

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE

MINISTRE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA RECHERCHE
SCIENTIFIQUE

UNIVERSITE SAAD DAHLEB BLIDA 1



INSTITUT D'AERONAUTIQUE ET DES ETUDES SPATIALES

Département de Construction Aéronautique



Projet de fin d'études

En vue de l'obtention du diplôme de Master en

Aéronautique

Spécialité : Avionique



Thème

Reconnaissance vocale robuste utilisant l'apprentissage
en profondeur pour interaction homme-drone

Présenté par :

- ❖ BARKA MOHAMMED YACINE
- ❖ AISSAOUA NESRINE

Encadré par :

- ❖ Dr *KHEIREDDINE CHOUTRI*
- ❖ Professeur *MOHAND LAGHA*

IAES
2021 - 2022

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

Remerciements

Au terme de ce travail, on saisit cette occasion pour exprimer nos vifs remerciements au bon dieu qui nous a donné la force et la patience d'effectuer ce travail.

A nos chers encadreur : Dr.KHEIREDDINE CHOUTRI, P.MOHAND LAGHA.

Vous nous avez fait un très grand honneur d'encadré ce mémoire et de nous guider tout au long de son élaboration. Vos précieux conseils et votre disponibilité, nous ont beaucoup aidés lors de la réalisation de ce travail.

Veillez trouver ici le témoignage de notre plus grande estime et nos remerciements les plus sincères ainsi que l'assurance de notre respect.

A notre président de jury.

Nous vous remercions de l'honneur que vous nous faites en acceptant de présider ce jury de mémoire.

Veillez trouver ici l'expression de notre plus profond respect pour votre qualité d'enseignement. Soyez assuré de notre gratitude et de notre sympathie.

A nos membres de jury.

Nous tenons à vous remercier d'avoir accepté de participer au jury de notre mémoire, d'évaluer et d'enrichir ce travail. Nous vous remercions pour la qualité de votre enseignement durant nos études. Veillez trouver par ce travail le témoignage de notre reconnaissance et de notre profond respect.

Un spécial remerciement au corps professoral durant notre cursus d'étude pour la qualité de leur enseignement.

Dédicaces

Je dédie ce modeste travail accompagné d'un profond amour ;

Au meilleur des pères : Driss,

Aucune dédicace ne saurait exprimer l'amour, l'estime, le dévouement et le respect que j'ai toujours eu pour vous. Rien au monde ne vaut les efforts fournis jour et nuit pour mon éducation et mon bien être ;

A ma très chère mère : Mama,

Aucune dédicace ne saurait être éloquente pour exprimer ce que vous méritez pour tous les sacrifices, les prières et la bénédiction qui m'ont été d'un grand secours pour mener à bien mes études et tracer mon avenir ;

A ma chère sœur,

A qui je souhaite un avenir radieux plein de réussite et de joie et l'accomplissement de tout ce que tu souhaites ;

A tous mes amis et mes collègues,

*Particulièrement : B.IBRAHIM, B.MILOUD, S.BILEL, B.ISLAM, Z.KHALED.
Ils vont trouver ici le témoignage d'une fidélité et d'une amitié infinie ;*

A tous les amis de la famille,

Mes tantes Kenza, Sousouu, et spécialement ma tante Faiza et tout la famille ALÉM, merci pour votre soutien plus que précieux pendant tout mon cursus.

A ma meilleure amie,

*Je tiens à te remercier en particulier car tu as changé toute ma vie universitaire au point où je me sentais plus dans un endroit différent de chez moi, tu as fait de moi un membre de ta famille et un homme meilleure face à toutes les difficultés que j'ai eu,
Merci encore une fois pour ton aide que je n'oublierais jamais.*

Sans oublier ma binôme : A,Nesrine ,

Pour ton soutien moral, ta patience et ta compréhension tout au long de ce projet. Je vous souhaite une vie pleine de santé et de bonheur ;

MOHAMMED YACINE

Dédicaces

Je dédie ce modeste travail ;

Au meilleur papa au monde ;

Aucune dédicace ne saurait exprimer l'amour, l'estime, le dévouement et le respect que j'ai toujours eu pour vous. Merci pour le soutien que vous m'avez donné merci d'avoir cru en moi. Rien au monde ne vaut les efforts fournis jour et nuit pour mon éducation et mon bien être ;

A l'être le plus cher de ma vie ma mère,

Pour son amour, ses encouragements et ses sacrifices, a celle qui a attendu avec patience les fruits de sa bonne éducation et de ses dévouements

Merci MAMAN.

A mes chères sœurs et mon petit frère,

Pour leurs soutiens, leurs affections, elles étaient toujours à mes cotées aux moments les plus difficiles, ALIA ma confidente et MAYA.

A tous mes amis et mes collègues,

Particulièrement : B.IMANE, F.LISA, B.MANEL, et G.AICHA. Ils vont trouver ici le témoignage d'une fidélité et d'une amitié infinie ;

Au parent de mon cher binôme,

Je vous remercie pour tout le soutien et l'amour que vous nous a portez depuis le début de notre expérience.

Sans oublier mon binôme : B.YACINE,

Pour ton soutien moral, ta patience et ta compréhension tout au long de ce projet. Je vous souhaite mon frère une vie pleine de santé et de bonheur et de réussite ;

NESRINE

Résumé

La reconnaissance vocale se révèle de plus en plus performante grâce à l'intelligence artificielle et au Deep Learning. La révolution de la reconnaissance automatique de la parole est en marche.

L'objectif de ce travail est de contrôler le mouvement du drone en utilisant un système de reconnaissance vocale multilingue dans un environnement bruité.

Pour cela, un réseau neuronal profond (DNN) est entraîné à reconnaître la parole de l'utilisateur dans un environnement bruité, puis à générer la commande de contrôle souhaitée. Nous avons mené des expériences avec différent type de débruitage afin de comparer le niveau de reconnaissance. L'implémentation matérielle du système conçu prouve sa grande précision de reconnaissance et sa simplicité de contrôle.

Mots clé : Drone, interaction Homme-Drone, intelligence artificielle, environnement bruité

Abstract

Speech recognition is becoming increasingly powerful thanks to artificial intelligence and Deep Learning. The revolution in automatic speech recognition is underway.

The objective of this work is to control the movement of the drone using a multilingual speech recognition system in a noisy environment.

For this purpose, a deep neural network (DNN) is trained to recognise the user's speech in a noisy environment and then to generate the desired control command. We conducted experiments with different types of denoising to compare the level of recognition. The hardware implementation of the designed system proves its high recognition accuracy and control simplicity.

Keywords: Drone, human-drone interaction, artificial intelligence, noisy environment

الملخص

أثبت التعرف على الصوت أنه أكثر فاعلية بفضل الذكاء الاصطناعي والتعلم العميق. ثورة التعرف التلقائي على الكلام جارية

الهدف من هذا العمل هو التحكم في حركة الطائرة بدون طيار باستخدام نظام التعرف على الصوت متعدد اللغات في بيئة صاخبة

للتعرف على كلام المستخدم في بيئة صاخبة، ثم إنشاء (DNN) لهذا الغرض، يتم تدريب شبكة عصبية عميقة أمر التحكم المطلوب. أجرينا تجارب مع أنواع مختلفة من تقليل الضوضاء لمقارنة مستوى التعرف. يثبت تنفيذ الأجهزة للنظام المصمم دقته العالية في التعرف والتحكم البسيط

الكلمات المفتاحية: طائرة بدون طيار، تفاعل بين البشر والطائرات بدون طيار، ذكاء اصطناعي، بيئة صاخبة

CHAPITRE II : TECHNIQUES DRE CONNAISSANCES VOCALES

II.1.Introduction	19
II.2. Réseaux neuronaux profonds (DNN)	20
II.2.1. Définition	20
II.2.2. Architectures d'apprentissage profond	21
II.2.2.1. Réseau de neurones convolutifs	21
II.2.2.2. Caractéristiques de l'apprentissage profond	22
II.2.3. Fonctions d'activation	23
II.2.3.1. Sigmoidé	23
II.2.3.1.2. Tanh	24
II.2.3.1.3. Relu	24
II.3.Reconnaissance des commandes vocales à l'aide de l'apprentissage profond	25
II.3.1. Transformée de Fourier à court terme	26
II.4.Approches de la reconnaissance vocale multilingue	33
II.4.1. Portage	33
II.4.2. La reconnaissance inter linguistique	34
II.4.3. Reconnaissance multilingue simultanée de la parole	35
II.5.Reconnaissance multilingue des commandes vocales par apprentissage profond	36
II.5.1. L'environnement matériel	36
II.5.2.Implémentation	36
II.5.3. Développement de la base de données multilingue "arabe et amazigh"	37
II.5.4. Enregistrement de la base de données	37
II.5.5. Séparation des données de test, de validation et de formation	38
II.5.6.Entraînement du réseau	38
II.5.7.Comparaison entre la formation monolingue et la formation multilingue	39
II.5.8.Matrice de confusion	40
II.6. Conclusion	41

CHAPITRE III:	DEBRUITAGE DE LA COMMAND VOCALE	
III.1. Introduction		43
III.2 Le traitement de la parole		44
III.2.1. Description de la parole.		44
III.2.2. La notion du bruit		45
III.2.3 La réduction du bruit		46
III.2.4 Estimation du niveau de bruit		47
III.3. Les algorithmes de réduction de bruit		49
III.3.1 Base de la Transformée de Fourier rapide		49
III.3.2 Algorithme de calcul de la FFT		50
III.3.3 Le Spectrogramme		51
III.4 La soustraction spectrale		53
III.4.1 Le Principe		54
III.4.2 Application et résultat		55
III.5 Le Filtrage de Wiener		61
III.5.1 Le Principe		61
III.5.2 Application et résultat		62
III.6. Conclusion		67
CONCLUSION GENERALE & PERSPECTIVE		69
1. Conclusion générale		69
2. Perspectives		70
BIBLIOGRAPHIE		71

LISTE DES FIGURES

- Figure (I-1) : les principaux domaines de l'IDH.
- Figure (I-2) : Photographie de paysage à l'aide de drones.
- Figure (I-3) : Livraison par drones.
- Figure (I-4) : Surveillance des drones.
- Figure (I-5) : Le développement de l'utilisation d'interaction Homme-Drone au fil du temps.
- Figure (I-6) : Commande de drone avec le geste humain.
- Figure (I-7) : commande vocale d'un drone
- Figure (I-8) : Drone face au feu de forêts
- Figure (I-9) : Drone d'accompagnement sportif.
- Figure (II-1): Réseau Neuronal Profond.
- Figure (II-2): L'architecture basique de CNN.
- Figure (II-3): Taille du lot d'époch et itération.
- Figure (II-4): Effet du taux d'apprentissage sur la fonction de perte.
- Figure (II-5): Fonction d'activation Sigmoidale et sa dérivée.
- Figure (II-6): Fonction d'activation Tanh et sa dérivée.
- Figure (II-7): Fonction d'activation Relu et sa dérivée
- Figure (II-8): Transformée de Fourier à court terme.
- Figure (II-9): Un exemple de signal vocal d'entrée.
- Figure (II-10): Trames du signal vocal.
- Figure (II-11): Fenêtre de Hamming.
- Figure (II-12): Le signal en temporel/fréquentielle.
- Figure (II-13): échelle de fréquence Mel.
- Figure (II-14) : Banque de filtres dans l'échelle de fréquence Mel.
- Figure (II-15): Spectrogram du signal.
- Figure (II-16): Former le mode.
- Figure (II-17): Reconnaissance de commandes vocales utilisant l'apprentissage profond pour la classification audio.
- Figure (II -18): Une esquisse du scénario de portage.

Figure (II -19): Une esquisse du scénario translinguistique

Figure (II-20): Une esquisse du scénario du multilinguisme simultané.

Figure (II-21): Une vue d'ensemble de l'étape multilingue.

Figure (II-22): Le processus de reconnaissance vocale multilingue.

Figure (II-23) : Progression de la formation pour le jeu de données monolingue (anglais).

Figure (II-24) : Progression de la formation pour le jeu de données multilingue (Anglais, Arabe et Amazigh).

Figure (II-25) : Matrice de confusion pour un ensemble de données multilingues.

Figure (III-1):Description schématique de l'analyse temps/fréquence par la FFT.

Figure (III-2): La transformation des signaux en spectrogramme.

Figure (III-3) : Organigramme de filtrage.

Figure (III-4) : Signal en domaine temporelle de la commande « Yes » bruité par le son des moteurs.

Figure (III-5) : La Transformée de Fourier du signal de parole bruité.

Figure (III-6) : Le spectre d'amplitude de bruit.

Figure (III-7) : La représentation du bruit dans le temps.

Figure (III-8) : Le spectre d'amplitude du signal de parole débruité.

Figure (III-9) : Comparaison entre les graphes (avant/après) soustraction.

Figure (III-10) : Le signal filtré dans le domaine temporelle.

Figure (III-11) : Comparaison entre la commande bruitée et le bruit des moteurs.

Figure (III-12) : Signal en domaine temporelle bruité par le son des moteurs.

Figure (III-13) : La Transformée de Fourier du signal de parole bruité.

Figure (III-14) : Le signal filtré dans le domaine temporelle.

Figure (III-15) : Le spectre d'amplitude du signal de parole filtré.

Figure (III-16) : Comparaison entre les graphes (Avant/Après) filtrage en Temps.

Figure (III-17) : Comparaison entre les graphes (avant/après) filtrage en fréquence.

LISTE DES TABLEAUX

Tableau II.1 Caractéristiques de la base de données

Tableau II.2 : Le drone commande en Arabe et en Amzaigh

LISTE DES ABREVIATIONS

AM	Modèle Acoustique
ASR	Reconnaissance automatique de la parole
BCI	Interaction cerveau-ordinateur
BLE	Bleutooth Low Energy (basse énergie)
CNN	Réseau neuronal convolutif
CPU	Unité centrale de traitement
DNN	Réseaux neuronaux profonds
EEG	Electroencéphalographie
ESC	Contrôleur électronique de vitesse
FAA	Administration fédérale de l'aviation
FFT	Transformée de Fourier rapide
FS	Fréquence d'échantillonnage
GHZ	Gigahertz
GUI	Interface utilisateur graphique
HDI	Human Drone interaction
HRI	Human Robot Interaction (Interaction homme-robot)
ISM	Industriel, scientifique et médical.
KBPS	KiloBytes par seconde
LM	Modèle de langue
MBPS	Megabits par seconde
MFCC	Mel-frequency Cepstral Coefficient
NUI	Natural User Interface (Interface utilisateur naturelle)
MQTT	MQ Telemetry Transport (Transport de télémétrie MQ)
ReLU	Unité linéaire redressée
RF	Radiofréquence
SPI	Serial Peripheral Interface
SR	Reconnaissance vocale

STFT	Transformée de Fourier à court terme
UART :	Récepteur-émetteur asynchrone universel
UAV	Véhicule aérien sans nom
UI	Interface utilisateur
ULP	Ultra low power
USART	émetteur-récepteur synchrone et asynchrone universel
USB	Universal Serial Bus (bus série universel)
VR	Réalité virtuelle

INTRODUCTION GENERALE

Les drones sont des aéronefs capables d'effectuer des missions de vol sans pilote humain à bord. Avec les progrès techniques récents, les drones sont devenus une nouvelle forme de robot. Il est difficile de concevoir comment interagir avec ces robots volants.

Les drones changent notre vie quotidienne de manière extrêmement bénéfique. Ils nous facilitent la vie dans différents secteurs. Les drones sont une innovation incroyablement importante, et leurs applications potentielles impliquent qu'ils vont modifier nos vies en introduisant un tout nouveau domaine de technologie connu sous le nom d'interaction homme-drone(IHD). [1].

De nos jours, il est courant de voir de plus en plus de personnes n'ayant aucune connaissance préalable du sujet posséder un drone, que ce soit pour accomplir un objectif spécifique ou pour se divertir.

Cette situation pourrait encore être améliorée par l'introduction d'interfaces utilisateur naturelles (NUI) telles que les gestes du corps et les commandes vocales, qui ont été testées dans l'état de l'art dans [2].

La plupart des résultats semblent indiquer que la mise en œuvre d'une NUI facilite l'interaction entre l'homme et le drone.

L'objectif de ce mémoire est de résoudre la problématique de contrôler un Robot (Drone) avec une commande vocale au bruit de l'environnement (environnement bruité).

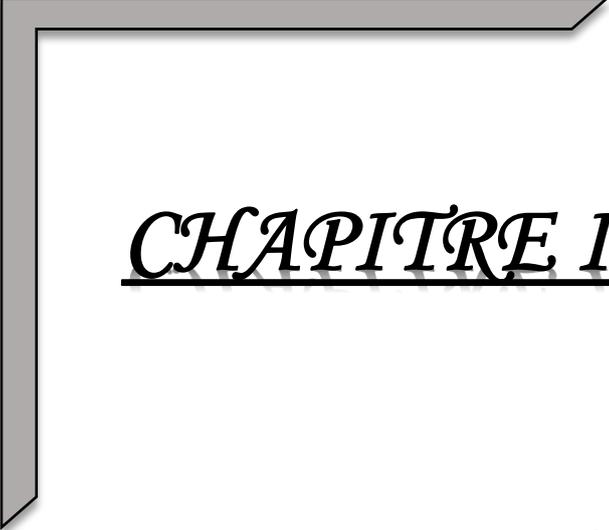
Pour cela, nous sommes particulièrement intéressés à la conception d'un quad-copter à commande vocale. Notre travail est divisé en quatre chapitres :

- Nous introduisons le sujet de l'interaction homme-drone dans le premier chapitre afin de comprendre la valeur des drones dans nos vies et comment s'engager avec eux.
- Le deuxième chapitre couvre quelques concepts de base de la reconnaissance vocale, ainsi que le réseau neuronal profond, qui a montré

une amélioration significative dans l'extraction et la reconnaissance des caractéristiques de la parole et la reconnaissance vocale multilingue et les étapes à suivre pour créer la base de données multilingue arabe et amazighe.

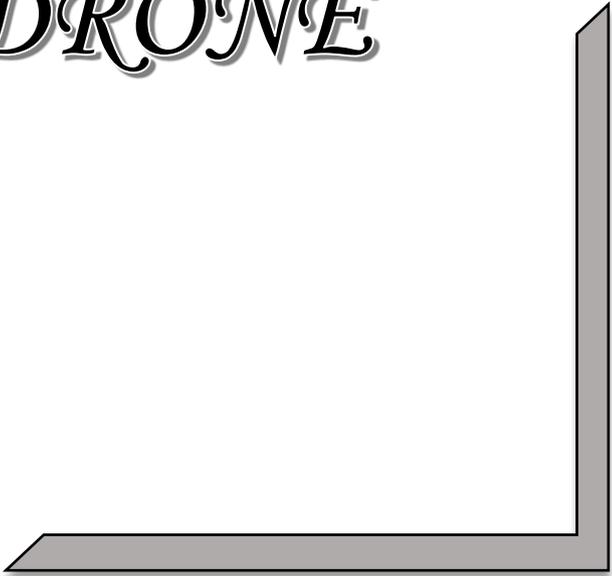
Nous la testons sur une variété de personnes afin de voir les performances de ce jeu de données.

- Le troisième chapitre consiste sur la reconnaissance vocale dans un environnement qui contient du bruit, c à dire comment commander ou contrôler notre drone par la commande vocale dans un environnement réelle (le bruit).
- Nous terminons notre travail par une conclusion qui résume les résultats et propose des suggestions pour les travaux futurs afin de maintenir la continuité et la performance du sujet suggéré.



CHAPITRE I

INTERACTION
HOMME-DRONE



CHAPITRE I : INTERACTION HOMME-DRONE

I.1. Introduction

Le domaine de l'interaction homme-robot (IHR) est né de la recherche sur les robots visant à comprendre et à créer des interactions avec les utilisateurs humains au cours de la dernière décennie.

Avec les récents progrès technologiques, les drones sont apparus comme un nouveau type de robot qui a captivé l'intérêt de la recherche sur l'IHR, ce qui a donné naissance à un tout nouveau domaine de recherche sur l'interaction homme-drone (IDH). [1]

Le terme "drone" désigne simplement un véhicule aérien sans pilote. La technologie des drones est apparue après la Première Guerre mondiale, bien qu'ils aient été conçus pour des usages militaires. Les drones sont actuellement employés dans un large éventail d'applications, passant de la sphère militaire à la sphère civile. [2]

L'utilisation des drones a augmenté à un rythme exponentiel au cours des dernières années. Cette croissance rapide est à la fois excitante et effrayante.

D'une part, les drones ouvrent de nouvelles perspectives, avec des applications allant du divertissement à la livraison, en passant par l'assistance aux personnes ayant des besoins particuliers, le sport, l'agriculture et même le sauvetage.

D'autre part, l'utilisation des drones dans notre environnement comporte plusieurs risques. [3]

Il est donc important d'étudier le domaine de l'IDH (Interaction Homme-Drone) pour comprendre comment l'interaction entre les humains et les drones peut être étendue à plus de domaines d'utilisation.

1.2. Définition de l'IDH

L'interaction homme-drone est un domaine de recherche diversifié. Elle peut être définie comme un champ d'étude centré sur la compréhension et l'évaluation de la distance d'interaction et le développement de nouveaux cas d'usage. Les quatre principaux domaines de l'IDH sont illustrés dans la figure (I-1)

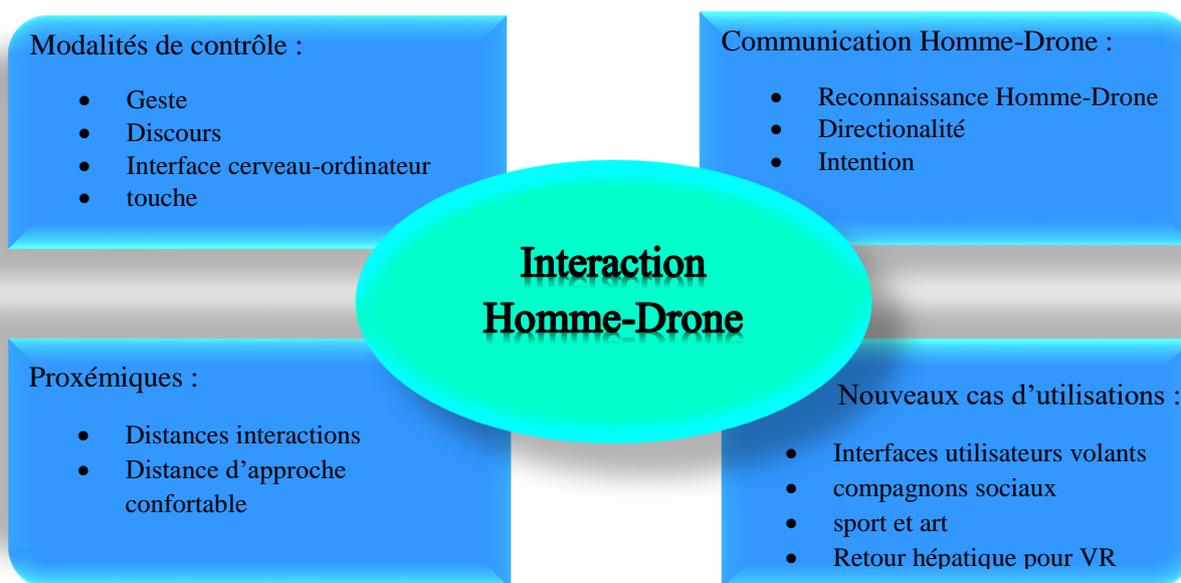


Figure (I-1) : les principaux domaines de l'IDH.

1.3. Recherche sur les interactions homme-drone

1.3.1. Rôle des humains dans les IDH

Lorsqu'ils interagissent avec un drone, les humains assument plusieurs responsabilités. Leur rôle est déterminé par l'application du drone et le degré d'autonomie.

Un humain peut agir en tant que "contrôleur actif", en contrôlant directement le drone à l'aide d'une interface de commande pour accomplir une tâche. Prenez la photo d'un paysage, par exemple, fig. (I-2)



Figure (I-2) : Photographie de paysage à l'aide de drones.

Un autre rôle est celui de "destinataire", dans lequel l'utilisateur n'utilise pas le drone mais bénéficie de son interaction. Prenons le cas d'un utilisateur qui reçoit un article livré par un drone.



Figure (I-3) : Livraison par drones.

Enfin, dans le cas des drones autonomes, il y a la tâche de "superviseur". Bien que la plupart des drones puissent voler de manière autonome, une personne est toujours nécessaire pour préprogrammer le comportement du drone (par exemple, planifier le vol) ou pour superviser le vol lui-même (par exemple, surveiller un vol pour une inspection autonome en temps réel).



Figure (I-4) : Surveillance des drones.

I.3.2. L'IDH dans le temps

L'IDH est un nouveau domaine. La figure suivante montre le nombre de publications par an sur Google Scholar utilisant l'interaction homme-drone dans la recherche pour montrer comment ce domaine s'est développé au fil du temps. Figure. (I-5).

Le degré de recherche dans ce domaine a dépassé toutes les attentes. L'université d'Australie du Sud collabore avec Dragonfly, une entreprise canadienne, pour créer un drone capable de détecter la température, le rythme cardiaque et la fréquence respiratoire d'un groupe, ainsi que la toux et les éternuements. Ses créateurs veulent employer cette technologie pour lutter contre le Covid-19. [5]



Figure (1-5) : Le développement de l'utilisation d'interaction Homme-Drone au fil du temps

I.3.3. Interfaces utilisateurs

L'interface utilisateur (IU) est le point où un ordinateur, un site Web ou une application interagit avec les humains. L'objectif d'une bonne IU est de rendre l'expérience de l'utilisateur simple et directe, en exigeant le moins de travail possible de la part de l'utilisateur pour obtenir le maximum de résultats souhaités. Les formats les plus importants sont GUI et NUI.

I.3.3.1. Interfaces utilisateur graphiques (GUI)

L'interface graphique, qui est encore utilisée aujourd'hui, génère une manière prévisible d'interaction en utilisant le WIMP (Windows, Icons, Menu, and Pointer).

L'interface graphique permet l'interaction avec le véhicule, ainsi que l'observation des états et de la dynamique, et aussi la présentation de vues et d'images graphiques pour aider l'utilisateur à comprendre le comportement interne du véhicule.

Les tâches suivantes sont souvent effectuées par l'opérateur à l'aide d'une interface graphique :

- Prédéfinir le comportement du drone (configuration du véhicule).
- Garder un œil sur le comportement du drone tout au long d'une mission.
- Recueillir des informations qui seront utilisées ultérieurement. [4]

I.3.3.2 Interfaces utilisateur naturelles (NUI)

Les interfaces utilisateur naturelles (NUI) constituent le niveau suivant dans l'évolution des interfaces utilisateur, par rapport aux interfaces utilisateur graphiques (GUI). [5]

Ces nouvelles approches permettent aux utilisateurs d'interagir avec les drones par le biais de gestes, de la parole, du toucher et même d'interfaces cerveau-ordinateur (ICO) comme l'électroencéphalographie (EEG) [3].

De nombreuses applications peuvent maintenant être trouvées dans notre vie quotidienne. Les assistants vocaux, Alexa et Siri, par exemple, répondent à une IUN activée par la voix [5].

L'objectif principal des interfaces utilisateur naturelles est de fournir une méthode de contrôle facile, décrite comme une interface qui fonctionne comme prévu par l'utilisateur. Les utilisateurs non experts peuvent interagir avec les interfaces utilisateur naturelles. [3]

I.3.3.3. Gestuelle

Lorsqu'il est demandé aux utilisateurs d'interagir avec un drone sans aucune instruction, les études démontrent que l'interaction gestuelle est prioritaire. Selon les recherches précédentes, il existe quatre critères de conception pour les gestes :

- 1) Les gestes doivent être naturels et simples à exécuter.
- 2) Les gestes doivent être connectés aux informations des photos enregistrées.
- 3) Il doit y avoir une séparation claire entre l'arrière-plan et le corps en mouvement.
- 4) Le traitement des données doit être effectué le plus rapidement possible.

Les caméras peuvent également être utilisées pour reconnaître les gestes. De nombreux drones sont déjà équipés d'une caméra embarquée qui peut être utilisée pour la reconnaissance des gestes sans qu'il soit nécessaire d'installer des capteurs supplémentaires lourds en charge utile.

Cette étude sur le contrôle gestuel montre qu'il s'agit d'une méthode de contrôle naturelle, qui présente les avantages de la facilité d'utilisation et de la réduction des périodes de formation. Elle a également l'avantage de ne pas obliger l'utilisateur à tenir des dispositifs externes, comme un joystick. Cependant, pour les applications qui exigent un contrôle délicat et précis, cette méthode n'est peut-être pas l'option idéale, car elle présente un temps de latence plus important et une moins bonne précision que les autres approches.



Figure (I-6) : Commande de drone avec le geste humain.

1.3.3.4. La parole

La parole est utilisée comme méthode d'interaction par 38% des utilisateurs américains et 58% des utilisateurs chinois, selon des études sur les interfaces utilisateur naturelles. La commande vocale est considérée comme plus simple que les autres approches, ce qui se traduit par des temps de formation plus courts.

Cependant, la reconnaissance vocale, comme le contrôle gestuel, peut ajouter des délais au système, ce qui limite ses utilisations.

En outre, le pilotage des drones par la parole pose des problèmes particuliers, car les hélices produisent un bruit fort qui peut diminuer la précision de la reconnaissance vocale.

Un autre problème est que si la reconnaissance vocale est effectuée à bord, les utilisateurs sont confinés à une interaction Co localisée car le drone doit être proche de l'utilisateur pour recevoir des instructions vocales. Ce problème ne concerne pas les systèmes où une station de contrôle au sol est utilisée pour décoder les commandes vocales et contrôler le drone.

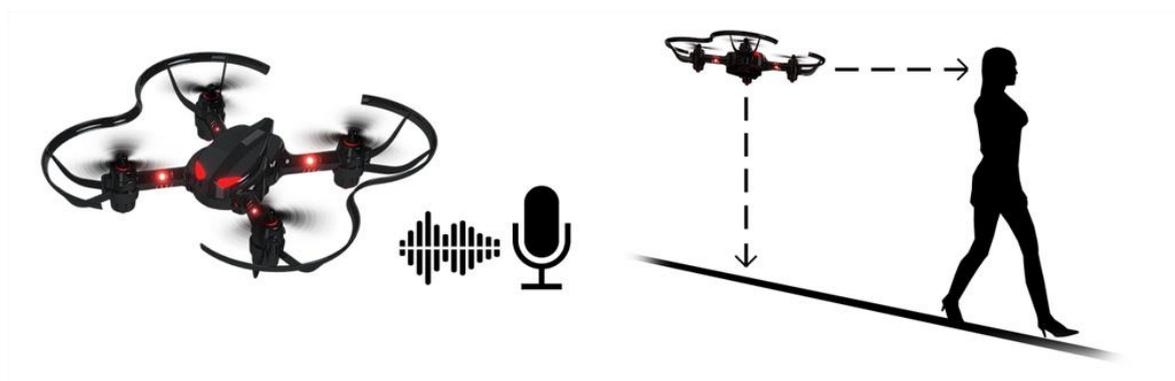


Figure (I-7) : Commande vocale d'un drone

1.3.3.5. Interaction cerveau-ordinateur (BCI)

Les dispositifs d'interface cerveau-ordinateur ont un large potentiel en tant que technologies d'assistance et méthodes de contrôle inédites.

Les chercheurs étudient l'utilisation d'interfaces cerveau-ordinateur (BCI) pour contrôler des aéronefs sans pilote (à voilure fixe) depuis 2010, et le premier projet de multi rotor contrôlé par le cerveau a été signalé en 2013. En 2016, l'université de Floride a organisé la première course de drones contrôlés par le cerveau, qui a été suivie d'une course à l'université d'Alabama.

Le pilote doit utiliser un casque BCI. Les plus populaires sont les casques d'électroencéphalographie (EEG), qui font fonctionner les drones via les signaux du cerveau.

Ces appareils utilisent des algorithmes d'apprentissage automatique pour analyser l'activité électrique du cerveau sur le cuir chevelu humain, qui est ensuite utilisée pour faire fonctionner des systèmes physiques grâce aux ondes cérébrales.

Par conséquent, les interactions avec les BCI sont actuellement limitées par rapport aux interfaces de commande classique, et des recherches supplémentaires sont nécessaires pour améliorer la fidélité et la fiabilité de ces systèmes avant qu'ils ne soient déployés au domicile des utilisateurs.

L'interaction mains libres et l'accessibilité pour les personnes handicapées seront possibles si la fiabilité et la précision des BCI atteignent des niveaux similaires à ceux des modalités de contrôle traditionnelles.

I.4. Communication homme-drone

I.4.1. Distance d'interaction

Pour une bonne connexion sociale, la distance d'interaction entre le drone et l'humain doit être prise en compte. [8] Dans l'expérience précédente, 37 % des utilisateurs américains sont restés dans l'espace intime du drone (45 cm), 47 % sont restés dans l'espace personnel (1,2 m) et les 16 % restants ont interagi dans l'espace social (3,7 m).

Cependant, les participants chinois se sont montrés plus à l'aise et plus proches les uns des autres : 50% dans l'espace intime, 38% dans l'espace personnel et 6% dans l'espace social. Une autre étude où les drones s'approchaient des utilisateurs à différentes hauteurs (1,80 m et 2,13 m) a conclu que la hauteur n'avait pas d'impact significatif sur la distance d'approche confortable. [7]

I.4.2. Retour d'information du drone

Des études ont précédemment exploré des méthodes permettant de reconnaître l'attention mutuelle entre un drone et ses utilisateurs, et une communication efficace pour éviter une mauvaise interprétation du système pouvant même conduire à des accidents.

Dans [8], le sujet était une comparaison de quatre gestes différents de reconnaissance du drone.

Les résultats montrent que les utilisateurs préfèrent la rotation dans l'axe de lacet pour indiquer la reconnaissance.

La capacité d'un drone à exprimer son intention aux utilisateurs a été étudiée dans [9], où l'expression concernait la manipulation de mouvements primitifs utilisant des trajectoires en arc et l'entrée et la sortie de profils de vitesse.

Les drones prototypés construits avec ces manipulations testent les tâches suivantes avec les participants : se rapprocher d'une personne (entrée et sortie faciles), éviter une personne (arc + entrée et sortie faciles), s'approcher d'un objet (anticipation) et s'éloigner d'un objet (arc + entrée et sortie faciles).

Les résultats ont révélé que les utilisateurs préfèrent travailler avec un drone en utilisant des trajectoires de vol manipulées plutôt que des trajectoires de base, une question de sécurité, d'interaction naturelle et intuitive.

I.4.3. Nouveaux cas d'utilisation

Les drones sont actuellement utilisés pour un large éventail d'applications, mais les chercheurs continuent d'explorer de nouvelles façons dont ces systèmes peuvent être utiles.

I.4.3.1. Interfaces utilisateur volantes

Cette sous-section présente des prototypes de drones conçus pour améliorer et ajouter de la mobilité aux interfaces utilisateur. Ces drones peuvent être utilisés pour contrôler des foules dans des situations d'urgence, fournir des informations et des conseils aux athlètes lors d'activités sportives, ou même servir de guide touristique pour des activités de plein air [10], [11].

Des travaux antérieurs ont exploré l'utilisation de deux drones comme écrans volants, l'un portant un projecteur et l'autre un écran de projection [12].

Cette approche peut être utilisée comme un nouveau modèle d'affichage public dans les environnements urbains, car elle permet à l'affichage d'attirer l'attention en s'approchant de l'utilisateur, en interagissant et en partant. La relation entre les drones était basée sur une relation maître-esclave, le drone projecteur

suivant la trajectoire du drone écran en utilisant des marqueurs visuels et la vision par ordinateur pour se positionner afin d'afficher correctement l'image.

Un octo-coptère (un hélicoptère à huit rotors) équipé d'un smartphone et d'un vidéoprojecteur a été utilisé avec succès pour afficher des images et des messages textuels sur des surfaces arbitraires [13]. À des fins d'évaluation, une expérience de vol a été réalisée en extérieur, affichant les messages reçus sur le mur d'un bâtiment. Le vol a duré 7 minutes et environ 40 personnes se trouvaient à 15 mètres de distance. Au cours de l'expérience, 23 messages au total ont été affichés. Les utilisateurs ont trouvé que le système était une expérience amusante capable d'attirer l'attention et ont considéré les cas d'utilisation comme des récits interactifs.



Figure (I-8) : Drone face au feu de forêts

I.4.3.2. Compagnons sociaux

Cette sous-section traite des prototypes de drones qui explorent l'interaction sociale avec les utilisateurs.

Ils peuvent être utilisés comme compagnons pour les personnes malvoyantes afin de leur fournir une aide à la navigation. [14] Cette étude envisage un système de drone qui se tient en attente sur un bracelet portable jusqu'à ce que son aide soit nécessaire. Un utilisateur aveugle commanderait le drone, et ce

dernier le guiderait jusqu'à ce que la cible soit atteinte. L'utilisateur peut suivre le drone grâce au retour auditif fourni par le son des hélices en rotation. Une fois la commande reçue, le drone calcule la distance jusqu'à l'emplacement cible et guide l'utilisateur en volant à une distance déterminée devant lui, en évitant les obstacles. Une connexion Bluetooth avec le bracelet permettrait au drone d'adapter sa distance et sa vitesse à l'utilisateur.

Les chercheurs ont également envisagé l'utilisation de drones comme agent de soutien dans un environnement propre [15].

Dans cette application, le drone trouve des déchets sur le sol, persuade les utilisateurs de les ramasser et les guide vers la poubelle la plus proche. Le drone disposait de différentes techniques de persuasion : visuelle, audio, et une combinaison des deux.

Bien que l'analyse des résultats n'ait pas révélé d'effet de la modalité d'interaction sur la conformité de l'utilisateur, d'autres facteurs ont été observés, tels que la culture du pays et le sexe de l'utilisateur.



Figure (I-9) : Drone d'accompagnement sportif

I.4.3.3. Retour h patique pour la r alit  virtuelle

Le terme "h patique" concerne l'utilisation des sensations tactiles dans les interfaces. Le retour h patique est la science et la technologie de la transmission et de la compr hension des informations par le sens du toucher.

Les syst mes actuels de RV peuvent fournir des exp riences visuelles et sonores immersives, mais il leur manque la capacit  de fournir un retour tactile. Comme les drones peuvent voler dans un espace 3D, ils peuvent  tre utilis s pour fournir un retour tactile en touchant l'utilisateur   n'importe quel endroit et   n'importe quelle vitesse pour offrir une exp rience ad quate.

De petits rotors quadruples ont  t  utilis s pour fournir un retour h patique dans les jeux de r alit  virtuelle [16].

Dans ce projet, des drones sont utilis s pour voler vers les utilisateurs   des vitesses variables pendant qu'ils sont immerg s dans un syst me d'environnement virtuel. Diff rents embouts peuvent  tre fix s au drone, en fonction de l'environnement virtuel, afin de fournir un retour ad quat. Le jeu prototypique consiste en une cit  maya dans la jungle. Les drones fournissent un retour d'information dans trois sc narios : ils jouent le r le de drones qui attaquent l'utilisateur, de fl ches tir es par des cr atures sur le joueur, et de briques et de bois qui tombent sur l'utilisateur lorsque les ruines s'effondrent.

I.5. Conclusion

Ces derni res ann es le d veloppement et la recherche de l'IDH m nent   de nouvelles utilisations, ce qui permet aux drones d' tre utilis s dans plusieurs domaines.

Les drones peuvent  galement  tre utilis s comme compagnons des personnes handicap es pour leur fournir une aide   la navigation.

Les deux cas d'utilisation envisag s sont les suivants : guider les utilisateurs vers un endroit sp cifique ou les aider   trouver des objets plac s   l'aide d'algorithmes de vision par ordinateur. Bien que le syst me ne soit pas encore op rationnel, une  tude pr liminaire a  t  r alis e avec un utilisateur aveugle. Les

participants ont réussi à suivre le drone comme prévu et ont fourni un retour positif sur l'idée du projet.

Les chercheurs ont également envisagé l'utilisation des drones comme agents pour favoriser un environnement propre. Dans cette application, le drone trouverait des déchets sur le sol, persuaderait les utilisateurs de les ramasser et les guiderait vers la poubelle la plus proche.

Dans un avenir proche, les drones seront largement utilisés dans les domaines de la publicité publique, des livraisons, du sport, des interventions d'urgence et pour augmenter les capacités humaines.

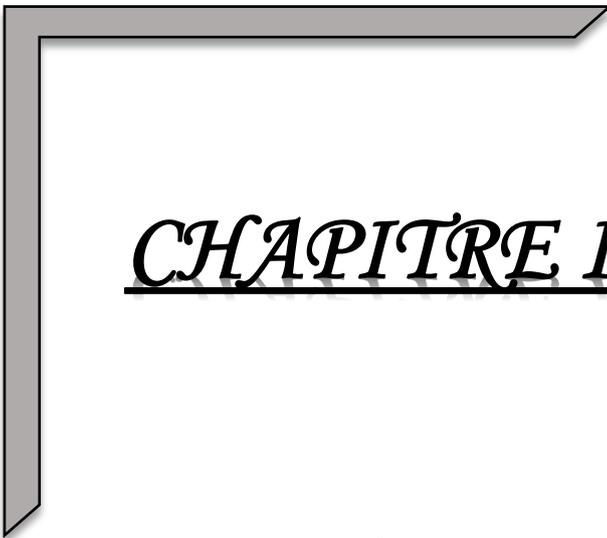
En outre, la popularité des drones augmentera lorsque nous comprendrons mieux comment la société accepte ces systèmes. Les futurs chercheurs pourraient donc y contribuer en étudiant la façon dont la société et les cultures perçoivent les drones.

« Sésame, ouvre-toi ! »

Cette phrase mythique n'est pas sans signification, car en dépit du trésor caché derrière la porte de pierre, une autre découverte s'ouvre à nous : La recherche en Reconnaissance Automatique de la Parole (RAP). Celle-ci ne cesse de s'étendre dans notre foyer en dépit de l'étonnement qu'on avait en regardant Bioman donné des ordres à son vaisseau.

Nous sommes cependant en dessous de la fiction étant donné la difficulté que nous avons encore à analyser un signal vocal complètement aléatoire.

Si dans un téléphone, on écoute les sons qu'émettent un Minitel, un fax ou un micro-ordinateur pour échanger des données, ils se présentent à nous comme un sifflement suraigu bourré de parasites : le message semble parfaitement inintelligible. A l'inverse, alors que notre propre langage nous paraît simple et clair, la machine, elle, n'y détecte rien de cohérent. [17].



CHAPITRE II

*TECHNIQUES DE
RECONNAISSANCES
VOCALES*



CHAPITRE II : TECHNIQUES DE RECONNAISSANCES VOCALES

II.1. Introduction

La reconnaissance de la parole est une méthode qui analyse les sons captés par un microphone et les transcrit en une série de mots que les machines peuvent comprendre.

Cependant, le résultat n'est pas fameux et dépend d'un certain nombre de facteurs. Les conditions favorables à la reconnaissance vocale impliquent une parole native, appartenant à un seul locuteur avec une diction correcte (ne présentant pas de pathologie vocale), enregistrée dans un environnement calme et silencieux, et basée sur un vocabulaire commun (mots connus par le système).

Lorsqu'il est confronté à des accents non natifs, à divers dialectes, à des locuteurs présentant une pathologie vocale, à des mots inconnus du système (généralement des noms propres) et à des signaux audio bruyants (faible rapport signal/bruit), les performances du système s'en ressentent. [18]

Ces dernières années, la communication vocale est devenue un élément de plus en plus essentiel de nos gadgets intelligents. Nous pouvons désormais activer notre smartphone par une simple commande vocale, indiqué à notre voiture où nous voulons aller, et même demander à un assistant vocal de passer une commande pour nous. Et ce n'est que la partie émergée de l'iceberg, car les applications sont très nombreuses.

Il y a eu récemment une augmentation du développement rapide de systèmes de reconnaissance automatique de la parole (ASR) de haute performance pour une variété de langues. Il a été démontré que les systèmes de reconnaissance de la parole construits avec des réseaux neuronaux profonds (DNN) offrent des avantages constants, en particulier pour les langues à faibles ressources. [19]

Dans notre travail on fait la base de données multilingue (en arabe et en amazigh).

Enfin, nous utiliserons cette simulation comme une application de drone.

II.2. Réseaux neuronaux profonds (DNN)

II.2.1. Définition

Un réseau neuronal profond (DNN) est un perceptron multicouche (MLP) avec de nombreuses couches cachées (généralement plus de deux). La figure (II-1) présente l'architecture d'un DNN avec une couche d'entrée, trois couches cachées et une couche de sortie. Cela lui permet de traiter les données de manière complexe, en utilisant des modèles mathématiques avancés.

- La couche d'entrée : prend les données de l'utilisateur et les envoie à la première couche cachée.
- Les couches cachées : utilisent nos entrées pour exécuter des calculs mathématiques.
- La couche de sortie : est chargée de renvoyer les données de sortie.

Chaque couche est reconnue pour extraire les données d'une manière unique. Dans la reconnaissance d'images, par exemple, la première couche recherche les bords, les lignes et autres caractéristiques. Deuxième couche : les yeux, les oreilles, le nez, et ainsi de suite.

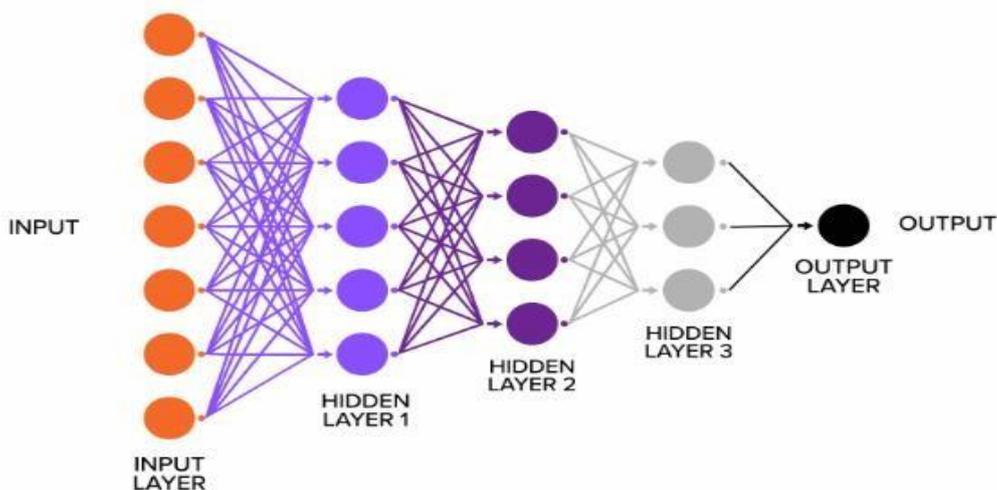


Figure (II-1): Réseau Neuronal Profond [20]

Les réseaux profonds, en général, sont toujours des réseaux neuronaux (formés par rétro-propagation, apprenant des abstractions hiérarchiques de l'entrée et optimisés à l'aide d'un apprentissage basé sur le gradient), mais avec des couches supplémentaires. [21]

II.2.2. Architectures d'apprentissage profond

II.2.2.1. Réseau de neurones convolutif

Pour la reconnaissance automatique de la parole, il existe une variété d'architectures d'apprentissage profond. Les CNNs sont un sous-type d'architecture profonde discriminante et ont montré des performances satisfaisantes dans le traitement de données bidimensionnelles. [22]

Le CNN, ou réseau neuronal convolutif, est un type d'architecture de réseau neuronal profond conçu pour des tâches spécifiques comme la classification d'images. Il possède certaines propriétés uniques qui le rendent utile pour le traitement de certains types de données, comme les images, l'audio et la vidéo.

Les CNN tirent leur nom du type de couches cachées qui les composent. Les couches cachées d'un CNN sont généralement constituées de :

- Couches convolutionnelles : c'est la plus importante, elle fonctionne en appliquant un filtre à un tableau de pixels d'image.
 - Mise en commun des couches : Réduit la taille de l'échantillon d'une certaine carte de caractéristiques, ce qui accélère également le traitement en réduisant le nombre de paramètres que le réseau doit traiter.
 - Les couches entièrement connectées : nous aident à classifier nos données.
 - La couche ReLu : sert de fonction d'activation, assurant la non-linéarité lorsque les données passent par les couches du réseau.

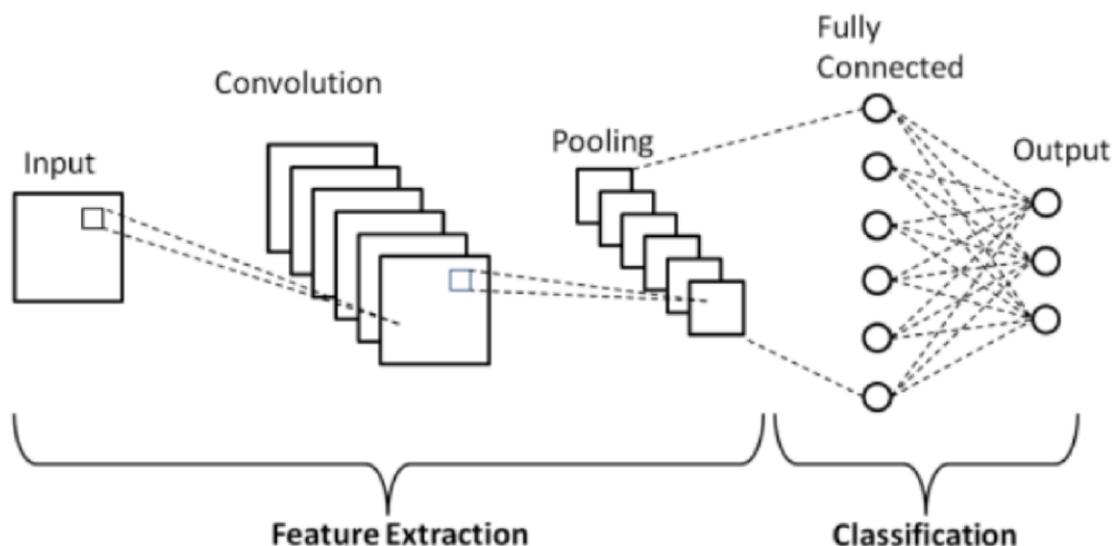


Figure (II-2): L'architecture basique de CNN [23]

II.2.2.2. Caractéristiques de l'apprentissage profond

La plupart du temps, il n'est pas pratique d'introduire toutes les données d'entraînement dans un algorithme en un seul passage. Une certaine terminologie est donc nécessaire pour améliorer la compréhension de la façon dont de plus petits morceaux de données sont utilisés. Comme le montre la figure (II-3).

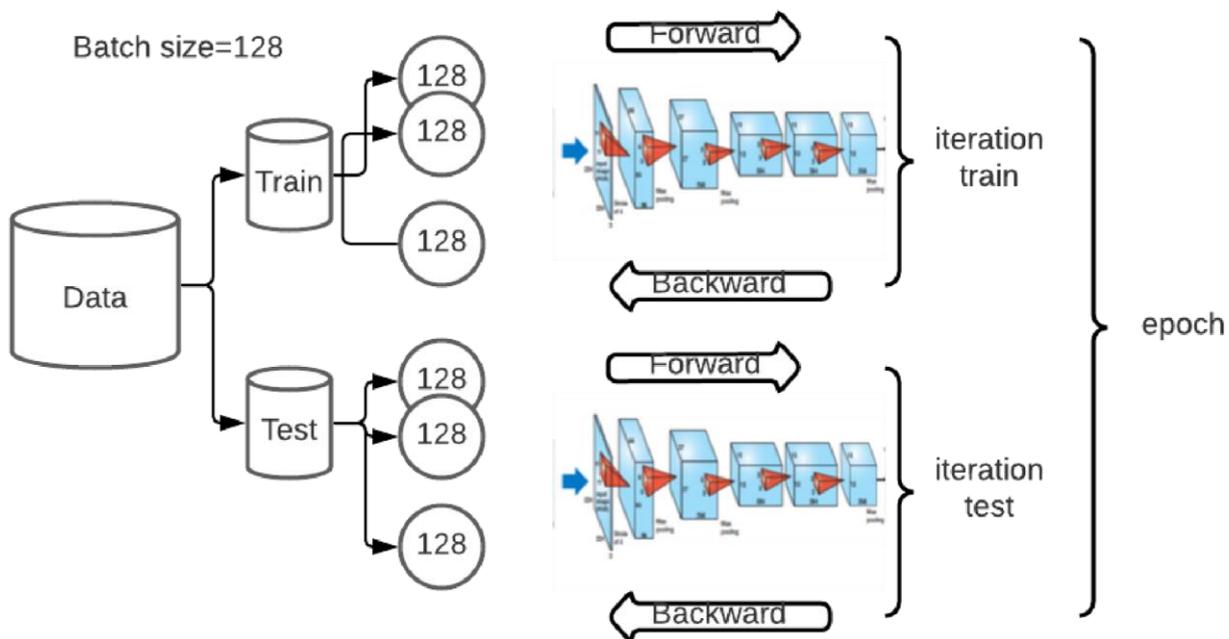


Figure (II-3): Taille du lot d'époque et itération.

- Époque : Le temps qu'il faut pour qu'un ensemble de données entier soit passé en avant et en arrière dans le réseau neuronal exactement une fois. Si l'ensemble des données ne peut pas être transmis à l'algorithme en une seule fois, il doit être divisé en mini-lots.

$$\text{Époque} = \text{iteration train} + \text{iteration test}$$

- Taille du lot : c'est le nombre total d'échantillons de formation présents dans un seul mini-lot.
- Itération : est mise à jour tout au long de la formation avec un gradient unique. Le nombre d'itérations est égal au nombre de lots requis pour une époque [24].

$$\text{Itération train} = \text{données train} / \text{taille du lot}$$

$$\text{Itération test} = \text{données test} / \text{taille du lot}$$

- Descente de gradient: processus itératif optimisation utilisé pour diminuer la fonction de perte dans le Deep Learning.
- Fonction de perte: indique la performance du modèle dans l'ensemble actuel de paramètres (poids et biais).

- Taux d'apprentissage: a un impact significatif sur l'efficacité de l'algorithme de descente de gradient. UN taux d'apprentissage très élevé peut faire en sorte que la fonction de perte commence à augmenter après quelques itérations, tandis qu'un taux modérément élevé fait en sorte que la perte plafonne à une valeur élevée après une diminution initiale rapide.

Un taux d'apprentissage très faible, par contre, peut être identifié par une lente diminution de la fonction de perte au cours des époques d'apprentissage.

Un bon taux d'apprentissage, d'autre part, combine une diminution rapide pendant les premières époques avec une valeur stable plus faible. [25]

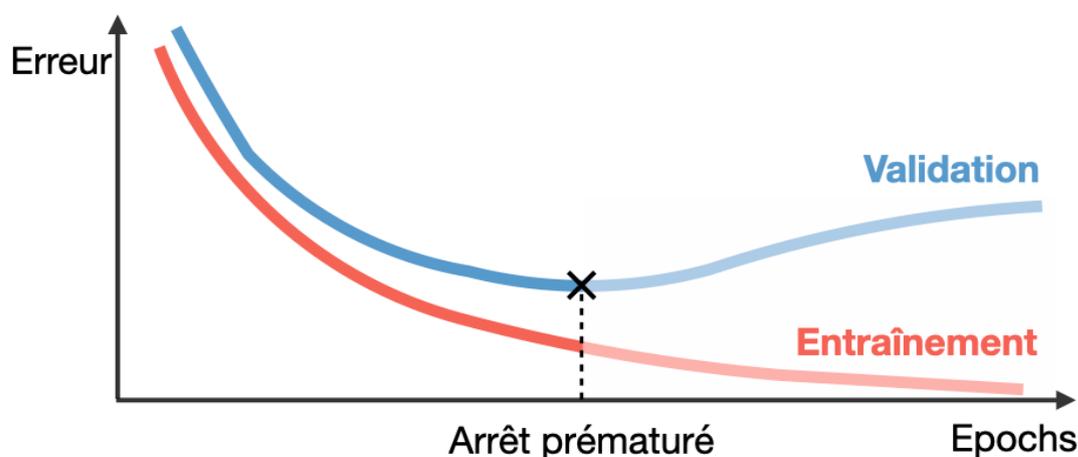


Figure (II-4): Effet du taux d'apprentissage sur la fonction de perte.

II.2.3. Fonctions d'activation

II.2.3.1. Sigmoïde

Pour diverses raisons, la fonction sigmoïde est une activation utile.

Comme la montre le graphique de la figure, elle fonctionne comme une fonction d'écrasement continue, limitant sa sortie à la plage (0,1). Elle est comparable à la fonction échelon, mais sa dérivée est lisse et continue, ce qui la rend excellente pour les méthodes de descente de gradient.

Elle est également centrée sur le zéro.

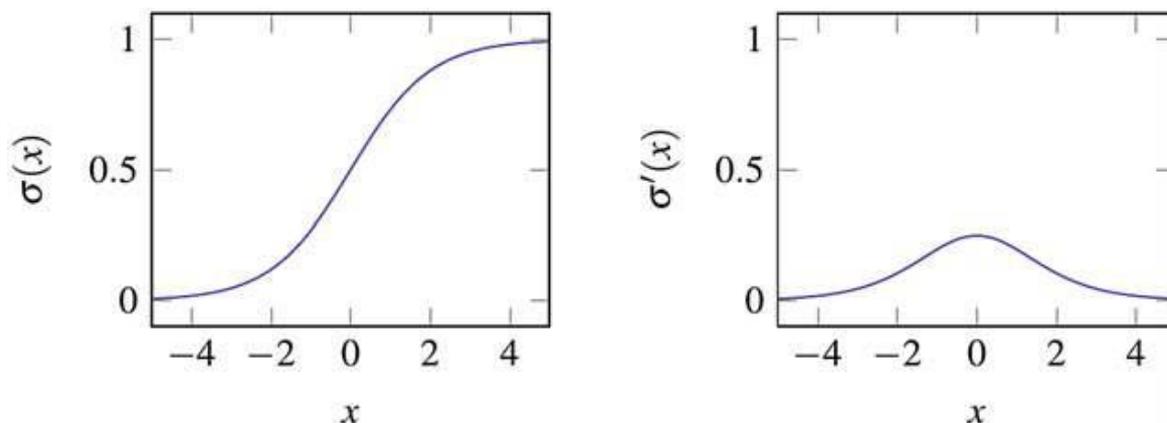


Figure (II-5): Fonction d'activation Sigmoide et sa dérivée.[26]

II.2.3.2. Tanh

Une autre fonction d'activation populaire est la fonction tanh. Elle sert également de fonction d'activation carrée, avec une sortie limitée à l'intervalle $(-1, 1)$.

Comme elle est centrée sur zéro, la fonction tanh élimine l'un des problèmes associés à la non-linéarité sigmoïde. Cependant, nous avons toujours le même problème de saturation du gradient aux extrêmes de la fonction.

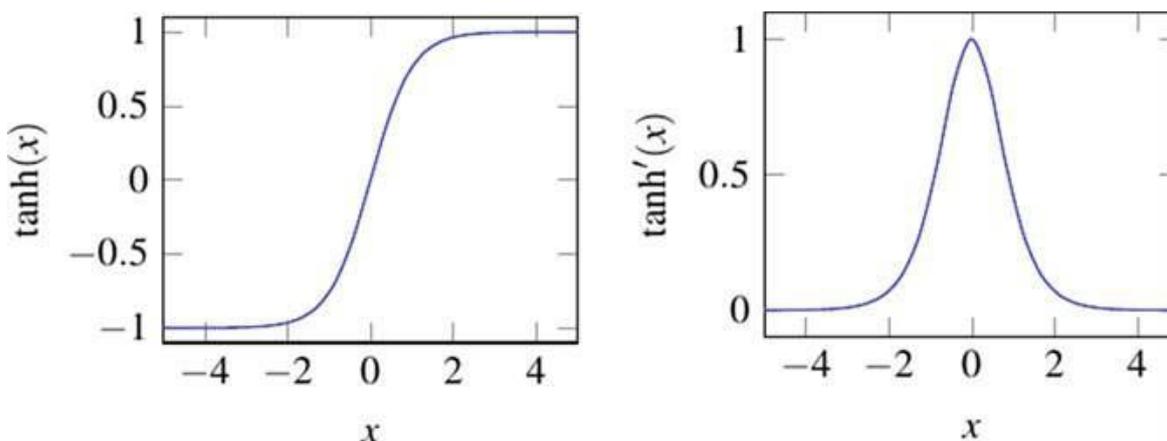


Figure (II-6): Fonction d'activation tanh et sa dérivée.[26]

II.2.3.3. Relu

L'unité linéaire rectifiée (ReLU) est une fonction d'activation simple et rapide qui est couramment utilisée dans la vision par ordinateur. Cette fonction de base a gagné en popularité en raison de sa convergence plus rapide. Par rapport à la sigmoïde et à la tanh, cela peut être dû à son gradient non saturant dans la direction positive.

La fonction ReLU est nettement plus rapide en termes de calcul, en plus de sa convergence plus rapide. Les fonctions sigmoïdes et tanh nécessitent des exponentielles, qui prennent beaucoup plus de temps qu'une simple opération max. [26]

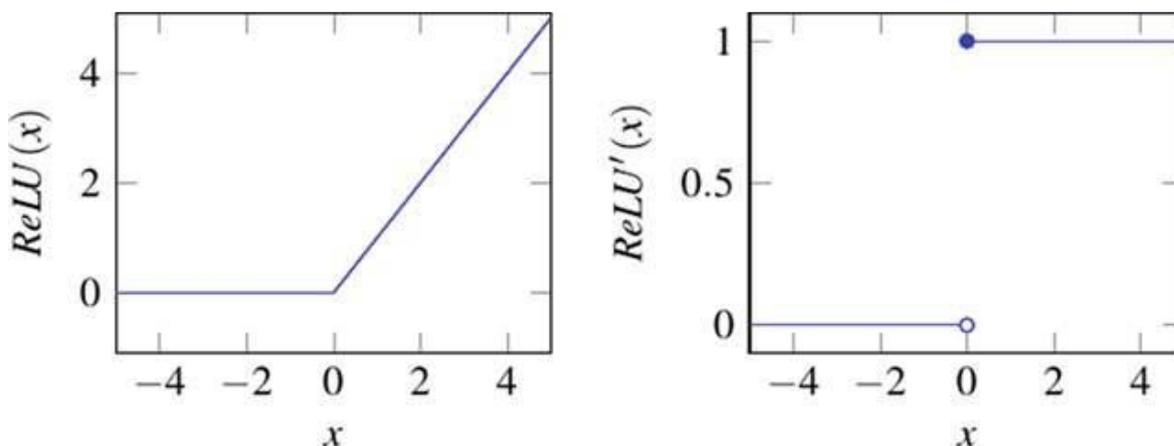


Figure (II-7): Fonction d'activation Relu et sa dérivée.[26]

II.3.Reconnaissance des commandes vocales à l'aide de l'apprentissage profond

Le réseau neuronal profond a démontré une grande amélioration dans l'extraction et la reconnaissance des caractéristiques de la parole.

Avec l'essor de l'apprentissage profond, les couches initiales des réseaux profonds ont fondamentalement remplacé l'extraction de caractéristiques. Les transformations temps-fréquence, telles que la transformée de Fourier à court terme (STFT), peuvent être utilisées comme représentations du signal pour les données d'entraînement dans les modèles d'apprentissage profond.

Les réseaux de neurones convolutifs sont couramment utilisés pour les données d'image et peuvent apprendre à partir de représentations de signaux 2D fournies par des transformations temps-fréquence. [27].

Les CNN profonds peuvent être formés en empilant un CNN avec un DNN entièrement connecté ou avec un ou plusieurs CNN où ils obtiennent un succès robuste pour la reconnaissance des images et de la parole.

Jeu de données Google Speech command

Les équipes de Tensor Flow et de l'AIT ont collaboré pour construire le jeu de données des commandes vocales de Google. La collection comprend 65 000 clips d'une seconde. Chaque clip contient l'un des 30 mots différents prononcés par des milliers de sujets différents.

Les clips ont été enregistrés dans des environnements réalistes avec des téléphones et des ordinateurs portables.

Les 8 mots de commande qui sont les plus utiles dans un environnement robotique:

- Oui
- Non
- Haut
- En bas
- Gauche
- Droite
- Marche
- Arrêt

Il est organisé en plus de 30 catégories, comprenant des instructions telles que "stop" et "up", ainsi que d'autres éléments tels que des numéros et des noms. Il y a environ 2400 enregistrements dans chaque dossier. [28]

Une fois les données obtenues, nous les convertissons dans le domaine temps-fréquence à l'aide de spectrogrammes. Ensuite, comme l'illustre la figure (II-6), nous utilisons ces spectrogrammes comme entrée pour nos réseaux neuronaux convolutifs profonds.

II.3.1. Transformée de Fourier à court terme

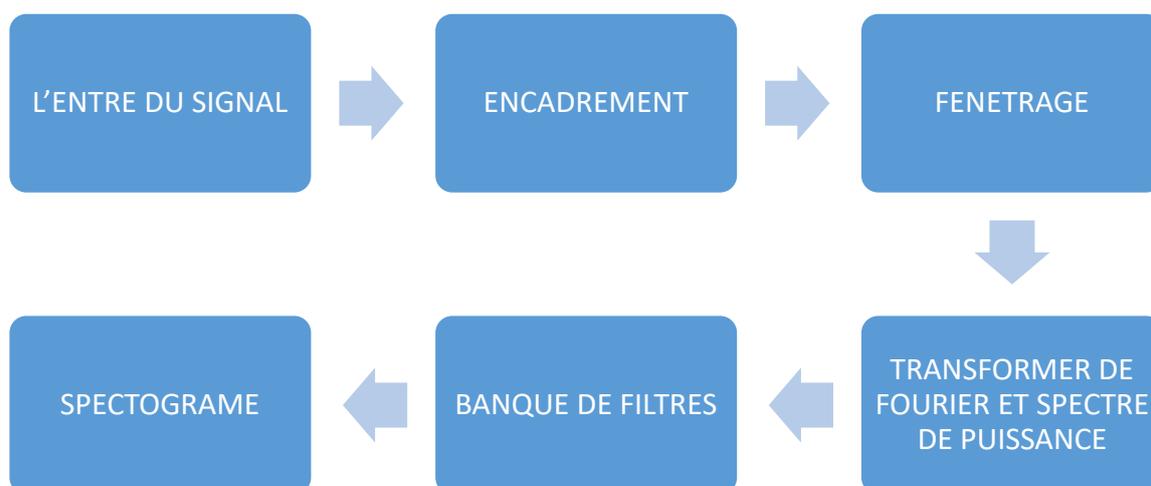


Figure (II-8): Transformée de Fourier a court terme.

➤ **Signal entrée:**

Un signal analogique étant un signal continu variant dans le temps, la tension instantanée du signal change continuellement avec la pression des ondes sonores dans une transmission audio analogique. Le signal vocal d'entrée est enregistré au format (.wav), comme indiqué sur le schéma. Ce fichier (.wav) sera essentiel pour les diverses modifications nécessaires à l'extraction des caractéristiques des sources audio.

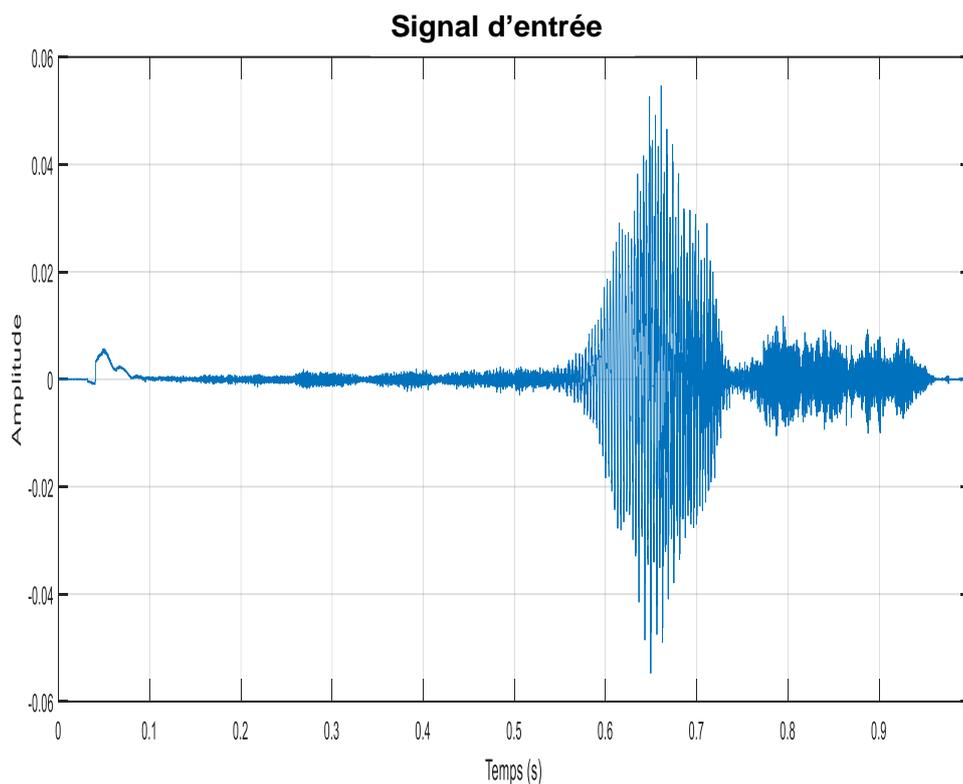


Figure (II-9): Un exemple de signal vocal d'entrée.

➤ **Cadrage :**

Un signal audio est en constante évolution. Pour faciliter les choses, nous supposons que le signal audio ne varie pas de manière significative sur de courtes échelles de temps (quand nous disons qu'il ne change pas, nous voulons dire statistiquement). C'est pourquoi nous encadrons le signal dans des cadres de 20-40 ms.

La figure (II-9) lustre les trams du signal audio.

Si le cadre est beaucoup plus court, nous n'avons pas assez d'échantillons pour obtenir une estimation spectrale fiable. Si elle est plus longue, le signal change trop au cours de la trame.

Le nombre de trames est généralement de 256 (en puissance de 2) car lors du calcul de la FFT, il serait simple que le nombre de trames soit en puissance de 2.

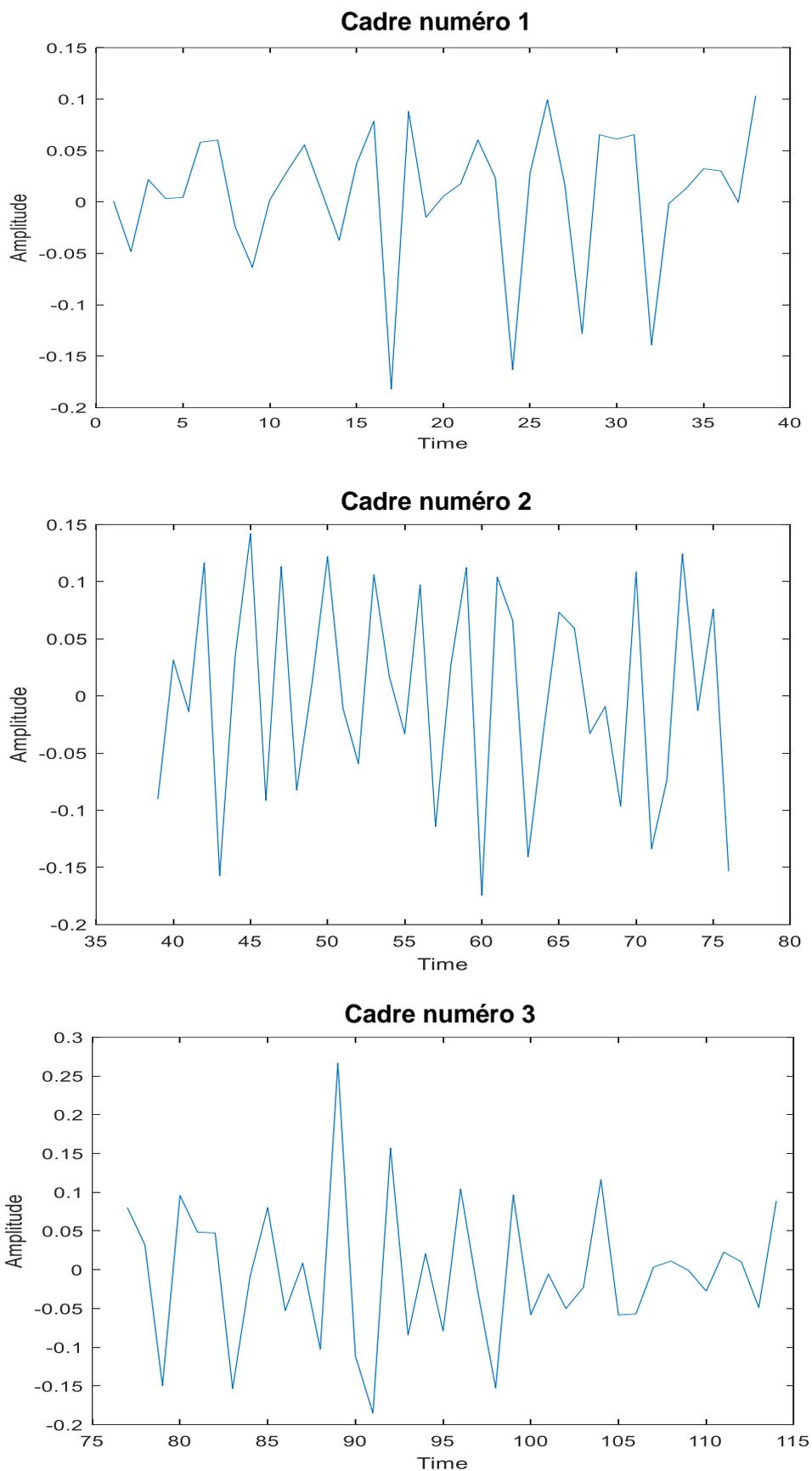


Figure (II-10): Trames du signal vocal.

➤ **Fenêtrage :**

L'étape suivante après le cadrage est le fenêtrage de chaque image individuelle. L'objectif du fenêtrage est de réduire les discontinuités du signal au début et à la fin de chaque image.

La fonction de fenêtrage est définie comme $W(n)$, avec n compris entre 0 et $N-1$. La longueur de l'image est indiquée par la lettre N . Le résultat du fenêtrage est le signal donné par l'équation (1).

$$Y(n) = x(n) \times w(n), \quad 0 < n < N-1 \quad (\text{II-1})$$

La fenêtre de Hamming a été considérée car le paramètre "side-lobe" y est bon. La forme de la fonction de fenêtre de Hamming est illustrée à la figure (II-10).

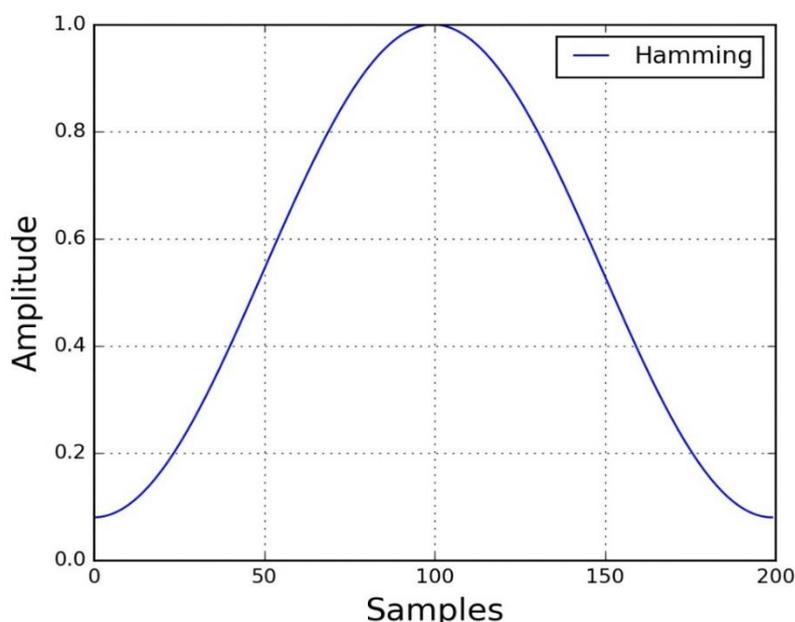


Figure (II-11): Fenêtre de Hamming.

Bien qu'il existe d'autres fenêtres comme la fonction fenêtre triangulaire, la fonction fenêtre rectangulaire, la fenêtre de Hamming présente les caractéristiques gaussiennes, qui ont la forme :

$$W(n) = 0,54 - 0,46 \cos \left(2 \times \pi \times n / (N-1) \right), \quad 0 < n < N-1 \quad (\text{II-2})$$

➤ **Transformée de Fourier et spectre de puissance:**

Ensuite, chaque trame est soumise à une transformée de Fourier rapide. Par conséquent, chaque trame de N échantillons doit être représentée dans le domaine fréquentiel en la transformant du domaine temporel au domaine fréquentiel à l'aide de la transformée de Fourier rapide, qui est largement utilisée en ingénierie, en mathématiques et en sciences à de nombreuses fins.

Le domaine fréquentielle est plus efficace que le domaine temporel pour les signaux audio, car la fréquence d'un individu particulier est un meilleur moyen de représenter cette personne que l'amplitude du signal.

La figure (II-12) illustre à la fois le domaine temporel et le domaine fréquentielle du signal.

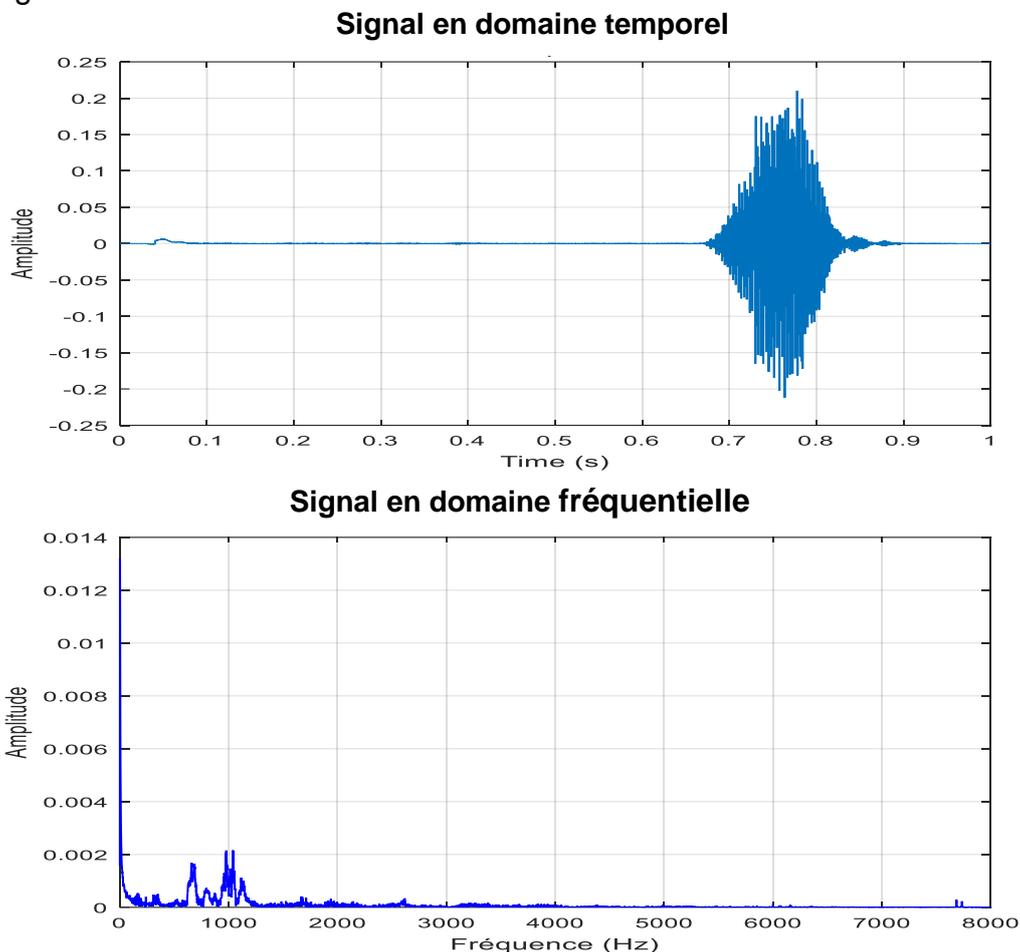


Figure (II-12): Le signal en temporel/fréquentielle.

Et ensuite, pour calculer le spectre de puissance, nous utilisons l'équation suivante :

$$P = \frac{|FFT(xi)|^2}{N} \quad (II-3)$$

Où x_i , est la i -ème trame du signal x .

➤ **Banc de filtres à échelle de Mel :**

Mel est une abréviation de mélodie, et la mélodie est très liée au concept de hauteur.

L'échelle de Mel relie la fréquence perçue, ou hauteur, d'un son pur à sa fréquence réelle mesurée. À basse fréquence, l'être humain est nettement plus apte

à détecter d'infimes variations de hauteur qu'à haute fréquence. En incluant cette échelle, nos caractéristiques se rapprochent davantage de ce que les humains entendent.

La formule utilisée pour calculer la fréquence de Mels pour toute fréquence est la suivante :

$$Mel(f) = 2595 \times \log(1 + f/700) \quad (II-4)$$

Mel (f) : la fréquence (mel) et f la fréquence (HZ)

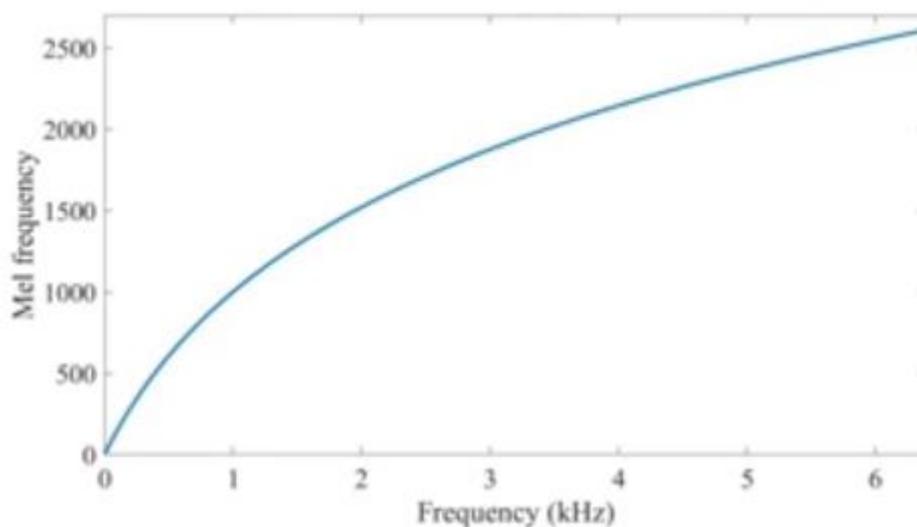


Figure (II-13): échelle de fréquence Mel.

L'étape finale du calcul des bancs de filtres consiste à appliquer des filtres triangulaires, généralement 40 filtres, Comme l'illustre la figure (II-13), chaque filtre du banc de filtres est triangulaire, avec une réponse de 1 à la fréquence centrale et décroissant linéairement vers 0 jusqu'à atteindre les fréquences centrales des deux filtres voisins, où la réponse est 0.

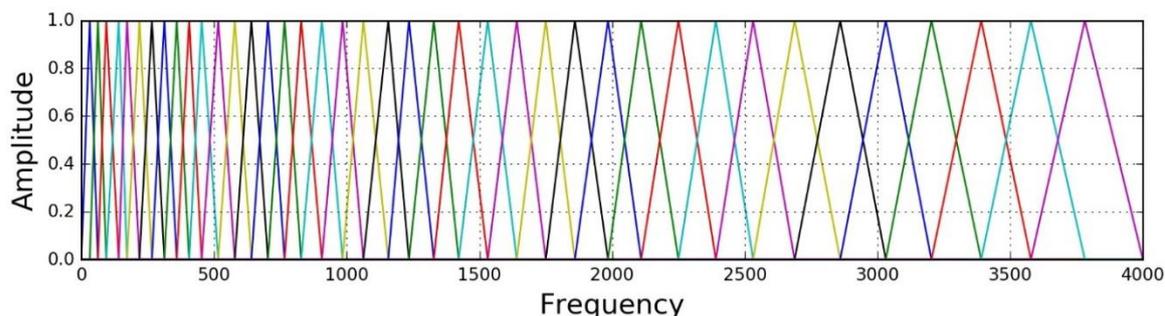


Figure (II-14) : Banque de filtres dans l'échelle de fréquence Mel.

Après avoir appliqué le banc de filtres au spectre de puissance du signal, nous avons obtenu le spectrogramme suivant : [29]

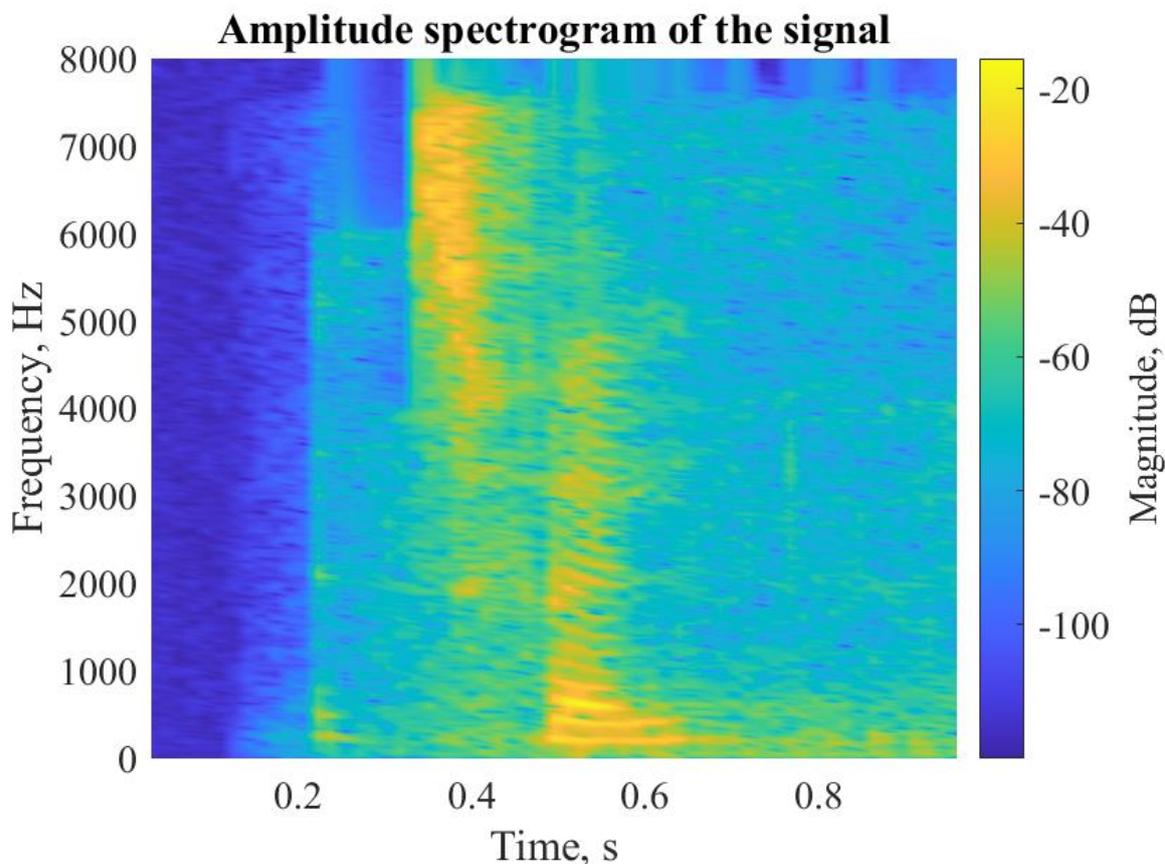


Figure (II-15): Spectrogram du signal.

Le spectrogramme d'un signal décrit son spectre dans le temps et est similaire à l'image d'un signal. Le temps est représenté sur l'axe des x, tandis que la fréquence est représentée sur l'axe des y. C'est comme si nous prenions le spectre à différents moments dans le temps et que nous les couvions ensemble en un seul graphique. Il utilise une variété de couleurs pour représenter l'amplitude ou l'intensité de chaque fréquence : plus la couleur est vive, plus l'énergie du signal est élevée.

Nous utilisons maintenant ces spectrogrammes comme entrée de notre réseau neuronal profond pour former et tester le modèle, comme le montrent les figures (II-16) et (II-17) [26].

➤ **Entraînement :**

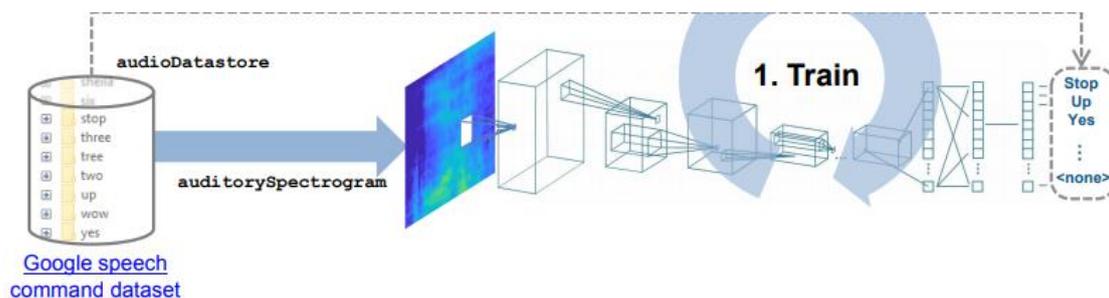


Figure (II-16): Former le mode. [26]

➤ **Test de reconnaissance :**

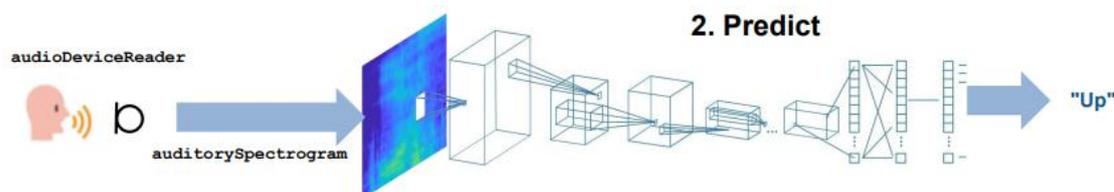


Figure (II-17): Reconnaissance de commandes vocales utilisant l'apprentissage profond pour la classification audio. [26]

II.4. Approches de la reconnaissance vocale multilingue

En fonction de l'objectif de l'application, nous décrivons trois techniques de reconnaissance vocale multilingue : le portage, la reconnaissance multilingue croisée et la reconnaissance vocale multilingue simultanée.

Ces différentes techniques sont déterminées par l'objectif de l'application, c'est-à-dire quelles langues seront reconnues à un moment donné et combien. En outre, la technique utilisée est déterminée par les données d'entraînement disponibles en termes de langue parlée, de locuteur et de paramètres d'enregistrements.

II.4.1. Portage

Le portage est la première méthode de reconnaissance vocale multilingue. Un système de reconnaissance vocale créé pour une langue est porté vers une autre langue pour être utilisé dans cette langue.

Le système de reconnaissance est destiné à la vente de la nouvelle langue, et les données d'entraînement sont uniquement pour la nouvelle langue.

L'ancien et le nouveau système linguistique sont distincts, comme le montre la figure (II-18).

Les algorithmes et les principes de la nouvelle langue sont dérivés du système de reconnaissance de la langue originale, et seuls de petits ajustements sont apportés aux algorithmes afin d'obtenir des performances optimales dans la nouvelle langue.

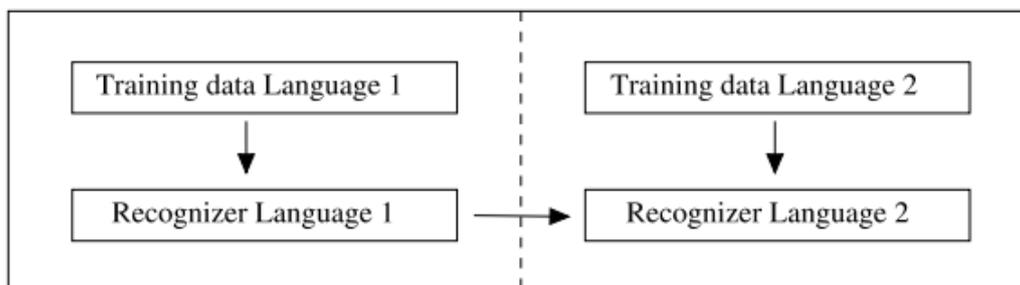


Figure (II -18): Une esquisse du scénario de portage.

II.4.2. La reconnaissance inter linguistique

Cette méthode atteint le même objectif que la méthode de portage. La différence par rapport à la technique précédente est qu'il n'y a pas assez de matériel d'entraînement pour entraîner le reconnaissant dans la nouvelle langue. Par conséquent, il faut développer des techniques permettant d'utiliser du matériel d'entraînement des caractéristiques acoustiques dans la reconnaissance inter linguistique.

IL faut déterminer les langues qui seront utilisées pour l'entraînement du dispositif de reconnaissance.

IL est nécessaire d'identifier les langues qui permettent d'obtenir les meilleures performances de reconnaissance dans les nouvelles langues.

Il est nécessaire de choisir une relation entre les langues utilisées pour l'entraînement et la langue à reconnaître.

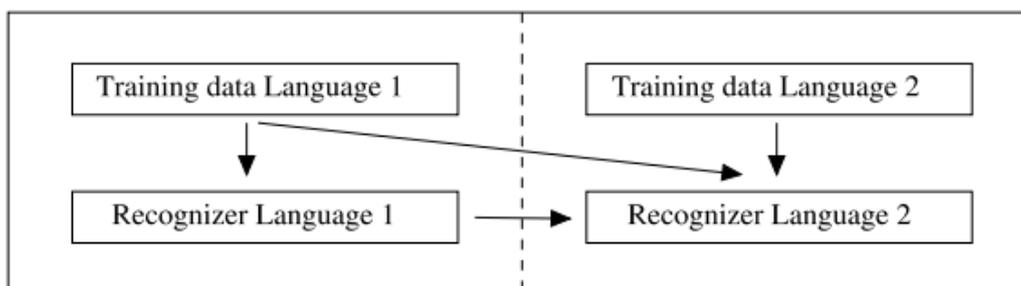


Figure (II -19): Une esquisse du scénario translinguistique

II.4.3. Reconnaissance multilingue simultanée de la parole

La technique de reconnaissance multilingue simultanée constitue le troisième groupe d'approches. Cette technique permet aux applications de reconnaître la parole dans plusieurs langues en même temps. La technologie n'a aucun moyen de savoir quelle langue est parlée.

La figure (II -20) est un dessin approximatif de cette méthode. Chaque langue possède son propre ensemble de données d'entraînement. Il n'existe donc qu'un seul reconnaissant pour toutes les langues concernées.

Pour la reconnaissance vocale multilingue simultanée, il existe deux stratégies principales : l'identification explicite de la langue et l'identification implicite de la langue.

La première approche utilise le signal vocal pour identifier la langue. Le système de reconnaissance vocale pour la langue spécifiée est engagé lorsque la langue est identifiée, et l'énoncé est reconnu.

L'avantage de cette stratégie est qu'elle produit des résultats équivalents à la reconnaissance monolingue, pour autant que l'étape d'identification de la langue soit effectuée correctement.

L'autre méthode consiste en une identification de la langue qui se fait de manière implicite.

Les mots de toutes les langues concernées peuvent être reconnus par la distribution de modèles de langue. Il est possible de passer d'une langue à l'autre. Les mots reconnus peuvent être utilisés pour déterminer la langue parlée. Les mêmes modèles acoustiques peuvent être utilisés pour toutes les langues incluses dans la stratégie.

En outre, au lieu d'un groupe de modèles linguistiques monolingues partageant un nœud de début et de fin commun, un seul modèle linguistique multilingue peut être utilisé. La technique optimale dans le cadre de cette approche peut varier en fonction des langues et des données disponibles. S'il y a peu de données pour une langue, par exemple, les unités acoustiques peuvent être partagées entre les langues. Les unités multilingues peuvent améliorer les performances si les langues sont comparables ou si les locuteurs comprennent des non-natifs.

En revanche, si les langues sont Similaires, il peut être plus avantageux de les garder aussi distinctes que possible afin d'éviter toute confusion. [31]

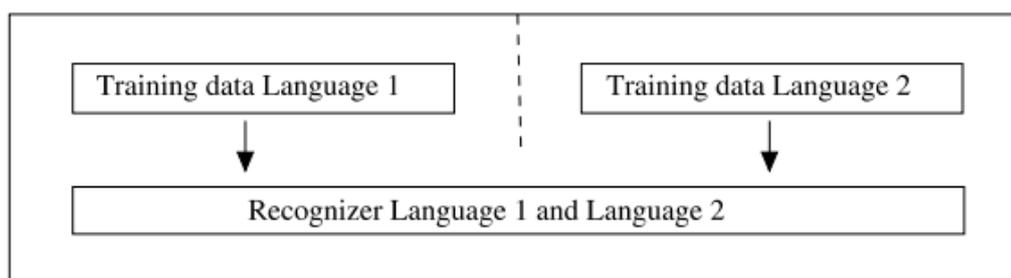


Figure (II-20): Une esquisse du scénario du multilinguisme simultané.

II.5.Reconnaissance multilingue des commandes vocales par apprentissage profond



Figure (II-21): Une vue d'ensemble de l'étape multilingue [31].

II.5.1. L'environnement matériel

L'évolution de l'environnement matériel est caractérisée par :

- Système d'exploitation : Windows 10 Professionnel
- Processeur : Intel® Xeon® CPU ES_1650 v2 @ 3.50 GHz
- Mémoire: 64 GB

II.5.2. Implémentation

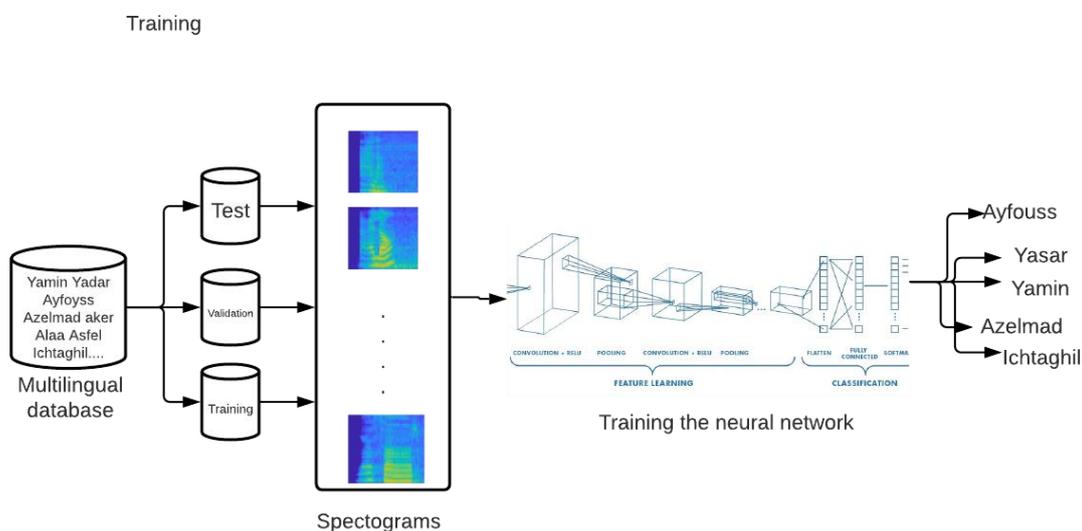


Figure (II-22): Le processus de reconnaissance vocale multilingue [32].

II.5.3. Développement de la base de données multilingue "arabe et amazigh"

Dans cette partie, nous allons présenter les différentes phases de la réalisation de notre base de données à l'aide de captures d'écran.

II.5.4. Enregistrement de la base de données

Pour cette première étape, nous choisissons de créer la base de données à l'aide d'un programme simple dans MATLAB R2020 avec les paramètres suivants.

Les enregistrements ont été enregistrés à l'aide d'un ordinateur via un simple programme MATLAB avec une fréquence d'échantillonnage de 16 KHz, au format (.wav), comprenant 12 commandes (six pour chaque) et des mots inconnus. Chaque dossier contient environ 15 000 enregistrements.

Tableau (II-1) : Caractéristiques de la base de données

	Base de données arabe	Base de données Amazigh
Fréquence d'échantillonnage (Fs)	16KHz	16KHz
Format audio	.wav	.wav
Haut-parleurs	12(5M + 7F)	10(3M + 7F)
Nombre de commandes	6	6
Nombre de mots (Inconnu)	20	
Taille des commandes	15000	15000
Échantillons de bruits de moteur	16	

La base de données des enregistrements multilingues est rassemblée car une commande telle que "up" a la même signification pour le programme même si elle lui est adressée dans différentes langues.

Le tableau suivant résume les commandes du drone en arabe et en amazigh.

Table (II-2) : Le drone commande en Arabe et en Amzaigh

La commande en anglais	La commande en arabe	La commande en Amazigh	Action
UP	Aala	Oussawen	Augmenter l'altitude du drone
DOWN	Asfel	Oukser	Diminuer l'altitude du drone
RIGHT	Yamin	Ayfouss	Déplacer le drone vers la droite
LEFT	Yasar	Azelmad	Déplacer le drone vers la gauche
ON	Ichtaghil	Akker	Allumez les moteurs
OFF	Tawakef	Ekhsi	Arrêtez les moteurs

II.5.5. Séparation des données de test, de validation et de formation

Nous séparons les intervenants entre les ensembles de formation, de validation et de test. Ainsi, pour trouver efficacement les meilleures valeurs pour notre algorithme, la meilleure approche consiste à diviser notre ensemble de données en trois ensembles indépendants :

- Un ensemble de données d'entraînement pour notre algorithme (80% de l'ensemble de données).
- Un ensemble de données de validation pour évaluer notre algorithme d'entraînement que l'algorithme d'entraînement n'observe pas. (10% de l'ensemble de données)
- Un jeu de données de test pour l'évaluation de l'algorithme final. (10% de l'ensemble de données)

Ensemble de données d'apprentissage : Un ensemble d'exemples utilisés pour ajuster le modèle.

Ensemble de données de validation : L'échantillon de données utilisé pour fournir une évaluation non biaisée d'un modèle ajusté sur l'ensemble de données d'apprentissage tout en ajustant les hyper paramètres du modèle. L'évaluation devient plus biaisée à mesure que les compétences de l'ensemble de données de validation sont incorporées dans la configuration du modèle.

Ensemble de données de test : L'échantillon de données utilisé pour fournir une évaluation impartiale de l'ajustement final du modèle sur l'ensemble de données d'entraînement.

II.5.6. Entraînement du réseau

L'entraînement est la partie la plus difficile du Deep Learning car nous avons besoin d'un grand ensemble de données et d'une grande puissance de calcul.

L'entraînement d'un réseau est un problème d'optimisation classique. Il y a une fonction de coût qui exige que le réseau produise des sorties, aussi proches que possible de celles prescrites, puis un algorithme pour trouver les valeurs des

ponds Du réseau qui minimisent la fonction de coût (c'est ce qu'on appelle la rétro-propagation).

II.5.7.Comparaison entre la formation monolingue et la formation multilingue

La figure (II-24) représente la formation avec seulement la base de données anglaise (monolingue).

La figure (II-25) représente l'entraînement avec les bases de données anglaise, arabe et amazighe (multilingue).

Pour notre entraînement, nous utilisons l'optimiseur Adam avec une taille de mini-batch de 128. Nous nous entraînons pendant 25 époques et réduisons le taux d'apprentissage d'un facteur 10 après 20 époques.

Dans Matlab, il suffit d'appeler cette fonction train.

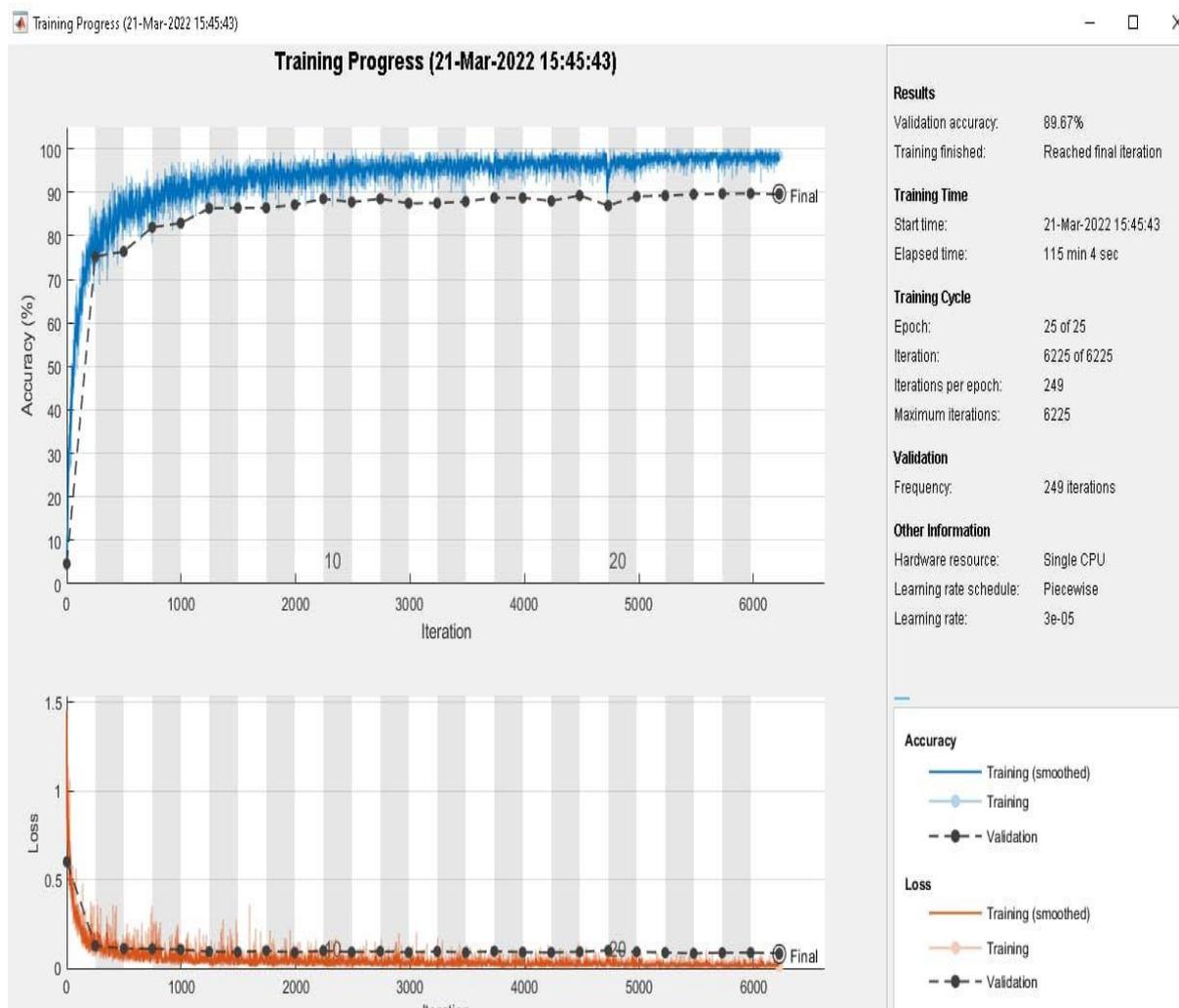


Figure (II-24) : Progression de la formation pour le jeu de données monolingue (anglais).

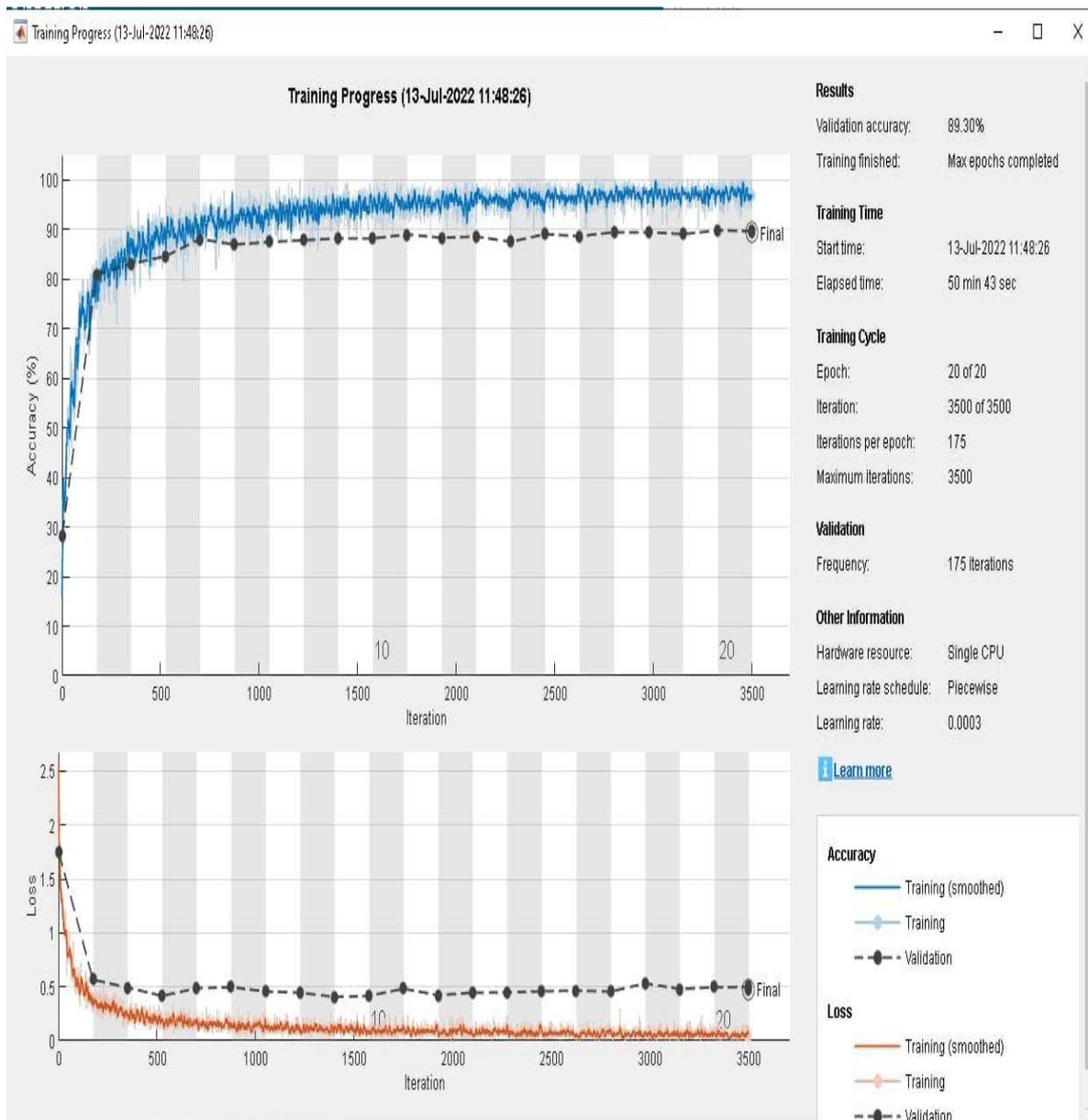


Figure (II-25) : Progression de la formation pour le jeu de données multilingue (Anglais, Arabe et Amazigh).

II.5.8. Matrice de confusion

Vous pouvez également être plus analytique et utiliser une matrice de confusion pour évaluer la performance du réseau sur les données de test, comme le montrent les figures (II-26).

Le tableau peut nous suggérer des choses à améliorer. Par exemple, dans la matrice de confusion pour le jeu de données multilingues, six des vrais "RIGHT", "YAMIN" ou "AYFOUSS" ont été reconnus à tort comme "LEFT", "YASSAR" ou "AZELMAD" à cause des sons qui ont été confondus.

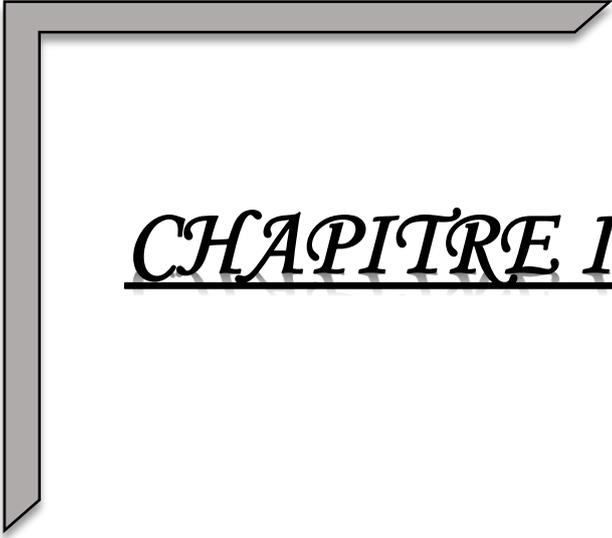


Figure (II-26) : Matrice de confusion pour un ensemble de données multilingues.

II.6. Conclusion

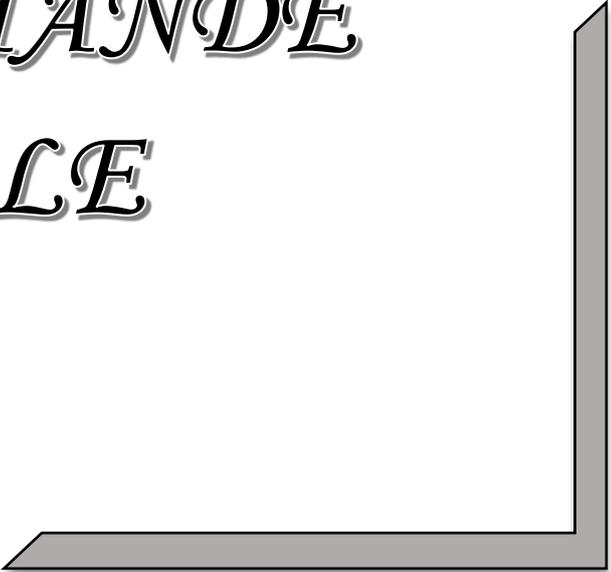
Le traitement de la parole est essentiel dans tout système vocal, qu'il s'agisse de la reconnaissance automatique de la parole (ASR), de la reconnaissance du locuteur ou de toute autre chose... À titre d'exemple, les coefficients cepstraux de fréquence Mel (MFCC) ont été des caractéristiques populaires pendant longtemps, mais les banques de filtres ont récemment gagné en popularité.

Avec l'essor de l'apprentissage profond, les couches initiales des réseaux profonds ont essentiellement remplacé l'extraction de caractéristiques, mais surtout pour les données d'image.



CHAPITRE III

DEBRUITAGE DE
LA COMMANDE
VOCALE



CHAPITRE III : DEBRUITAGE DE LA COMMANDE VOCALE

III.1 Introduction

Notre recherche vise à développer des "drones de communication" qui peuvent interagir naturellement avec les humains et les aider dans leurs quotidiennes.

Des études antérieures en robotique ont mis l'accent sur l'intérêt de l'incarnation des robots, montrant l'efficacité de l'expression faciale, le regard et les gestes. Récemment, plusieurs robots pratiques ont été développés, comme des outils thérapeutiques, des outils d'orientation dans les musées et des divertissements, Depuis les robots élargissent leur domaine de travail dans notre vie quotidienne. Donc la communication basée sur le langage est indispensable, afin d'exploiter pleinement la présence humaine.

Cependant, l'une des difficultés concerne la reconnaissance de la parole dans les environnements bruyants. La technologie actuelle a une bonne performance dans la reconnaissance d'énoncés formels dans des environnements non bruyants, mais les performances se dégradent dans des conditions plus réalistes.

Plusieurs chercheurs se sont récemment efforcés de résoudre ce problème, appelé "audition robotisée" La plupart de ces travaux font appel à la technologie des réseaux de microphones, pour localisation et la séparation des sources sonores, avant la reconnaissance de la parole [34].

Dans ce chapitre, nous développons notre système ASR (reconnaissance automatique de la parole), pour qu'il tienne compte de l'influence du bruit présent dans l'environnement réel et précisément le bruit engendré par la combinaison (Moteurs+ Hélices).

Pour cela nous allons préparer un bloc de filtrage de la parole pour but d'atténuer le bruit avant d'injecter la commande vocal dans le system ASR afin de faciliter la reconnaissance pour notre robot.

III.2 Le traitement de la parole

Le traitement de signal est un domaine de l'électronique pouvant s'appliquer à n'importe quel phénomène physique à condition que ce dernier puisse être converti, au moyen de capteurs adéquats, en signal électrique.

Le signal de parole, est une représentation électrique, par l'intermédiaire d'un microphone, de mots, phrases ou textes prononcés par un individu qui consiste nos command vocal pour le Drone.

Ce signal de parole est souvent altéré par des phénomènes perturbateurs, appelés bruit électrique, qui rendent l'information sonore dégradée voire incompréhensible du coup cette élément il va démineur le potentielle de reconnaissance de notre système [35].

III.2.1. Description de la parole.

Par définition, la parole c'est : « La faculté d'exprimer la pensée par le langage articulé. »

➤ Caractéristiques

Comme tous les autres sons arrivant aux oreilles, c'est une onde de pression. Elle varie suivant les personnes (hommes, femmes ou enfants), suivant l'intensité (énervement, chuchotement)... La parole est décrite dans plusieurs domaines :

- Le domaine temporel
- L'enveloppe, qui contient l'énergie du signal
- La structure fine, qui contient les différentes variations du signal
- Le domaine fréquentiel :
- La fréquence fondamentale, F_0 , et ses harmoniques
- Les formants, F_1 , F_2 , ...

Des méthodes existent pour trouver ces différentes informations. L'enveloppe temporelle peut être obtenue en calculant le module de la transformée

de Hilbert. La F_0 se calcul par une autocorrélation du signal, Et les formants par la méthode de la LPC (Linear Predicting Coding) [39].

➤ **Fréquence fondamentale**

La fréquence fondamentale dans le signal de parole « représente » la vibration des cordes vocales (le signal glottique). Ce signal est composé de la fondamentale et de ses harmoniques. Elle caractérise la personne qui parle :

- Pour un homme : $\approx 100Hz$
- Pour une femme : $\approx 200Hz$
- Pour un enfant : 300 à 400Hz

III.2.2. La notion du bruit

Le bruit continuellement présent autour de nous représente une caractéristique fondamentale de notre environnement. Par ailleurs, le bruit, structure sous la forme du langage nous permet la communication, l'expression, la socialisation, etc. mais, comme toute bonne chose, il est dangereux d'en abuser ! Trop de bruit nuit à notre santé physique et mentale. Le bruit dans l'environnement (également appelé bruit résidentiel ou bruit domestique) est défini comme le bruit émis par toutes les sources sauf le bruit sur le lieu de travail industriel. Les sources principales de bruit à l'intérieur sont les systèmes de ventilation, les machines de bureau, les appareils ménagers et le voisinage [36]. De manière beaucoup plus générale, on désigne par bruit, tout signal nuisible de nature aléatoire qui se superpose au signal utile, porteur de l'information. Dans les domaines de l'électronique, de nombreux phénomènes physiques entachent les signaux). [37]

De ce fait, le traitement des signaux ou les informations utilise la méthode et technique de filtrage. Ce qui nécessite une opération courante et très importante en traitement du signal. Le filtrage à un but essentiel d'améliorer la qualité du signal en rejetant la composante perturbatrice tout en conservant la partie porteuse de l'information utile. [38]

III.2.3 La réduction du bruit

La base de la technique de réduction de bruit dans un système monovoie contient deux principales techniques. La première est la « Soustraction Spectrale » et la seconde est le « filtrage de Wiener ». La réduction de bruit monovoie est effectuée à partir d'un seul microphone, il n'y aura alors qu'un seul signal à traiter. Les démonstrations qui vont suivre seront décrites de la même façon pour les deux méthodes. Dans tous les cas, une amélioration de la compréhension de la parole est nécessaire, ces algorithmes sont faits pour améliorer la parole par rapport au reste du signal [40].

Réduire le bruit dans un système se joue sur l'amélioration de la qualité des signaux de parole bruités tout en minimisant la perte d'intelligibilité pouvant être causée par les traitements effectués sur ces signaux à partir d'un signal acquis.

Au départ, on peut considérer que le signal qui arrive au microphone est composé d'un signal utile qui est la parole et d'un bruit qui est ce que l'on doit atténuer. Le but d'un tel algorithme dans l'aide auditive ou bien dans l'implant cochléaire est d'améliorer le rapport signal sur bruit en sachant que le signal est considéré comme étant de la parole et que le reste est un bruit.

On considère le signal arrivant au microphone comme un mixage entre un signal utile et un bruit :

$$s(t) = u(t) + b(t) \quad (3.1)$$

Ou :

t : représente le temps.

$b(t)$: représente la composante perturbatrice.

$s(t)$: Le signal sonore reçu par le microphone.

$u(t)$: Le signal utile à la reconnaissance.

Cette équation (3.1) doit être modifiée pour traiter le signal. Il doit être échantillonné. De ce fait, l'équation devient alors :

$$S(n) = U(n) + B(n) \quad (3.2)$$

Ou :

n : Représente le nombre d'échantillon

$B(n)$, $U(n)$ et $S(n)$: Représente les composantes de l'équation (3.1) échantillonner.

III.2.4 Estimation du niveau de bruit

Les algorithmes d'estimation du niveau de bruit sont très importants pour le filtrage du signal audio. Il s'agit d'estimer la densité spectrale de bruit, c'est-à-dire à la fois le niveau sonore et la répartition spectrale.

Dans ce type d'approche, le bruit est considéré comme stationnaire ou quasi-stationnaire, c'est-à-dire que les statistiques du bruit de fond varient lentement par rapport à celles du signal source.

La parole est constituée d'une alternance de sons et de silences, l'estimation du bruit peut donc se faire sur les périodes de silence. On suppose alors que le bruit de fond conserve les mêmes statistiques en dehors de ces périodes. Pour détecter les zones de silences et les zones de sons, on utilise un détecteur d'activité vocale basé à la fois sur le niveau sonore et le contenu spectral.

L'estimation du niveau de bruit peut également se faire de manière continue. On considère approximativement que toute hausse instantanée du niveau au-dessus de la valeur moyenne estimée du bruit témoigne de la présence de signal.

[41]

La première étape consiste à transformer le signal temporel par l'intermédiaire de la FFT en nombre complexe et plus particulièrement le module. Il est très rare d'utiliser la phase du signal car l'amplitude suffit habituellement. L'équation (3.2) devient alors :

$$S(f) = U(f) + B(f) \quad (3.3)$$

Et en module :

$$|S(f)| = |U(f)| + |B(f)| \quad (3.4)$$

Où :

f : représente les fréquences

S : représente le spectre du signal capté par le microphone

U : représente le spectre du signal utile

B : représente le bruit

Le calcul du bruit se fait généralement sur la densité spectrale de puissance (DSP) qui est obtenue en calculant l'énergie par fréquence du signal (théorème de Parseval).

$$P_s(f) = |S(f)|^2 \quad (3.5)$$

Où :

$P_s(f)$: représente la DSP du signal

L'équation devient alors :

$$P_s(f) = P_u(f) + P_b(f) \quad (3.6)$$

Une fois le module du bruit considéré, il faut essayer de l'atténuer dans la trame correspondante, il existe plusieurs méthodes (classiquement utilisées) pour exécuter ce débruitage. Les deux principaux filtres en monovoie seront présents dans les sous sections suivantes.

III.3 Les algorithmes de réduction de bruit

L'aspect principal de la réduction de bruit est justement d'améliorer le signal mais par la même occasion de garder l'aspect dit : temps réel. En effet, si le débruiteur enlevé parfaitement le bruit dans le signal capté par le microphone, mais qu'il lui faut environ une seconde de traitement pour l'effectuer alors il ne sert à rien car il est impossible de donner la commande au robot avec des latences aussi grandes. De plus si l'atténuateur permet d'augmenter le rapport signal sur bruit de la trace capté par les microphones, au final, le signal est toujours d'étérioré

La méthode pour d'ébruiter un signal peut être séparé en deux étapes : La première consiste à estimer le bruit dans le signal et la seconde à rehausser le signal utile (la parole) ou atténuer le bruit. Généralement ces méthodes sont effectuées dans le domaine spectral par l'intermédiaire de la transformée de Fourier rapide (FFT, Fast Fourier Transform). [41]

III.3.1 Base de la Transformée de Fourier rapide

La Transformée de Fourier (TF) permet de passer d'un espace temporel à une dimension fréquentielle. Cette fonction est basée (Fourier) sur la décomposition d'un signal en somme de sinusoides (ou cosinus) qui sont des périodes multiples de la fréquence fondamentale (fréquence de base). Les coefficients de chaque fréquence représentent le niveau d'énergie

La formule de la TF est :

$$G(f) = \int_{-\infty}^{+\infty} g(t) \cdot e^{-j2\pi ft} dt \quad (3.7)$$

Où :

t : représente le temps

f : représente les fréquences

$g(t)$: représente la fonction temporelle

Attention cependant au repliement après l'échantillonnage pour respecter le théorème de Shannon, $f_{\max} \leq f_e / 2$, qui indique que la fréquence maximale du signal d'entrée ne doit pas dépasser la fréquence d'échantillonnage divisée par deux.

En pratique, la TFD n'est appliquée que sur une partie du signal et non sur le total. Cette partie est appelée « Trame ». Une fenêtre de pondération est ensuite appliquée pour éviter les discontinuités au début et à la fin de la trame. Introduisant dans le spectre du signal des informations non existantes

III.3.2 Algorithme de calcul de la FFT

Le but de l'algorithme de la transformée de Fourier rapide est d'optimiser le temps de calcul de la TFD. En effet, la TFD classique demande énormément de calcul. Il faut N multiplications complexes et $N - 1$ additions et il y a N composantes à calculer. Ceci conduit à N^2 multiplications complexes et à $N(N - 1)$ additions complexes. Il a fallu rechercher un algorithme qui permette de diminuer fortement ce nombre d'opération pour optimiser le temps de calcul dans les processeurs.

Le principe de la FFT repose sur des séparations des échantillons deux à deux du signal (échantillons pairs et impairs). On a au départ une trame de taille N , il y a alors $k = 2r$ éléments de rang pair et $k = 2r + 1$ de rang impair. On peut alors décomposer en deux TFD

$$G_n(k) = \sum_{r=0}^{\frac{N-1}{2}} g_{2r}(n) \cdot e^{-j \frac{2\pi n(r)}{N}} + \sum_{r=0}^{\frac{N-1}{2}} g_{2r+1}(n) \cdot e^{-j \frac{2\pi n(2r+1)}{N}} \quad (3.8)$$

Ou :

$$G_n(k) = \sum_{r=0}^{\frac{N-1}{2}} g_{2r}(n) \cdot e^{-j \frac{2\pi nr}{N}} + e^{-j \frac{2\pi nr}{N}} \sum_{r=0}^{\frac{N-1}{2}} g_{2r+1}(n) \cdot e^{-j \frac{2\pi nr}{N}} \quad (3.9)$$

Où le premier terme représente les rangs pairs et le second les rangs impairs. Ceci représente la décomposition de la TFD d'ordre 2. On peut généraliser en prenant :

$$G_n = G_n^e + e^{-j\frac{2\pi n}{N}} G_n^o \quad (3.10)$$

Le calcul se fait ensuite sur le résultat de l'ordre deux pour obtenir le rang 4 et ainsi de suite. Les temps de calculs sont alors de $N = 2^p$ avec $p = \log_2(N)$. Ils sont fortement diminués. Le coût de traitement de l'algorithme est de $N \cdot \log_2(N)$ additions et $(N/2) \cdot (\log_2(N) - 1)$ multiplications

Ainsi la FFT conduit à un gain de temps de calcul non négligeable. Il conviendra ensuite d'interpréter le résultat. En effet, le résultat étant sous forme complexe, on peut alors facilement calculer le module et la phase des fréquences correspondantes à la trame [40].

III.3.3 Le Spectrogramme

L'intérêt du spectrogramme est de pouvoir représenter le spectre en évoluant dans le temps. Le nom scientifique de la fonction mathématique associée à cet outil, plus communément appelé « Spectrogramme », est la Transformée de Fourier à Court Terme (TFCT). Ce nom provient de l'analyse effectuée sur des fenêtres de support temporel fini. Une autre dénomination de cette représentation est « sonagramme ». Il s'agit d'une marque déposée Kay Electronics.

Le principe du spectrogramme est de « découper » le son en trames. Pour chacune de ces trames on calcule une transformée de Fourier comme le schématise la figure (III-1). Ce spectre est alors représenté à un temps correspondant à celui du centre de la fenêtre, sous forme d'un code de couleur. [41]

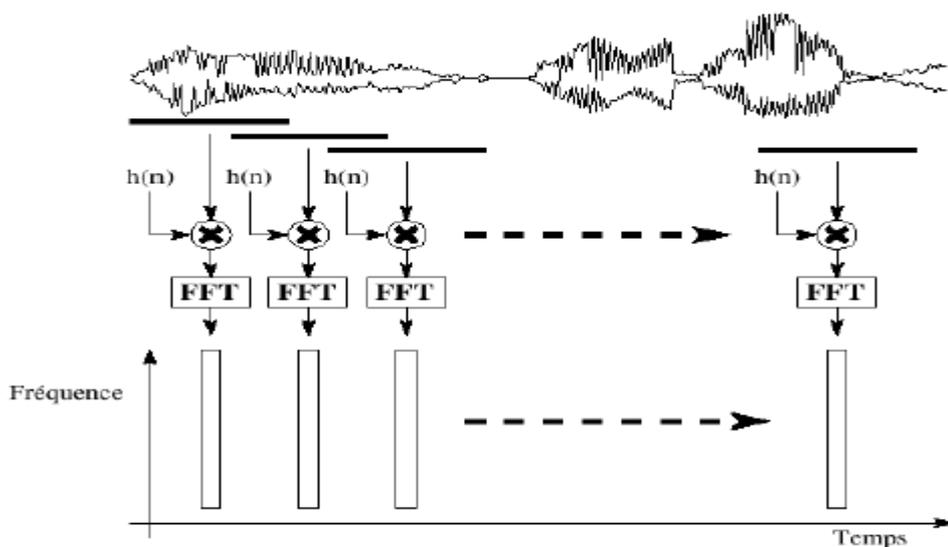


Figure (III-1):Description schématique de l'analyse temps/fréquence par la FFT.

[41]

Ou : la valeur $h(n)$ correspond à la trame découpé pour but de transformation

Une fois que nous avons les données, nous voulons les transformer en spectrogrammes Mel. La figure (III-1) montre un échantillon aléatoire de trois fichiers dans l'ensemble de données et le spectrogramme est produit à partir de ceux-ci.

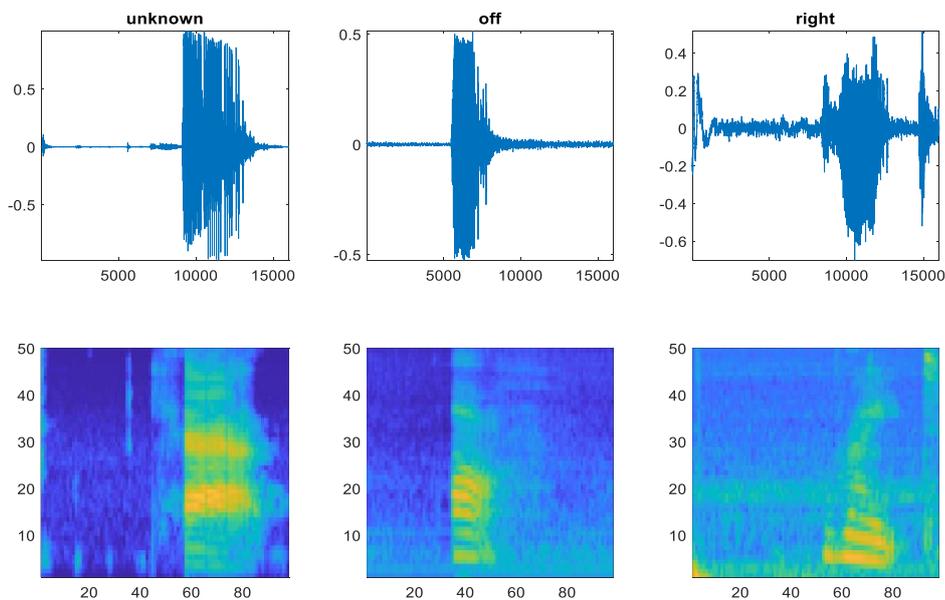


Figure (III-2): La transformation des signaux en spectrogramme.

III.4 La soustraction spectrale

La soustraction spectrale est le débiteur de plus ancien. Elle a été introduite par Boll [42]. Comme son nom l'indique, elle effectue son travail dans le domaine spectral et a pour principe de soustraire le bruit estimé au signal. L'estimation du bruit se fait sur plusieurs trames d'acquisition ($\approx 300ms$).

$$P_x(f) = P_U(f) + P_B(f) - EP_B(f) \quad (3.11)$$

Ou :

$EP_B(f)$: représente l'estimation du bruit

$P_x(f)$: représente la DSP après la soustraction spectrale

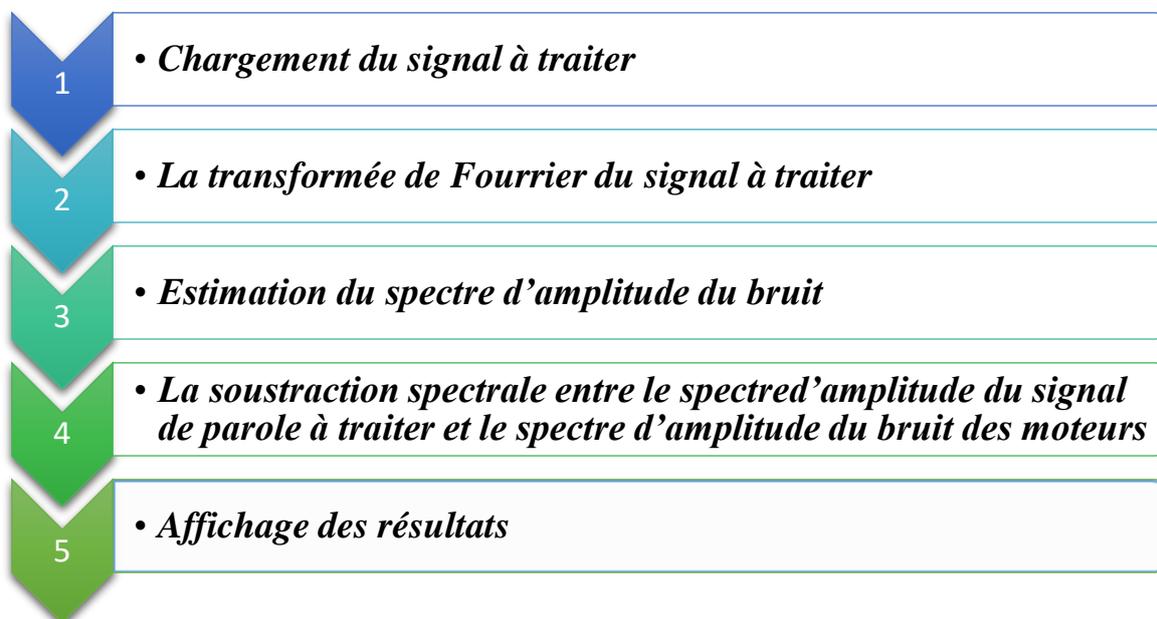


Figure (III-3) : Organigramme de filtrage.

III.4.1 Le Principe

Il existe deux versions de base pour la soustraction spectrale se différenciant par l'amplitude ou la puissance.

$$|X(f)| = |S(f)| - |EB(f)| \quad (3.12)$$

Dans ce cas, il s'agit de la « soustraction spectrale d'amplitude ». Mais le plus souvent comme indiqué dans le paragraphe précédent, la soustraction se fait au niveau de la puissance.

$$|X(f)|^2 = |S(f)|^2 - |EB(f)|^2 \quad (3.13)$$

Le problème de ces deux équations (3.12) et (3.13), est que le second terme peut être négatif. On peut le rendre positif en changeant de signe ou bien en l'annulant comme dans l'équation (3.14). C'est l'amélioration que l'on peut proposer.

$$|X(f)|^2 = \begin{cases} |S(f)|^2 - |EB(f)|^2 & \text{si } |S(f)|^2 > |EB(f)|^2 \\ 0 & \text{sinon} \end{cases} \quad (3.14)$$

➤ Pratiquement

En prenant en compte la complexité de l'estimation du bruit des moteurs en temps réel on penche vers une autre alternative ou on va remplacer la valeur du bruit estimé ($EP_B(f)$) dans (3.11) par un ensemble de mesure de bruit prélevé directement de notre drone, avec ces échantillons on peut déduire la forme spectrale du bruit des moteurs à plusieurs niveaux grâce à un algorithme de FFT

L'équation (3.11) devient alors :

$$P_X(f) = P_U(f) + P_B(f) - MP_B(f) \quad (3.15)$$

Ou :

$MP_B(f)$: Représente la mesure du bruit

Dans le cas où : $MP_B(f) \approx P_B(f)$

On obtient :

$$P_x(f) = P_U(f) \quad (3.16)$$

Et ça représente le cas parfait qui est pratiquement inatteignable.

Le passage dans le domaine temporel se fait par une iFFT. La phase du signal d'entrée est gardée, une estimation du bruit de la phase s'avère être une tâche très compliquée

➤ **Les mesures**

Pour prélever un maximum d'échantillons varié de bruit depuis notre drone, on à procéder en tenant compte de deux paramètres :

-La portée du microphone (La distance entre la source du bruit et le capteur d'enregistrement)

-La puissance des moteurs au décollage de notre drone (La puissance du bruit).

III.4.2 Application et résultat

La technique de filtrage par la méthode de la soustraction spectrale comporte les étapes suivantes :

➤ **Première étape :**

Cette étape consiste à charger le signal de parole bruité par le bruit des moteurs $S(t)$, pour cela nous avons utilisé la commande « Audio Read ». Figure (III-4)

On prend l'exemple de la commande « Yes »

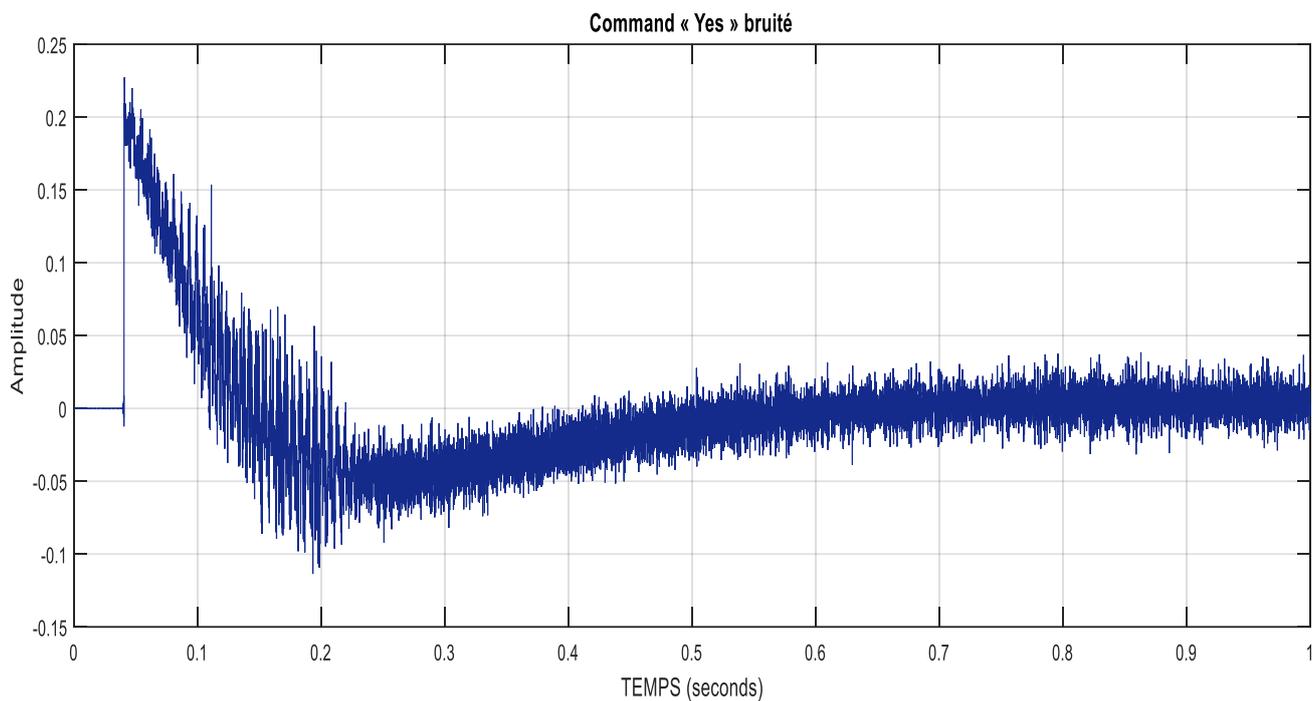


Figure (III-4) : Signal en domaine temporelle de la commande « Yes » bruité par le son des moteurs.

L'application de la transformée de Fourier au signal de parole de la figure (III-4) nous a donné l'allure de la Figure (III-5).

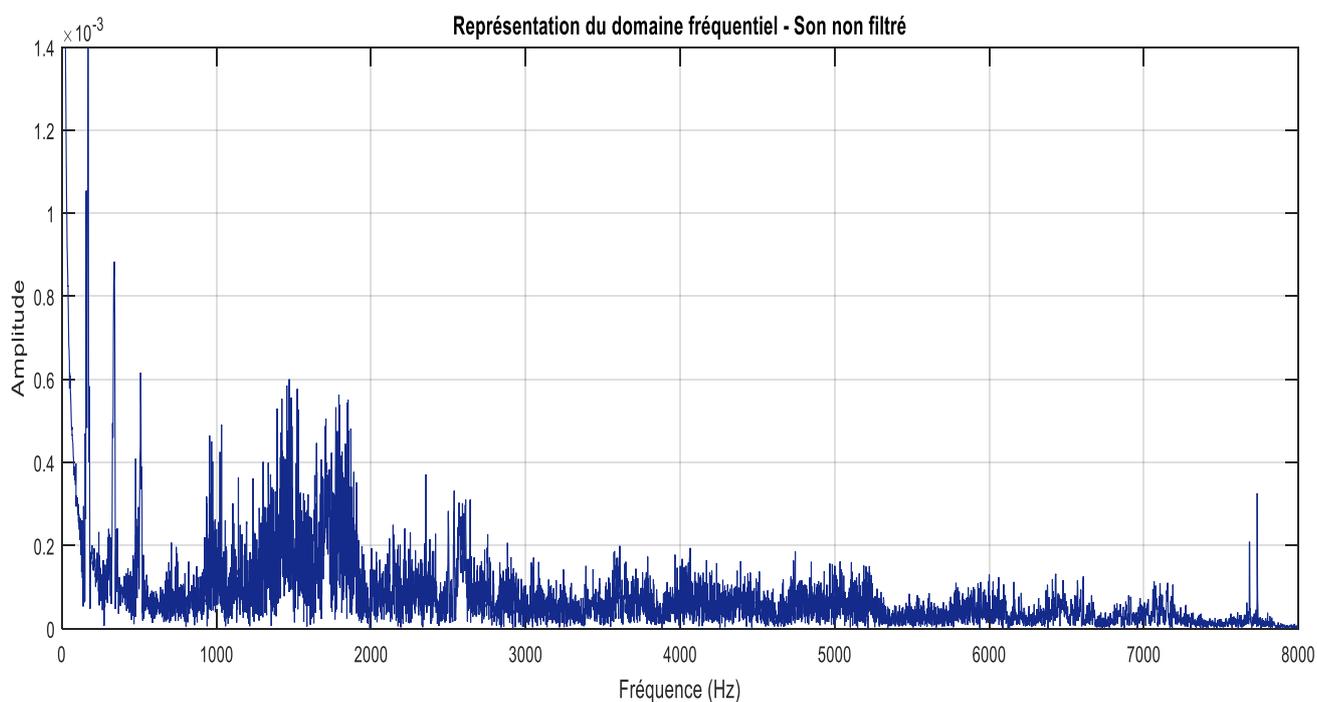


Figure (III-5) : La Transformée de Fourier du signal de parole bruité.

La figure (III.5) montre que le spectre d'amplitude du signal de parole, $S(f)$ Contient un bruit, ayant une amplitude importante sur toute la longueur du signal.

➤ **Seconde étape :**

Cette étape consiste à trouver le spectre d'amplitude du bruit $B(f)$ à une fréquence supérieure à 20Khz, pour cela nous allons utiliser l'algorithme de la FFT sur les mesures de bruit effectuer afin d'afficher le spectre d'amplitude de bruit.

Vu qu'on a les mesures du bruit à notre disposition on profite à projeter ce signal en domaine temporelle Figure (III-7) pour une analyse plus simple.

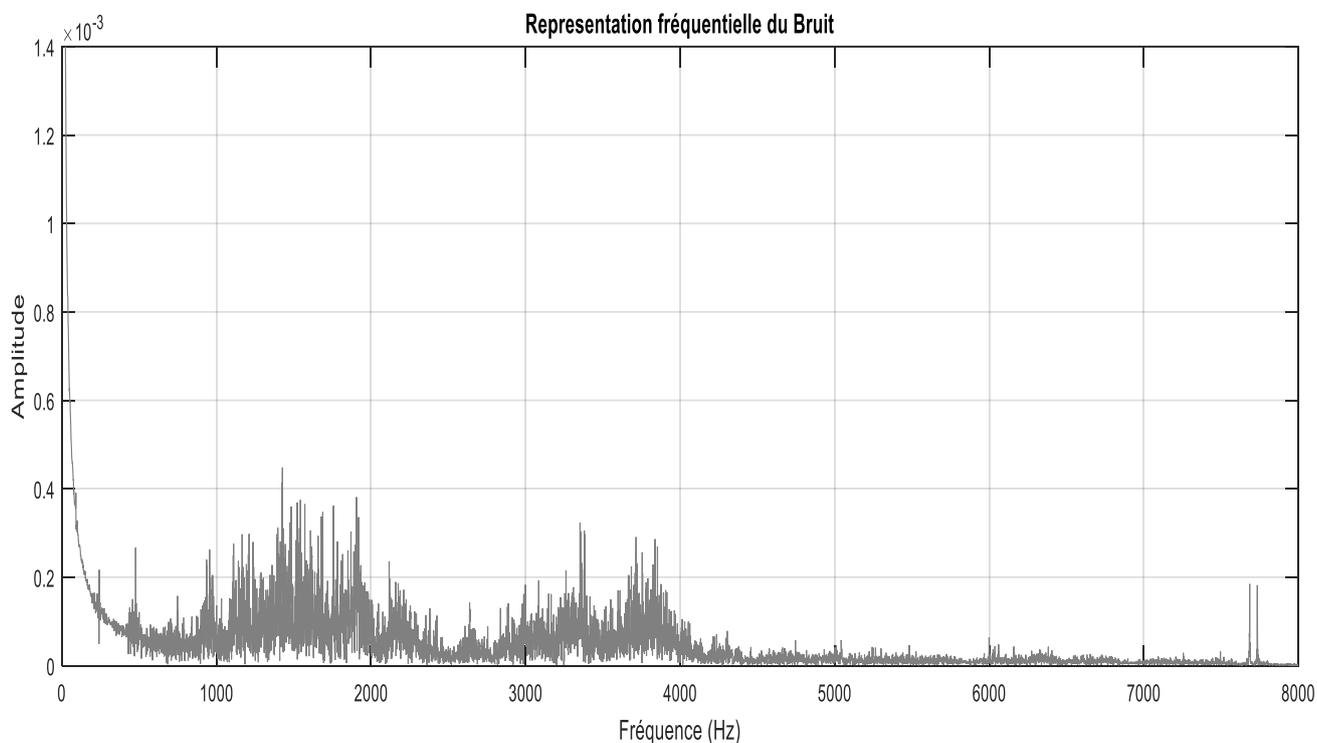


Figure (III-6) : Le spectre d'amplitude de bruit.

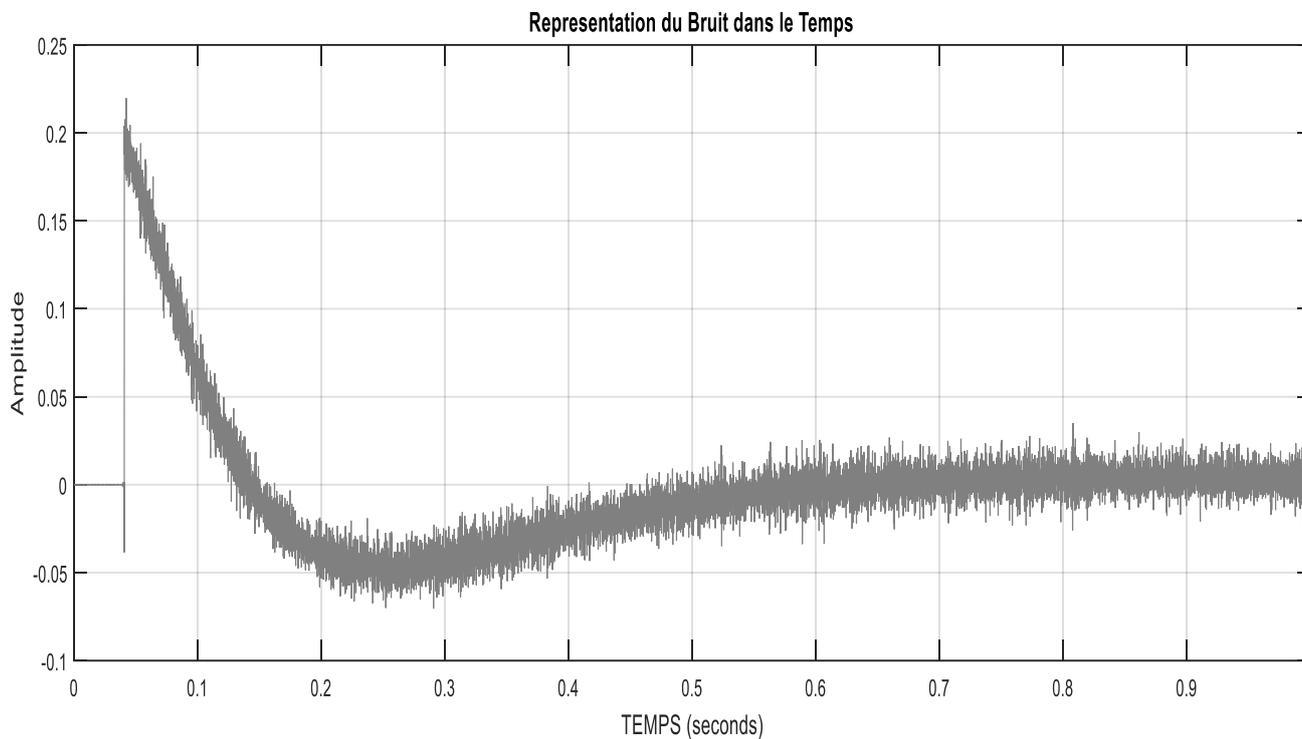


Figure (III-7) : La représentation du bruit dans le temps.

➤ **Troisième étape :**

Cette étape consiste à faire la soustraction spectrale entre le spectre d'amplitude du signal bruité $S(f)$ et le spectre d'amplitude du bruit des moteurs $B(f)$ afin d'obtenir le spectre d'amplitude du signal de parole filtré $U(f)$

L'application de la soustraction spectrale au signal de parole nous a donné le graphe de la figure (III.8)

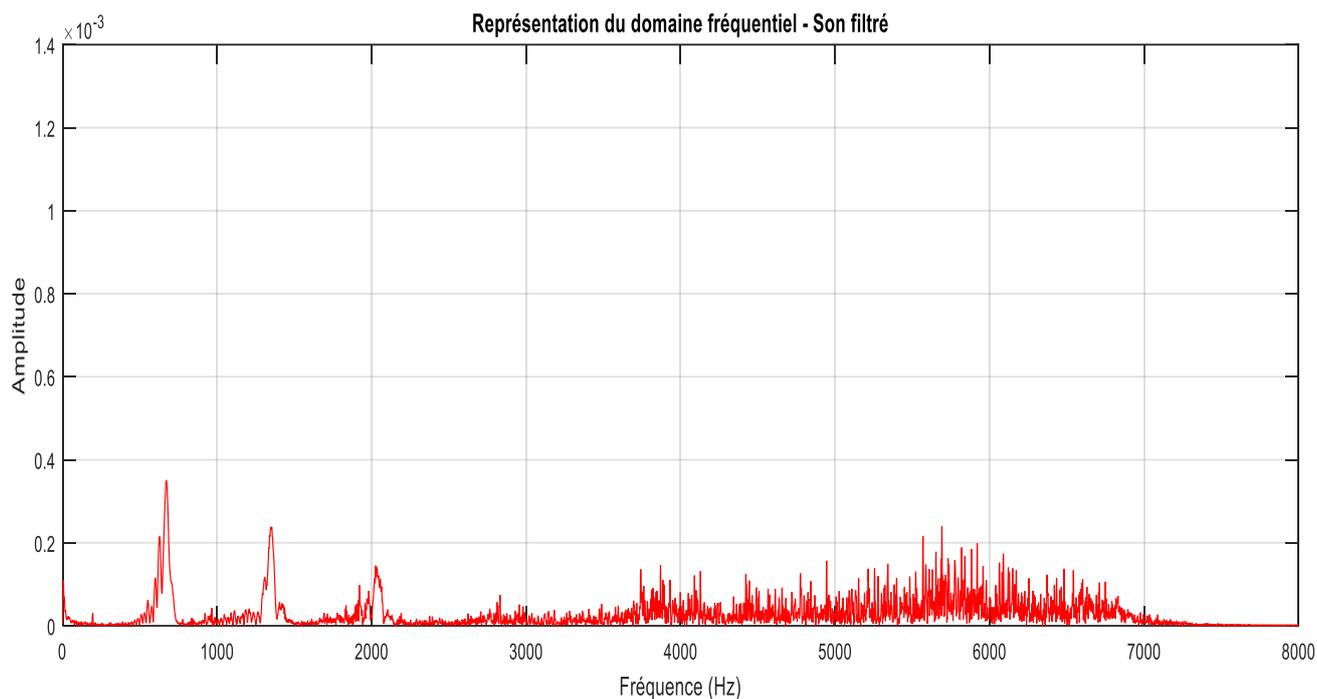


Figure (III-8) : Le spectre d'amplitude du signal de parole débruité.

La figure (III.8) montre que le spectre d'amplitude du signal de parole est fortement atténué ce qui signifie que le bruit est éliminé.

Pour montre le degré d'atténuation on remet les deux signaux au même graphe dans la figure (III-9).

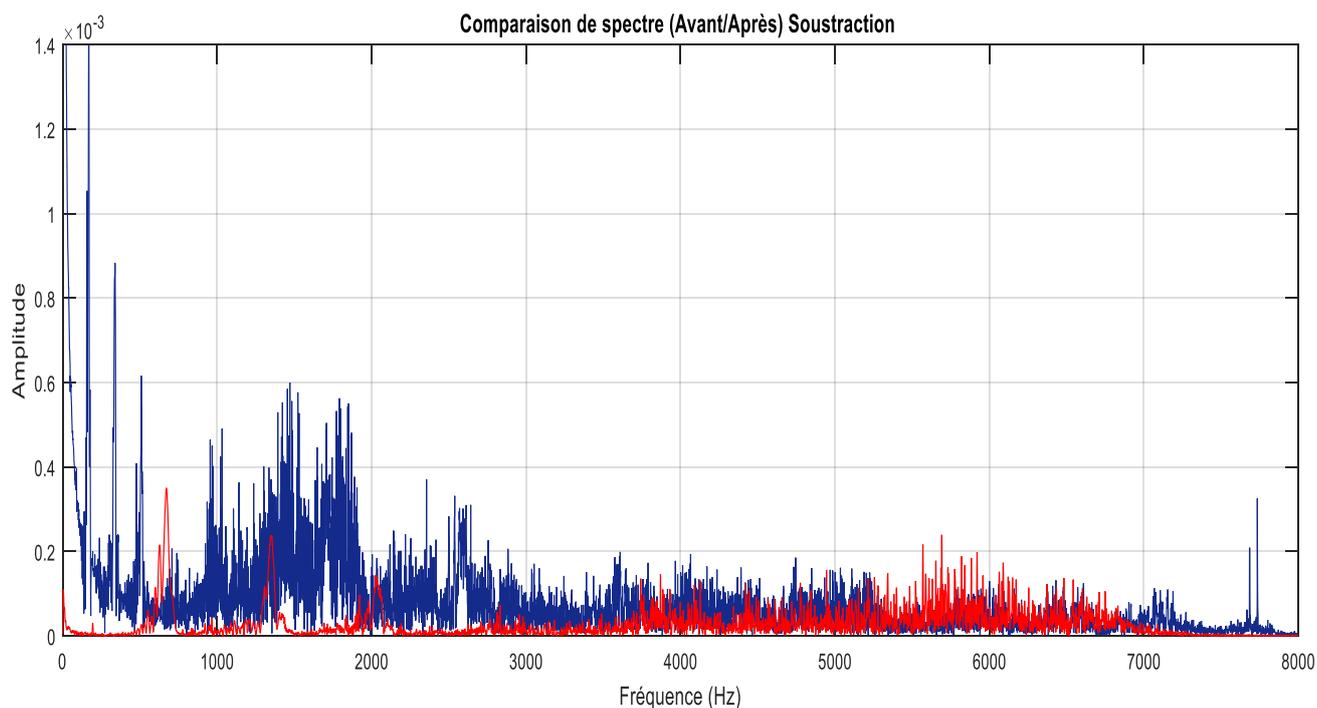


Figure (III-9) : Comparaison entre les graphes (avant/après) soustraction

➤ **Quatrième étape :**

Reconstruire le signal $U(t)$ à partir du résultat de la soustraction avec la iFFT pour obtenir le signal utile à la reconnaissance vocale figure (III-10).

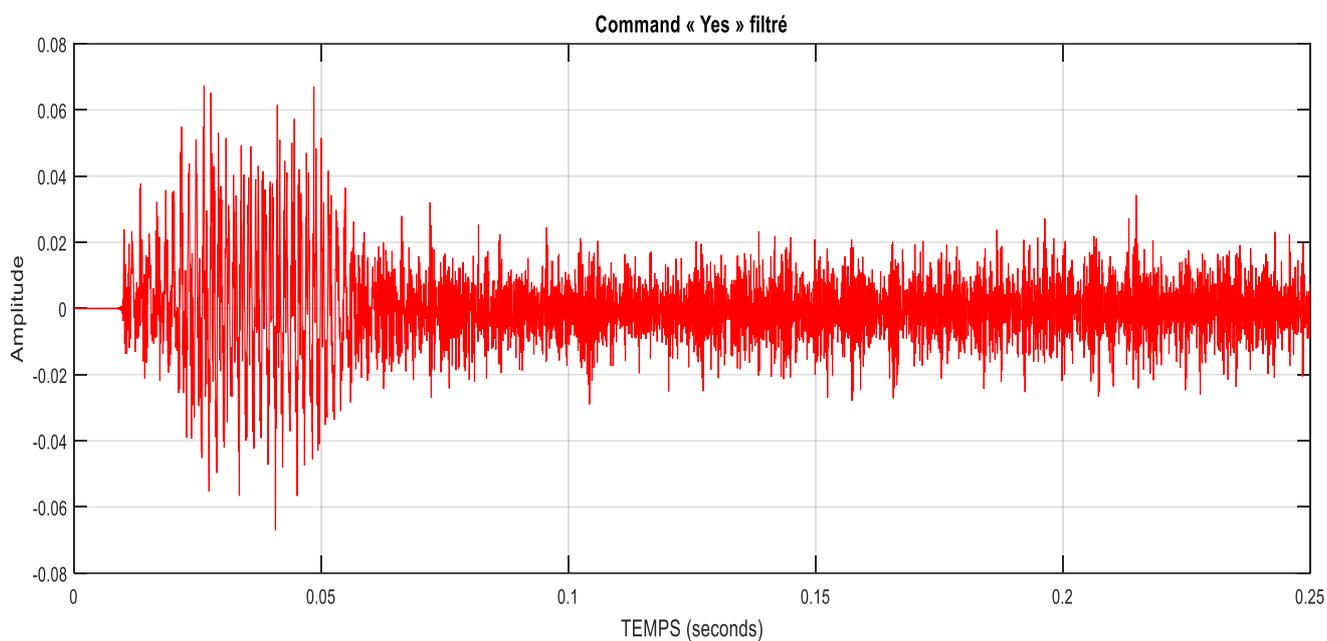


Figure (III-10) : Le signal filtré dans le domaine temporelle

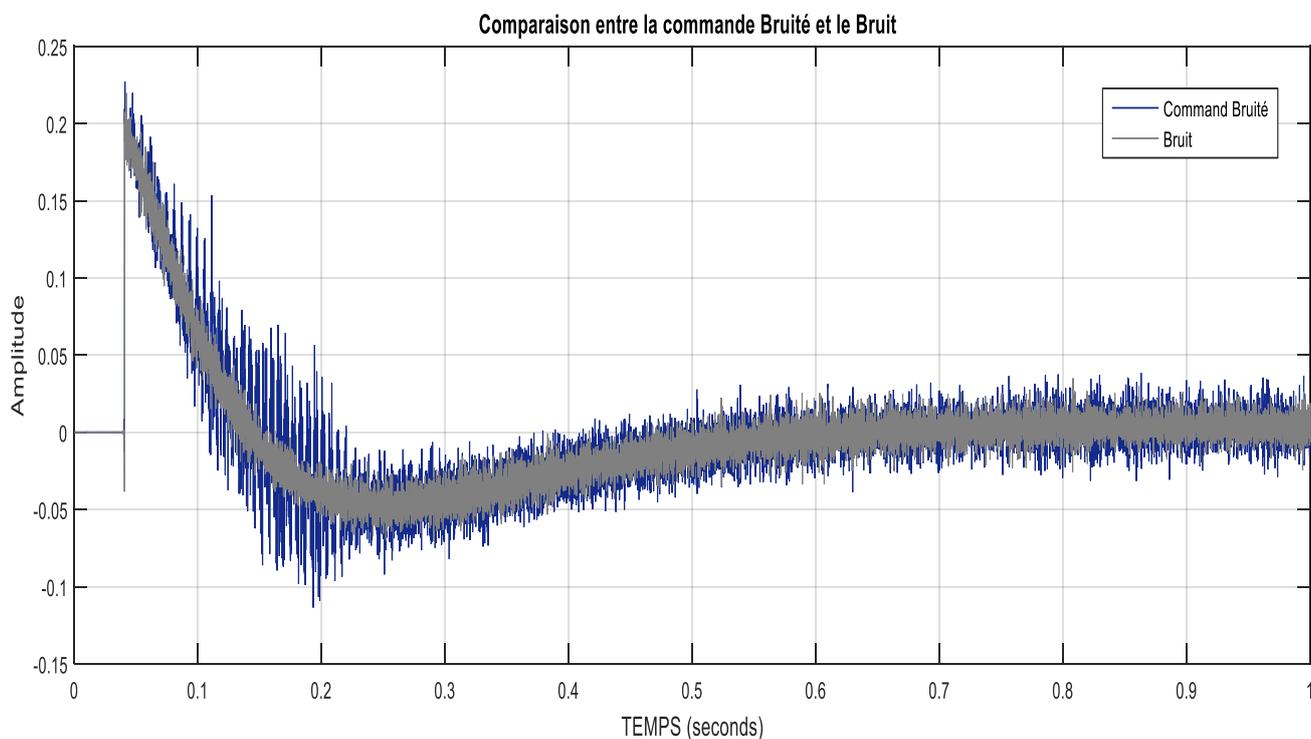


Figure (III-11) : Comparaison entre la commande bruitée et le bruit des moteurs.

La figure (III-11) montre la différence entre le signal $S(t)$ et $B(t)$ qui est notre signal utile $U(t)$.

On remarque que le signal $S(t)$ est affecté par le bruit ou il prend la même forme, et il débord en amplitude surtout dans la partie [0,1 à 0,3] seconds, c'est dans cette partie où la présence de notre signal utile est bien visible.

III.5 Le Filtrage de Wiener

Le filtrage de Wiener a été introduit à la fin des années 60 [43] pour essayer d'améliorer la qualité de la trace recueillie dans les potentiels évoqués. Le problème du type de filtrage proposé est qu'il n'est pas applicable sur une moyenne d'acquisition mais pour chaque trace. Doyle [44], propose une modification pour pouvoir l'adapter à la moyenne (le calcul dans ce cas est beaucoup plus rapide), il n'y a pas besoin de filtrer chaque trace. Cependant, il faut considérer que le bruit est stationnaire dans ce cas. [44]

Tout comme la soustraction spectrale, les calculs sont effectués dans le domaine fréquentiel. La DSP est calculée comme précédemment dans l'équation (3.11). Le système peut être isolé entre le bruit et le signal utile. Ce qui revient à dire que le rapport $P_S(f) / P_B(f)$ doit être maximisé pour obtenir le signal utile. Le filtre de Wiener est défini de la façon suivante :

III.5.1 Le Principe

La DSP du bruit est prise dans les périodes de silence. La DSP du signal utile est quand elle est calculée sur chaque trame d'acquisition.

$$W(f) \frac{P_S(f)}{P_S(f) + P_B(f)} = \frac{1}{1 + \frac{P_B(f)}{P_S(f)}} \quad (3.17)$$

Sous cette forme :

$$W(f) = \frac{RSB_{prio}(f)}{1 + RSB_{prio}(f)} \quad (3.18)$$

Le filtrage de Wiener est un problème d'estimation où on dispose d'une connaissance a priori sur le paramètre à estimer. Cette connaissance se présente sous la forme de données probabilistes. Typiquement on veut estimer un signal noyé dans un bruit et on sait que le signal est a priori centre, blanc, etc.

En inférence statistique, lorsque l'on prend en compte une connaissance probabiliste sur le paramètre à estimer on parle d'estimation bayésienne. [45]

La remarque de ce filtre, est que si le bruit est très bien estimé alors le signal transmis sera directement le signal utile. Le filtrage de Wiener est le filtrage optimal au sens du minimum de l'erreur quadratique moyenne (MEQM). Il adapte le rapport signal sur bruit pour chaque trame traitée. Cependant, le régime du bruit doit être stationnaire et non transitoire car l'estimation ne sera pas bonne dans ce second cas. Ce type de filtrage est utilisé en général derrière une soustraction spectrale ou autre réducteur de bruit pour améliorer le rapport signal sur bruit de la trace.

III.5.2 Application et résultat

Pour commencer le filtrage de Wiener va être utilisé dans ce cas pour filtrer le command vocal avant de l'avoir injecté dans le (ASR)

➤ **Première étape :**

C'est pratiquement la même première étape déjà vu dans la soustraction spectrale ou il s'agit de charger le signal de parole bruité par le bruit des moteurs $S(t)$, en utilisant la commande « Audio Read » Figure (III-12)

On prend l'exemple de la commande « Up » cette fois.

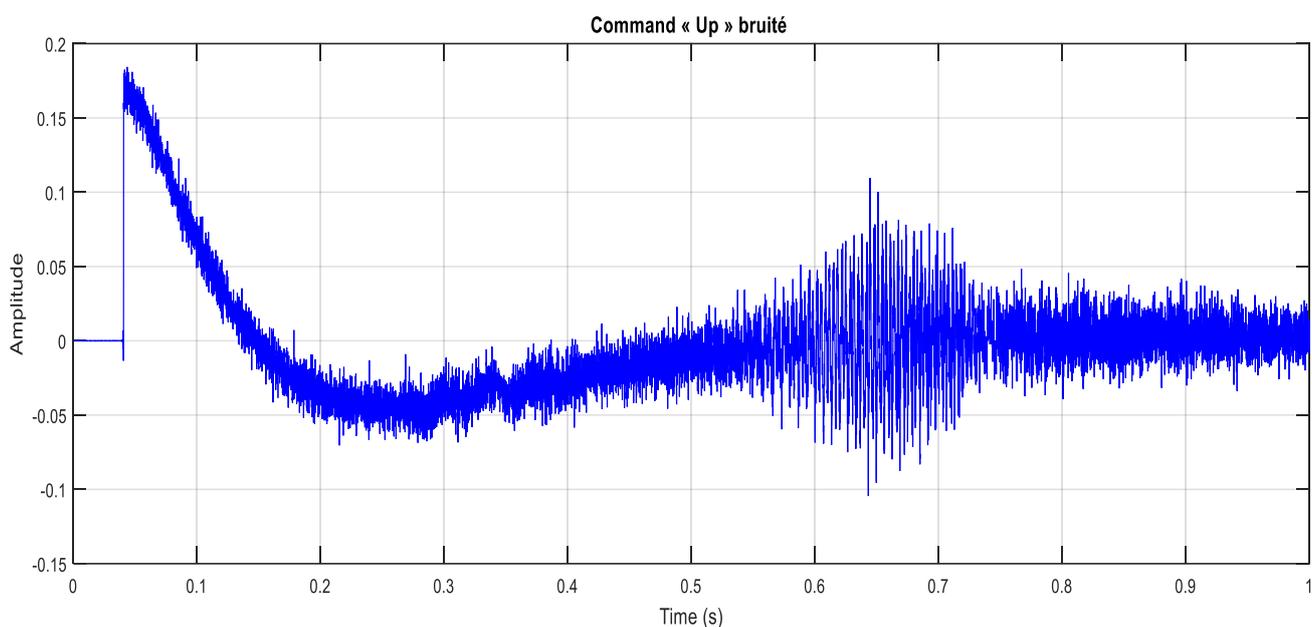


Figure (III-12) : Signal en domaine temporelle bruité par le son des moteurs.

L'application de la transformée de Fourier au signal de parole de la figure (III.12) nous a donné l'allure de la Figure (III-13).

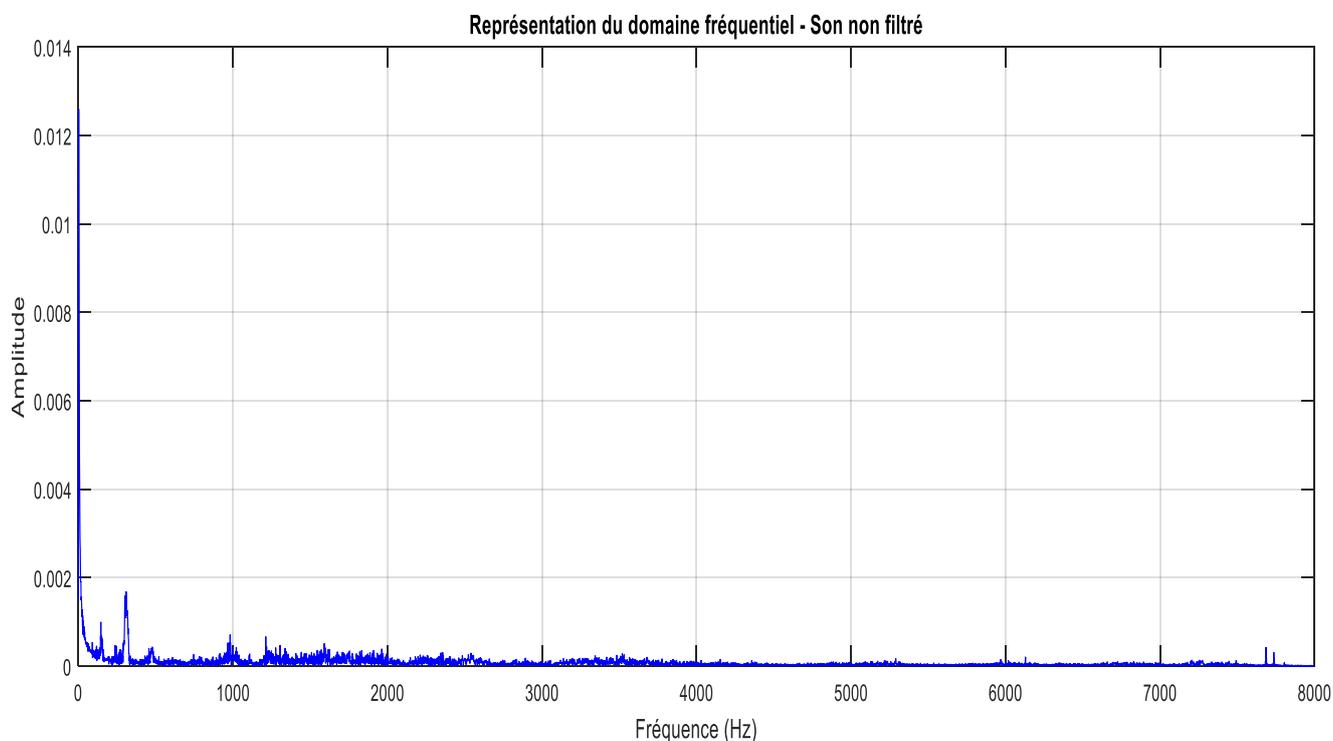


Figure (III-13) : La Transformée de Fourier du signal de parole bruité.

La figure (III-13) montre que le spectre d'amplitude du signal de parole, $S(f)$ Contient un bruit, ayant une amplitude importante sur toute la longueur du signal.

Cette étape est tout simplement la préparation du signal au traitement donc elle est essentiel dans tout type de filtrage ou traitement

➤ **Seconde étape :**

Cette étape consiste à faire passer le signal dans le filtre de wiener dans la fonction « r » sur MATLAB pour s'approcher le maximum du spectre d'amplitude du signal de parole débruité $U(f)$.

Le résultat de cette partie de filtrage de wiener nous donne les graphes suivant :

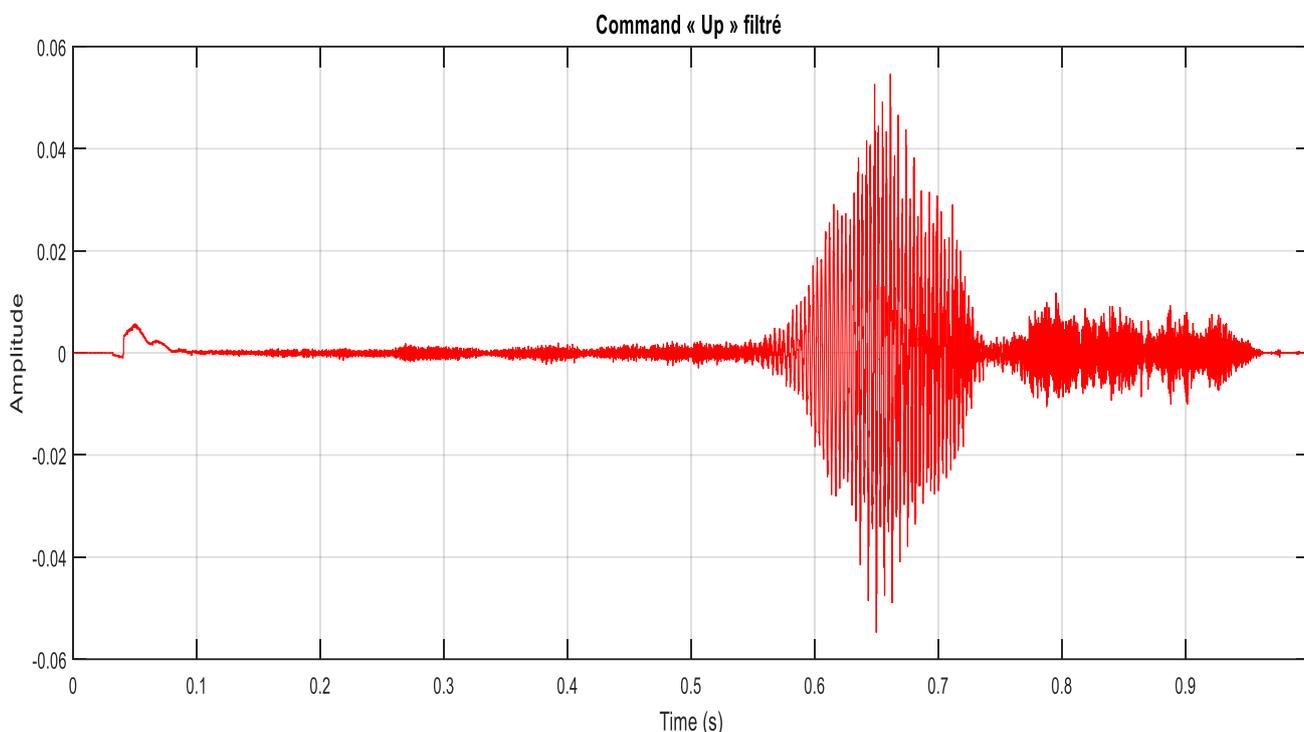


Figure (III-14) : Le signal filtré dans le domaine temporelle.

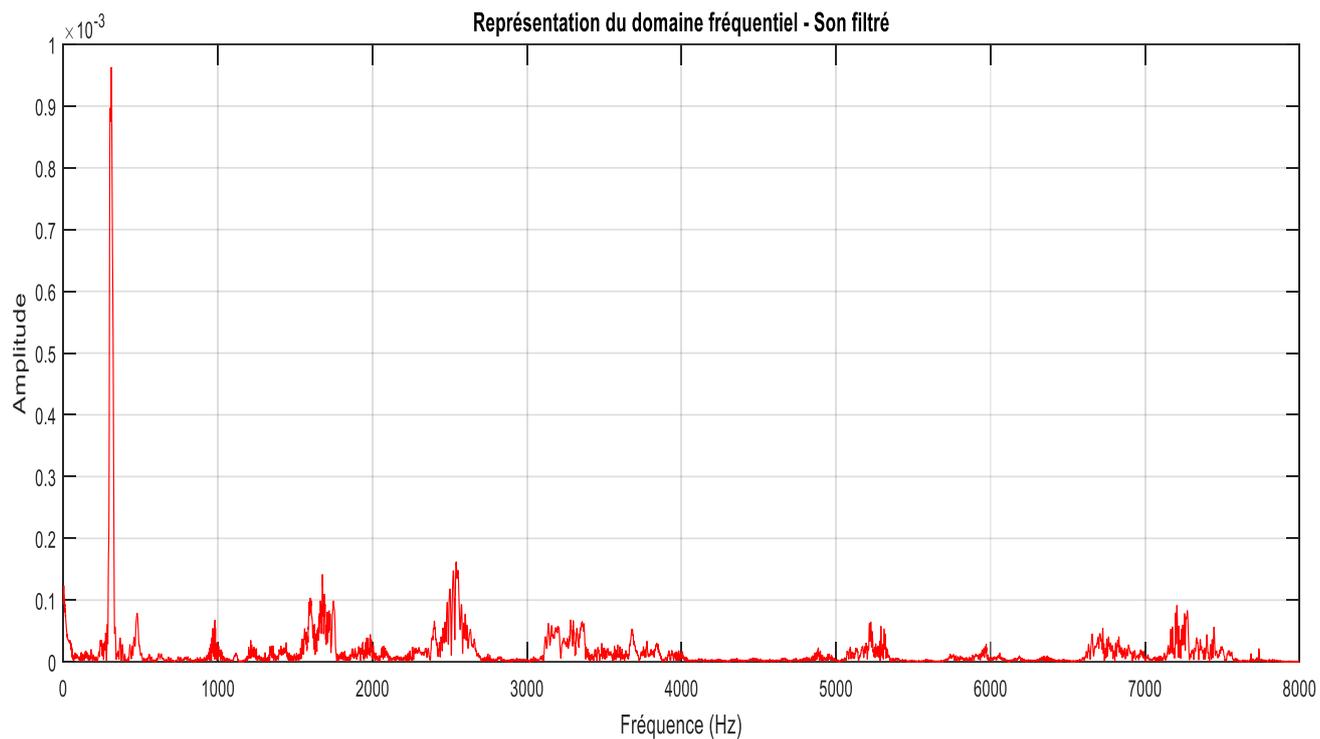


Figure (III-15) : Le spectre d'amplitude du signal de parole filtré.

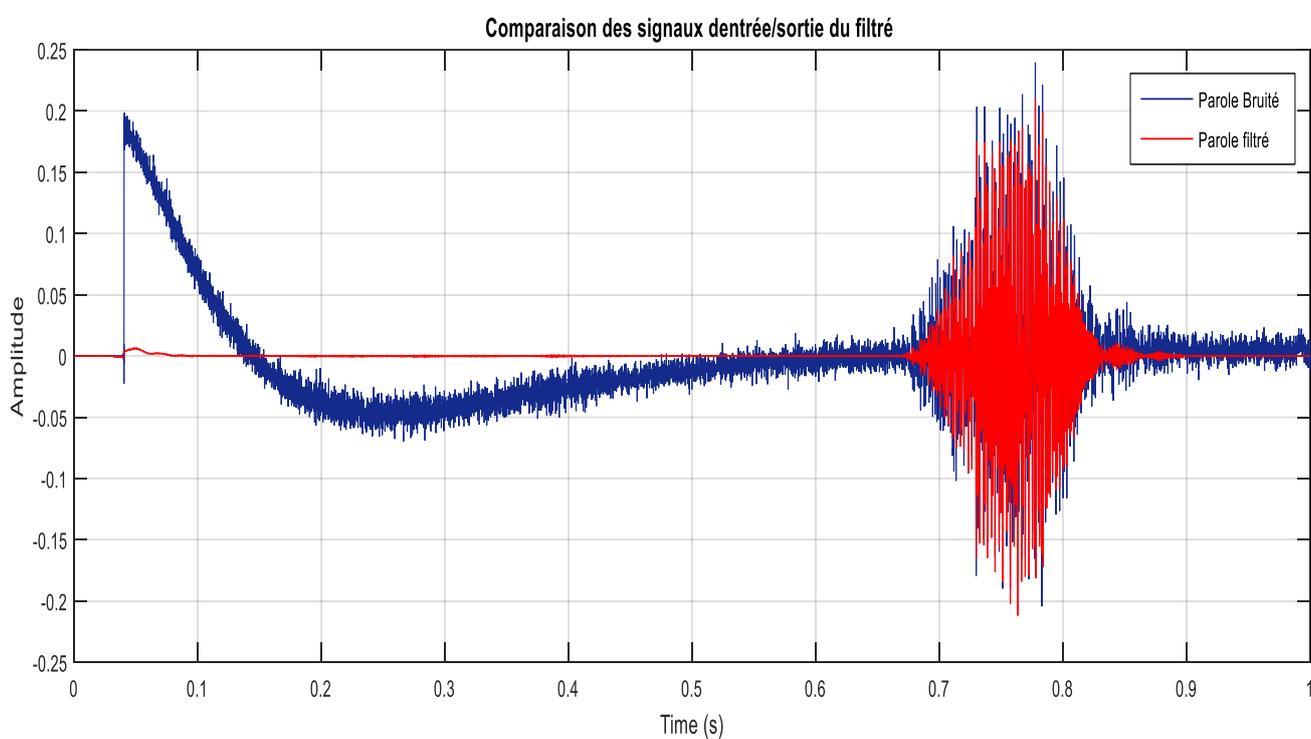


Figure (III-16) : Comparaison entre les graphes (Avant/Après) filtrage en Temps.

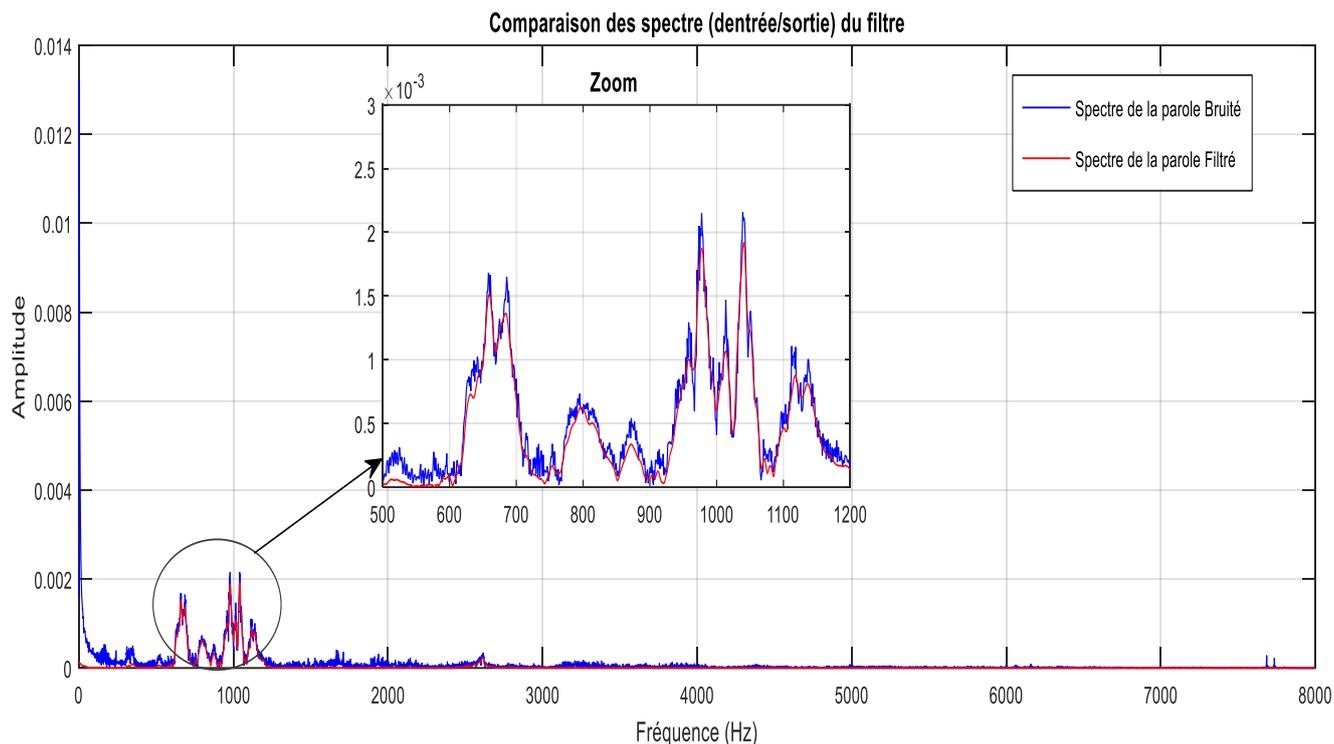


Figure (III-17) : Comparaison entre les graphes (avant/après) filtrage en fréquence.

La figure (III-17) montre la différence entre le spectre $S(f)$ et $U(f)$ en amplitude ou le premier pique qui est le plus important (Bruit) est complètement enlevé et que le spectre de la parole bruité est fortement atténué pour arriver au spectre $U(f)$

La partie zoomer montre que le spectre $S(f)$ a été lissé pour obtenir $U(f)$

On remarque dans la figure de comparaison (III-16) que le signale $S(t)$ constituer totalement du bruit à partir de l'instant zéro jusqu'au 0.6 second là ou $U(t)$ est clairement visible avec une amplitude atténuer partiellement et même à la fin du signal utile on constate la présence de bruit encore une fois ce qui signifie que notre signal est noyé dans le bruit.

III.6 conclusion

Le signal de parole, est la représentation électrique, par l'intermédiaire d'un microphone, de mots, phrases ou textes prononcés par un utilisateur dans le cadre de traitement mono-voix, lors de son acquisition, ce signal est perturbé par des bruits divers rendant sa reconnaissance impossible.

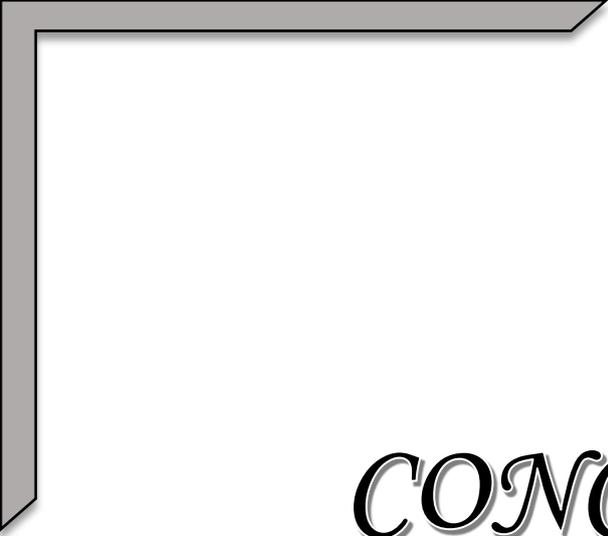
Ce signal étant non stationnaire, les méthodes classiques du traitement du signal ne sont pas adaptées pour son débruitage.

Par l'application des différentes méthodes (soustraction spectrale, filtrage de Wiener) aux signaux bruités par les différents bruits et surtout les bruits des moteurs nous sommes arrivés à une atténuation de bruit du signal perturbé.

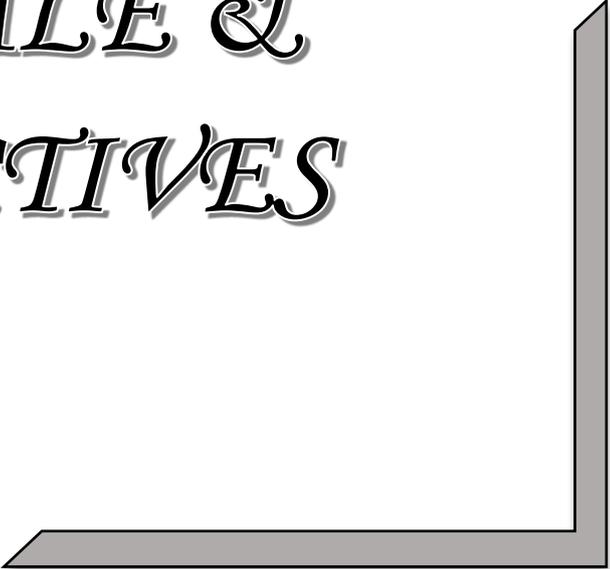
L'utilisation des méthodes citées nous a permis de faire la comparaison entre les deux, on a constaté que le filtrage Wiener est très efficace parce qu'il a la capacité de filtrer le signal audio même si le bruit ciblé est complètement inconnu.

En revanche la soustraction spectrale c'est une méthode basée sur la détermination du bruit à la perfection pour pouvoir extraire le signal utile, et cette particularité c'est ce qui nous intéresse dans notre recherche pour améliorer la robustesse de notre système de reconnaissance vocale.

Parce que le bruit ciblé c'est celui des moteurs de notre drone donc il est prêt à être étudié et utilisé pour le filtrage.



*CONCLUSION
GENERALE &
PERSPECTIVES*



CONCLUSION GENERALE & PERSPECTIVES

1. Conclusion générale

L'objectif de notre travail était une réponse à la question posée dans l'introduction générale, ainsi dans ce travail nous nous sommes intéressés à établir en arabe et en amazigh un système de reconnaissance de commande vocale multilingue afin de l'utiliser comme application pour contrôler les mouvements des drones dans un environnement bruité.

Pour ce faire, nous avons d'abord décrit les types de drones et l'interaction entre les humains et les drones pour pouvoir comprendre ce type d'interaction, et mettre en valeur l'impact de cette technologie dans la vie quotidienne.

Nous avons déduit a travers les multiples façons d'interagir avec un robot quel qu'il soit drone ou autre, que la commande vocale est la technique la plus utile jusqu'à présent pour transmettre les commandes humaines, et aussi pour avoir une exploitation totale de l'utilisateur et sont drone ce dernier doit avoir les mains libre tout en gardant le robot sous contrôle.

Ensuite, nous sommes passés à la reconnaissance vocale avec l'aide de l'apprentissage en profondeur sur « MATLAB » pour déterminer les termes à utiliser pour interagir avec le drone et programmer ces mouvements suite à chaque commande reconnu.

En derniers lieu nous avons mis en place un bloc de filtrage pour réduire le bruit des moteurs précisément à l'entrée du (ASR), avec cette approche nous avons réussi à augmenter la robustesse de notre system de reconnaissance, désormais le bruit des moteurs de notre drone n'affecte plus la reconnaissance. Vu qu'il est carrément réduit voir éliminer comme montrent les résultats.

Dans le cadre d'une recherche plus poussée, il serait intéressant d'élargir la base de données pour augmenter son efficacité. Ainsi, nous suggérons de créer un site web pour enregistrer les commandes d'un grand nombre de personnes afin de couvrir tous les dialectes existants. Un système de confirmation des commandes devrait compléter cette application. Nous pourrions même nous diriger vers un champ lexical, ou simplement vers d'autres commandes complexes à d'autres fins.

2. Perspectives

Ce travail nous a permis d'enrichir nos connaissances sur le système de reconnaissance vocale robuste, et d'étudier les différentes techniques de filtrage les plus efficaces dans un system monovoie, et d'explorer le domaine de la commande des UAV nous espérons que ce travail sera bénéfique pour l'institut, ces nouvelles techniques représente un domaine de recherche prometteur, nous proposons alors pour de futurs travaux les axes suivants :

- ❖ Travailler avec les réseaux de microphone pour pouvoir rivaliser avec la quantité de bruit qui augmente dans notre environnement
- ❖ Implémentation de la structure proposée sur des robots connecter.
- ❖ Réaliser la même étude de reconnaissance avec une modélisation mathématique du bruit environnant.

BIBLIOGRAPHIE

- [1] Karjalainen, K., & Romell, A. (2017). *Human-Drone Interaction: Drone as a companion? An explorative study between Sweden and Japan* (Master's thesis).
- [2]<https://www.digitalistmag.com/digital-economy/2019/11/05/are-drones-changing-way-we-live-06201367/>
- [3] Tezza, D., & Andujar, M. (2019). The state-of-the-art of human–drone interaction: A survey. *IEEE Access*, 7, 167438-167454.
- [4] Fernandez, R. A. S., Sanchez-Lopez, J. L., Sampedro, C., Bavle, H., Molina, M., & Campoy, P. (2016, June). Natural user interfaces for human-drone multi-modal interaction. In *2016 International Conference on Unmanned Aircraft Systems (ICUAS)* (pp. 1013-1022). IEEE.
- [5] Yam-Viramontes, B. A., & Mercado-Ravell, D. (2020, September). Implementation of a Natural User Interface to Command a Drone. In *2020 International Conference on Unmanned Aircraft Systems (ICUAS)* (pp. 1139-1144). IEEE.
- [6] Funk, M. (2018). Human-drone interaction: let's get ready for flying user interfaces!. *Interactions*, 25(3), 78-81.
- [7] <https://insideunmannedsystems.com/draganfly-selected-to-develop-pandemic-drone/>
- [8] Yeh, A., Ratsamee, P., Kiyokawa, K., Uranishi, Y., Mashita, T., Takemura, H., ... & Obaid, M. (2017, October). Exploring proxemics for human-drone interaction. In *Proceedings of the 5th international conference on human agent interaction* (pp. 81-88).
- [9] Duncan, B. A., & Murphy, R. R. (2013, August). Comfortable approach distance with small unmanned aerial vehicles. In *2013 IEEE RO-MAN* (pp. 786-792). IEEE.
- [10] Jensen, W., Hansen, S., & Knoche, H. (2018, April). Knowing you, seeing me: investigating user preferences in Drone-Human acknowledgement. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (pp. 1-12).
- [11] Szafir, D., Mutlu, B., & Fong, T. (2014, March). Communication of intent in assistive free flyers. In *Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction* (pp. 358-365).

- [12] Szafir, D., Mutlu, B., & Fong, T. (2015, March). Communicating directionality in flying robots. In *2015 10th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (pp. 19-26). IEEE.
- [13] Nozaki, H. (2014). Flying display: a movable display pairing projector and screen in the air. In *CHI'14 Extended Abstracts on Human Factors in Computing Systems* (pp. 909-914).
- [14] Schneegass, S., Alt, F., Scheible, J., & Schmidt, A. (2014, June). Midair displays: Concept and first experiences with free-floating pervasive displays. In *Proceedings of The International Symposium on Pervasive Displays* (pp. 27-31)
- [15] Schneegass, S., Alt, F., Scheible, J., Schmidt, A., & Su, H. (2014). Midair displays: Exploring the concept of free-floating public displays. In *CHI'14 Extended Abstracts on Human Factors in Computing Systems* (pp. 2035-2040).
- [16] Scheible, J., Hoth, A., Saal, J., & Su, H. (2013, June). Displaydrone: a flying robot based interactive display. In *Proceedings of the 2nd ACM International Symposium on Pervasive Displays* (pp. 49-54).
- [17] Avila, M., Funk, M., & Henze, N. (2015, October). Dronenavigator: Using drones for navigating visually impaired persons. In *Proceedings of the 17th International ACM SIGACCESS Conference on Computers & Accessibility* (pp. 327-328).
- [18] Orosanu, L. (2015). *Reconnaissance de la parole pour l'aide à la communication pour les sourds et malentendants* (Doctoral dissertation, Université de Lorraine).
- [19] Tong, S., Garner, P. N., & Boulard, H. (2017). An investigation of deep neural networks for multilingual speech recognition training and adaptation. In *Proc. of INTERSPEECH* (No. CONF).
- [20] <https://smartboost.com/blog/deep-learning-vs-neural-network/>
- [21] Kamath, U., Liu, J., & Whitaker, J. (2019). *Deep learning for NLP and speech recognition* (Vol. 84). Cham: Springer.
- [22] Liu, W., Wang, Z., Liu, X., Zeng, N., Liu, Y., & Alsaadi, F. E. (2017). A survey of deep neural network architectures and their applications. *Neurocomputing*, 234, 11-26.
- [23] https://www.researchgate.net/figure/Schematic-diagram-of-a-basic-convolutional-neural-network-CNN-architecture-26_fig1_336805909
- [24] <https://srdas.github.io/DLBook/>
- [25] Varma, S., & Das, S. (2018). Introduction to deep learning.
- [26] Kamath, U., Liu, J., & Whitaker, J. (2019). *Deep learning for NLP and speech recognition* (Vol. 84). Cham: Springer.
- [27] <https://www.mathworks.com/discovery/feature-extraction.html>
- [28] Warden, P. (2017). Launching the speech commands dataset. *Google Research Blog*.

- [29] Fayek, H. (2016). Speech processing for machine learning: Filter banks, mel-frequency cepstral coefficients (mfccs) and what's in-between.
- [30]<https://www.matlabexpo.com/content/dam/mathworks/mathworks-dot-com/images/events/matlabexpo/us/2018/master-class-deep-learning-for-signals.pdf>
- [31] Ueblér, U. (2001). Multilingual speech recognition in seven languages. *Speech*
- [32]<https://www.matlabexpo.com/content/dam/mathworks/mathworks-dot-com/images/events/matlabexpo/us/2018/master-class-deep-learning-for-signals.pdf>
- [33]Fayek, H. (2016). Speech processing for machine learning: Filter banks, mel-frequency cepstral coefficients (mfccs) and what's in-between.
- [34] K. Nakadai, D. Matsuura, H. G. Okuno, and H. Kitano, Applying Scattering Theory to Robot Audition System, IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS2003), pp. 1147-1152, 2003.
- [35] C. T. Ishi, H. Ishiguro, N. Hagita: "Evaluation of prosodic and voice quality features on automatic extraction of paralinguistic information," accepted to IROS 2006.
- [36] W. Herbordt, H.Buchner, S.Nakamura, and W.Kellermann, "Application of a double-talk resilient DFT-domain adaptive filter for bin-wise stepsize controls to adaptive beamforming," Proc.IEEE-EURASIP Workshop on Nonlinear Signal and Image Processing, pp. 175.181, May 2005.
- [37] S. Matsuda, T. Jitsuhiro, K. Markov, and S. Nakamura, "ATR Parallel Decoding Based Speech Recognition System Robust to Noise and Speaking Styles," IEICE Trans. Inf. & Syst., vol. E89-D, No. 3, pp. 989-- 997, 2006.
- [38] F.K. Soong, W.K. Lo, and S. Nakamura, "Generalized Word Posterior Probability (GWPP) for Measuring Reliability of Recognized Words," Proc. SWIM2004, 2004
- [39] R. Boite, H. Boulard, T. Dutoit, J. Hancq, and H. Leich. Traitement de la parole. P U Polytec Rom, 2000.
- [40] Arnaud Jeanvoine " Intérêt des algorithmes de réduction de bruit dans l'implant cochléaire : Application à la binauralité " 2013
- [41] M.tahon "le CNAM, CPDA traitemen de signal " Laboratoire d'Acoustique, Conservatoire National des Arts et Métiers 2 rue Conté, 75003 Paris, 2015.
- [42] Steven F. Boll. Suppression of acoustic noise in speech using spectral subtraction. IEEE, 27(2):113 – 120, April 1979.
- [43] DO. Walter. A posteriori "wiener filtering" of average evoked responses. Electroencephalogr Clin Neurophysiol, Suppl 27 :61+, 1968.

[44] DJ. Doyle. Some comments on the use of wiener filtering for the estimation of evoked potentials. *Electroencephalogr Clin Neurophysiol*,38(5) :533–4, May 1975.

[45] Filtre de Wiener Maurice Charbit ;26 juin 2002

