

- Enceinte bas de gamme ;
- Quatre relais électromagnétique ;
- Un module de carte SD ;
- Un servo moteur;

## III.1/ Etudes et choix des organes du système vocal

### a) Arduino ATMEGA 2560

L'Arduino Méga est une carte microcontrôleur basé sur l'ATmega1280 . Il dispose de 54 broches numériques d'entrée / sortie (dont 14 peuvent être utilisées comme sorties PWM), 16 entrées analogiques, 4 UART (ports série matériels), une MHz oscillateur en cristal de 16, d'une connexion USB, une prise d'alimentation, d'une embase ICSP et un bouton de réinitialisation.

Il contient tout le nécessaire pour soutenir le microcontrôleur, il suffit de le connecter à un ordinateur avec un câble USB ou avec un adaptateur AC-DC ou batterie pour commencer.


*Figure 25: La carte Arduino MEGA 2560 [24]*

*Tableau 1: caractéristiques de la carte arduino ATMEGA 2560*

| Microcontrôleur | ATmega2560 |
|---|---|
| Tension de fonctionnement | 5V |
| Tension d'entrée (recommandé) | 7-12 V |
| Tension d'entrée (recommandé) | 6-20 V |
| Digital I /O Pins | 54(dont 15fournissent sortie PWM) |
| Broches d'entrée analogiques | 16 |
| DC Courant I /O Pin | 40 Ma |
| Courant DC pour 3.3 Pin | 50 mA |
| Mémoire Flash | 256 Ko (ATmega 2560) dont 8 Kb utilisé par Boot Loader |
| SRAM | 8Kb (ATmega 2560) |
| EEPROM | 4Kb (ATmega 2560) |
| Fréquence d'horloge | 16 MHz |

### b) Relais électromagnétique

C'est un organe électrique permettant de dissocier la partie puissance à la commande. Il permet l'ouverture / fermeture d'un circuit électrique par un second circuit complétement isolé (isolation galvanique) et pouvant avoir des propriétés différentes. Les bobines des relais nécessitent un fort courant pour s'exciter, c'est le rôle de l'ULN2803 qui comporte huit amplificateurs à bases de Darlington. Il est alimenté par une tension continue de 12V (tension de fonctionnement des bobines des relais). *[23]*

Dans le cadre de ce mémoire, nous utiliserons la carte de relais 8 canaux suivant:

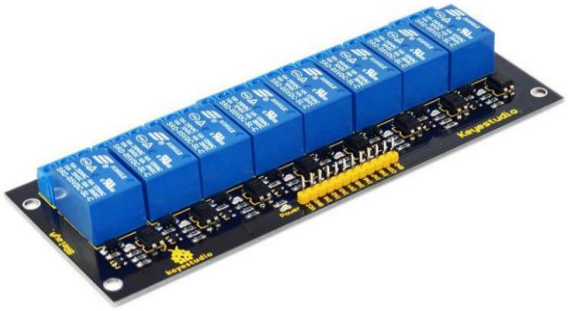

Figure 26: Illustratif d'une carte de relais électromagnétique 8 canaux

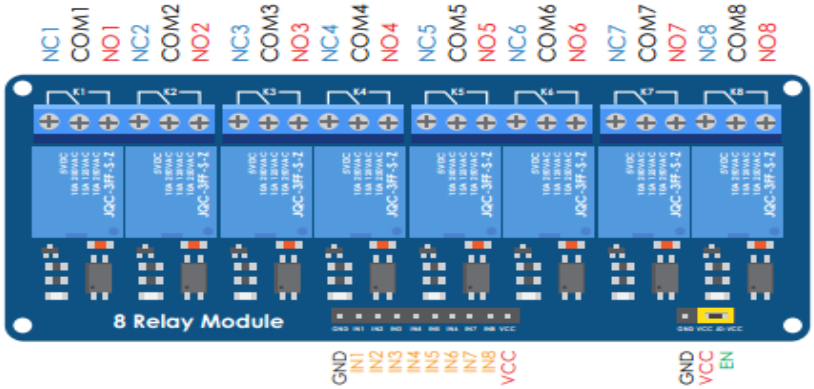


Figure 27: Repérage carte de relais électromagnétique 8 canaux

### c) Module de reconnaissance vocale Geetech

Ce module de reconnaissance vocale Geeetech peut reconnaître jusqu'à 15 instructions vocales et convient à la plupart des cas impliquant le contrôle vocal.
Il reçoit les commandes de configuration ou répond via l'interface de port série. Avec ce module, nous pouvons contrôler les voitures ou d'autres appareils électriques par la voix.

Il peut stocker 15 morceaux d'instruction vocale. Ces 15 pièces sont divisées en 3 groupes, avec 5 dans un groupe. Nous devons d'abord enregistrer les instructions vocales groupe par groupe. Après cela, nous devrions importer un groupe par commande série avant de pouvoir reconnaître les 5 instructions vocales au sein de ce groupe. Si nous devons implémenter des instructions dans d'autres groupes, nous devons d'abord importer le groupe. Ce module est indépendant du locuteur. Si une personne parle l'instruction vocale à la place d'une autre, il peut ne pas identifier l'instruction. Il est à noter cette indépendance de haut-parleur, il exige strictement un microphone.

**Caractéristiques**

- Tension: 4.5 à 5.5 V
- Courant: <40Ma
- Interface numérique: interface UART de niveau TTL 5 V
- Interface analogique: connecteur de microphone mono-canal de 3.5mm + interface de broche de microphone
- Taille: 30mm x 47.5mm
- Précision de la reconnaissance: 99% (dans un environnement idéal)

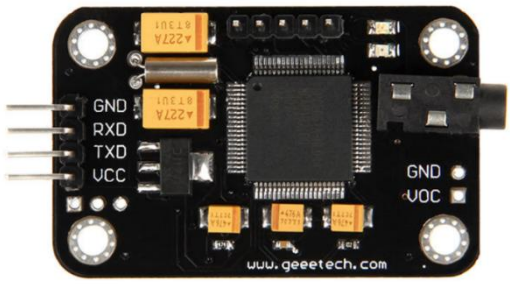


*Figure 28: Module de reconnaissance vocale Geetech [24]*

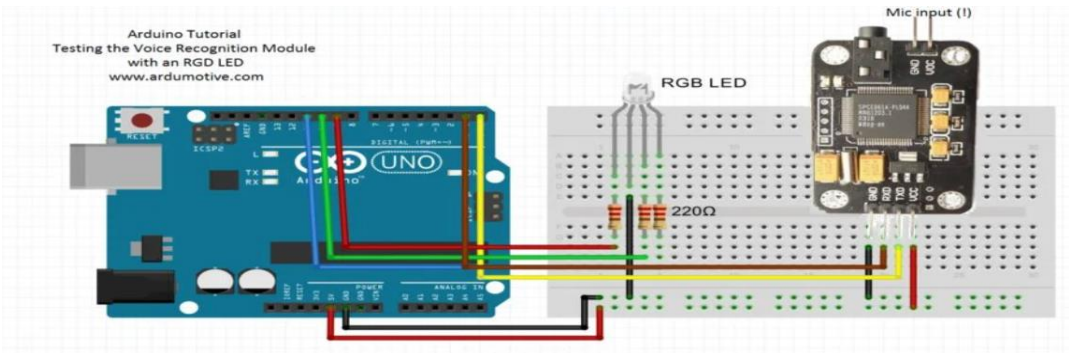


*Figure 29: Branchement avec la carte Arduino UNO[24]*

**d) Microphone**

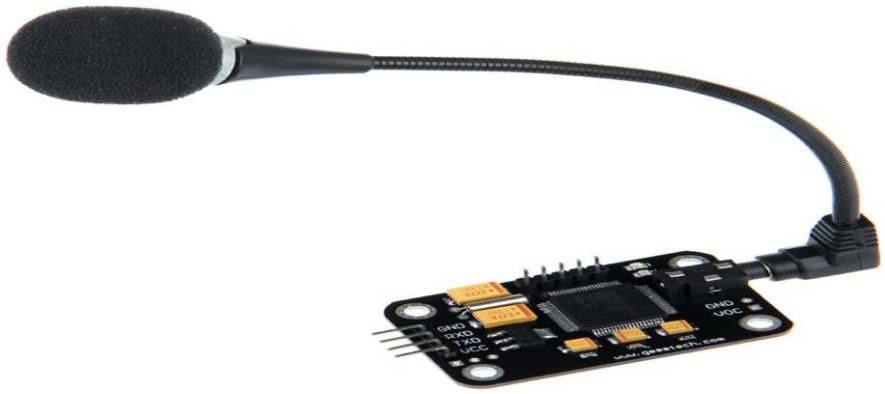


*Figure 30: microphone pour Arduino*

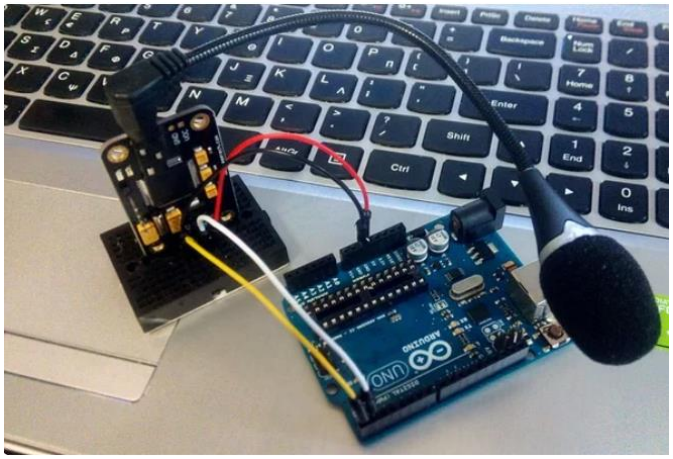

*Figure 31: câblage sur Arduino*

## III.2/ Plateforme de programmation Arduino

L'interface de l'IDE Arduino est plutôt simple, il offre une interface minimale et épurée pour développer un programme sur les cartes Arduino. Il est doté d'un éditeur de code avec coloration syntaxique et d'une barre d'outils rapide. Ce sont les deux éléments les plus importants de l'interface, c'est ceux que l'on utilise le plus souvent. On retrouve aussi une barre de menus, plus classique qui est utilisé pour accéder aux fonctions avancées de l'IDE. Enfin, une console affichant les résultats de la compilation du code source, des opérations sur la carte, etc…



*Figure 32: Vue de l'environnement IDE Arduino*

Le langage Arduino est inspiré de plusieurs langages. On retrouve notamment des similarités avec le C, le C++, le Java et le Processing. Le langage impose une structure particulière typique de l'informatique embarquée.

- La fonction « **setup** » contiendra toutes les opérations nécessaires à la configuration de la carte (directions des entrées sorties, débits de communications série, etc…).
- La fonction « **loop** » elle, est exécutée en boucle après l'exécution de la fonction setup. Elle continuera de boucler tant que la carte n'est pas mise hors tension, redémarrée (par le bouton reset). Cette boucle est absolument nécessaire sur les microcontrôleurs étant donné qu'ils n'ont pas de système d'exploitation. En effet, si l'on omettait cette boucle, à la fin du code produit, il sera impossible de reprendre la main sur la carte Arduino qui exécuterait alors du code aléatoire.

# CONCLUSION

Il était question dans cette partie de concevoir la partie matérielle de notre système embarqué en passant de la mise en place de la méthodologie au dimensionnement du hardware et du software. Dans la suite il sera question de présenter la simulation des réseaux de neurones sous Matlab.

# CHAPITRE3 : RESULTATS ET INTERPRETATION

## INTRODUCTION

Afin de valider nos méthodes d'intelligence artificielle, nous devons passer par une simulation des réseaux de neurones par le logiciel Matlab 2018a. Le choix du logiciel Matlab n'est pas fortuit. Ce logiciel offre une large gamme d'outils appelés toolbox qui permet de simuler et de valider nos méthodes d'intelligence artificielle.

## I. SIMULATION DU SYSTEME NEURONAL

La particularité de notre système embarqué est qu'il prend en compte le genre (masculin ou féminin). Ceci dit, il est question pour nous de montrer comment former un modèle d'apprentissage profond qui détecte la présence de commandes vocales dans le son. Le système utilise le jeu de données des commandes vocales pour entraîner un réseau neuronal convolutif à reconnaître un ensemble donné de commandes et de classifier les différents états selon le mot prononcé. Pour cela, nous allons former un réseau à partir de zéro, nous enregistrons d'abord l'ensemble de données, nous les convertissons au format WAV, enfin nous allons Reconnaître les commandes avec un réseau préformé et détecter les commandes en utilisant le flux audio du microphone.

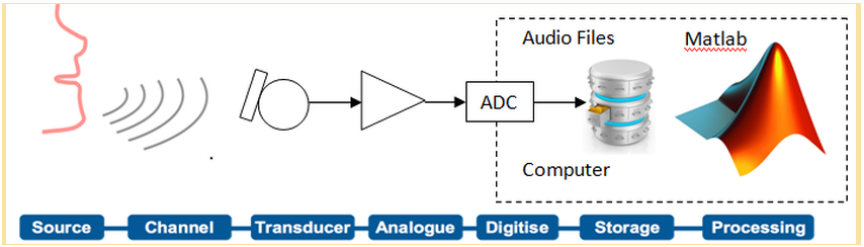


*Figure 33: Schéma fonctionnel d'un traitement de signal audio/verbal typique*

Il est important de noter qu'en utilisant Matlab, nous avons défini une fréquence d'échantillonnage appelée « fs » et une durée appelée Duration. Nous avons eu besoin pour la phase de pré-traitement de données sur Matlab afin de se débarrasser des espaces vides où aucun discours n'a été prononcé, afin d'améliorer les taux prédictions.

## I.1/ Prédiction du temps de réaction et architecture des réseaux de neurones

### a) Pré-traitement de données

Au sein de Matlab se trouve une boîte à outils de réseau neuronal qui offrait de nombreuses fonctions nécessaires pour mettre en œuvre tout type de réseau de neurones.

Dans cette présente étude, nous avons enregistré dix échantillons différents des mots (**go, hello, no, stop, yes, bienvenue**) Cinq échantillons de chaque mot ont été mis de côté pour tester le réseau neuronal après qu'il ait été construit (données d'entrées), les cinq autres échantillons de chaque mot seront utilisés comme données de formation (données de sorties).

### b) Analyse spectrale de la parole à l'aide de la fonction MFCC

Les MFCC (Mel-Frequency Cepstrum Coefficients) sont des fonctionnalités populaires extraites des signaux vocaux pour une utilisation dans les tâches de reconnaissance. Dans le modèle de source-filtre de la parole, les MFCC sont censés représenter le filtre (voie vocale).

Le réseau préformé utilise des spectrogrammes auditifs comme entrée. Nous devons d'abord convertir la forme d'onde de la parole en un spectrogramme auditif. Nous allons donc appliquer les fonctions de MFCC sur chacun de nos échantillons d'enregistrement fournis en entrée, nous allons montrer les étapes avec l'enregistrement du mot bienvenu stocké dans le fichier bienvenue.wav.

- Représentation spectrale du signal

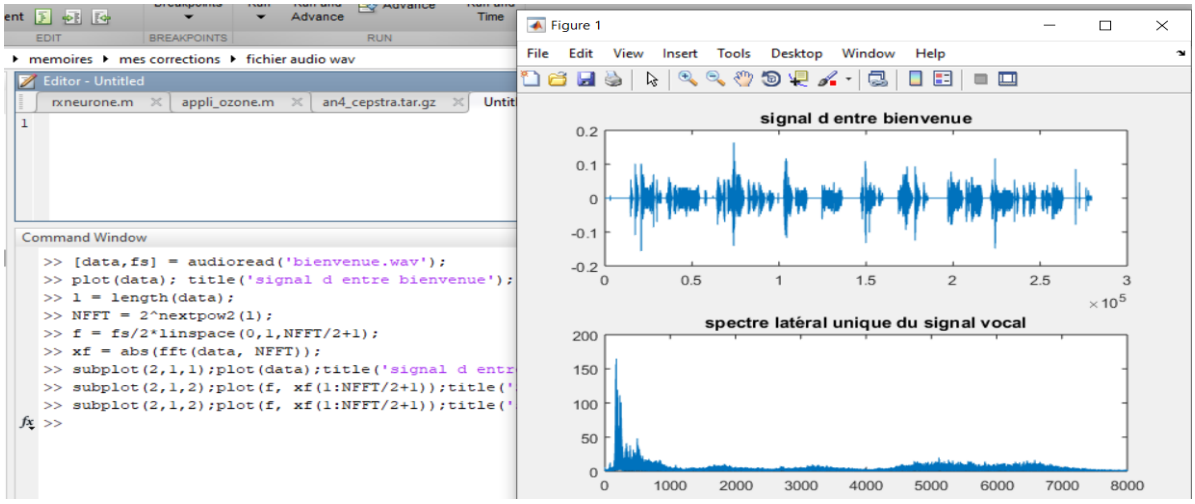


*Figure 34: signal vocal d(origine et transformé de fourrier*

- Diagramme spectral de puissance

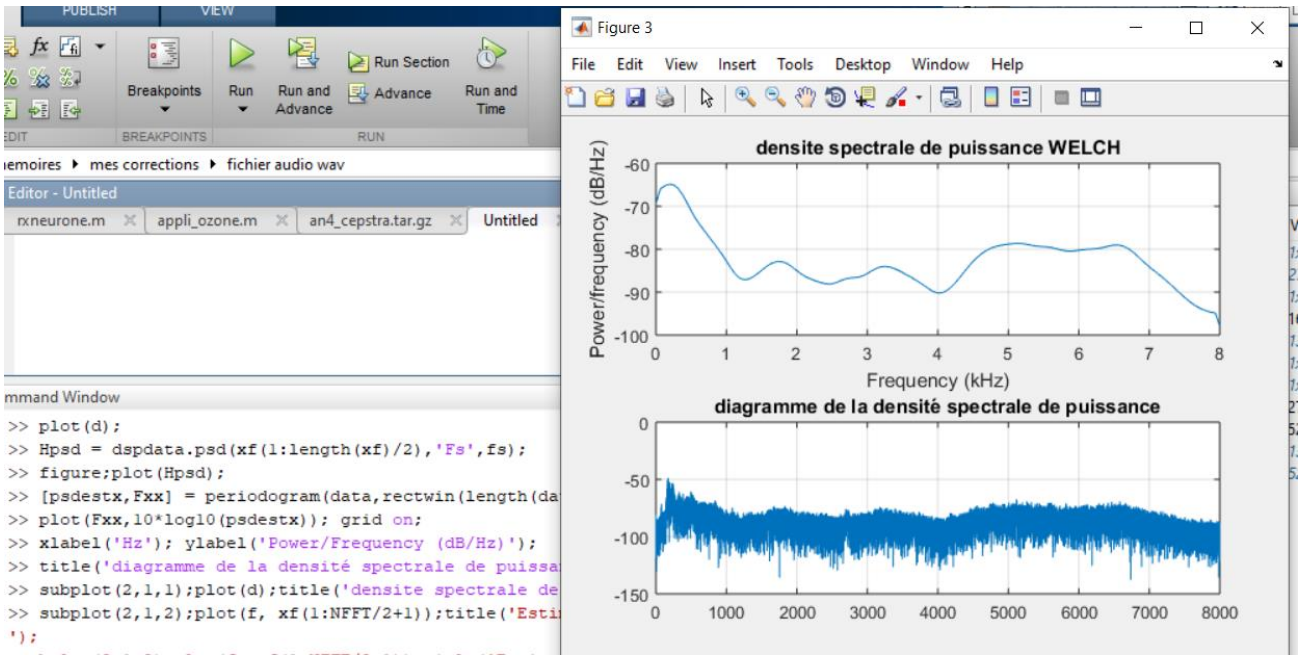


*Figure 35: diagramme spectrale de puissance*

- Cadrage, fenêtrage et pré-accentuation du signal

Cette phase consiste à enlever le bruit du signal

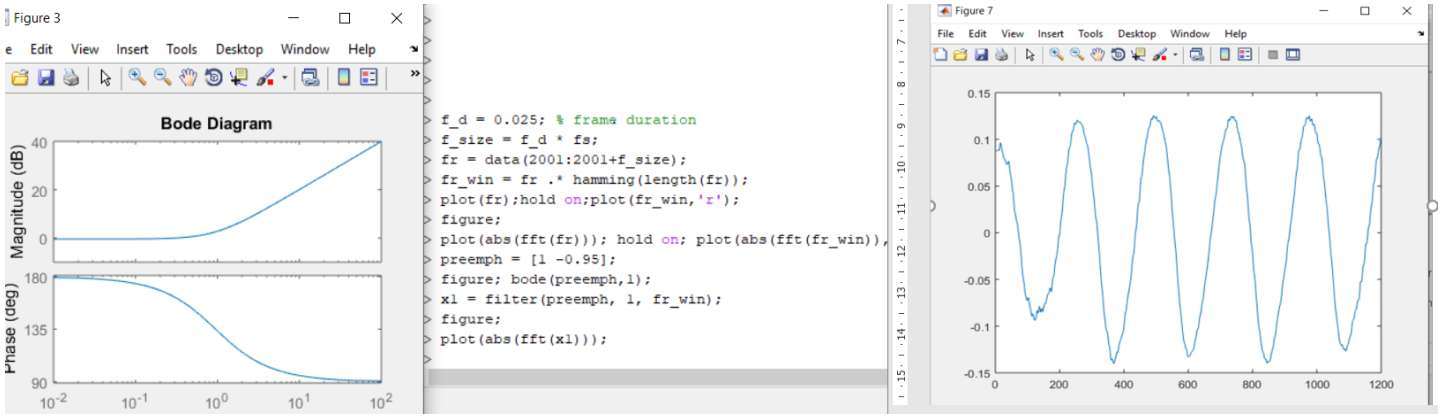


*Figure 36: fenêtrage et pré-accentuation du signal*

- Analyse de la voix/non fracturé/silence et suppression du silence de la voix

Cette phase consiste à enlever les temps morts ou la voix n'existe pas dans notre signal vocal, ceci a l'aide de la fonction d'autocorrélation. La figure 1 ci-dessous recadre le signal, la figure 2 applique sur le signal recadré la fonction d'autocorrélation, et la figure 3 analyse le signal en enlevant le silence

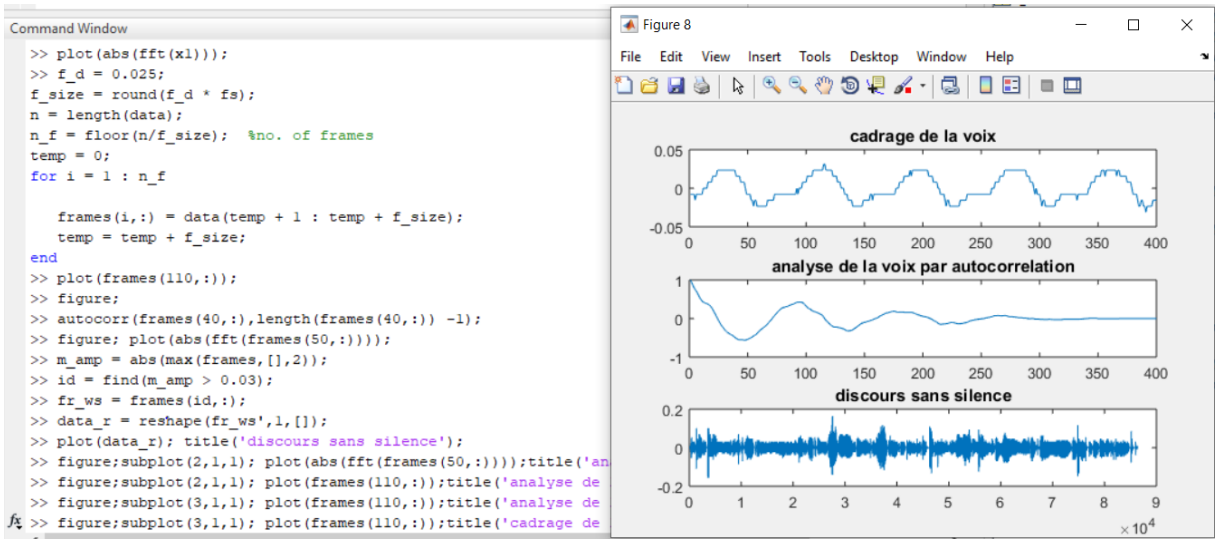


*Figure 37: cadrage, autocorrélation, signal sans silence*

- Rapport énergétique pré-accentué et suppression du silence

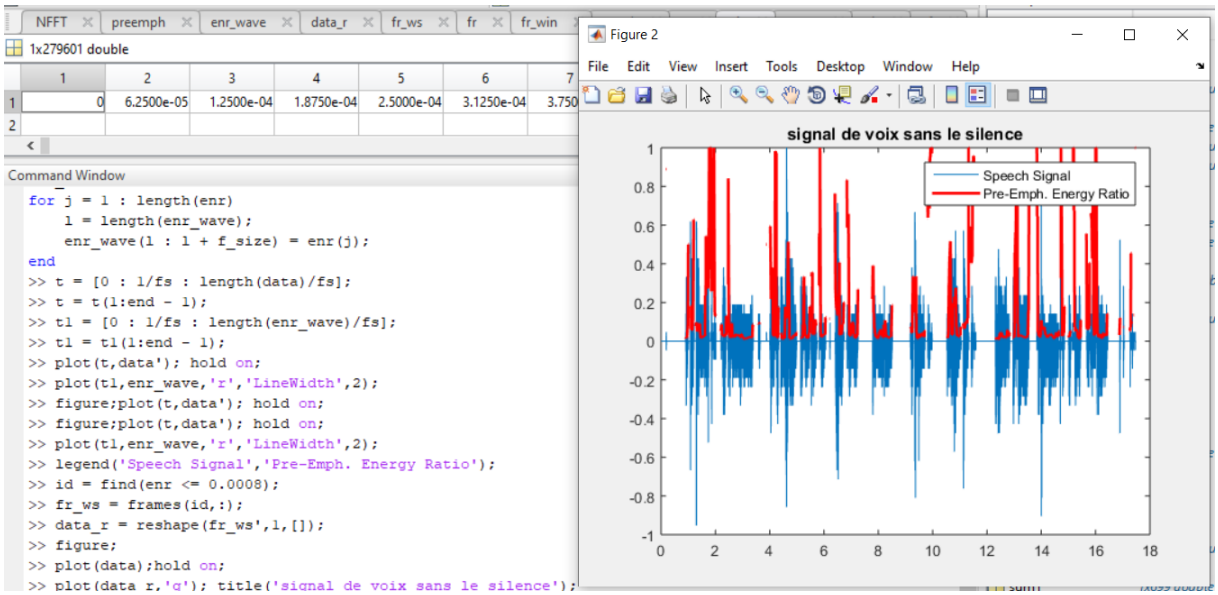


*Figure 38:signal final sans silence*

Nous allons repéter ces étapes d'analyse du signal avec les fonctions MFCC pour chacun de nos enregistrements cités ci-dessus.

### c) Construction du réseau de neurones

La dernière partie de la préparation avant la construction du réseau de neurones a consisté à créer la formation matrice et matrice cible. Nous avons formé une matrice de taille 130 x 25, ce qui correspond à 130 caractéristiques pour 25 mots, en prenant le vecteur MFCC de chaque mot et en les combinant dans une matrice. Ensuite, nous avons fait une matrice cible de taille 5 x 25, soit cinq classes de 25 mots.

Tous les détails de notre réseau de neurones étant décidé, nous avons entré ces données dans Matlab afin de construire le réseau. Nous avons défini les entrées du réseau comme étant la matrice de formation, et les cibles du réseau comme étant les matrice cible. Nous avons une variable appelée hiddenLayerSize et avons attribué aux 90 neurones cachés que nous avons décidé d'utiliser. En utilisant la fonction patternnet, un réseau neuronal de feedforward standard servant à la classification des entrants en fonction des classes cibles a été créée.

```
%    trainMatrix - input data.
%    targetMatrix - target data.

inputs = trainMatrix;
targets = targetMatrix;

% Create a Pattern Recognition Network
hiddenLayerSize = 90;
myNetwork = patternnet(hiddenLayerSize);
```

*Figure 39: creation du reseau de neurone*

Ensuite, nous avons défini la répartition de nos données d'entrées pour la formation, la validation et les tests. Nous avons utilisé une norme division de 70 % des données pour la formation, 15 % des données pour la validation et 15 % des données pour les tests.

```
    evaluate:       outputs = myNetwork(inputs)

>> myNetwork.divideMode='sample';%mode de division de chaque échantillon
>> myNetwork.divideParam.trainRatio=70/100;
>> myNetwork.divideParam.valRatio=15/100;
>> myNetwork.divideParam.testRatio=15/100;%données de test
>>
```

*Figure 40:Mise en place de la division des données*

Nous avons ensuite défini l'algorithme de formation que nous voulons que notre réseau neuronal mette en œuvre, qui est l'algorithme de Levenberg-Marquardt. Nous avons utilisé la fonction d'erreur quadratique moyenne standard pour évaluer la performance du réseau afin de commencé la formation.

```
>> myNetwork.trainFcn='trainlm';% Levenberg Marquardt
>> myNetwork.performFcn='mse';%  la fonction d'erreur quadratique moyenne
>> myNetwork.plotFcns= {'plotperform','plottrainstate','ploterrhist','plotregression','plotfit'};%  lis
>> [myNetwork,tr] = train(myNetwork,x,y); %  entrainnement des données
```

*Figure 41; code d'entrainement des données*

La formation du réseau a pris en moyenne trois minutes à chaque fois que nous avons essayé. Le site a nécessité entre 10 et 20 itérations avant d'être optimisé par l'algorithme de Levenberg-Marquardt .

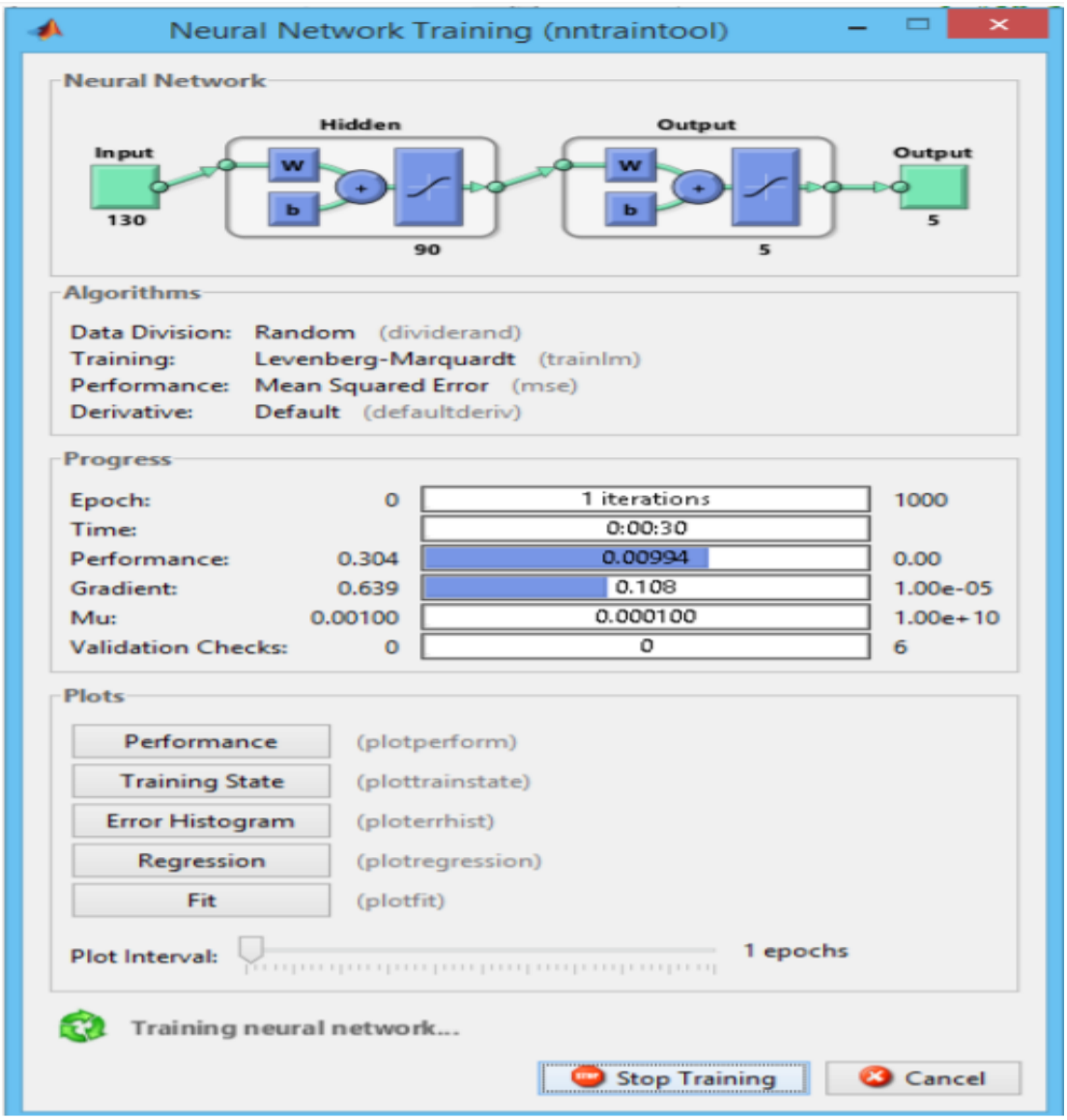


*Figure 42: entrainement du réseau de neurone*

### d) Test du réseau de neurone

Une fois le réseau créé et formé, nous avons écrit une fonction dans Matlab qui permettrait de tester le discours des capacités de reconnaissance.

Cette fonction a utilisé les paramètres testSoundFile et myNetwork pour lire dans un wav et le tester dans le réseau. Après avoir lu le fichier, nous avons extrait les caractéristiques en utilisant l'algorithme des coefficients cepstraux de fréquence propre, puis nous avons simulé le réseau de neurone créé avec l'algorithme du nouveau vecteur d'entrée. En utilisant les if-statements, nous avons produit la classe que le réseau de neurones pensait être le mot entré.

```
function [] = testNetwork( testSoundFile , myNetwork)
%UNTITLED2 Summary of this function goes here
%   Detailed explanation goes here
fileName = testSoundFile;
myNetwork1 = myNetwork;
testSound = wavread(fileName, 2000);
TestSound = mfcc(testSound);
TestSound = TestSound';
TestSound = TestSound(:)';
TestSound = TestSound';
ResultMatrix = sim(myNetwork1, TestSound);
Class1 = round(ResultMatrix(1));
Class2 = round(ResultMatrix(2));
Class3 = round(ResultMatrix(3));
Class4 = round(ResultMatrix(4));
Class5 = round(ResultMatrix(5));

if Class1 == 1
    display('The word is Go');
elseif Class2 == 1
    display('The word is Hello');
elseif Class3 == 1
    display('The word is No');
elseif Class4 == 1
    display('The word is Stop');
elseif Class5 == 1
    display('The word is Yes');
end
end
```

*Figure 43: test de la fonction de neurone*

En utilisant la fonction de test que nous avons écrite dans Matlab, nous avons testé les 5 échantillons de chaque mot dans le réseau de neurones. Ces mots tests n'ont pas été utilisés dans la formation du réseau, et donc pourrait révéler si le réseau a correctement classé les mots.

```
>> testNetwork('go_test1.wav', myNetwork);
The word is Go
>> testNetwork('go_test2.wav', myNetwork);
The word is Go
>> testNetwork('go_test3.wav', myNetwork);
The word is Go
>> testNetwork('go_test4.wav', myNetwork);
The word is Go
>> testNetwork('go_test5.wav', myNetwork);
The word is Go
```

*Figure 44: Résultat de test pour "go"*

```
>> testNetwork('hello_test1.wav', myNetwork);
The word is Hello
>> testNetwork('hello_test2.wav', myNetwork);
The word is Hello
>> testNetwork('hello_test3.wav', myNetwork);
The word is Hello
>> testNetwork('hello_test4.wav', myNetwork);
The word is Hello
>> testNetwork('hello_test5.wav', myNetwork);
The word is Hello
```

*Figure 45: Résultat de test pour "hello"*

```
>> testNetwork('yes_test1.wav', myNetwork);
The word is Yes
>> testNetwork('yes_test2.wav', myNetwork);
The word is Yes
>> testNetwork('yes_test3.wav', myNetwork);
The word is Yes
>> testNetwork('yes_test4.wav', myNetwork);
The word is Yes
>> testNetwork('yes_test5.wav', myNetwork);
The word is Yes
```

*Figure 46:Résultat de test pour "yes"*

```
>> testNetwork('stop_test1.wav', myNetwork);
The word is Stop
>> testNetwork('stop_test2.wav', myNetwork);
The word is Stop
>> testNetwork('stop_test3.wav', myNetwork);
The word is Stop
>> testNetwork('stop_test4.wav', myNetwork);
The word is Stop
>> testNetwork('stop_test5.wav', myNetwork);
The word is Stop
```

*Figure 47: Résultat de test pour "stop"*

## II. Présentation du prototype de l'assistant vocal

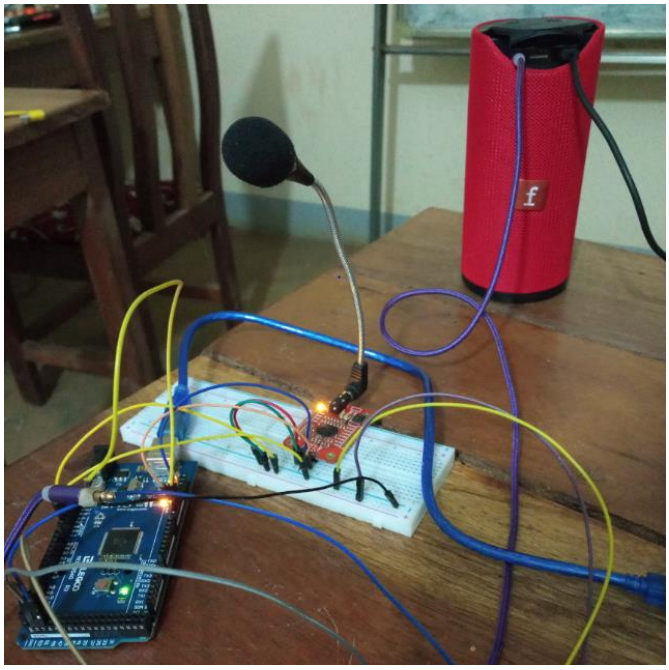

*Figure 48: phase d'apprentissage de notre système*

*Figure 49prototype du système d'assistance vocal*

# III. Interprétation des résultats

## III.1/ Discussion

Lors de l'utilisation d'un réseau perceptron multicouche à action anticipée avec la formation Levenberg-Marquardt pour classer les mots Go, Hello, No, Stop, Yes, et bienvenue en fonction de leurs coefficients de cepstre de fréquence propre, les mots testés ont été reconnus sans erreur. Ces résultats ont été très positifs et démontrent les capacités des réseaux de neurones en matière de modélisation de reconnaissance. Étant donné que le réseau a correctement identifié tous les mots tests qui ont été saisis, les résultats étaient ce que nous espérons réaliser. Cependant, étant donné les limites que nous avons fixé à ce projet, ils peuvent avoir modifié les résultats finaux de manière positive.

Lors de la conception du système de reconnaissance vocale, nous avons limité le vocabulaire à 5 commandes simples qui ont été enregistrées et parlées uniquement par notre voix. Nous avons dû isoler manuellement les mots afin qu'il n'y ait pas de confusion quant au moment où un mot se terminait et où un autre commençait. Avec tous ces limitations, cela réduit la difficulté que le système de reconnaissance vocale doit subir et réduit donc le risque que le réseau fasse une erreur. Une fois les premiers tests terminés, nous avons décidé d'essayer un test de

reconnaissance en temps réel en utilisant Matlab pour recevoir un mot parlé, nous avons automatiquement traité les coefficients de fréquence propre et simulé le réseau de neurones avec eux. Comme nous nous y attendions, le réseau neuronal n'a pas pu prendre une décision complète car il n'y avait pas de classification qui puisse arrondir à un.

Cela pourrait être dû au fait que même si le mot a été enregistré de manière isolé avec peu ou pas de bruit de fond, il y a trop de place autour du mot qui doit être découpé. Le fait d'avoir autant d'espace vide autour du mot a modifié les coefficients et finalement a fait que le mot est passé inaperçu.

## III.2 Difficultés rencontrées

Lors de la mise en place de notre système de reconnaissance vocale, nous avons rencontré quelques difficultés :

- Lorsque nous avons commencé à concevoir le système de reconnaissance vocale, il nous a semblé être un défi de taille de la tâche en raison de tous les éléments fondamentaux que nous avons détaillé. Afin de réduire la complexité du système et nous assurer que nous étions en mesure de le compléter, nous avons réduit le système pour reconnaître cinq des mots isolés et se concentrer sur l'audio pré-enregistré au lieu de la reconnaissance en temps réel.
- Lors de l'enregistrement des échantillons audio pour former et tester le réseau neuronal, nous avons utilisé à l'origine un taux d'échantillonnage de 8000. Nous avons essayé d'utiliser un taux élevé afin de récupérer suffisamment de coefficients pour effectuer une reconnaissance efficace. Toutefois, lors de la formation du réseau neuronal, le nombre de prédictions par mot était beaucoup plus élevé et a fait tomber Matlab dans le "Out of Memory". Nous avons donc réduit le taux d'échantillonnage de nos fichiers wav ce qui a permis au réseau de se former dans un délai acceptable et faire une bonne reconnaissance.

# CONCLUSION ET PERSPECTIVES

Notre travail présenté porte sur est la conception et la réalisation d'un système intelligent d'assistance vocal pour les personnes à mobilité réduite, afin d'apporter une solution aux problèmes que ces personnes rencontrent au quotidien.

Dans un premier temps, nous avons présenté une nouvelle approche pour la prise de décision adaptative a l'habitat grâce à l'intelligent artificielle. Après avoir exposé l'état actuel des recherches dans ce domaine, nous avons détaillé le fonctionnement de notre approche reposant sur l'apprentissage supervisé. Une grande partie de notre mémoire a été consacrée à la compréhension des mécanismes sous-jacents à une telle méthode d'apprentissage et à son utilisation dans le cadre d'un habitat intelligent. Après la consultation d'une bibliographie liée aux réseaux de neurones convolutifs et à l'apprentissage profond et des coefficients de spectres ceptraux MFCC, nous avons utilisé MATLAB 2018a pour modéliser et entrainer notre système à reconnaitre les mots et les voix. Nous avons par la suite utilisé le microcontrôleur Arduino ATMéga 2560 et un module de reconnaissance vocal, pour l'entrainement des données et la réalisation en temps réel de notre système.

Selon les résultats obtenus, on peut conclure que le système Réseau de Neurone convolutif fait apparaître un bon compromis entre la caractérisation et l'efficacité des calculs. Sa robustesse, sa rapidité et la précision de ses sorties lui permettent de donner des décisions correctes et d'éviter les cas d'indécisions, par le biais de l'apprentissage.

À l'avenir, nous aimerions mettre en œuvre un algorithme capable de reconnaitre la voix d'une personne et ensuite reconnaitre la personne exacte à qui appartient ladite voix. Grâce à des recherches plus approfondies sur l'analyse audio, nous espérons améliorer les fonctionnalités que nous avons essayé de reconnaître, de manière à pouvoir résoudre des problèmes plus difficiles comme la reconnaissance des phonèmes pouvant altérer la voix d'un individu, c'est-à-dire que le système doit être capable de reconnaitre lorsqu'une personne est contente, anxieuse, fâchée, angoissée ou en danger juste à l'aide de la voix prise sur des conditions sus énoncées et d'effectuer une tâche précise par rapport à cela : comme par exemple appeler la police ou alerter un proche automatiquement au cas où la voix de la personne lui signalerait que celle-ci est en danger.

# REFERENCES BIBLIOGRAPHIQUES

**[1] Rihana Jamadar, Eram Malim, Shaikh Aamir, Ansari Abdulhai,** « Internet of Things Based Home Automation » , *International Journal of Science & Engineering Development Research* (*IJSDR*) vol .2, no. 4, April 2017.

**[2] A. Edip**, « Siri, Alexa, Google Home, M… Les assistants numériques, révolution ou gadget dangereux ? », 2017.

**[3] M. Fréjus,** « *Construire la coopération Homme-IA*. Document interne. Saclay : EDF R&D », 2017.

**[4] H. Guillaud**, « Vers une éthique pour l'intelligence artificielle ? », 2016.

**[5] F. Lefevre**, « Interaction vocale & intelligences artificielles : état des lieux et opportunités », 2017.

**[6] M. Luria, G. Hoffman & O. Zuckerman, Comparing Social Robot, «** Screen and Voice Interfaces for Smart-Home Control ». In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (CHI '17). ACM, New York, NY, USA, pp. 580-628, 2017.

**[10] W. Minker & F. Néel**, « Développement des technologies vocales. Le travail humain », vol. 65, no. 3, pp. 261-287, 2002.

**[11] C. Nass & S. Brave**, « Wired for Speech – How Voice Activates and Advances the Human–Computer Relationship. *MIT Press »*, 2007.

**[12] P. Pestanes & B. Gautier**, « Essor des assistants vocaux intelligents : nouveau gadget pour votre salon ou fenêtre d'opportunité pour rebattre les cartes de l'économie du web ? », *Wavestone*, 2017.

**[13] F. Portet, M. Vacher & S. Rossato**, « Les technologies de la parole et du TALN pour l'assistance à domicile des personnes âgées : un rapide tour d'horizon (Quick tour of NLP and speech technologies for ambient assisted living) [in English] », *Atelier ILADI 2012: Interactions Langagières pour personnes Agées Dans les habitats Intelligents*, Jun 2012, Grenoble, France. pp.3-16, 2012.

**[14] Portolan, N., Nael, M., Renoullin, JL & Naudin**, « Will we speak to our TV remote control in the future? In: *Proceedings of the 17th international symposium on human factors in telecommunication »*, 1999, Copenhagen.

**[15] H. Soronen, M. Turunen & J. Hakulinen**, « Voice Commands in Home Environment - a Consumer Survey. In *Proceedings of Interspeech 2008 »*, pp. 2078-2081, 2008.

**[16] K. Whitenton**, « Voice Interaction UX: Brave New World...Same Old Story », 2016. *Nielsen Norman Group.com*,https://www.nngroup.com/articles/voice-interaction-ux/ View publication

**[17] M.C. Amara Korba, D. Messadeg, R. Djemili, H. Bourouba**. « Robust Speech Recognition Using Perceptual Wavelet Denoising and Mel-frequency Product Spectrum Cepstral Coefficient Features », Informatica Journal, Vol. 32, No 3, pp. 283-288, 2008.

**[18] J.-P. Haton, J.-M. Pierrel, G. Pérennou, J. Caelen et J.-L. Gauvain**. « Reconnaissance automatique de la parole », 239 p, Collection AFCET – Dunod informatique, Dunod, 1991.

**[19] Y. lecun,** "Une procédure d'apprentissage pour réseaux à seuil asymétrique". In proc. Cognitiva, pp. 599-604, 1985.

**[20] M. Metahri Mohammed el Habib et Mlle Abdelli Sela** *« Smart House». Grade de Master 2 ; Université ABOU* BEKR BELKAID ; 2017.

**[21] Krama Abdebasset et Gougui Abdemoumen**, *« Etude et réalisation d'une carte de contrôle par Arduino via le système Androïde».* Grade de Master 2; UNIVERSITE KASDI MERBAH OUARGLA*;* 08 juin 2015.

**[22] BENTABET Abdel Hamid**, *« Conception et réalisation d'un protocole domotique» ;* Grade de master 2; Université ABOU BEKR BELKAID ; 2012.

**[23] Hamid Hamouchi**, *« Conception et réalisation d'une centrale embarquée de la domotique: Smart Home».* Grade de master 2 ; Université MOHAMMED V ; 6 Juillet 2015.

**[24] TOM Babette**, *« Implémentation d'un système de contrôle domotique».* Grade de master 2 ; Université de Liège; 2012-2013.

**[25] J.W. Picone**, "Signal modeling techniques in speech recognition", Proc. IEEE, Vol. 81, No. 9, pp. 1215-1247, 1993.

**[26] Amana, Mbita ; Menye ; Nsi** *« Commande à distance de l'éclairage et des prises dans une maison par Smartphone: Smart House».* Grade de DIPET I ; Université de Douala; 6 Juin 2016.

**[27 ]BASTIEN Maurice,** *« Reconnaissance vocale de mots clés».* Deep Learning ; Octobre 2008.

**[28] C. S. Myers et L. R. Rabiner,** « Connected digit recognition using a level building DTW algorithm ». IEEE Transactions on Acoustics, Speech and Signal Processing, Vol. 29, pp 351-363, 1981.

**[29] Le Cam Quentin**, Tricha M"hamed, *« Interface Androïde pour la consultation des données envoyées par un capteur» ;* 2011-2012.

**[30] GOTRONIC**, *« Guide de mise en marche du module Bluetooth HC-05».* Cyber Networks.

**[31] Terminale STMG SIG,** « débuter avec app inventor »; projet avancé 2013-2014.

**[32]** https://fr.wikipedia.org/wiki/Assistant_personnel_intelligent

**[33]** https://fr.wikipedia.org/wiki/Apprentissage_automatique

**[34]** https://theconversation.com/deep-learning-des-reseaux-de-neurones-pour-traiter-linformation-76055

**[35]** http://marketing-digital.audencia.com/en/machine-learning-vs-deep-learning-difference/

**[36]** https://fr.mathworks.com/discovery/neural-network.html

**[37] Eric Davalo Patrick Naim**, « *DES RESEAUX DE NEURONES* »,2011

**[38] Hyan-Soo Bae , Ho-Jin Lee, Suk-Gyu** Lee « Voice Recognition Based on Adaptive MFCC and Deep Learning » , 2016

**[39] Alexis Brenon** « *Modèle profond pour le contrôle vocal adaptatif d'un habitat intelligent* », Communauté université Grénoble Alpes, 25 Mai 2016.

**[40] Zied Elloumi, Benjamin Lecouteux, Olivier Galibert, Laurent Besacier** « *Prédiction de performance des systèmes de reconnaissance automatique de la parole à l'aide de réseaux de neurones convolutifs»,* 25 Mai 2016.

**[41] G. E. Dahl, D. Yu, L. Deng, and A. Acero**, « *Context-dependent pre-trained deep neural networks for large-vocabulary speech recognition Audio, Speech, and Language Processing* », IEEE Transactions on, vol. 20, no. 1, pp. 30–42, 2012.

**[42] G. Hinton, L. Deng, D. Yu, G. E. Dahl, A.-r. Mohamed et al**., « *Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups,* » Signal

**[43] Anjan Chatterjee , «** Artificial Intelligence based IoT Automation: Controlling devices with Google and Facebook **», 2018**

**[44] Sameer Tuteja , *UG Student,*** *«* Expanding Voice Commands for Visually ImpairedInterfacing Google Home with Arduino using Webhooks », I*nternational Journal of Research in Advent Technology, Vol.7, No.4, April 2019*

*[45]*

# ANNEXE

## Code Arduino basé sur la reconnaissance vocale

```
#include <SoftwareSerial.h>
#include "VoiceRecognitionV3.h"
#include "SD.h"
#include "TMRpcm.h"
#include <SPI.h>
#include <Servo.h>
#define SD_ChipSelectPin 53
TMRpcm tmrpcm;
VR myVR(10,11);    // 10:RX 11:TX, you can choose your favourite pins.
int const rtv = 25; int const rven = 22; int const rmou = 23;  int const rlam = 24;
uint8_t records[7]; // save record
uint8_t buf[64];
#define sam (0) #define lampe (1) #define vent (2) #define porte (3) #define moux (4) #define tele (5)
void printSignature(uint8_t *buf, int len){
int ab=0  ; int ac=0; int ad=0; int ae=0 ;
void setup() {
  myVR.begin(9600);
   tmrpcm.speakerPin = 46;
  Serial.begin(115200);
  Serial.println("Elechouse Voice Recognition V3 Module\r\nControl LED sample");
  pinMode(rtv, OUTPUT);  pinMode(rven, OUTPUT);  pinMode(rlam, OUTPUT);  pinMode(rmou,
OUTPUT);  digitalWrite(rtv, HIGH);  digitalWrite(rven, HIGH);  digitalWrite(rlam, HIGH);
  digitalWrite(rmou, HIGH);
   while(1) ; }
```

```
  if(myVR.load((uint8_t)sam) >= 0){
   Serial.println("salomé loaded");}
   if(myVR.load((uint8_t)lampe) >= 0){
   Serial.println("lampe loaded");}
  if(myVR.load((uint8_t)vent) >= 0){
   Serial.println("ventillateur loaded");  }
  if(myVR.load((uint8_t)moux) >= 0){
   Serial.println("moulinex loaded");  }
  if(myVR.load((uint8_t)tele) >= 0){
   Serial.println("télé loaded"); }
   if (!SD.begin(SD_ChipSelectPin)) {
    return; }
void loop(){
  int ret;
  ret = myVR.recognize(buf, 50);
  if(ret>0){
   switch(buf[1]){
     case sam :
      Serial.println("salomé");
       tmrpcm.play("1.wav") ;
      break;
     case lampe :
      ac=~ac;
        if(ac==0){
      digitalWrite(rven, HIGH);
      Serial.println("ventillateur");
       tmrpcm.play("5.wav")  }
         else {
          digitalWrite(rven, LOW);
        Serial.println("ventillateur") ;
```

```
      tmrpcm.play("4.wav") ; }
    break;
     case porte:
     Serial.println("porte");
      tmrpcm.setVolume(6);
      tmrpcm.play("6.wav");
     myservo.attach(9);  // attaches the servo on pin 9 to the servo object
     for ( int i = 0; i<90;i++){
       myservo.write(i);              // sets the servo position according to the scaled value
          delay(15);
     }
break;
     case moux:
      ae=~ae;
        if(ae==0){
     digitalWrite(rtv, HIGH);
      Serial.println("tele");
     tmrpcm.play("11.wav") ; }
       else {
      Serial.println("tele");
      digitalWrite(rtv, LOW);
      tmrpcm.setVolume(6);
      tmrpcm.play("10.wav"); }
     break;
    default:
     Serial.println("Record function undefined");
     break;
  }
  printVR(buf)
  } } }
```