N° d'ordre :
N° de série :

RÉPUBLIQUE ALGÉRIENNE DÉMOCRATIQUE ET POPULAIRE
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

---

UNIVERSITE ECHAHID HAMMA LAKHDAR - EL OUED
FACULTÉ DES SCIENCES EXACTES
Département D'Informatique

Mémoire de Fin D'étude
Présenté pour l'obtention du Diplôme de

MASTER ACADEMIQUE

Domaine : **Mathématique et Informatique**
Filière : **Informatique**
Spécialité : **Systèmes Distribués et Intelligence Artificielle**

Présenté par :

- **GUEDIRI Abderrahmane**
- **HAKKOUM Abdelaziz**

Thème

Proposition d'un Algorithme de segmentation de la voie arabe à base de spectrogrammes

Soutenue le 20-06-2019 Devant le jury:

Dr.	YAKOUB Mohamed Amine	MCA	Président
M.	BALI Mouadh	MAA	Rapporteur
Dr.	ZAIZ Faouzi	MAA	Encadreur

Année Universitaire: 2018/2019

Proposition d'un algorithme de segmentation de la voie arabe à base de spectrogrammes

GUEDIRI Abderrahmane
hakkoum Abdelaziz

1er juillet 2019

Dédicace

Avant tous, je remercie dieu le tout puissant de m'avoir donné le courage et la patience pour réaliser ce travail malgré toutes les difficultés rencontrées.

Je dédie ce modeste travail À celui qui m'a orienté et m'a pris les secrets de la vie : « mon Père ».

À celle qui m'a ouvert les portails et m'a donné la tendresse et le courage.

À celle qui endeuillée pour me rendre heureuse.

À celle qui attend chaleureusement ce jour : "ma chère Mère".

À Mes frères et Mes sœurs.

À tout mes oncles, tantes, leurs conjoints ainsi que leurs enfants.

À mes meilleurs amis.

À mon binôme .

À tous les étudiants de la faculté Informatique surtout les étudiants de la 2éme année Master promotion 2019. . . ..

À tous les habitants d' EL OUED

Remerciements

A Dieu, le tout puissant, nous rendons grâce pour nous avoir donné santé, patience, volonté et surtout raison..

En premier lieu, je tiens à remercier "Dr. ZAIZ Faouzi" notre promoteur pour sa serviabilité, sa disponibilité, ses remarques et ses orientations constructives qui nous ont été utiles tout au long de notre projet.

Je remercie également les membres qui nous ont fait l'honneur de participer, au jury de ce mémoire..

Enfin, je remercie tous ceux qui ont, de prés ou de loin, contribué à la concrétisation de ce travail trouvent ici ma gratitude et reconnaissance..

Résumé

L'utilisation des ordinateurs dans divers domaines de la vie est devenue de plus en plus importante. Peut-être n'a-t-il jamais eu un développement égal à celui de l'informatique au cours des 30 dernières années, en particulier dans le domaine de l'intelligence artificielle, qui a fait des progrès remarquables. Plus précisément, la reconnaissance et le traitement de la parole.

Dans notre mémoire, nous allons concentrer sur la reconnaissance automatique du son capturé par le microphone qui doit être converti en texte par le système. Pour effectuer ce processus, nous allons suivre plusieurs étapes (pré-traitement, segmentation, extraction de caractéristiques, classification et reconnaissance) afin d'atteindre le résultat final. Les étapes de segmentation, d'extraction de caractéristiques et de classification sont les plus importantes de notre procédure .

Malheureusement, les méthodes utilisées dans la Division pour un discours dans une langue étrangère (anglais, français, etc.) ne sont pas entièrement compatibles avec l'arabe en raison de la différence radicale entre cette dernière et les autres langues. Par conséquent, nous devons proposer de nouveaux moyens et méthodes pour résoudre ce problème, et nous l'avons fait.

La méthode que nous avons proposée a donné des résultats prometteurs pour la division et en peu de temps par rapport aux autres méthodes.

Mots Clés: Segmentation, Classification, Reconnaissance automatique de la parole , Segmentation très efficace.

Abstract

The use of computers in various fields of life has become increasingly important. Maybe it has never been any equal development to the one of computer science in the last 30 years, especially in the field of artificial intelligence, which has made remarkable progress. Specifically, Speech Recognition and Treatment field.

In our memory, we will focus on the automatic recognition of the sound captured by the microphone which is to be converted into text by the system. To perform this process we will take several steps (Preprocessing, Segmentation, Feature Extraction, Classification and Recognition) in order to reach the final result. Segmentation, Feature Extraction and Classification stages are the most important in our procedure.

Unfortunately, the methods used in the Division for a speech in a foreign language (English, French, etc.) are not entirely compatible with Arabic because of the radical difference between this latter and other languages. Therefore, we have to propose new ways and methods to address this problem, and we have.

The method we proposed gave promising results for the segmentation and in a short time compared to other methods.

Keywords: Segmentation, Classification,Automatic speech recognition, very efficient division.

الملخص

اتتزايد أهمية استخدام الحاسبة في شتى مجالات الحياة، ولعله لم يحدث من قبل أن تطور علم من العلوم كما تطور علم الحاسبات خلال السنوات الثلاثين الأخيرة وخاصة في مجال الذكاء الصناعي الذي عرف تقدما باهرا وخاصة في مجال التعرف على الكلام ومعالجته.

سنركز في هذه الأطروحة على التعرف التلقائي للصوت المأخوذ من خلال الميكروفون والذي يتم تحويله الى نص من قبل الجهاز، وللقيام بهذه العملية سنمر بعدة خطوات (قبل المعالجة، التقسيم، استخراج الخصائص، التصنيف والتعرف) بغية الوصول الى النتيجة النهائية، وركزنا في عملنا هذا على مرحلة التقسيم واستخراج الخصائص والتصنيف وهي الأهم في العملية.

لسوء الحظ الأساليب المتبعة والمستخدمة في تقسيم الكلام للغات الأجنبية (الإنجليزية، الفرنسية،... الخ) لا تتوافق كليا مع اللغة العربية نظرا للاختلاف الكبير بين هذه الأخيرة وغيرها من اللغات ، لذلك توجب علينا اقتراح طرق وأساليب جديدة لمعالجة هذا المشكل، وهو ما قمنا به .

الطريقة التي قمنا باقتراحها أعطت نجاعة كبيرة في التقسيم في زمن وجيز مقارنة بالطرق الأخرى.

كلمات مفتاحية: التقسيم، التصنيف ، التعرف الالي للكلام، نجاعة كبيرة في التقسيم .

Table des matières

Dédicace ii

Remerciements iii

Résumé iv

Table des matières vii

Acronymes xi

Introduction 1

1 La reconnaissance automatique de paroles 2
Introduction . . . 2
1.1 Historique . . . 2
1.1.1 La naissance . . . 2
1.1.2 Les premiers résultats . . . 2
1.1.3 La changement de l'orientation vers les langues connues . . . 2
1.2 Généralités sur la parole . . . 2
1.2.1 Son analogique . . . 3
1.2.2 Son numérique . . . 3
1.2.3 Convertisseur analogique / numérique . . . 3
1.3 Bruit . . . 4
1.3.1 Bruits additifs . . . 4
1.3.2 Bruits convolutionnels . . . 4
1.3.3 Bruits physiologiques . . . 4
1.4 Outils d'acquisition . . . 5
1.4.1 Microphones . . . 5
1.4.2 Cartes son . . . 6
1.5 Représentation du signal . . . 8
1.5.1 Audiogramme . . . 8
1.5.2 Spectrogramme . . . 8
1.6 Système de reconnaissance automatique de paroles SRAP . . . 9
1.6.1 Définition . . . 9
1.6.2 Les avantages de la reconnaissance automatique de la parole . . . 9
1.6.3 Applications de la reconnaissance de la parole . . . 9
1.6.4 Approche de la reconnaissance . . . 10
1.6.5 La reconnaissance selon le mode d'élocution . . . 11
1.6.6 Schéma général d'un système de reconnaissance de parole . . . 11
Conclusion . . . 13

2 Traitement des paroles 14
Introduction . . . 14
2.1 Caractéristiques de signal de parole . . . 14
2.1.1 L'énergie d'un son (intensité) . . . 14
2.1.2 La fréquence fondamentale . . . 15
2.1.3 Le timbre . . . 15

2.2 Présentation de langue arabe . . . 16
2.3 Les opérations de traitement . . . 17
2.3.1 La numérisation . . . 17
2.3.2 L'échantillonnage . . . 18
2.3.3 La Quantification . . . 18
2.3.4 Le Codage . . . 18
2.4 Généralité de la segmentation . . . 19
2.4.1 Définition de la segmentation . . . 19
2.4.2 Paramétrisation du segment . . . 20
2.4.3 Approches de la segmentation . . . 20
2.4.4 Les méthodes de segmentation . . . 20
2.4.5 Les problèmes liés à la segmentation . . . 21
2.5 la classification . . . 21
2.5.1 Les phases de la classification . . . 21
2.5.2 Exemples de méthode de classification et segmentation . . . 21
Conclusion . . . 22

3 Conception et mise en œuvre **23**
Introduction . . . 23
3.1 Mise en œuvre du système . . . 23
3.1.1 Description des étapes de système . . . 24
Conclusion . . . 30

4 Résultats et discussion **31**
Introduction . . . 31
4.1 Choix du langage de programmation . . . 31
4.2 Interface du système . . . 32
4.3 Fonctionnalité du système . . . 32
4.4 Test et bilan . . . 33
Conclusion . . . 33

Conclusion **34**

Liste des figures

1.1 Exemple d'une chaîne analogique [17]. . . . 3
1.2 Exemple d'une chaîne numérique [16]. . . . 3
1.3 Conversion analogique numérique. [7]. . . . 4
1.4 Processus d'acquisition d'un signal. [11]. . . . 5
1.5 Fonctionnement d'un microphone dynamique. . . . 5
1.6 Exemple d'une carte son. [7]. . . . 7
1.7 Audiogramme de signaux de parole. [12]. . . . 8
1.8 Spectrogramme et évolution temporelle [12]. . . . 9
1.9 Système de reconnaissance automatique de parole 12
1.10 Schéma de principe d'un système de reconnaissance de la parole continue [5]. . . . 12

2.1 Intensité de son. . . . 15
2.2 Le timbre de son. . . . 16
2.3 Enregistrement numérique d'un signal [2] 17
2.4 La convertisseur analogique numérique. . . . 18
2.5 un signal échantillonné[4]. . . . 18
2.6 un signal quantifié[4]. . . . 19
2.7 Exemples de segmentation de parole continue. . . . 20

3.1 Architecture générale de notre système 24
3.2 détermination de la fenêtre. . . . 26
3.3 le déplacement de la fenêtre. . . . 26
3.4 Extraction des caractéristiques de spectrogramme 28

4.1 L'interface utilisateur globale du système 32

Liste des tableaux

1.1 Tableau résumé des caractérstiques des microphones. [6]. 6
1.2 Exemples d'applications de la reconnaissance de la parole [13] 10

2.1 Les fréquences fondamentales pour l'homme,femme et enfant. 15
2.2 Classification des sons arabes selon le mode d'articulation . [12]. 17

4.1 Résultats de segmentation et classification. 33

Acronymes

RAP	Reconnaissance Automatique de Parole
DSP	Digital Signal Processor
DAC	Digital to Analog Converter
FFT	Fast Fourier transform
K-PPV	K Plus Proches Voisins
SP	Speech Processing
SRAP	Système de Reconnaissance Automatique de la Parole
TF	Transformée de Fourier

Introduction

La reconnaissance automatique de la parole (RAP) est un domaine de recherche actif depuis plus de cinq décennies. Malheureusement, malgré l'incroyable évolution des ordinateurs et des connaissances, la reconnaissance automatique de la parole donne toujours des résultats trop loin de l'idéal qu'on aurait pu en attendre.

Le développement continu de la technologie et le système de reconnaissance de la parole idéal n'existe pas encore, et petit à petit La reconnaissance automatique de la parole commence à équiper certains téléphones, certains ordinateurs, en déterminant certains mots clefs, permettent de réaliser les tâches demandées, Les systèmes de reconnaissance de la parole sont également utilisés comme interface de dialogue homme-machine .....etc .

D'une façon générale, le traitement automatique de la parole utilise des sources de connaissance ainsi que de la quantité et la qualité des données utilisées pour l'apprentissage des modèles. Ainsi, les systèmes nécessitent d'importants volumes de données pour l'estimation des modèles acoustiques et linguistiques. En effet, malgré l'évolution continue des techniques mises en œuvre dans les SRAP, l'amélioration des modèles acoustiques reste souvent liée à l'augmentation des quantités de données d'apprentissage

Dans cette étude nous proposons une contribution qui se présente par une méthode de segmentation de la voie arabe en syllabe à base de spectrogrammes. La méthode est testée et comparée avec d'autres méthodes déjà réalisées, et elle permet d'avoir un très bon taux de segmentation en syllabes.

L'organisation générale du mémoire est articulé en 4 chapitres comme suit :

Le premier chapitre présente un état de l'art sur les systèmes de reconnaissance de la parole, dans laquelle nous allons donner un aperçu sur l'architecture d'un système de reconnaissance de la parole.

Le deuxième chapitre illustre et exposer les différentes approches, méthodes et technique pour les deux phases « segmentation et extraction de caractéristique ».

Le troisième chapitre est consacré à la conception et la mise en œuvre du système qui décrit en détail la méthode de segmentation proposée.

Le chapitre quatrième montre les résultats obtenus par notre méthode ainsi qu'une comparaison,discussion et évaluation de la méthode.

Enfin, nous terminons le travail par une conclusion sur les résultats obtenus par la méthode proposée, et proposons quelques perspectives qui nous croyons utiles et nécessaires pour améliorer et rendre le processus de segmentation efficace pour la parole arabe

Chapitre 1

La reconnaissance automatique de paroles

Introduction

Le système de reconnaissance automatique de parole (**SRAP**) est un système qui permet à une machine d'extraire le message oral contenu dans un signal de parole.
Nous nous sommes intéressés dans ce chapitre à un système RAP pour traiter les paroles arabes . Dans un premier temps, nous allons essayer de donner un historique et une idée générale sur la reconnaissance automatique de parole , de montrer les avantages de cette dernière , ainsi que les méthodes utilisées pour créer un systéme **SRAP** efficace.

1.1 Historique

Ce chapitre va montrer l'évolution de la reconnaissance automatique de la parole depuis ses débuts jusqu'à nos jours.

1.1.1 La naissance

Les premiers essais ont consisté en la mise en place des machines capables de comprendre des mots humains à la fin de l'année 40 .C'était au ministère de la Défense aux Etats-Unis afin d'interpréter les lettres russes qui ont interceptées. En dépit de tous les grands efforts et le coût élevé , les résultats obtenus ne sont pas ceux attendus.

1.1.2 Les premiers résultats

Dans les années soixante , Les chercheurs ont porté leur attention sur la reconnaissance des mots isolés, Et ont essayeé d'utiliser cette reconnaissance dans les mini-applications telles que :

- la commande vocale.
- la dictée vocale.[1]

Ces applications ont été à l'origine de cette technique.

1.1.3 La changement de l'orientation vers les langues connues

À l'heure actuelle,la plupart des chercheurs tentent de guider leurs applications vers la compréhension des langues parlées et des phrases complexes. Pour cette tendance, Il faut trouver des solutions pour les problèmes linguistiques face à l'application. Enfin , nous voyons que l'accès à une application puissante

de la reconnaissance semblable à celle de l'être humain est devenu possible, avec le développement rapide de cette technologie

1.2 Généralités sur la parole

D'un point de vue physique, un son est une énergie qui se propage sous forme de vibrations dans un milieu compressible (l'eau, l'air, les matériaux solides), produit a partir d'une source sonore et capté par un récepteur sensible, il se propage à une certaine vitesse dans un milieu élastique (340 m/s dans l'air à

15 °C), appelée aussi «célérité», On parle alors de pression acoustique. Plus cette pression acoustique est forte et plus l'on entend le son fortement, la propagation du son diminue avec la distance. Ceci est dû à l'amortissement du système[5].On distingue deux types de son : le son analogique et le son numérique.

1.2.1 Son analogique

Lorsque le son est capté à partir d'un microphone, ce dernier transforme l'énergie mécanique (la pression de l'air exercée sur sa membrane), en une variation de tension électrique continue. Ce signal électrique dit « analogique » pourra ensuite être amplifié,et envoyé vers un hautparleur dont la fonction est inverse, voir la figure (1.1). Le son analogique est généralement fixé sur des supports comme les bandes magnétiques, K7 audio..etc. Le problème rencontré au son analogique de faite il n'est pas traitable par l'ordinateur[17, 16].

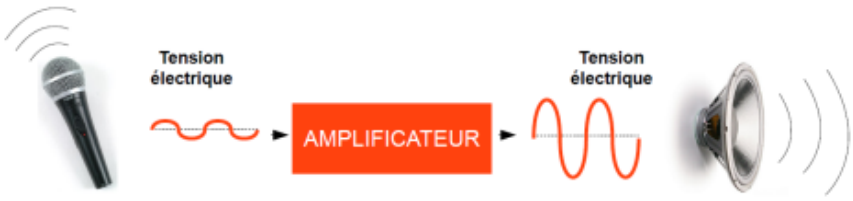

Figure 1.1: Exemple d'une chaîne analogique [17].

1.2.2 Son numérique

Avec l'informatique, lorsque ce même signal électrique est capturé à partir du micro,il est converti en une suite binaire (0,1), on parle alors de numérisation du signal. C'est la carte son qui s'en charge,ce processus est appelé numérisation, qui consiste donc à passer d'un signal continu (une variation de tension électrique) en une suite de valeurs mesurées à intervalles réguliers (discontinu) comme l'indique la figure (1.2), ce dernier est enregistré sur un support numérique tel que le disque dure [16].

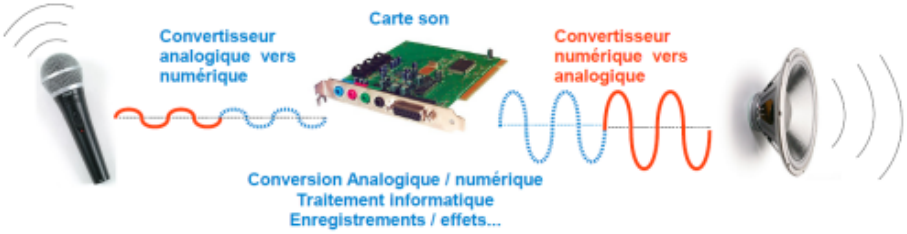

Figure 1.2: Exemple d'une chaîne numérique [16].

1.2.3 Convertisseur analogique / numérique

Un convertisseur analogique / numérique (CAN) est un appareil électronique qui se charge de convertir les tensions (un signal analogique continue) en chaînes de nombres binaires (un signal numérique discret) à chaque période de l'horloge d'échantillonnage d'une façon périodique. Les nombres binaires sont stockés sur un support d'enregistrement numérique sorte de mémoire, il s'agit de données multimédia (1.3)[7].

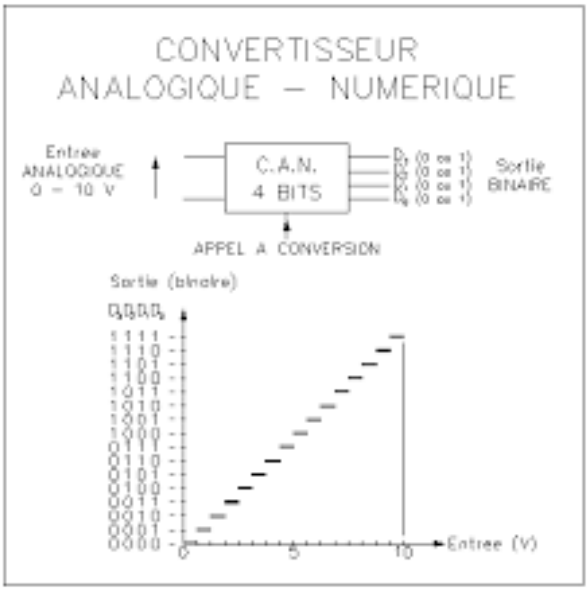

Figure 1.3: Conversion analogique numérique. [7].

1.3 Bruit

On appelle bruit tout phénomène perturbateur ou un signal nuisible qui se superpose au signal utile en un point quelconque d'une chaine de mesure ou d'un système de transmission. Il constitue donc une gêne dans la compréhension du signal utile ou a l'interprétation d'un signal, qui est dans notre cas, la parole. En physique, en acoustique et en traitement du signal, bien que le bruit soit, par nature, aléatoire, il possède certaines caractéristiques statistiques, spectrales ou spatiales.on distingue trois types de bruit [5, 8]

1.3.1 Bruits additifs

Les bruits additifs sont dus à la multiplicité des systèmes de communication dans un même environnement. Plusieurs émetteurs et plusieurs receveurs pouvant être confinés dans un même espace, les messages de tous les émetteurs peuvent donc se trouver en concurrence sur une même voie sans que les récepteurs possèdent un mécanisme infaillible pour isoler le message qui leur est destiné. L'émetteur et le récepteur peuvent aussi se trouver en présence d'un ou de plusieurs équipements générant un bruit de fond de force variable[5]

1.3.2 Bruits convolutionnels

Les bruits convolutionnels (ou multiplicatifs) sont dus à la distorsion induite par la voie de communication. Ils résultent de la mauvaise qualité d'un ou de plusieurs éléments de support du signal ou, tout simplement, de son étroitesse en bande passante. La qualité de la transmission varie cependant très peu au cours d'une même communication. De manière plus générale, le bruit convolutionnel est présent dans toute application de RAP par l'intermédiaire du microphone utilisé pour la saisie de la voix ou lorsque le microphone utilisé pour l'enregistrement est placé assez loin du locuteur. Ou bien dépend aux milieux d'enregistrement qui sont de mauvaise qualité et peuvent provoquer des phénomènes de réverbération[5]

1.3.3 Bruits physiologiques

D'autres bruits peuvent également être considérés dans le domaine de la RAP car ils sont spécifiques à l'être humain lors de sa phase de production de parole. Dans un environnement bruité la personne essaie, lui, de s'adapter aux conditions sonores rencontrées en modifiant sa méthode de production de parole. Ce qu'on appel l'effet Lombard. Cette accentuation de la voix pose cependant un problème majeur aux systèmes de RAP car les spectres de tous les phonèmes peuvent être modifiés ce qui a pour effet de nettement amoindrir les taux de reconnaissance[5]

1.4 Outils d'acquisition

En informatique, le système d'acquisition de données représente l'interface entre le capteur et l'ordinateur. Ce système, composé de circuit imprimé et de logiciel, permet de recueillir automatiquement les informations analogiques provenant du capteur. En vu de faire rentré et stocké ces informations dans un support de stockage (CD, disque dur,. . . etc.) ce dernier peut être vu comme un système d'acquisition de données, figure(1.4)

Figure 1.4: Processus d'acquisition d'un signal. [11].

1.4.1 Microphones

Le microphone est un appareil qui capture les modifications des vibrations sonores dans l'air Et les transforme en une tension électrique.
Donc , le travail de base du microphone est la conversion de diverses vibrations acoustiques en impulsions électriques,et leur enregistrement dans l'ordinateur [21].

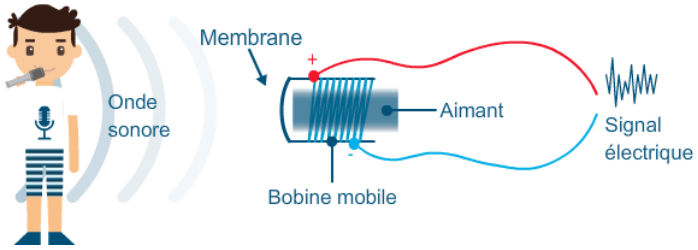

Figure 1.5: Fonctionnement d'un microphone dynamique.

1.4.1.1 Type des microphones

Il existe plusieurs types de microphone :
- à charbon.
- à condensateur.
- à magnétostriction.
- à électrodynamique.
- à électronique.
- à thermique.
- à ionique...etc.voir table 1.1

On prend le microphone à condensateur comme exemple,parmi ses propriétés :
- plus performant parmi les microphones disponibles.
- petite taille.
- simple construction.

Type	Utilisations	Avantages	Inconvénients	Qualité	Prix
Charbon	Téléphone	Forte puissance de sortie	Souffle, Bande passante réduite	Médiocre	peu coûteux
Dynamique	Toutes	Sensibilité élevée Robustesse, Directivités variés,Courbe de réponse étendue		De médiocre à excellent	peu coûteux à très coûteux
Ruban	Studio et intérieur	Très grande fidélité, Sensibilité moyenne	Usage réservé à l'intérieur très fragile	assez bon à excellent	moyen à coûteux
Statique	Toutes	Haute fidélité, nécessité de forte amplification	Nécessite une source de tension de polarisation	Excellent	très coûteux
Électret	Toutes	Comme statique, mais sans polarisation	Craint parfois la chaleur et l'humidité	excellent	coûteux à très coûteux
Piézo	Toutes	Bonne sensibilité, assez fidèle, tension de sortie élevée	Comme électret	de médiocre à assez bon	peu coûteux

Tableau 1.1: Tableau résumé des caractérstiques des microphones. [6].

1.4.1.2 Problèmes liées aux microphones

le microphone est un organe de capture sensible à la moindre variation de pression, il peut aussi capter des ondes sonores latérales, qui constituent un bruit à la parole originale destiné à être exploiter, qui va constituer par la suite une défaillance au capture, engendrant une perturbation au traitement du signal projeté, en plus de cela, on peut constater l'influence du microphone lui-même sur la qualité de signale transformé, tout à fait normal, due à la nature de chaque microphone et son principe de fonctionnement (défaillance matérielles)[17].

1.4.2 Cartes son

Une carte son est une carte d'extension d'ordinateur. La principale fonction de cette carte est de gérer tous les sons émis pour les envoyer vers les haut-parleurs ou reçus par l'ordinateur. Elle se présente sous la forme d'un périphérique que l'on peut connecter à l'ordinateur sur un bus PCI, PCI Express, PCMCIA (pour ordinateur portable), USB ou Firewire (bus informatique).[21].

1.4.2.1 Les principaux éléments d'une carte

1. **Le processeur spécialisé DSP (digital signal processor) :** Il fait tous les traitements numériques du son.
2. **Le convertisseur digital-analogique DAC (digital to analog converter) :** Il fait la conversion des données sonores de l'ordinateur en signal analogique .
3. **Le convertisseur analogique-numérique appelé ADC**
4. **Les connecteurs d'entrées-sorties externes :** parmi eux :
   — Une entrée microphone (notée parfois Mic).
   — Une interface MIDI : permettant de connecter des instruments de musique .
5. **Les connecteurs d'entrées-sorties internes :** parmi eux :
   — Connecteur CD-ROM / DVD-ROM.
   — Connecteur pour répondeur téléphonique[20].

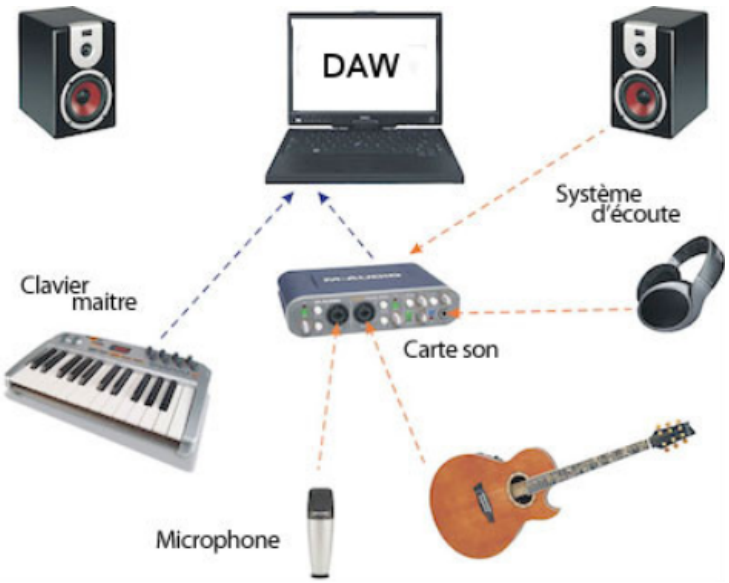

Figure 1.6: Exemple d'une carte son. [7].

1.4.2.2 Rôle d'une carte son

Une fois le signal analogique, issu du microphone arrive à l'entrée MIC de la carte son, il doit passer par un circuit de conditionnement, qui permet l'amplification et le filtrage de ce signal, après quoi la conversion Analogique-Numérique est effectuée, dans le but de rendre l'information récupérée, traitable par le système numérique (microordinateur). Cette conversion comprend l'échantillonnage, la quantification et le codage. Après la conversion Analogique-Numérique, la carte son passe à la mémorisation des données numérisées dans un espace mémoire ou tampon (buffer) sous forme de valeurs numérique. Ces données seront présentées par des vecteurs comportant une série de chiffres. On utilise ce genre de mémorisation plusieurs fois pour un même mot prononcé selon le choix de la taille du dictionnaire voulu, attribuée à l'apprentissage des données.

1.4.2.3 Différents types des cartes son

Il existe une multitude de types de cartes son, on cite les cartes suivantes :

1. Cartes son multimédia non professionnelle
   elles sont des interfaces audio-numériques grand public, plus adaptées à sortir le son vers les hauts parleurs qu'à véritablement enregistrer quelque chose hormis une conversation par internet ou à faire fonctionner un logiciel d'apprentissage de langue[19]
   - **Carte son intégrée :** C'est celle que l'on trouve par défaut sur la carte mère, que ce soit un ordinateur de forme tour, desktop ou portable. Souvent basée autour d'un chipset de type AC'97 ou Realtek [19].
   - **Carte son multimédia(interne) :** Carte interne de type PCI ou ISA (pour les modèles les plus anciens) elle s'installe dans un boitier d'ordinateur, tour ou desktop, jamais dans un portable. Le modèle phare, dont on emploie parfois le nom comme terme générique est la soundblaster[19].
   - **Carte son multimédia(externe) :** Certains constructeurs de cartes multimédia comme Hercules ou Soundblaster proposent des cartes multimédia avec un boitier externe pour faciliter les branchements. En général elles sont de qualité supérieures aux cartes multimédia internes et souffrent moins des rayonnement électroniques que l'on rencontre à l'intérieur des boitiers d'ordinateur. Orientées grand-public[19].
2. Cartes son professionnelles :
   les véritables professionnels en studio n'ont généralement pas d'interface audio-numérique à proprement parler. Ils ont tout un tas d'appareils distincts qui remplissent bien mieux, et pour beaucoup plus cher, toutes les fonctions de nos cartes son. ce genre d'équipement a montrer une qualité supérieure[19].
   - **Carte son pro interne :** Carte interne de format PCI ou ISA (pour les plus anciennes) elle ressemble furieusement à une bonne vieille Soundblaster[19].
   - **Carte son pro externe :** Beaucoup plus pratiques à l'usage, compatibles avec les ordinateurs portables, ces interfaces externes ont le vent en poupe[19]..

1.4.2.4 Différentes limitations des cartes son

Comme chaque dispositif électronique, la carte son présente une multitude de différences qui caractérise chaque génération matérielle et logicielle, en plus le domaine où dont elle est destiné à être utiliser, soit professionnel ou non, et chaque firme pratique sa propre philosophie industrielle qui reflète la qualité fournie pour chaque type de cartes son, par conséquence on remarque la différence entre la qualité des types du son produit par le processus de conversion A/N , en plus de l'impacte de logiciels utilisés.

1.5 Représentation du signal

Le signal de la parole est un vecteur acoustique porteur d'informations d'une grande complexité, il est représentable sous plusieurs formes, tout dépend l'utilité de différentes composantes fréquentielles du signal et la signification sur le plan perceptuel, on cite les représentations suivants :

1.5.1 Audiogramme

L'échantillonnage transforme le signal à temps continu x(t) en signal à temps discret X(nT e) défini aux instants d'échantillonnage, multiples, entiers de la période d'échantillonnage Te, celle-ci est elle-même l'inverse de la fréquence d'échantillonnage fe. Pour ce qui concerne le signal vocal, le choix de fe résulte d'un compromis. Son spectre peut s'étendre jusque 12 kHz [10, 12]. La figure (1.7) représente l'évolution temporelle, ou audiogramme du signal vocal pour les mots : (**بسم الله الرحمان الرحيم**).

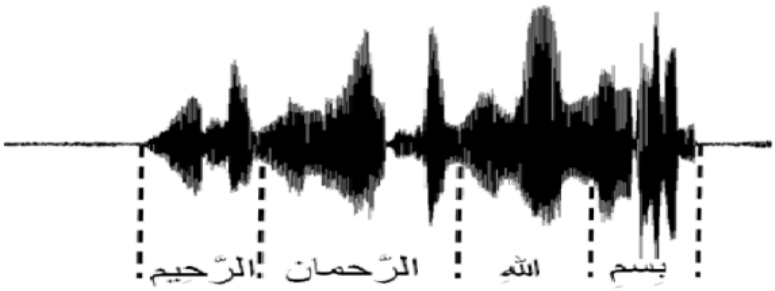

Figure 1.7: Audiogramme de signaux de parole. [12].

1.5.2 Spectrogramme

Il est souvent intéressant de représenter l'évolution temporelle du spectre à court terme d'un signal, sous la forme d'un spectrogramme. Le spectrogramme permet de mettre en évidence les différentes composantes fréquentielles du signal à un instant donné, une transformée de Fourier rapide étant régulièrement calculée à des intervalles de temps rapprochés [5, 8]. L'amplitude du spectre y apparaît sous la forme de niveaux de gris dans un diagramme en deux dimensions temps-fréquence [10]. On parle de spectrogramme à large bande ou à bande étroite selon la durée de la fenêtre de pondération. Les spectrogrammes à bande large sont obtenus avec des fenêtres de pondération de faible durée (typiquement 10 ms), ils mettent en évidence l'enveloppe spectrale du signal, et permettent par conséquent de visualiser l'évolution temporelle des formants, voir figure (1.8) .Les périodes voisées y apparaissent sous la forme de bandes verticales plus sombres. Les spectrogrammes à bande étroite sont moins utilisés. Ils mettent plutôt la structure fine du spectre en évidence : les harmoniques du signal dans les zones voisées y apparaissent sous la forme de bandes horizontales [7, 12].

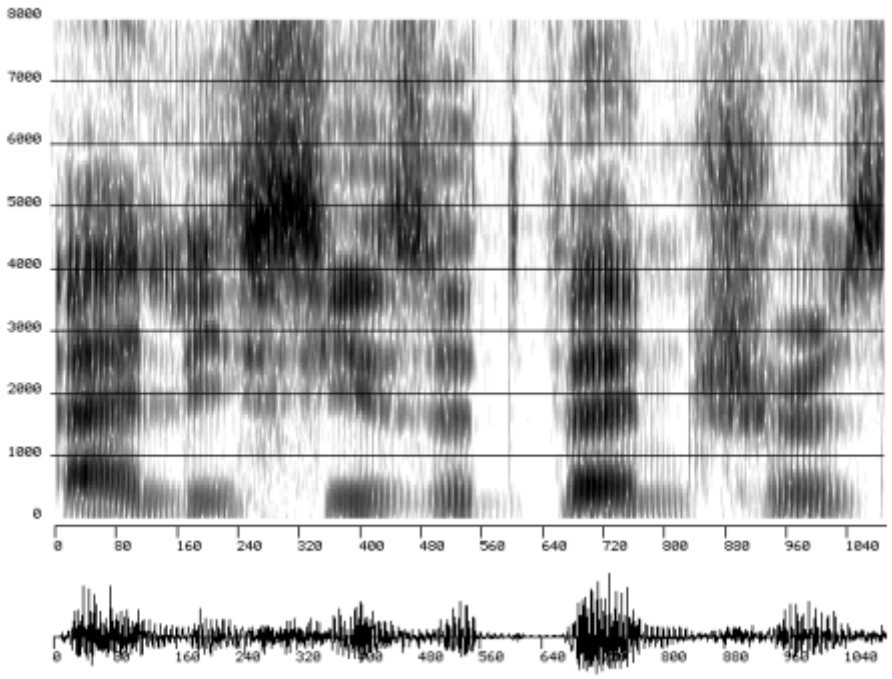

Figure 1.8: Spectrogramme et évolution temporelle [12].

1.6 Système de reconnaissance automatique de paroles SRAP

1.6.1 Définition

Dans l'informatique Il existe deux domaines du traitement automatique de la parole. Le premièr est la reconnaissance automatique de la parole , Le second est la synthèse vocale. La reconnaissance automatique donne la possibilité à la machine de comprendre et de traiter les informations fournies oralement. à la différence de la synthèse vocale qui consiste à transformer l'entrée du texte par l'utilisateur humain en parties sonores prononcés

1.6.2 Les avantages de la reconnaissance automatique de la parole

Les avantages de la reconnaissance de la parole sont nombreux. Elle libère complètement l'usage de la vue et des mains, et laisse l'utilisateur libre de ses mouvements. La vitesse de transmission des informations est supérieure, dans la RAP à celle que permet l'usage du clavier. Enfin tout le monde ou presque sait parler, alors que peu de gens sont à l'abri des fautes de frappe et d'orthographe, etc.
Ces avantages sont tellement importants que l'on trouve déjà sur le marché des dispositifs d'utilisation limitée, mais néanmoins efficaces. Citons certaines applications qui ont déjà vu le jour :

- Saisie vocale de données ;
- Donne des ordres tout en pilotant une automobile ou un avion ;
- Aide aux handicapés ;
- Chambre d'hôpital avec possibilités de commandes vocales pour le malade ;
- Commande vocale de machines ou de robots ;
- Commande vocale d'une montre portable, etc.

1.6.3 Applications de la reconnaissance de la parole

Les applications de la reconnaissance automatique de parole sont nombreuses. Elles existent là où la parole peut remplacer ou compléter une interface existante pour communiquer avec une machine, par exemple, pour accéder à un service ou contrôler une fonctionnalité d'un équipement. La parole s'impose parfois comme le seul mode de communication, par exemple dans les applications mains-libres où l'utilisateur ne touche pas l'équipement. Le tableau 1.2 présente quelques exemples d'applications démontrant l'intérêt de la reconnaissance automatique de parole.

Domaine	Applications
Téléphonie	Automatisation de transactions téléphoniques (ex : opérations bancaires), selfservice téléphonique pour l'accès à des services d'information (ex : consultation des bulletins météorologiques), etc.
Automobile	Contrôle mains-libres des équipements tels que la radio, le conditionnement dans le système de navigation, le téléphone sans fil (ex : voice dialing), les systèmes télématiques, etc.
Multimédia	Logiciels de dictée vocale, interaction vocale dans les logiciels pédagogiques (ex : apprentissage des langues) et ludiques (ex : jeux vidéo), etc.
Médical	Aide aux personnes handicapées.
Industriel	Contrôle vocal de machines, application pour la gestion de stocks, etc.

Tableau 1.2: Exemples d'applications de la reconnaissance de la parole [13]

1.6.4 Approche de la reconnaissance

Il existe deux approches permettant d'aborder la reconnaissance de la parole : l'approche globale et l'approche analytique. Elles se distinguent essentiellement par la nature et par la taille des unités abstraites qu'elles s'efforcent de mettre en correspondance avec le signal de parole.

1.6.4.1 Approche globale

Dans l'approche globale, l'unité de base sera le plus souvent le mot considéré comme une entité globale, c'est à dire non décomposée. L'idée de cette méthode est de donner au système une image acoustique de chacun des mots qu'il devra identifier par la suite. Cette opération est faite lors de la phase d'apprentissage, où chacun des mots est prononcé une ou plusieurs fois. Cette méthode a pour avantage d'éviter les effets de coarticulation, c'est à dire l'influence réciproque des sons à l'intérieur des mots. Elle est cependant limitée aux petits vocabulaires prononcés par un nombre restreint de locuteurs (les mots peuvent être prononcés de manière différente suivant le locuteur).[15]

1.6.4.2 Approche analytique

L'approche analytique, qui tire parti de la structure linguistique des mots, tente de détecter et d'identifier les composantes élémentaires (phonèmes, syllabes, ...). Celles-ci sont les unités de base à reconnaître. Cette approche a un caractère plus général que la précédente : pour reconnaître de grands vocabulaires, il suffit d'enregistrer dans la mémoire de la machine les principales caractéristiques des unités de base.

Dans l'approche globale, l'unité de base est le mot : le mot est considéré comme une entité indivisible. Une petite phrase, de très courte durée, peut aussi être considérée comme un mot.

Dans l'approche analytique, on tente de détecter et d'identifier les composantes élémentaires de la parole que sont les phonèmes.

Pour la reconnaissance de mots isolés à grand vocabulaire, la méthode globale ne convient plus car la machine nécessiterait une mémoire et une puissance considérable pour respectivement stocker les images acoustiques de tous les mots du vocabulaire et comparer un mot inconnu à l'ensemble des mots du dictionnaire. Il est de plus impensable de faire dicter à l'utilisateur l'ensemble des mots que l'ordinateur a en mémoire. C'est donc la méthode analytique qui est utilisée : les mots ne sont pas mémorisés dans leur intégralité, mais traités en tant que suite de phonèmes. Mais la méthode analytique a un grand inconvénient : l'extrême variabilité du phonème en fonction du contexte (effets de la coarticulation).[15]

1.6.4.3 Principe général de la méthode globale et analytique

Le principe est le même que ce soit pour l'approche analytique ou l'approche global, ce qui différencie ces deux méthodes est l'entité à reconnaître : pour la première il s'agit du phonème, pour l'autre du mot. On distingue deux phases :

— **La phase d'apprentissage**

Un locuteur prononce l'ensemble du vocabulaire, souvent plusieurs fois, pour créer en machine le dictionnaire de références acoustiques. Pour l'approche analytique, l'ordinateur demande à l'utilisateur d'énoncer des phrases souvent dépourvues de toute signification, mais qui présentent l'intérêt de comporter des successions de phonèmes bien particuliers.

— **La phase de reconnaissance**

Un locuteur prononce un mot du vocabulaire. Ensuite la reconnaissance du mot est un problème typique de reconnaissance de formes. Tout système de reconnaissance des formes comporte toujours les trois parties suivantes :
— Un capteur permettant d'appréhender le phénomène physique considéré (dans notre cas un microphone) ;
— Un étage de paramétrisation des formes (par exemple un analyseur spectral) ;
— Un étage de décision chargé de classer une forme inconnue dans l'une des catégories possibles.

1.6.5 La reconnaissance selon le mode d'élocution

Le mode d'élocution caractérise la façon dont on peut parler au système. Il existe quatre modes d'élocution distincts :

1.6.5.1 Reconnaissance de mots isolés

La segmentation d'un message parlé en ses constituants élémentaires est un sujet difficile. Pour l'éviter, de nombreux projets de la RAP se sont intéressés à la reconnaissance de mots prononcés isolement. La reconnaissance des mots isolés ou tous les mots prononcés sont supposés être séparés par des silences de durée supérieure à quelques dixièmes de secondes, se fait essentiellement par l'approche globale[9].

1.6.5.2 Reconnaissance de la parole continue

Bien que les méthodes les plus adaptées à la reconnaissance de la parole continue sont les méthodes analytiques, plusieurs tentatives ont été faites pour la généralisation des méthodes de reconnaissance globales. Ces systèmes dont l'étape de décodage acoustico-phonélique est fondamentale s'articulent le plus souvent sur le niveau lexical[9].

1.6.6 Schéma général d'un système de reconnaissance de parole

Les systèmes de reconnaissance de parole modernes reposent sur une architecture séquentielle et modulaire (figure 1.9), dans laquelle une onde acoustique de parole est mesurée et analysée afin d'en extraire son contenu linguistique sous la forme d'un ou d'une séquence de mots.

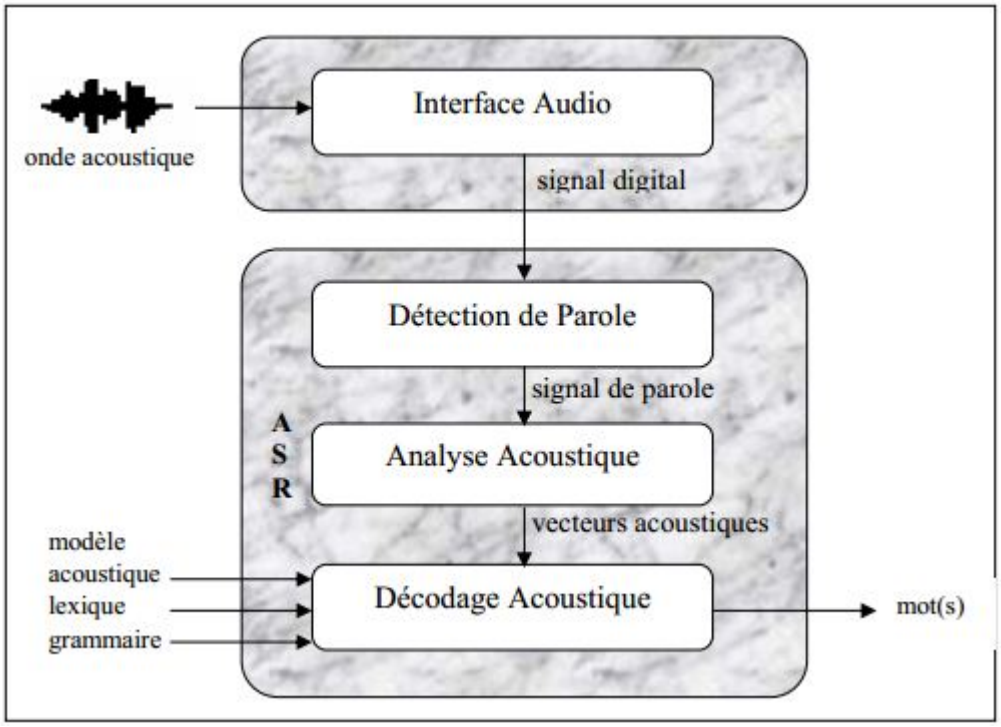

Figure 1.9: Système de reconnaissance automatique de parole .

Le schéma de principe (voir la figure 1.10) d'un système de reconnaissance de la parole continue fait apparaître une succession de modules réalisant les différentes étapes du processus de reconnaissance Tout d'abord, un module acoustique extrait les caractéristiques physiques du signal, destinées au module phonétique, qui reconnaît les phonèmes prononcés, c'est à dire les sons élémentaires de la langue, en faisant appel à un dictionnaire de phonèmes. Ensuite le module lexical et le module phonologique souvent confondus, reconnaissent les mots, et utilisant pour cela un lexique des mots autorisés par l'application considérée, ainsi que des règles phonologiques décrivant les assemblages possibles de phonèmes dans la langue. Les modules syntaxiques et sémantiques, qui n'en font parfois qu'un, reconnaissent la phrase, en faisant appel à une description des règles de grammaire, et du sens des mots autorisés pour l'application considérée. Un niveau supplémentaire, appelé prosodique, portant sur la mélodie, le rythme, l'intensité du discours orale, intervient en parallèle avec les autres modules et fournit les informations qu'il extrait du niveau acoustique (hauteur, intensité, rythme) à tous les autres niveaux.

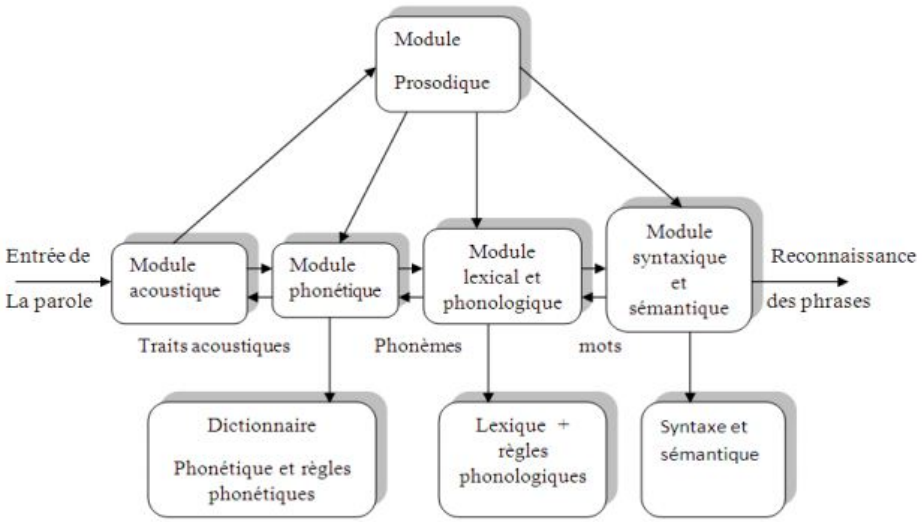

Figure 1.10: Schéma de principe d'un système de reconnaissance de la parole continue [5].

Conclusion

Dans ce chapitre nous avons vu un bref historique et les éléments de base des systèmes de reconnaissance automatique de la parole spécifiquement pour remplacer Les moyens de communication existants (homme-machine),et après nous avons énuméré les avantages et les applications de ce système (pour le grand public ou pour des personnes handicapées ...etc).

Chapitre 2

Traitement des paroles

Introduction

La reconnaissance automatique de la parole **ASR** (Automatic speech recognition) est un domaine de traitement de la parole concerné aux transformations de la parole en texte. ASR nous aide à rédiger des messages texte libres à l'intention des amis et des personnes sourdes ou malentendantes, afin qu'ils puissent interagir avec les communications parlées.
ASR fournit également un cadre pour la compréhension de la machine. Le langage humain devient interrogeable et exploitable, donnant aux développeurs la possibilité de dériver des analyses avancées telles que l'identification du locuteur ou l'analyse des sentiments.
Pour cette raison, Le traitement de la parole SP (Speech processing) est devenu comme une partie du domaine informatique pour améliorer la communication homme- machine.

On va voir dans ce chapitre Les étapes nécessaires dans le traitement de la parole et les moyens les plus utilisés dans chaque étape.

2.1 Caractéristiques de signal de parole

Parmi de ses caractéristiques, en tant qu'onde longitudinale, qui sont la fréquence, la longueur d'onde, et la vitesse de propagation qui dépend du milieu matériel de propagation.le signal électrique résultant est le plus souvent numérisé. Il peut alors être soumis à un ensemble de traitements statistiques qui visent à en mettre en évidence les traits acoustiques : sa fréquence fondamentale, son énergie, et son spectre. Chaque trait acoustique est lui-même intimement lié à une grandeur perpétuelle : pitch, intensité, et timbre. L'opération de numérisation, requiert successivement : un filtrage de garde, un échantillonnage,et une quantification. La fréquence de coupure du filtre de garde, la fréquence d'échantillonnage, le nombre de bits et le pas de quantification sont respectivement notés fc, fe, b, et q [7, 8]. Le son a par conséquent d'autres caractéristiques, qui sont :

2.1.1 L'énergie d'un son (intensité)

C'est la qualité qui fait distinguer un son fort d'un faible. L'intensité est liée à la pression de l'air en amont du larynx, qui fait varier l'amplitude des vibrations sonores. Souvent l'énergie observée dans un segment voisé est plus importante que celle observée dans un segment non voisé[7].

Exemple :
120 dB : L'avion au moment du décollage.
35 dB : maison dans quartier calme.

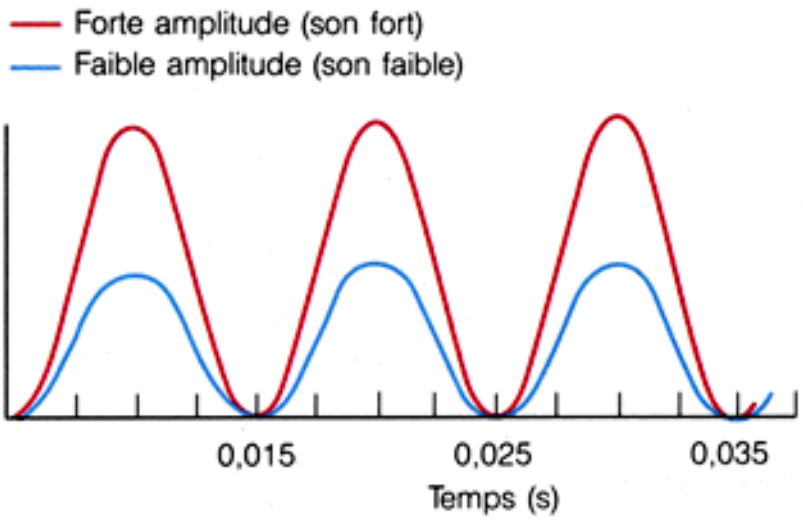

Figure 2.1: Intensité de son.

2.1.2 La fréquence fondamentale

La fréquence fondamentale F0 (ou pitch) joue un rôle important dans la parole. C'est elle qui véhicule une grande partie de l'information prosodique. L'intensité de la voix et les durées successives des syllabes complètent ces informations, elle peut faire ressortir bien des caractéristiques du locuteur, mais participe aussi à la caractérisation de la langue elle-même, par la manière dont elle est utilisée pour différencier les divers éléments syntaxiques comme les énoncés (interrogatifs, exclamatifs ou déclaratifs), l'importance de certains mots, ou bien même pour caractériser les différences lexicales entre les mots. Elle s'étend approximativement de 70 à 600 Hz pour l'homme, la femme et l'enfant (voir tableau 2.1), cette différence est due à la différence de la taille des cordes vocales ; les hommes adultes ont généralement une voix plus grave et des cordes vocales plus longues, soit entre 17 et 25 mm. Celles des femmes se situent entre 12,5 et 17,5 mm[7, 14].

Sexe/Age	**Plage de la fréquence fondamentale**
Homme	de 70 à 250 Hz
Femme	de 150 à 400 Hz
Enfant	de 200 à 600 Hz

Tableau 2.1: Les fréquences fondamentales pour l'homme,femme et enfant.

2.1.3 Le timbre

Le timbre est un ensemble de fonctionnalités et de caractéristiques qui permettent la différenciation entre les sons. Il est lié à des vibrations et à des résonances sortant du nez et de la gorge...etc pendant la prononciation, et qui déterminent le timbre de chaque son[4].

L'objectif principal de cette caractéristique est la distinction entre deux sons de même hauteur et de même intensité.

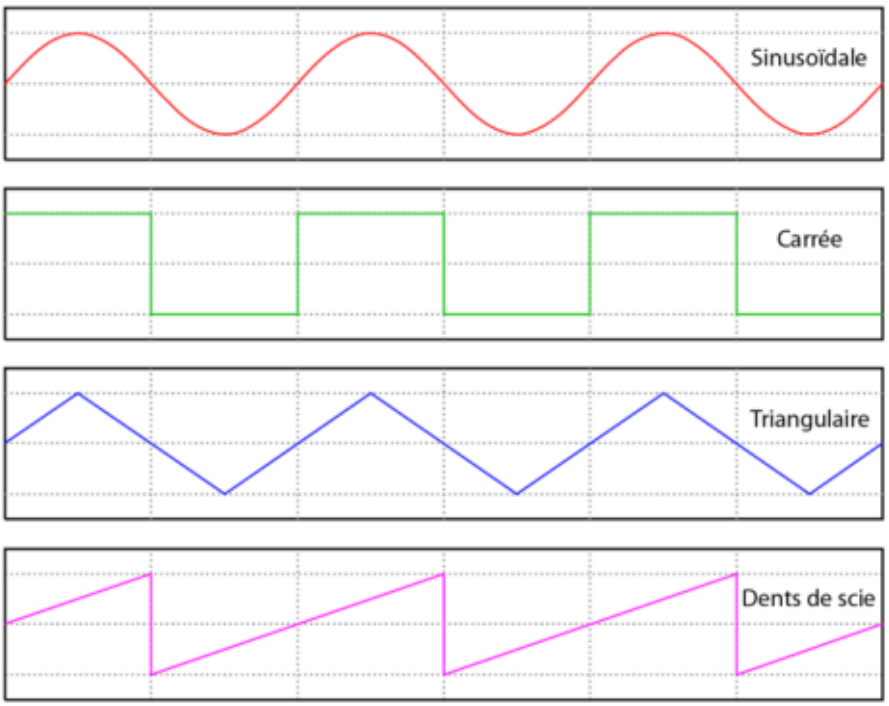

Figure 2.2: Le timbre de son.

2.2 Présentation de langue arabe

L'arabe est une langue parlée par plus de 250 millions de personnes classé sixième mondialement. Elle est une langue officielle dans moins 22 pays. C'est aussi la langue de référence pour plus d'un milliard de musulmans. Comme son nom l'indique.
la langue arabe est la langue parlée à l'origine par le peuple arabe. C'est une langue sémitique (comme l'hébreu, l'araméen et le syriaque). Au sein de cet ensemble, elle appartient au sous groupe du sémitique méridional. Le développement de la langue arabe a été associé à la naissance et la diffusion de l'islam. L'arabe s'est imposée, depuis l'époque arabo-musulmane, comme langue religieuse mais plus encore comme langue de l'administration, de la culture et de la pensée, des dictionnaires, des traités des sciences et des techniques. Ce développement s'est accompagné d'une rapide et profonde évolution (en particulier dans la syntaxe et l'enrichissement lexical). Elle a un alphabet de vingt-huit lettres, dont vingt-cinq représentent des consonnes et trois représentent les voyelles longues (و \ ي \ ا).

Chaque lettre apparaît souvent en quatre formes selon qu'elle soit en début, en milieu ou en fin de mot, ou isolée[16, 12]. Dans la phonologie l'alphabet arabe est classé selon des consonnes et des voyelles.

- **Les consonnes :** Une consonne est un phonème dont la prononciation se caractérise par une obstruction totale ou partielle en un ou plusieurs points du conduit vocal. Elle est généralement précédée ou suivie d'une voyelle [16].
- **Les voyelles :** Lors de la prononciation des voyelles, l'air émis par les vibrations des cordes vocales passe librement à travers le conduit. On distingue trois types de voyelles[16] :
  1. les voyelles courtes «" ُ", " ِ", " َ"»
  2. les voyelles longues «" و ", " ي ", " ا "»
  3. les semi-voyelles « sekune et tanwin ».

le tableau (2.2) illustre les différentes classes des sons arabes classés selon le mode d'articulation.

Mode d'articulation							
voyelle orale	semi voyelle	fricative	son combiné	occlusive	Nasale	Littérale liquide	Vibrant
ـَ	و	ف	ج	ب	م	ل	ذ
ـُ	ي	د		ض	ن		
ـِ		ث		ر			
أ		ظ		ل			
أو		س		ت			
إي		ز		ط			
		خ		ك			
		ح		ف			
		ص					
		ش					
		ع					
		غ					
		ه					

Tableau 2.2: Classification des sons arabes selon le mode d'articulation . [12].

2.3 Les opérations de traitement

Parmi les types d'opérations de traitement possibles et successifs sur la signal de parole , Nous voulons parler de :

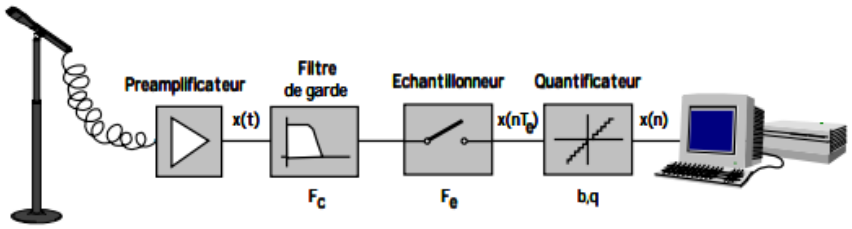

Figure 2.3: Enregistrement numérique d'un signal [2] .

2.3.1 La numérisation

La numérisation est une opération du traitement , dont le rôle est la conversion des informations de type complexe :

- texte
- image
- audio et vidéo
- signal électrique...etc

les informations de type numérique sont traitables par l'ordinateur.

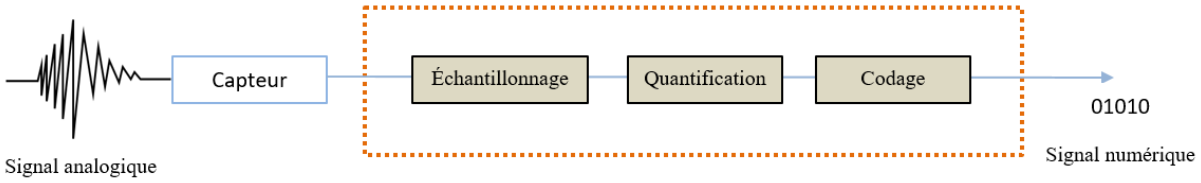

Figure 2.4: La convertisseur analogique numérique.

2.3.2 L'échantillonnage

L'échantillonnage consiste à choisir une partie de chaque ensemble d'échantillons à étudier...
Il prend les valeurs du signal à intervalles réguliers, et d'une façon systématique. Afin de produire une série de valeurs discrètes.

La fréquence d'échantillonnage est d'une extrême importance dans ce processus, Lors de la grande fréquence il donne de bons résultats, Et pendant, la petite il donne des résultats inexacts.

En mathématiques, L'échantillonnage est utilisé pour changer une fonction f (x) à valeurs continues en une fonction f' (x) discrète constituée par l'ensemble de valeurs a(x) aux instants d'échantillonnaget = kx avec k est un constant[4].

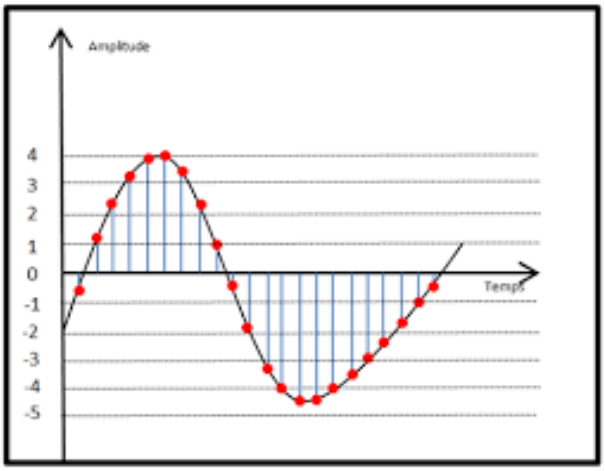

Figure 2.5: un signal échantillonné[4].

2.3.3 La Quantification

La quantification est le processus de détermination d'une valeur spécifique à Un ensemble limité de valeurs, Afin de faciliter le traitement par ordinateur.
Cette étape est de donner des valeurs rapprochées aux vraies valeurs des échantillons.

L'échelle de quantification est la plage qui fait cette approximation.
Le bruit de quantification est l'erreur systématique résultant du rapprochement de ces valeurs à la valeur spécifique Grâce au travail de la quantification[4].

2.3.4 Le Codage

C'est la représentation binaire des valeurs quantifiées qui permet le traitement du signal sur machine.De façon générale un codage permet de passer d'une représentation des données vers une autre.

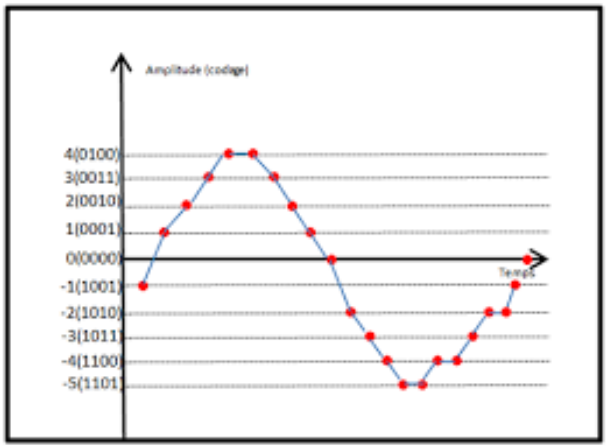

Figure 2.6: un signal quantifié[4].

2.4 Généralité de la segmentation

Pour bien comprendre la segmentation, il est évident de la définir et de connaitre ses différentes paramétrés :

2.4.1 Définition de la segmentation

La segmentation de la parole est une opération nécessaire dans le traitement de la parole, qui consiste à découper le signal en segments assez extrêmement homogènes pouvant être transcrits en unités de base (phonème, syllabe. . .). Ou on trouve ces unités variées selon la nature du segment considéré (figure 2.7). Il existe plusieurs types de segmentation selon la taille du segment traité, on cite quelques types de segment du plus court au segment le plus long [3] :

- **Segment en voisé/non-voisé :** Les sons voisés se résulte par la vibration des cordes vocales. Les voyelles sont généralement voisées, cependant les consonnes peuvent être voisées ou non .

- **Segment en phonèmes :** Cette technique consiste à délimiter la continuité acoustique d'un signal à une séquence de segments d'un ensemble discret et fini d'éléments, qui est l'alphabet phonétique de la langue (exemple : le mot "arab" en le divise en " a,r,a et b").

- **Segment en syllabes :** La syllabe est l'unité structurante de la langue, elle est décomposée en 3 parties : l'attaque,le noyau et la coda. On trouve quelquefois une difficulté de segmentation d'une phrase en syllabes à cause de la caractéristique facultative des consonnes.

- **Segment en mots :** il est difficile de segmenter un message en ses constituants élémentaires. Pour résoudre cette complexité, on se base à la reconnaissance de mots prononcés isolés qui sont séparés par des silences de durée supérieure à quelques dixièmes de secondes, elle applicable par l'approche globale.
- **Segment en locuteurs et tours de parole :** La segmentation selon le locuteur apparait pour résoudre l'ambiguïté entre plusieurs locuteurs, il s'agit de segmenter en tours de parole pour chaque locuteur.

- **Segment en groupes inter-pausaux :** segments délimités par deux pauses silencieuses.

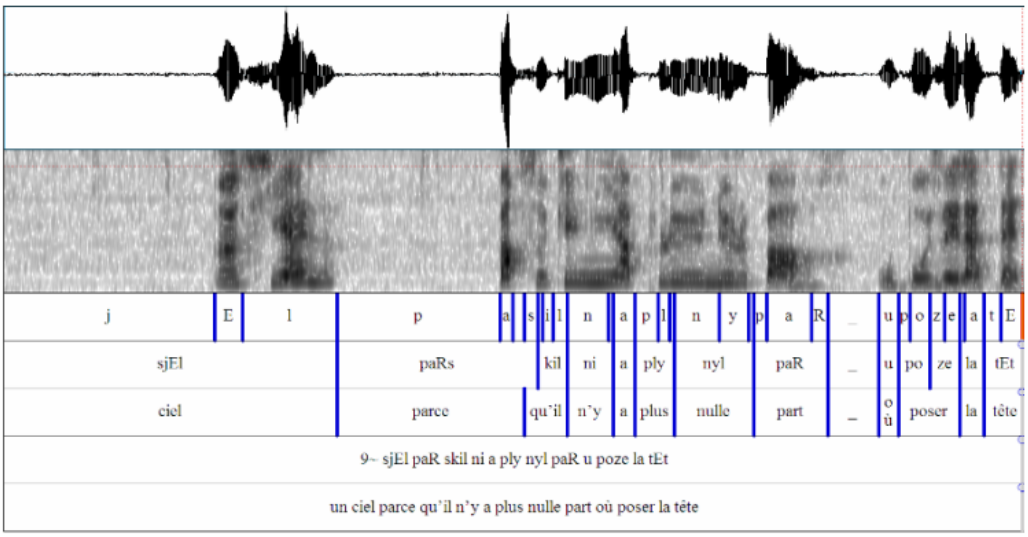

Figure 2.7: Exemples de segmentation de parole continue.

2.4.2 Paramétrisation du segment

Pour chaque segment un vecteur de paramètres (traits acoustiques) est extrait, ces paramètres peuvent être[5] :

- **pertinents :** Extraits de mesures suffisamment fines, ils doivent être précis mais leur nombre doit rester raisonnable (éliminer la redondance des données) afin de ne pas avoir de coût de calcul trop important dans le module du décodage.
- **discriminants :** Ils doivent donner une représentation caractéristique des sons de base et les rendre facilement séparables.
- **robustes :** Ils ne doivent pas être trop sensibles à des variations de niveau sonore ou à un bruit de fond.

2.4.3 Approches de la segmentation

Il existe trois approches permettant d'aborder la reconnaissance de la parole, Nous voulons parler de :
— L'approche globale.
— l'approche analytique.
— l'approche hybride (statistique).

2.4.3.1 L'approche globale

Cette approche basée sur la décomposition de la parole continue la plupart du temps en utilisant un décodage acoustico-phonétique utilisé par des modules de niveau linguistique[12].

2.4.3.2 l'approche analytique

Cette approche basé sur l'identification globale d'un mot ou une phrase en utilisant la notion de comparaison de l'identifient avec des références enregistrées[12].

2.4.3.3 l'approche hybride (statistique)

L'introduction des méthodes statistiques basées sur des modèles de Markov à donner une intégration pour la reconnaissance de la parole continue et le traitement des grands vocabulaires[12].

2.4.4 Les méthodes de segmentation

Les méthodes de la segmentation acoustique de la parole se divisent en deux grandes classes de méthodes selon leur démarche au connaissance à priori du contenu linguistique de ce signal ou non [18]. Tel que :

1. **Segmentation sans contrainte linguistique :**

Ce sont les méthodes de segmentation du signal de paroles Sans connaissance préalable du contenu de ce signal. Lorsqu'elles travaillent sur la division d'une manière générale et sans une étude du contenu linguistique ou autre .

2. **Segmentation avec contrainte linguistique :**

Ce sont les méthodes de segmentation qui sont basées sur l'utilisation de certaines contraintes linguistiques et cognitifs, Afin de diviser le signal de paroles[3].

2.4.5 Les problèmes liés à la segmentation

On peut les diviser en 3 groupes principaux :

1. **Les problèmes liés au locuteur :**
   - **la variabilité inter locuteur :**
     - Age et sexe...etc
     - type d'élocution .
   - **la variabilité intra locuteur :**
     - Les conditions psychologiques (stress, émotion).
     - Les conditions physiques (fatigue, rhume).
2. **Les problèmes liés au contexte :**
   - Discours non clair.
   - La vitesse d'élocution.
3. **Les problèmes liés à l'environnement :**
   - le bruit .
   - le mouvement des choses .

2.5 la classification

le traitement de la parole est un domaine de recherche scientifique renouvelé. Lorsque , un groupe de chercheurs ont effectué plusieurs études pour apprendre les meilleures méthodes qui donnent une meilleure classification.

La classification est un processus d'intelligence artificielle qui travaille pour que le comportement de la machine soit plus intelligent

2.5.1 Les phases de la classification

Donc Comme nous avons noté dans la précédente figure, Il y a deux phases principales dans le processus de la classification qui :

1. **L'apprentissage :** La phase d'apprentissage est la phase la plus importante dans la reconnaissance , elle est intéressé de la création du vocabulaire (dictionnaire linguistiques).
2. **Test ou Décision :** La décision est la deuxième étape de reconnaissance,Lorsque la machine fonctionne pour Déterminer le modèle le plus proche dans le dictionnaire à La nouvelle entrée. cette estimation doit être en temps court que possible.

Comme nous avons dit précédemment Il y un facteur appelé le taux de classification qui permet de mesurer la performance de la classification[16].

2.5.2 Exemples de méthode de classification et segmentation

1. **Classification Bayésienne :** Les réseaux bayésiens sont des réseaux qui nous permettent de trouver de nouveaux états de décision probabiliste à partir de connaissances incertaines. Ils résultent de la combinaison entres des domaines scientifiques : La probabilistes et la théorie de graphes.

$$P(B|A,c) = \frac{(P(B|c)P(A|B,c)}{(P(A|c)}$$

tels que :
— c est un contexte
— A,B deux événements
— P(B|A,c) est la probabilité de B si A est vrai .
— P(B|c) est la probabilité a priori de l'événement B
— P(A|B,c) est la probabilité de A si B est vrai
— P(A|c) est la normalisation

2. **méthode k-ppv (K-Plus Proches Voisins) :** Cette méthode de classification est basée de façon globale sur l'approximation . de sorte qu'elle vise à trouver pour Élément d'entrée l'élément le plus proche dans le dictionnaire

Conclusion

Dans ce chapitre nous avons vu les méthodes de segmentation de la parole les plus connues et ces classifications
On peut dire que depuis une décennie, les techniques de traitement de la parole ont connu plusieurs grandes révolutions.

Chapitre 3

Conception et mise en œuvre

Introduction

3.1 Mise en œuvre du système

Notre travail vise à développer une application capable de faire l'Acquisition d'un signal de parole arabe enregistré par l'utilisateur ,ensuite faire un ensemble de traitements afin de transformer le dernier en spectrogramme et le segmenter, pour faciliter l'extraction des vecteurs acoustiques ,et classifier les lettres .

Le système commence par l'acquisition et le prétraitement du spectrogramme de signal de la parole (eventuellement réalisées par les bibliothèques speech recognition et pylab), ensuite ,applique l'algorithme de segmentation proposé pour extraire les différentes spectrogramme de chaque lettre .Par la suite ,extraire un vecteur de caractéristiques (histogramme) de chaque lettre
Finalement ,ces vecteurs sont passés à au classificateur FLC afin de construire la base(file.csv :Contient les vecteurs de chaque lettres) pour connaitre la classe de chaque lettre dans la phase de décision (Figure 3.1) .

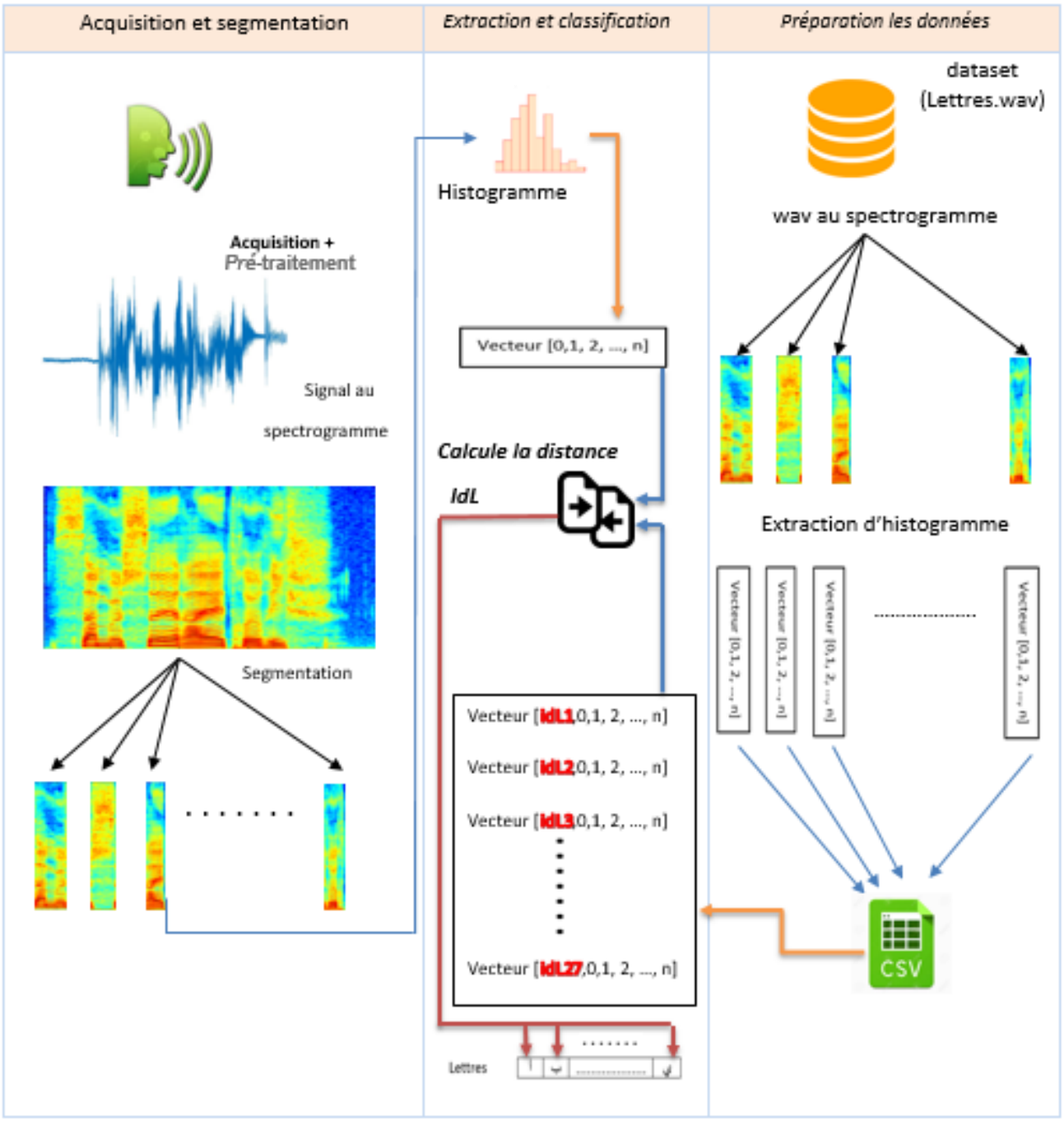

Figure 3.1: Architecture générale de notre système .

3.1.1 Description des étapes de système

Pour bien comprendre notre travail, il est évident de décrire l'ensemble des étapes réalises au cours de cette application en détaillant chaque phase.

3.1.1.1 Acquisition

La première phase de notre système ,c'est l'acquisition de signal vocal enregistré par l'utilisateur via un microphone ,ainsi la carte son de micro-ordinateur permet de réaliser les étapes de traitement comme la numérisation l'échantillonnage et la quantification. ces étapes peuvent être réalisées par l'utilisation de la bibliothèque nommé **speech recognition**
Pour commencer l'acquisition nous avons besoin de lancer un thread , L'algorithme suivant représente la méthode record() de ce thread :

algorithme d'acquisition

```
Fonction acquisition (stream)
 Début
    frames=[]
    Tanque(utilisateur entre voix)
        data=stream.read(CHUNK)  //lire les données
        frames.append(data)
    FinTanque
Fin
```

3.1.1.2 Pré-traitement

cette phase consiste à utiliser des filtres pour améliorer la qualités du signal acquis pour qu'il soit prêt à la phase suivante. Ces filtres permettant de diminuer les effets indésirables influant sur la qualité du signal de la parole prévenant des outils d'acquisition.

3.1.1.3 transformation le signal au spectrogramme

nous pouvons résumer la transformation le signal au spectrogramme dans ces trois étapes

1. Divisez le signal en segments de même longueur. Les segments doivent être suffisamment courts pour que le contenu fréquentiel du signal ne change pas sensiblement dans un segment. Les segments peuvent ou non se chevaucher.
2. Fenêtre chaque segment et calculer son spectre pour obtenir la transformation de Fourier à court terme.
3. Affichez segment par segment la puissance de chaque spectre en décibels. Représentez les magnitudes côte à côte sous forme d'image avec une palette de couleurs dépendante de la magnitude.

algorithme de transformation le signal au spectrogramme

```
Fonction signal_spect(fille_wav,path_save)
   Début
    wav = wave.open(fille_wav, 'r')  // Lire le fichier audio
    sound_info = pylab.fromstring(wav, 'int16')  // Convertir les informations en un
    frame_rate = wav.getframerate() // Extraire les dimensions de l'audio
    pylab.specgram(sound_info, Fs=frame_rate) // Convertir l'audio en spectrogramme
    out=path_save
    pylab.savefig(out,transparent=True) //Enregistrer spectrogram sous forme d'image
   Fin
```

3.1.1.4 Méthode de segmentation proposée :

Cette phase a pour but d'extraire les segments acoustiques de spectrogramme nécessaires pour la phase de décision.

1. Lire l'image
2. Tracer une fenêtre avec la même largeur de l'image, la longueur à partir de 0 à 150

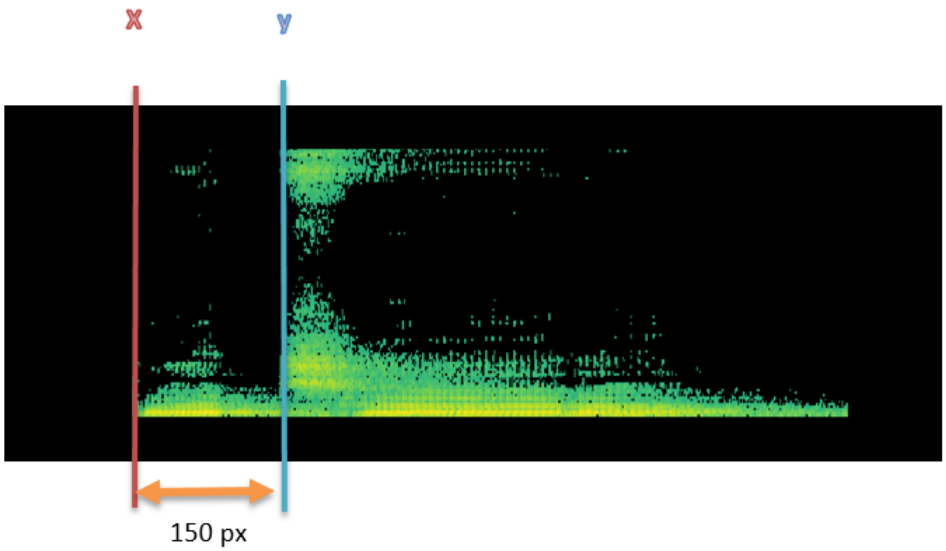

Figure 3.2: détermination de la fenêtre.

3. Extraire les dimensions de l'image.
4. Faites correspondre la fenêtre à l'image et coupez la partie de 0 à 100 pour qu'elle soit la première lettre.
5. nous décalons de 100 fois les dimensions de la fenêtre et comparons la longueur de l'image.

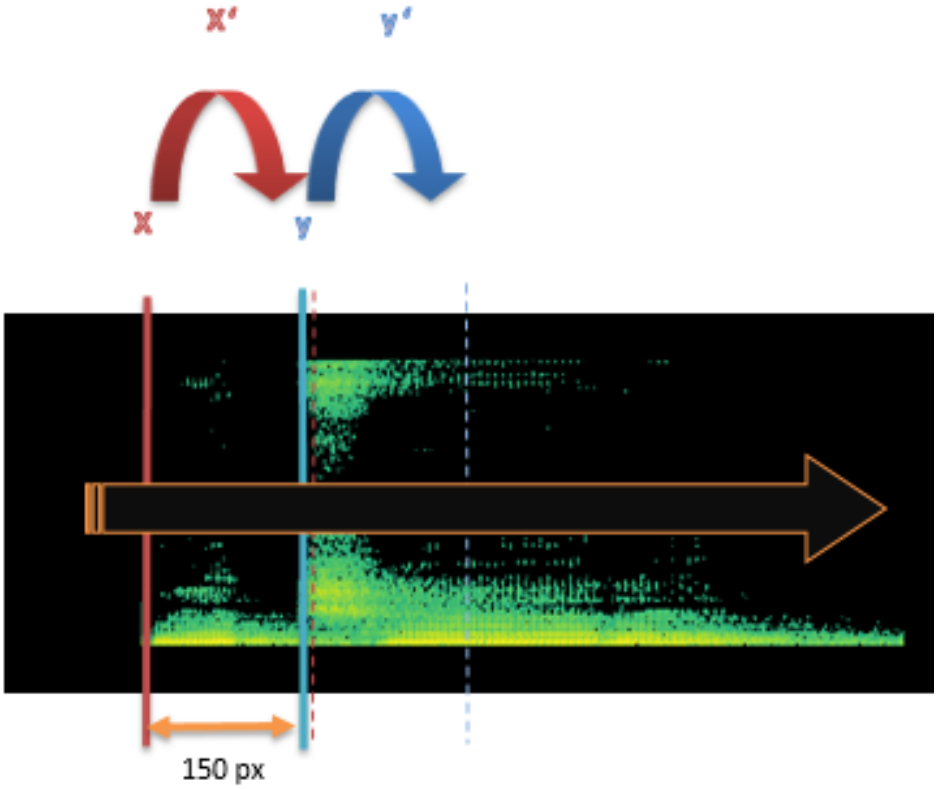

Figure 3.3: le déplacement de la fenêtre.

6. Retournez à l'étape 3

algorithme de segmentation

```
Fonction segmentation (img)
  liste_segment=[]
  x=117
  y=267
  Début
      Tanque(x < 600)
          t=img[0:432,x :y]
          liste_segment.append(t)
          x=x+100
          y=y+100
      FinTanque
    return liste_segment
  Fin
```

3.1.1.5 Extraction des caractéristiques :

dans cette étape, nous nous sommes concentrés sur les caractéristiques de couleur du spectrogramme, la meilleure façon de les fournir est de mettre en œuvre des histogrammes.

- extraire les valeurs de couleur (RGB) de chaque pixel de chaque segment
- Calculer le nombre de pixels de chaque couleur (RGB)
- Fusionner les tables résultantes

L'algorithme proposé pour extraire l'histogramme :

algorithme d'extraction l'histogramme

```
Fonction extract_vecteur(img_spect,mask)
    Début
     hist = cv2.calcHist([img_spect],
     [0, 1, 2], mask, self.bins,
     [0, 180, 0, 256, 0, 256]) // Calcule l'histogramme
     hist = cv2.normalize(hist).flatten() // normalizer les Valeurs d'histogramme Dans
     return hist
    Fin
```

3.1.1.6 Classification :

Dans cette étape Nous allons premièrement sélectionner les lettres extraites de l'image et après On calcule la distance entre (le vecteur) extraite précédemment Et le fichier CSV sauvegardé que nous avons déjà extrait

- **Pré-classification :**
  Nous allons transformer les données (dataset) sauvegardées de tous les lettres de type (.wav) aux spectrogramme et après on les sauvegardées sous forme d'image de type (.PNG), et après nous allons extraire les vecteurs (déjà expliqué) et les sauvegardés sous forme de fichier de type (.csv)

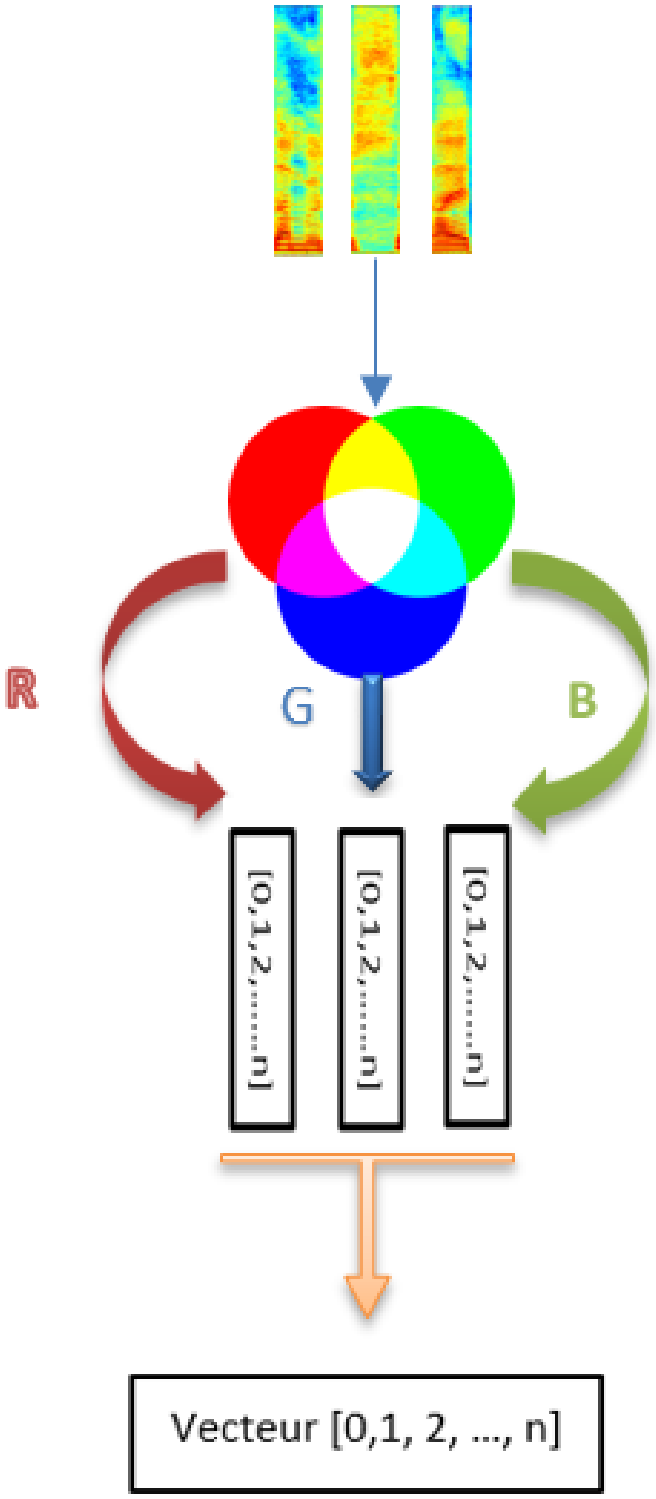

Figure 3.4: Extraction des caractéristiques de spectrogramme .

algorithme de transformation des données

```
Fonction transform_data(path_data)
    Début
        y=0
        ID=1
        vecteur=0
        output = open("out.csv", "w")
        Pour(fille a path_data\*.wav)
            signal_spect(fille,"Model\Z{}.png".format(y))
            y=y+1
        FinPour

        Pour(i=0 a 27)
                file = 'Model/Z{}.png'.format(i)
                img=cv2.imread(file)
        hsv = cv2.cvtColor(img, cv2.COLOR_BGR2HSV)//Convertir le type de couleur en hsv
        lower_yellow = np.array([15,50,50])
        upper_yellow = np.array([80, 255, 255])
        mask = cv2.inRange(hsv, lower_yellow, upper_yellow) // appliquer mask entre deux niv
                vecteur=extract_vecteur(img,mask)
                vecteur= [str(f) for f in vecteur]
                output.write("%s,%s\n" % (ID, ",".join(vecteur))) //Ajouter au fichier csv
                ID=ID+1
        FinPour
    Fin
```

Remarque :

— chaque vecteur sauvegardé avec un id
— chaque id représente une lettre
cette étape se fait une seule fois

algorithme de classification

```
Algorithme:classification
 leter=["أ","ب","ت","ث","ج","ح","خ","د","ذ","ر","ز","س","ش","ص","ط","ظ","ع","غ","ف",'
 x=117
 y=267
 Début
   acquisition (stream).start //Commencer l'enregistrement
   acquisition (stream).stop //Arrêter l'enregistrement
   signal_spect("sound.wav","recherch.png")//Convertir l'audio en spectrogram et sauv
    img=cv2.imread("recherch.png")

    hsv = cv2.cvtColor(img, cv2.COLOR_BGR2HSV)//Convertir le type de couleur en hsv
    lower_yellow  = np.array([15,50,50])
    upper_yellow  = np.array([80, 255, 255])
    mask = cv2.inRange(hsv, lower_yellow , upper_yellow ) // Proposition mask entre c
    Tanque(x < 600)
                t=img[0:432,x :y] //Coupé partie de la image
                vectour=extract_vecteur(t,mask) //Extrait des propriétés(histogram)
                results = recherche(vectour) //Trouver les vecteurs les plus proches
                hist,ID=results .get_min()
                mot=mot+leter[ID] //Choisissez une lettre
                x=x+100
                y=y+100
      FinTanque
 Fin
```

Conclusion

Dans ce chapitre, nous avons décrit premièrement L'architecture générale proposée de notre système de reconnaissance automatique de la parole et après nous avons montré Les différentes phases de système précisément la segmentation Que nous avons utilisé une méthode innovante en utilisant le spectrogramme en vue de trouver un SRAP optimal pour la langue arabe.

Chapitre 4

Résultats et discussion

Introduction

Après avoir achever la phase de conception du système proposé tout en détaillant les différentes étapes suivis et leurs algorithmes réalisés de chacune, il est évident de démonter la concrétisation du travail. Pour cela nous allons consacré ce chapitre à décrire l'environnement de la programmation, l'interface globale du système avec les fonctionnalités offertes. Ensuite nous allons exposer les résultats obtenus avec une comparaison des différents types d'enregistrements effectués.

4.1 Choix du langage de programmation

Nous avons choisi le langage Python comme environnement de programmation, car Python est le langage favori d'un très grand nombre de développeurs comparé à d'autres langages comme Java, PHP, C++, etc..

- Python est facile à apprendre
- Python est un langage de choix, c'est-à-dire à usage général.
- Python permet de créer plus de fonctions avec moins de lignes de code
- Python est multiplateforme et open source.

Il existe plusieurs outils de développement visuels (IDE,Editor),tels que :

- Eclipse + PyDev. Category : **IDE**.
- Visual Studio Code. Category : **IDE**.
- PyCharm. Category : **IDE**.
- Sublime Text. Category : **Editor**.
- Atom. Category : **Editor**.
- GNU Emacs. Category : **Editor**.

4.2 Interface du système

figure ci-dessous ,présente l'interface utilisateur globale du système

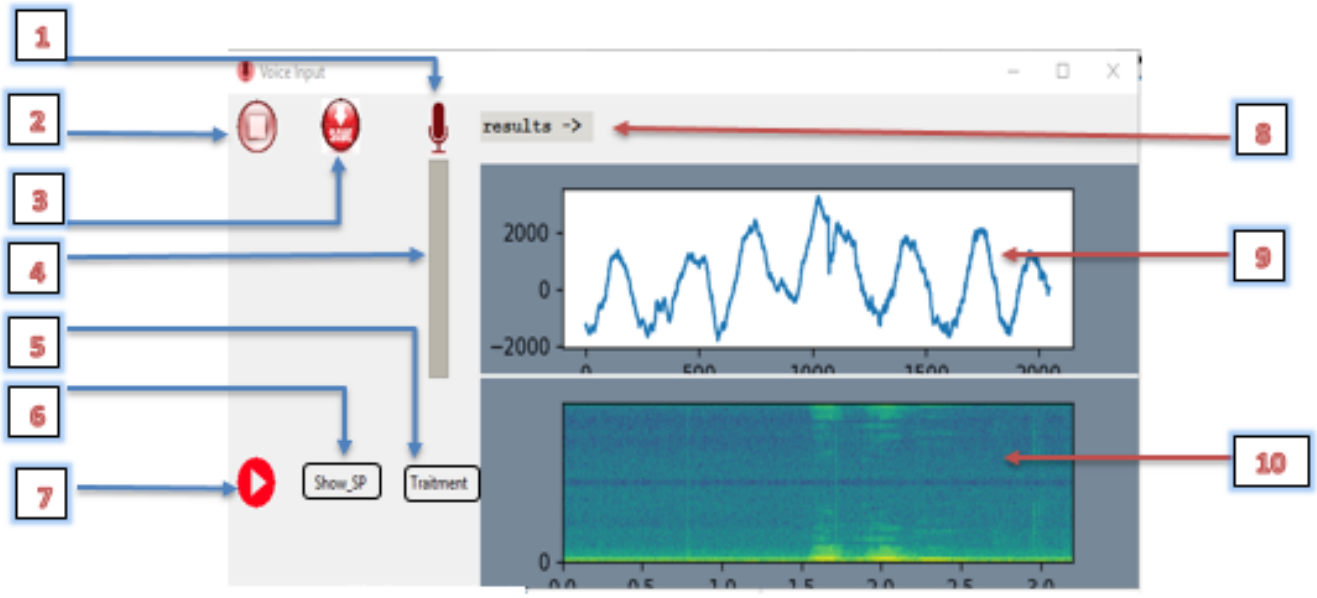

Figure 4.1: L'interface utilisateur globale du système .

1. Enregistrer.
2. stop.
3. sauvegarder.
4. l'amplitude d'un signal.
5. segmenter et classifier.
6. afficher spectrogramme.
7. lecture audio.
8. afficher résultat.
9. signal temporel
10. spectrogramme de parole.

4.3 Fonctionnalité du système

Notre application permet d'enregistrer la parole et la segmenter, pour faire cela :

1. L'utilisateur clique sur le bouton " enregistrer" et commence sa parole.
2. Lorsqu'il veut compléter sa parole, nous cliquons sur le bouton "stop",
3. pour sauvegarder la parole enregistrée, cliquons sur le bouton "sauvegarder" .
4. pour écouter la parole enregistrée, cliquons sur le bouton "lecture audio".
5. pour afficher le spectrogramme de parole enregistrée, nous cliquons sur le bouton "afficher spectrogramme" .
6. pour segmenter et classifier la parole enregistrée, nous cliquons sur le bouton "traitement"(afficher le résultat on 8)

4.4 Test et bilan

Afin d'essayer d'estimer le degré de performance de notre programme,Nous l'avons testé sur trois mots :

— بتر
— اكل
— سماء

Après la segmentation, nous avons obtenu les résultats suivants :

— ب ت ر ا
— ا ك ك ل
— ص س م اء

Maintenant, nous allons essayer d'appliquer cette méthode sur un ensemble de test dont on a pris premièrement les mêmes que les mots précédents prononcé dix (10) fois par trois (3) personnes de différent âge et sexe. On a évalué le degré de segmentation et classification de chaque énoncé a partir de nombre des lettres correctes que nous avons obtenus.
Enfin, on a calculé le taux de segmentation et classification (TSC) correcte de chaque personne. Le tableau suivant résume des résultats :

mots / **personne**	بتر	أكل	سماء	**TSC**
teste 1	5	6	4	0.50
teste 2	5	7	5	0.56
teste 3	6	5	7	0.60

Tableau 4.1: Résultats de segmentation et classification.

Conclusion

Nous avons présenté dans ce chapitre l'interface de notre système avec les différentes fonctionnalités, ainsi que les résultats de notre méthode proposée.
Les résultats obtenus sont très satisfaisants et ouvrent d'autres perspectives dans le domaine de la reconnaissance de la parole

Conclusion

L'objectif de la reconnaissance vocale est de convertir l'audio en texte. Cette technologie est largement appliquée dans notre vie. Google Assistant et Amazon Alexa sont quelques-uns des exemples qui prennent notre voix en entrée et la convertissent en texte pour comprendre notre intention.

Jusqu'à aujourd'hui, La reconnaissance du son et la reconnaissance de la voie arabe en particulier présente un défi très grand, malgré les efforts et les travaux intensifs réalisés dans ce domaine, aucun système RAP n'est jugé fiable.

Dans ce travail ont été intéresser à présenter le signal de la parole sous forme de spectrogramme, parce que à partir d'un spectrogramme, vous pouvez visualiser beaucoup des caractéristiques que le signal.

Suite à la réalisation de notre système nous avons rencontrée beaucoup des difficultés et l'un des problèmes critiques est le manque de données adéquates sur le volume de formation surtout dans la phase de training, Cela mène à une sur-adaptation ou à la difficulté de traiter des données invisibles. Malgré ça nous nous sommes adaptés avec ce problème-là pour réaliser ce système.

Pour valider notre approche, nous avons effectué des tests avec différents locuteurs selon le sexe, l'âge sur des mots et des phases. Les résultats obtenus sont encourageants et montre la robustesse de notre méthode.

Comme perspective, nous voyons qu'il est possible d'améliorer cette méthode par compilation beaucoup de données.

Bibliographie

[1] J. ALLEGRE – *Approche de la reconnaissance automatique de la parole*, vol. 13, 2013.

[2] F. Benaissa – *Modèles de markov caches appliqués a la reconnaissance automatique de la parole », thèse*, 1999.

[3] — , *Segmentation non supervisée d'un flux de parole en syllabes, mémoire*, 2012.

[4] R. Benammar – *Traitement automatique de la parole arabe par les hmms : Calculatrice vocale*, 2012.

[5] A. Bendahmane – *Cours de traitement automatique de la parole*, vol. 1, 2014.

[6] D. Benoit – *Les microphones*, 1997.

[7] K. Berbache – *Modèles de markov cachés : Application a la reconnaissance automatique de la parole, mémoire*, 2014.

[8] S. E. Berchaoua – *Reconnaissance de la parole arabe par les supports vecteurs machines (svm), mémoire*, 2014.

[9] O. Douib – *Reconnaissance automatique de la parole arabe par cmu sphinx 4, mémoire*, 2013.

[10] T. Dutoit – *Introduction au traitement automatique de la parole*, vol. 1, 2000.

[11] S. Z. B. D. Elhak – *Proposition d'une méthode de segmentation adaptative de la parole arabe*, vol. 71, 2016.

[12] L. Lazli-Boukhalfa – *Système neuro-markovien basé sur la fusion de données floues et génétiques : Application pour la reconnaissance automatique de la parole, thèse*, 2006.

[13] A. Lotfi – *Un système hybride ag/pmc pour la reconnaissance de la parole arabe*, 2005.

[14] — , *Modélisation ar et arma de la parole pour une vérification robuste du locuteur dans un milieu bruité en mode dépendant du texte, mémoire*, 2013.

[15] — , *Reconnaissance automatique de la parole arabe par cmu sphinx 4*, 2013.

[16] T. S. et M. OumElhana – *Proposition d'un modèle de descripteur structurel pour la voix arabe, application saisie des notes, mémoire*, 2015.

[17] N. eddine Mesbahi – *Conception et réalisation d'un système de pilotage d'un véhicule par commande vocale,memoire*, 2011.

[18] S. Nefti – *Segmentation automatique de corpus de parole continu dédie à la synthèse vocale »,thèse*, 2007.

[19] Netophonix – *Forum wiki découverte*, juillet 2010,consulté le 10 avril 2016.

[20] J.-F. PILLOU – *Carte son*, 2015.

[21] L. A. Simard – *Les microphones*, 2014.