République Algérienne Démocratique et Populaire

Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

Université Akli Mohand Oulhadj de Bouira

Faculté des Sciences et des Sciences Appliquées

Département d'Informatique

Mémoire de Master 2

en Informatique

Spécialité

Ingénierie des Systèmes d'Information et de Logiciel (ISIL)

Thème

Un dataset pour la reconnaissance de la parole arabe pour les locuteurs Tamazight pour le projet : Fabrication de systèmes de reconnaissance vocale arabe pour les locuteurs Tamazight

---

Encadré par

— M. OUKAS Nourredine.

Réalisé par

— Mlle CHABI Tiziri

— Mlle SARI Tilelli

2022/2023

Remerciements

Nous remercions le DIEU de nous avoir accordé la patience, la santé et le courage pour réaliser ce travail.

On tient à remercier nos très chers parents que nul remerciement, aucun mot ne pourrait exprimer à leur juste valeur la gratitude et l'amour qu'on vous porte. On met entre vos mains. le fruit de longues années d'études, de votre amour de votre tendresse, de longs jours d'apprentissage. Votre éducation votre soutien et votre encouragement nous a toujours donné de la force pour persévérer et pour prospérer dans la vie. Chaque ligne de cette thèse chaque mot et chaque lettre vous exprime la reconnaissance, le respect, l'estime et le merci d'être nos parents que Dieu vous garde.

Nous tenons à remercier notre encadreur OUKAS Noureddine pour ses conseils et ses encouragements qui nous ont permis de réaliser ce travail.

Nous tenons également à remercier les membres du jury pour l'honneur qu'ils nous ont fait en acceptant de juger ce travail, et d'avoir consacrer leurs temps pour la lecture et venir proclamer sa valeur scientifique.

Nous tenons à remercier tous ceux qui ont pris le temps de lire cette recherche, et à qui nous espérons avoir apporté des connaissances utiles.

Nous aimerions également exprimer notre profonde gratitude à tous ceux qui ont fait du bénévolat et contribué à la création de (CollectVoice), la base de données vocale arabe pour les locuteurs amazighs.

Enfin, nos sincères sentiments vont à tous ceux qui ont contribué de près ou de loin à la réalisation de ce projet.

Dédicaces

Je dédie ce mémoire :

A mes très chers parents, qui m'ont donné la vie et m'ont soutenu depuis toujours. Que Dieu les protège.

A mes sœurs Fatima, Samia, et Samira et mon frère Massi.

A mes très chères amies Amel, Lydia.

A ma chère binôme Telili.

Et à toute ma famille.

CHABI Tiziri

Dédicaces

À mes chers parents (mon père Mansour, ma mère Sadia), pour leurs sacrifices et leur soutien continu. Je vous dédie cette humble réalisation.

A mes frères (Djilali, Belkacem, Amar, Belaid, Chabane, Amazigh) qui m'ont toujours soutenu pour avancer.

A ma sœur Faiza (la femme de mon frère).

À mes amies Sonia, Lydia, Amel, Zohra et Imane qui ont partagé avec moi des moments de joie et de pression.

À ma chère binôme Tiziri.

SARI Telili

ملخص

يعد التعرف التلقائي على الكلام مجالًا نشطًا جدًا، من الأبحاث التي أنتجت تقنيات تم اعتمادها في العديد من المجالات مثل الصحة، الخدمات العامة والواجهات البشرية والآلية، لا سيما مع ظهور الشبكات العصبية وتقنيات التعلم العميق.

هذا المشروع يسلط الضوء على أهمية تطور أنظمة التعرف التلقائي على الكلام والتقنيات الحديثة في هذا المجال. كما تبرز أهمية توفير مجموعات بيانات عربية عالية الجودة لتدعم تطبيقات التعرف على الصوت في الأجهزة الذكية باللغة العربية.

الهدف الرئيسي لهذا العمل هو جمع مجموعة بيانات جديدة مصممة خصيصًا للتعرف على الكلام العربي من قبل الناطقين باللغة الامازيغية، مما يعزز التنوع اللغوي في البيانات المتاحة.

وتتضمن هذه العملية مراقبة الجودة للبيانات، مما يساهم في تحسين دقة نماذج التعرف على الكلام العربي باستخدام هذه المجموعة الجديدة من البيانات، يمكن تدريب وتقييم نظم التعرف على الكلام العربي المصممة خصيصًا للناطقين باللغة الامازيغية.

الكلمات الدالة : الشبكات العصبية المتكررة ، التعلم العميق، التعرف التلقائي على الكلام، مجموعة بيانات، التعرف على الصوت.

Abstract

Automatic Speech Recognition (ASR) is a highly active research field that has produced technologies adopted in various domains such as healthcare, public services, and human-machine interfaces, notably with the emergence of neural networks and deep learning techniques.

This final-year project underscores the significance of the evolution of Arabic speech recognition (ASR) systems and modern techniques in this domain. It also emphasizes the importance of providing high-quality Arabic datasets to support speech recognition applications on Arabic language-enabled smart devices.

The primary objective of this work is to gather a new dataset specifically designed for Arabic speech recognition by Amazigh speakers, thus enhancing the linguistic diversity of available data. This process includes data quality control, which contributes to improving the accuracy of Arabic speech recognition models.

With the aid of this new dataset, Arabic ASR systems tailored for Amazigh speakers can be trained and evaluated.

Keywords : Recurrent Neural Networks, Deep Learning, Automatic Speech Recognition, Dataset, Speech Recognition.

Résumé

La reconnaissance automatique de la parole RAP est un domaine de recherche très actif qui a produit des technologies adoptées dans de nombreux domaines comme la santé, les services publiques et les interfaces homme machine, notamment avec l'émergence des réseaux de neurones et des techniques d'apprentissage profond.

Ce projet de fin d'étude souligne l'importance de l'évolution des systèmes de Reconnaissance Automatique de la Parole (RAP) et des techniques modernes dans ce domaine. Il souligne également l'importance de fournir des ensembles de données arabes de haute qualité pour soutenir les applications de reconnaissance vocale sur les appareils intelligents en langue arabe.

L'objectif principal de ce travail est de collecter un nouvel ensemble de données spécialement conçu pour la reconnaissance de la parole arabe par des locuteurs amazighs, ce qui renforce la diversité linguistique des données disponibles. Ce processus inclut un contrôle de la qualité des données, ce qui contribue à améliorer la précision des modèles de reconnaissance de la parole arabe.

À l'aide de ce nouvel ensemble de données, les systèmes de RAP arabe conçus spécifiquement pour les locuteurs amazighs peuvent être formés et évalués.

Mots clés : Réseaux de neurones Récurrents, Apprentissage profond, Reconnaissance automatique de la parole, Jeu de données, Reconnaissance vocale.

Table des matières

Table des figures

Liste des tableaux

Liste des abréviations

AMS Arabe Moderne Standard

AC Arabe Classique

AD Arabe Dialactal

RAP Reconnaissance Automatique de la Parole

QASR Qcri Aljazeera Speech Resource

ESCWA Cross-lingual Code Switching Corpus

ADI17 Arabic Dialect Identification

AMCASC Algerian Modern Colloquial Arabic Speech Corpus

MGB Multi-Genre BroadCast

TSV tab seperated values

ASR Arabic speech recognition

APK Android Package

UML Unified Modeling Language

Introduction générale

La reconnaissance automatique de la parole (RAP) est une technologie en constante évolution qui trouve des applications dans divers domaines tels que la communication homme-machine, la transcription de discours, la traduction en temps réel et bien d'autres. L'RAP permet de convertir les signaux audio contenant des paroles en texte écrit, offrant ainsi une interface efficace entre les êtres humains et les systèmes informatiques.

L'un des défis majeurs de l'RAP réside dans la variabilité des accents et des dialectes, qui peuvent grandement influencer la précision et la performance globale des systèmes de reconnaissance. L'arabe est une langue qui est parlée par des millions de personnes dans le monde, mais il est important de reconnaître que la manière dont elle est prononcée peut varier considérablement en fonction de la région et de la communauté linguistique.

Lorsque des locuteurs tamazight, une population berbère distincte, parlent en arabe, leur prononciation et leur accent peuvent différer sensiblement de ceux des locuteurs arabophones. Cette différence soulève un défi majeur pour les systèmes de reconnaissance vocale, qui sont souvent optimisés pour les variétés standard de l'arabe.

Notre objectif pour ce projet est de créer un ensemble de données de reconnaissance de la parole spécifiquement adapté à l'arabe tel qu'il est parlé par des locuteurs tamazight, tout en développant une nouvelle application pour simplifier la collecte de données vocales. Cette initiative vise à pallier le manque d'ensembles de données vocales représentatifs pour cette variante linguistique de l'arabe, ce qui entrave la capacité des chercheurs à développer des systèmes de reconnaissance vocale automatique précis et performants pour cette communauté linguistique.

Notre mémoire est présenté en quatre chapitres décrit comme suit :

- **Le premier chapitre : Reconnaissance automatique de la parole**
  Dans ce chapitre, nous explorerons les principes de base de la reconnaissance automatique de la parole, en mettant l'accent sur les défis posés par la diversité linguistique et les accents dans le contexte de la parole arabe pour les locuteurs tamazight.

- **Le second chapitre : Architecture de Dataset**
  Dans ce chapitre, nous aborderons la collecte, la gestion et la structure des données de notre ensemble de données.

- **Le troisième chapitre : Collecte des données**
  Dans ce chapitre, se concentre sur la méthodologie de collecte des données, en détaillant le processus de développement de l'application utilisée pour collecter les échantillons vocaux.

- **Le dernier chapitre : Analyse et Statistiques de Dataset**
  Dans ce chapitre, nous procéderons à une analyse approfondie de notre ensemble de données de reconnaissance de la parole arabe pour les locuteurs Tamazigh. Nous explorerons les caractéristiques clés de l'ensemble de données, y compris le nombre d'échantillons vocaux, les informations sur l'âge et le sexe des locuteurs, ainsi que la structure du fichier tab seperated values (TSV) du dataset. Cette analyse est essentielle pour comprendre la composition et la distribution des données.

Enfin, nous terminons par une conclusion générale.

Chapitre 1

Reconnaissance automatique de la parole

Contents

1.1 Introduction

Les ordinateurs sont de plus en plus capables de comprendre et de répondre au langage humain. L'un des domaines de recherche les plus importants dans ce domaine est la reconnaissance de la parole qui est la capacité d'un ordinateur à reconnaître les mots prononcés et à les convertir en texte.

La reconnaissance de la parole est une tâche complexe en raison de la variabilité de la parole humaine. Le même mot peut être prononcé de nombreuses façons différentes en fonction de l'accent, du dialecte et de l'émotion de la personne qui parle.De plus la parole peut être perturbée par du bruit.

Malgré ces défis, la reconnaissance de la parole est une technologie en pleine croissance Elle est utilisée dans une variété d'applications notamment les assistants vocaux les systèmes de traduction automatique et les systèmes de dictée.

1.2 Présentation de La langue Arabe

L'arabe est une langue sémitique et l'une des plus anciennes langues du monde [23].Actuellement c'est l'une des langues les plus parlées dans le monde, avec environ 372 millions de locuteurs natifs [17].

En termes de nombre de locuteurs, l'arabe est classé cinquième dans le monde, après l'anglais le mandarin, l'hindi, l'espagnol et le français[20].

À travers des siècles, c'est la langue arabe qui a permis aux parlers natifs des pays arabes de se communiquer et de partager leurs cultures à travers le monde. Surtout, lors de l'avènement de l'Islam, elle est devenue la langue sacrée du Coran en exerçant des influences irrésistibles sur les peuples pour convertir à cette nouvelle religion. De plus, la langue arabe a recueilli des progrès étourdissants dans des domaines divers tels que la culture, la science grâce à la l'expansion territoriale de l'empire musulmane qui a fait de cette langue, une langue d'administration et de rédaction de manuscrits et de livres. Ainsi, il faut noter que le passage de l'arabe classique qui est ciblé en tant qu'une langue du Coran à l'arabe standard moderne (AMS) était fait à travers l'existence de la diversité au niveau des populations arabophones et ces cultures à travers des siècles. À son tour, le Arabe Moderne Standard (AMS) représentant la langue officielle utilisée dans les communautés et la presse a été influencé par des spécificités historiques et culturelles des populations appartenant au monde arabe en donnant naissance à l'arabe dialectal.
La langue arabe peut être classée en trois catégories principales[20] :

- **Arabe littéraire ancien ou Arabe Classique (AC) :** Cette appellation désigne la langue arabe dans sa forme la plus classique et la plus ancienne. Cela concerne essentiellement tout le patrimoine culturel médiéval parvenu par écrit : le texte coranique, la poésie ancienne, la philosophie, l'histoire, etc. La nature et l'origine de cette langue de la littérature antéislamique ont donné lieu à une évolution qui a abouti à l'apparition d'un arabe dit moderne ou standard .
- **Arabe moderne standard (AMS) :** D'une manière générale, l'arabe standard ou l'arabe contemporain est le résultat de l'interaction entre l'arabe classique et les dialectes. Dans le monde arabe, l'arabe moderne standard (AMS) est la langue des médias de la vie intellectuelle et de la littérature. En outre, il représente la forme de l'arabe universel enseignée dans les écoles du monde arabe et même utilisée à des conférences et des discussions formelles.
- **L'arabe dialectal Arabe Dialactal (AD) :** L'arabe dialectal est une forme extrêmement simplifiée de l'arabe classique et de l'arabe moderne. C'est la langue maternelle de chaque locuteur arabophone.

1.3 Variations linguistiques

La deuxième particularité linguistique de l'Algérie est que la langue parlée quotidiennement dans de nombreuses localités berbères (Kabyle, Chaoui, Mozabit, Touareg et Chleuh) est le tamazight, tandis que l'arabe est considéré comme la deuxième langue utilisée généralement à l'école ou pour communiquer avec des locuteurs non berbères. Dans de telles situations, il est connu que la langue maternelle de ces locuteurs influence leur prononciation de l'arabe en transférant les règles phonologiques de leur langue maternelle dans leur discours en arabe, créant ainsi des prononciations innovantes pour les sons arabes qui n'existent pas dans leur langue maternelle[10].

1.4 La reconnaissance automatique de la parole

La reconnaissance automatique de la parole, aussi connue sous le nom de reconnaissance vocale, est une branche de l'informatique qui se concentre sur l'analyse de la voix humaine fournie sous forme d'enregistrements numériques pour la transformer en texte, en respectant les règles de la langue étudiée. Les résultats obtenus peuvent ensuite être utilisés par la machine.
Solon [24], la Définition est la suivante : La reconnaissance vocale consiste à identifier et comprendre les déclarations ou commandes énoncées par la parole humaine, pour ensuite réagir en

conséquence. En tant qu'objet de recherche, la voix est au centre de ce processus, qui permet à la machine de reconnaître automatiquement le langage parlé humain grâce à un traitement du signal vocal et une reconnaissance des formes. »

1.4.1 Processus de la reconnaissance de la parole

Un Système de Reconnaissance Automatique de la Parole a pour objectif la transcription textuelle d'un signal de la parole. La figure 1.1. Montre les différentes étapes nécessaires à la reconnaissance d'un message m prononcé en entrée [15].

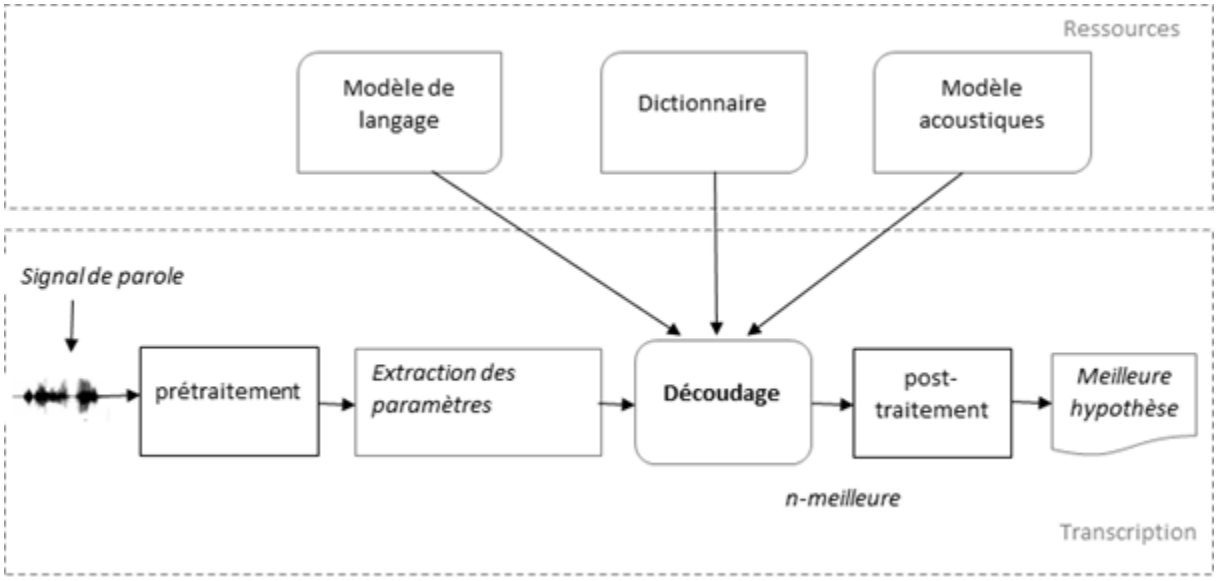

FIGURE 1.1 – Architecture d'un système de reconnaissance automatique de la parole [15]

- **Prétraitement :** vise à améliorer la qualité du signal vocal en le convertissant en numérique, en accentuant les signaux à hautes fréquences et en réduisant les basses fréquences. Elle inclut également la segmentation parole/non-parole pour éliminer les parties non pertinentes de l'enregistrement et éviter l'insertion de mots incorrects dans la reconnaissance de la parole[15].
- **Extraction de caractéristiques :** consistant à extraire des informations pertinentes des trames de parole sous forme de paramètres de caractéristiques ou de vecteurs. Les paramètres principaux incluent les coefficients de codage prédictif linéaire (LPC) et les coefficients de fréquence de Mel (MFCC). Ces paramètres sont privilégiés pour leur capacité à séparer la source sonore du filtre, leur modélisation analytique pratique, et leur efficacité prouvée dans les systèmes de reconnaissance vocale[15].

- **Modèle acoustique :** responsable de la majeure partie de la charge de calcul et des performances du système. Le modèle acoustique est développé pour détecter les phonèmes prononcés. Sa création implique l'utilisation d'enregistrements audio de la parole et de leurs scripts textuels, puis de les compiler en une représentation statistique des sons qui composent les mots[4].
- **Dictionnaire :** Pour fournir la prononciation de chaque mot dans une langue donnée, un lexique est développé. Diverses combinaisons de phonèmes sont définies à travers le modèle lexical afin de donner des mots valides pour la reconnaissance. Les réseaux neuronaux ont contribué au développement de cette tâche.
- **Modèle de langage :** générer des séquences de mots à partir de signaux vocaux. Ils utilisent des modèles linguistiques n-gramme qui prédisent la probabilité du mot suivant en fonction des mots précédents dans une phrase. Ces modèles sont complexes en raison du grand nombre de paramètres, et la rareté des données dans des domaines spécifiques complique leur construction[4].
- **Décodage :** Ceci est une combinaison des modèles précédents pour fournir la transcription textuelle la plus probable pour une déclaration vocale donnée.
- **Post-traitement :** produisent une liste n-best des meilleures hypothèses de séquences de mots, classées par score total. Le post-traitement consiste à organiser cette liste pour choisir les meilleures hypothèses parmi les n meilleures en utilisant des informations supplémentaires, telles qu'un modèle de langage d'ordre supérieur, afin de sélectionner les mots avec les scores les plus élevés comme sortie de reconnaissance[15].

1.5 Caractéristiques des systèmes de reconnaissance de la parole

Les caractéristiques d'un système de reconnaissance vocale sont les suivantes :

- **La précision :** est la mesure de la qualité d'un système de reconnaissance vocale. Elle est définie comme le pourcentage de mots correctement reconnus par le système. La précision est généralement mesurée par le taux d'erreur de mot , qui est le pourcentage de mots mal reconnus.

- **La vitesse :** est le temps nécessaire au système pour reconnaître un mot ou une phrase. La vitesse est importante pour les applications nécessitant une interaction rapide avec l'utilisateur, telles que les commandes vocales.

- **La flexibilité :** est la capacité du système à reconnaître une variété de voix et d'accents. La flexibilité est importante pour les applications qui doivent être utilisées par un large public.

- **La robustesse :** est la capacité du système à résister aux bruits et aux perturbations. La robustesse est importante pour les applications qui doivent être utilisées dans des environnements bruyants.

1.6 Application de la reconnaissance de la parole

La reconnaissance de la parole trouve un large éventail d'applications. Elle a été utilisée de manière efficace pour l'authentification des personnes dans les systèmes biométriques, la reconnaissance des émotions, le contrôle automatique dans les systèmes de voiture, l'automatisation résidentielle, la conversion de la parole en texte pour le sous-titrage, la robotique, la téléphonie mobile, la télématique, les jeux vidéo, l'évaluation de la prononciation, etc.

1.7 Dataset de la reconnaissance de la parole arabe

Reconnaissance automatique de la parole est un domaine très spécialisé qui nécessite des jeux de données spécifiques pour entraîner les modèles de reconnaissance vocale. Bien que de nombreux jeux de données pour RAP existent pour les langues les plus courantes, tels que l'anglais il y a moins de ressources disponibles pour les langues moins courantes. Pour l'arabe, il existe quelques jeux de données disponibles. Le tableau 1.1 décrit certains des jeux de données arabes pour la reconnaissance automatique de la parole.

Dataset	**Langue**	**Nombre de locuteurs**	**Sexe**	**Source**	**Durée**	**Référence**
Common Voice	AMS	1470	Homme et Femme	Enregistrement direct	154 heures	[7]
MGB-2	Multi-Dialect Broadcast News Arabic Speech	/	Homme et Femme	Programmes TV d'Aljazeera	1200 heures	[3]
MGB-3	Dialecte arabe (égyptien)	/	Homme et Femme	chaînes YouTube	16 heures	[2]
QASR	AMS	2000	Homme et Femme	Les données sont extraites de la chaîne Al Jazeera	2000 heures	[18]
ESCWA	The data includes intrasentential code alternation between Arabic and English	/	Homme et Femme	Recueilli lors des sessions de la Commission économique et sociale des Nations Unies pour l'Asie occidentale (CESAO) en 2019	2.8 heures	[6]
ADI17)	arabe dialectal	7205	Homme et Femme	Multimedia (YouTube)	3000 heures	[25]
AMCASC	Trois groupes de dialectes algériens	735	Homme et Femme	Conversations téléphoniques	Plus de 72 heures	[11]

TABLE 1.1 – Datasets de la reconnaissance de la parole arabe

1.8 Défis pour les systèmes de reconnaissance automatique de la parole arabe

Les systèmes RAP arabes sont confrontés à trois défis majeurs [1].

Premièrement, l'utilisation de textes non diacritisés comme matériel d'entraînement pose des problèmes tant pour la modélisation acoustique que pour la modélisation linguistique. L'entraînement de modèles acoustiques précis pour les voyelles arabes sans diacritiques est difficile et les mots non diacritisés peuvent avoir plusieurs sens, ce qui rend la modélisation linguistique moins prédictive.

Le deuxième défi réside dans l'existence de divers dialectes arabes qui sont principalement parlés et qui manquent de ressources écrites formelles. Le manque de données d'entraînement pour l'arabe conversationnel, en particulier l'arabe dialectal, est un obstacle majeur.

Le troisième défi est la complexité morphologique de l'arabe, qui pose des problèmes pour la reconnaissance automatique de la parole et la modélisation linguistique. La richesse de la morphologie arabe entraîne un grand nombre de formes de mots différentes, des taux élevés de mots hors vocabulaire et des espaces de recherche plus importants lors du décodage, ce qui ralentit le processus de reconnaissance.

1.9 Conclusion

La reconnaissance de la parole en arabe est une tâche complexe en raison des nombreuses caractéristiques linguistiques et phonétiques de la langue. Cependant, les avancées récentes dans les techniques d'apprentissage automatique et de traitement du langage naturel offrent de l'espoir pour surmonter ces défis. Grâce à la recherche et au développement continus, il est possible de réaliser des progrès significatifs dans la technologie de reconnaissance de la parole en arabe.

Chapitre 2

Architecture de Dataset

Contents

2.1 Introduction

Dans le domaine de l'apprentissage profond et de l'intelligence artificielle, les données sont cruciales pour le progrès et l'innovation. Les données sont le fondement essentiel qui soutient l'évolution des systèmes. L'amélioration requiert une ample diversité de données car elles alimentent l'apprentissage et le développement. Nous visons à rassembler une grande quantité de données auprès des locuteurs tamazight qui parlent en arabe, de divers horizons pour améliorer la reconnaissance vocale arabe via une application Android. Notre objectif est d'atteindre une précision supérieure et une efficacité optimale marquant une étape majeure vers une avancée significative dans ce domaine important.

2.2 Collecte de Données

2.2.1 Nature des données

Les données traitées sont des échantillons de parole humaine. La parole est une forme de communication complexe qui se manifeste sous forme de sons présentant une variété étendue de fréquences, différant d'une personne à l'autre. Elle est sujette à des changements basés sur des facteurs tels que l'état émotionnel de l'individu, qu'il soit fatigué, triste, ou s'exprime à voix basse ou forte. Les particularités linguistiques, telles que les dialectes et les accents, ajoutent un niveau supplémentaire de complexité, car ils peuvent altérer la prononciation et la sonorité des mots. En rassemblant une multitude conséquente de données vocales, il est envisageable d'enseigner à la machine à comprendre un large éventail de locuteurs

2.2.2 Sélection des locuteurs

La collecte de données vocales pertinentes nécessite une sélection soignée des locuteurs tamazight (Kabyle, Chaoui, Mozabit, Touareg, et Chleuh). Pour garantir une représentation variée de la population cible, il est important de choisir des locuteurs de différents âges, sexes, origines géographiques et niveaux d'éducation. Cette diversité permettra de créer un ensemble de données robuste, capable de traiter les variations naturelles de la parole et des accents.

2.2.3 Collecte des données

Pour créer un ensemble de données robuste pour la reconnaissance de la parole arabe chez les locuteurs Tamazight, nous avons développé une application Android conviviale et intuitive. L'objectif principal de cette application est de collecter des données essentielles pour améliorer

la précision du modèle de reconnaissance vocale. Les informations collectées comprennent l'âge l'adresse, la langue maternelle et le sexe de chaque utilisateur, des détails qui contribueront à une meilleure compréhension des caractéristiques linguistiques et des besoins spécifiques des locuteurs Tamazight.

Le processus de collecte de données implique également la lecture de phrases en arabe par les utilisateurs, qui sont des locuteurs Tamazight. Ces phrases ont été soigneusement sélectionnées pour couvrir une variété de contextes et de situations, permettant ainsi une couverture complète des diverses nuances de la langue arabe dans le contexte des locuteurs Tamazight.

Une attention particulière est accordée à la qualité audio des données collectées. Les utilisateurs sont encouragés à enregistrer les phrases dans un environnement calme et sans distraction afin de garantir la clarté et la précision des échantillons vocaux. Les données audio obtenues seront stockées au format WAV (Waveform Audio File Format), avec une résolution de 16 bits et une fréquence d'échantillonnage de 16 000 Hz. Ces paramètres sont optimaux pour la collecte de données vocales et permettront une analyse précise et détaillée.

2.2.4 Stockage des données

Nous avons opté pour une infrastructure de stockage robuste en utilisant la plateforme Firebase [1] de Google.

- **Firebase Cloud Firestore :** Nous avons choisi d'utiliser Firebase Cloud Firestore, une base de données NoSQL en temps réel, pour gérer les informations utilisateur telles que l'âge, l'adresse, et la langue maternelle. Chaque utilisateur est représenté par un document unique dans la collection "Utilisateurs", avec des champs correspondant aux données saisies. Cette approche garantit une structuration ordonnée et une facilité de gestion des informations.

- **Firebase Cloud Storage :** Pour stocker les enregistrements audio associés aux utilisateurs, nous faisons appel à Firebase Cloud Storage, un service de stockage cloud. Chaque enregistrement audio au mp3 est associé à l'identifiant de l'utilisateur, permettant ainsi une référence précise. Cela garantit une organisation efficace et une récupération aisée des données audio.

1. https ://firebase.google.com/

- **Sécurité et Confidentialité :** Nous prenons des mesures rigoureuses pour assurer la sécurité des données personnelles et audio. Les règles de sécurité de Firebase sont configurées pour permettre un accès restreint et spécifique aux utilisateurs autorisés. De plus, les données sont stockées dans un environnement sécurisé, conformément aux réglementations en matière de protection des données.

2.3 Contraintes juridiques relatives à la collecte de données personnelles

Dans le cadre de la collecte de données personnelles, il est essentiel de comprendre et de respecter les contraintes juridiques qui s'appliquent. Ces réglementations varient en fonction du lieu géographique et peuvent avoir un impact significatif sur la manière dont les données doivent être collectées et gérées, Parmi les principes à respecter :[14]

- **Consentement de la personne concernée :** La collecte de données personnelles ne peut être réalisée qu'après avoir obtenu le consentement libre, éclairé et sans ambiguïté de la personne concernée.

- **Objectif du traitement :** Les données personnelles ne peuvent être collectées que dans le but spécifique et légitime pour lequel elles ont été recueillies. Les personnes concernées doivent être informées de l'objectif du traitement au moment de la collecte des données.

- **Proportionnalité :** Les données personnelles collectées doivent être pertinentes et ne pas excéder ce qui est nécessaire par rapport à l'objectif du traitement.

- **Limitation de la durée de conservation :** Les données personnelles ne doivent être conservées que pendant la période nécessaire à l'atteinte de l'objectif du traitement.

- **Sécurité des données :** Les données personnelles doivent être traitées de manière à garantir leur sécurité et leur confidentialité.

2.4 Sélection des phrases de l'ensemble de données

Les phrases sont présentées dans un ordre aléatoire à travers une interface de présentation visuelle. Dans cette phase de collecte de phrases en arabe, nous avons utilisé une approche as-

sistée par ChatGPT afin de rassembler 100 phrases en arabe. ChatGPT, un modèle de langage développé par OpenAI, est capable de générer du texte dans différentes langues, dont l'arabe.

Pour obtenir les phrases, nous avons formulé une requête précise à ChatGPT, en spécifiant notre objectif d'obtenir 100 phrases en arabe. Grâce à son système d'intelligence artificielle et à son apprentissage basé sur de vastes corpus de texte, ChatGPT a généré les phrases en réponse à notre requête (voir la figure 2.1) :

FIGURE 2.1 – Méthode de collecte de phrases en arabe utilisant ChatGPT.

2.5 Structure des données

Les données audio sont stockées dans une collection, pour chaque phrase il existe un dossier qui contient toutes les URL des fichiers audio de cette phrase, avec chaque fichier individuel. Il est également lié à l'identifiant de l'utilisateur.

2.5.1 Champs de données

- **Identifiant_utilisateur (User_id) :** Il s'agit d'un champ de type chaîne de caractères qui est unique pour chaque utilisateur, et qui permet de suivre les enregistrements d'un utilisateur spécifique.

- **Âge_utilisateur (User_age) :** Champ de type chaîne de caractères représentant l'âge de l'utilisateur.

- **Genre_utilisateur (User_gender) :** Champ de type chaîne de caractères représentant le genre de l'utilisateur. Il peut être soit "Masculin" soit "Féminin".

- **Adresse :** Champ de type chaîne de caractères représentant l'adresse de l'utilisateur.

- **Langue_maternelle (Native_language) :** Champ de type chaîne de caractères représentant la langue maternelle de l'utilisateur. Il peut prendre l'une des valeurs suivantes : (Kabyle, Chaoui, Mozabit, Touareg et Chleuh).

- **URL_audio (Audio_url) :** Champ de type chaîne de caractères qui est unique pour chaque fichier audio et qui représente le lien URL exact de ce fichier dans la base de données.

- **Transcription :** Champ de type chaîne de caractères contenant la transcription du phrase enregistré.

2.5.2 Échantillons Vocaux Enregistrés

Chaque instance inclut le chemin d'accès au fichier audio (`audio_url`) ainsi que sa transcription correspondante. De plus, les attributs supplémentaires, tels que l'adresse de l'utilisateur, la langue maternelle de l'utilisateur, l'âge de l'utilisateur (*User age*), l'identifiant de l'utilisateur (*User id*), le genre (*User gender*) (voir la figure 2.1) :

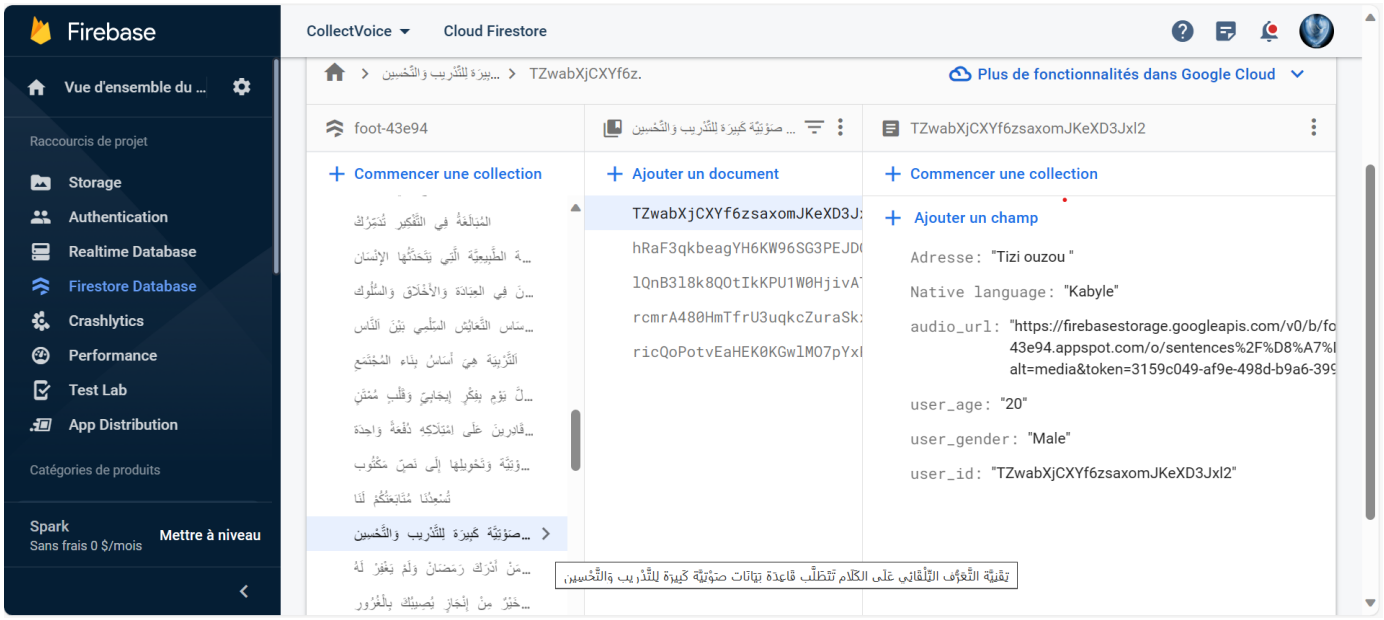

FIGURE 2.2 – Structure des données

2.6 Conclusion

Le processus de collecte et de préparation des données est une étape importante dans le développement d'un système de reconnaissance vocale arabe. En collectant et en pré-traitant soigneusement les données, nous pouvons améliorer la précision et l'efficacité du système.

Chapitre 3

Collecte des données

Contents

3.1 Introduction

En vue de la collecte des données, en utilisant la technologie Flutter, conjuguée à la plateforme Firebase de Google, pour privilégier la conception d'une application Android. Cette application a été rendue accessible sous la forme d'un fichier Android Package (APK) téléchargeable via MediaFire. Son interface utilisateur intuitive a été pensée de manière à faciliter l'expérience des utilisateurs. De plus, nous tablons sur la collaboration volontaire des locuteurs natifs de tamazight, qui contribueront à réunir les enregistrements audios. Pour vise à impacter de manière substantielle l'amélioration de la reconnaissance vocale en langue arabe.

3.2 Description de l'application

L'objectif principal de **CollectVoice** est de rassembler des données vocales arabes grâce à la participation volontaire des utilisateurs. Dès le lancement de l'application, les utilisateurs sont accueillis par une interface conviviale qui les guide tout au long du processus de collecte de données. La première étape consiste à fournir des informations de base, telles que l'âge, l'adresse la langue maternelle et le genre.

Une fois inscrits,les utilisateurs commencent à enregistrer des échantillons de parole en prononçant des phrases spécifiques fournis par **CollectVoice**. Avec une liste soigneusement élaborée de 100 phrases, l'application garantit une couverture complète de la langue arabe, capturant une large gamme de variations phonétiques, de dialectes et d'accents. Chaque phrase est présentée à l'écran, et les utilisateurs sont invités à enregistrer leur prononciation.

CollectVoice permet aux utilisateurs de revoir et de confirmer leurs enregistrements. Après la prononciation de chaque phrase, les utilisateurs ont le choix de confirmer ou de supprimer leur enregistrement. Cette fonctionnalité garantit que seules des données vocales précises et de haute qualité sont collectées.

Cet ensemble de données constitue une ressource précieuse pour les chercheurs, les linguistes et les développeurs travaillant sur la reconnaissance vocale, les assistants vocaux et les applications d'apprentissage automatique en arabe.

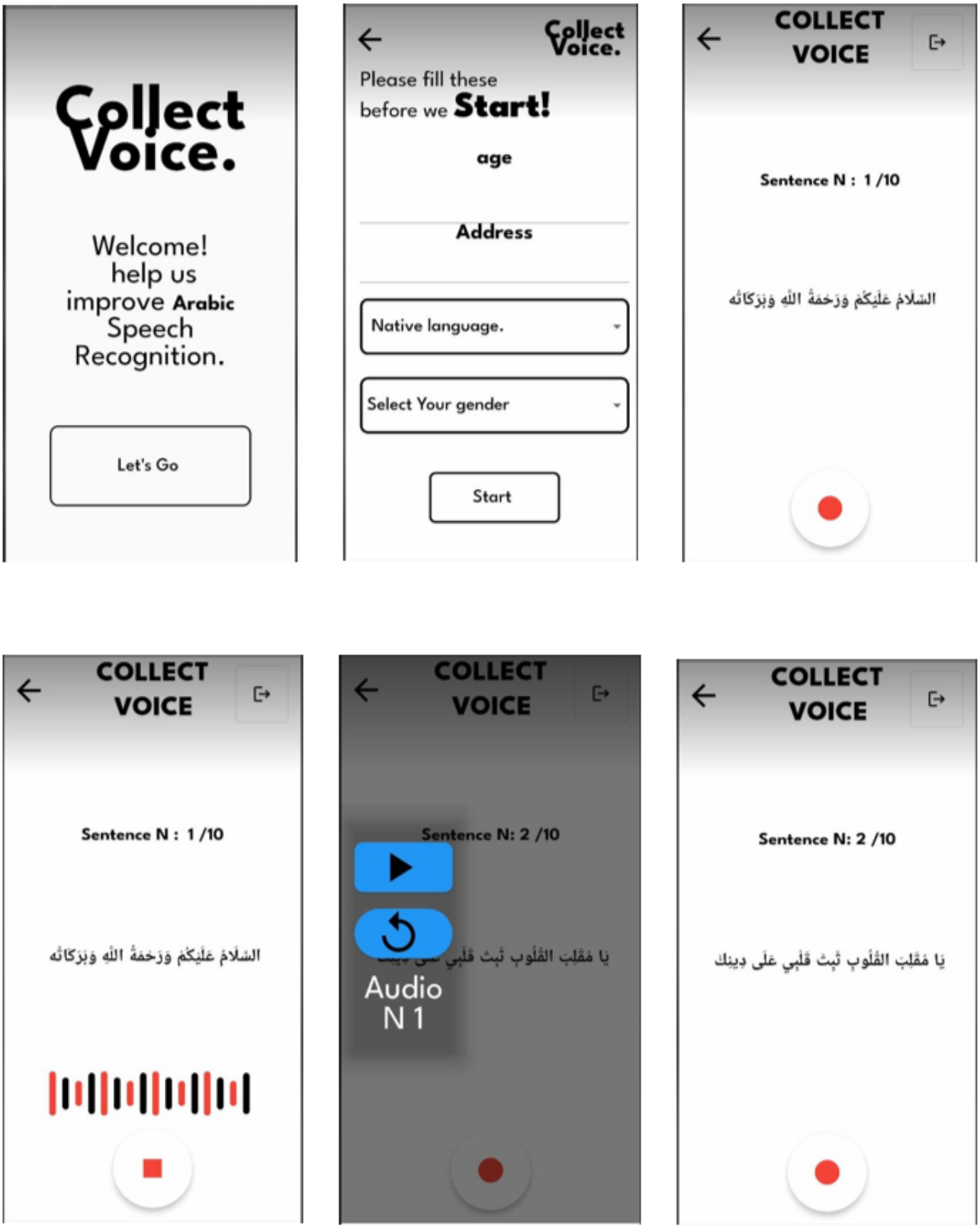

FIGURE 3.1 – Conception générale de **CollectVoice**

3.3 Structure de l'application

La modélisation d'une application représente une démarche très importante dans le développement de n'importe quel projet logiciel. Cette démarche permet de bien définir l'aspect fonctionnel du système. Nous présentons dans ce chapitre la modélisation de notre solution CollectVoice utilisant le langage UML qui s'est imposé comme une norme standard dans la conception orientée objets.

Unified Modeling Language (UML) permet de construire plusieurs modèles d'un système : certains montrent le système du point de vue des utilisateurs, d'autres montrent sa structure interne, d'autres encore en donnent une vision globale ou détaillée. Les modèles se complètent et peuvent être assemblés. Ils sont élaborés tout au long du cycle de vie du développement d'un système.

3.3.1 Diagramme général de cas d'utilisation

Un cas d'utilisation est une manière spécifique d'utiliser un système. Les acteurs sont à l'extérieur du système ; ils modélisent tout ce qui interagit avec lui. Un cas d'utilisation réalise un service de bout en bout, avec un déclenchement, un déroulement et une fin, pour l'acteur qui l'initie.[19]

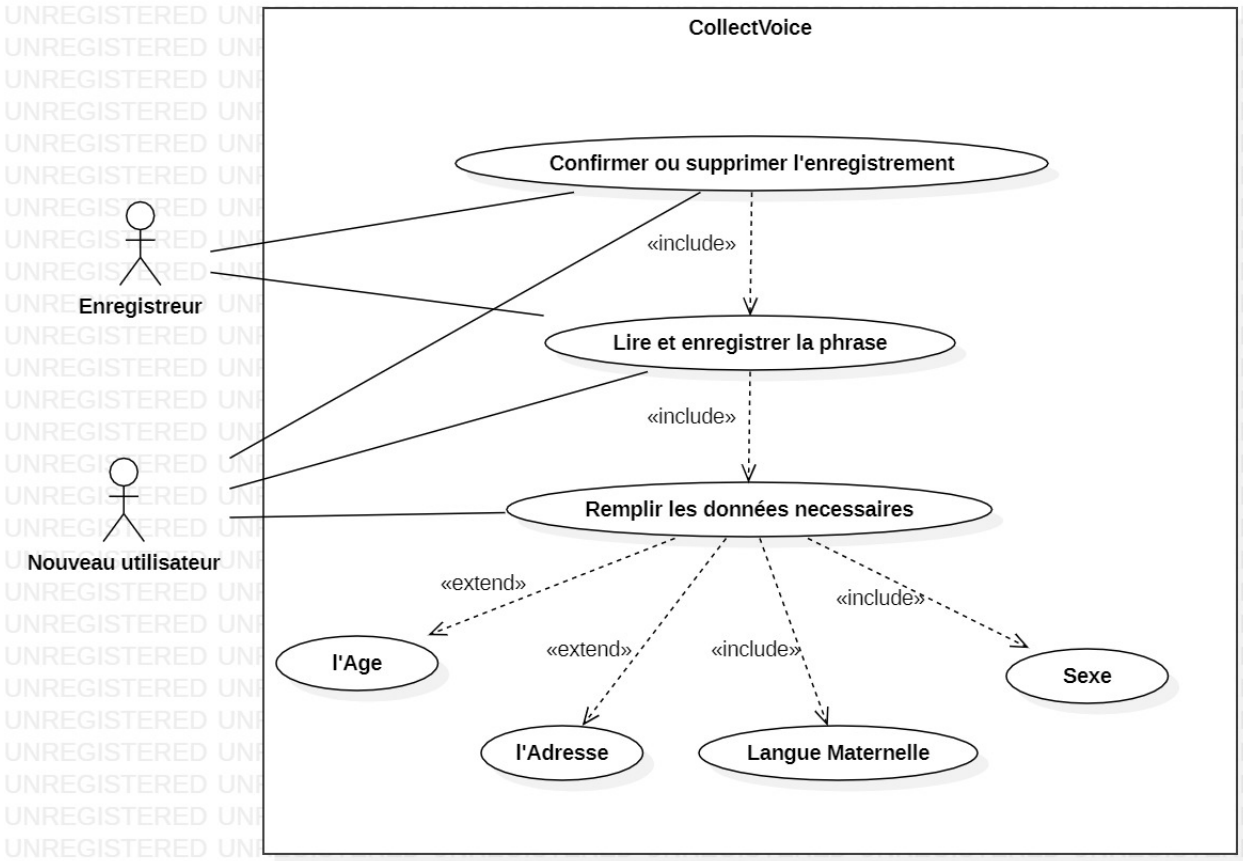

FIGURE 3.2 – Diagramme de cas d'utilisation

3.3.2 Diagramme de classe

Le diagramme de classe en montre la structure interne. Il permet de fournir une représentation abstraite des objets du système qui vont interagir ensemble pour réaliser les cas d'utilisation. Il s'agit d'une vue statique car on ne tient pas compte du facteur temporel dans le Comportement du système.[16]

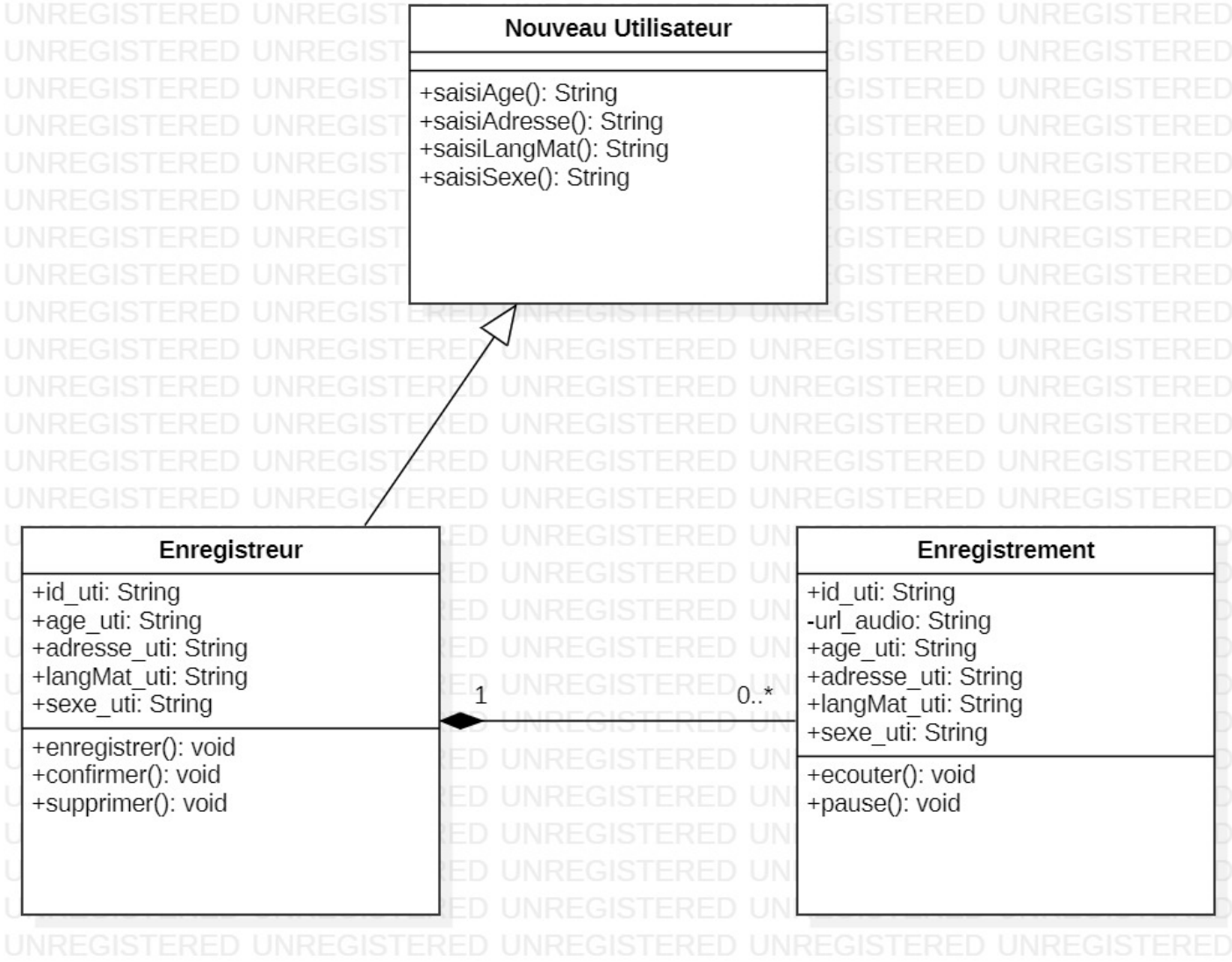

FIGURE 3.3 – Diagramme de classe

3.3.3 Diagramme de séquence

Les diagrammes de séquence montrent la séquence des interactions entre objets selon un point de vue temporel (chronologique). Ce diagramme permet de représenter les scénarios d'un cas d'utilisation, il permet de mieux visualiser la séquence des messages par une lecture de bas en haut. Un scénario est une instance d'un cas d'utilisation.[5]

- **Cas d'utilisation «Confirmer ou Supprimer l'enregistrement»**

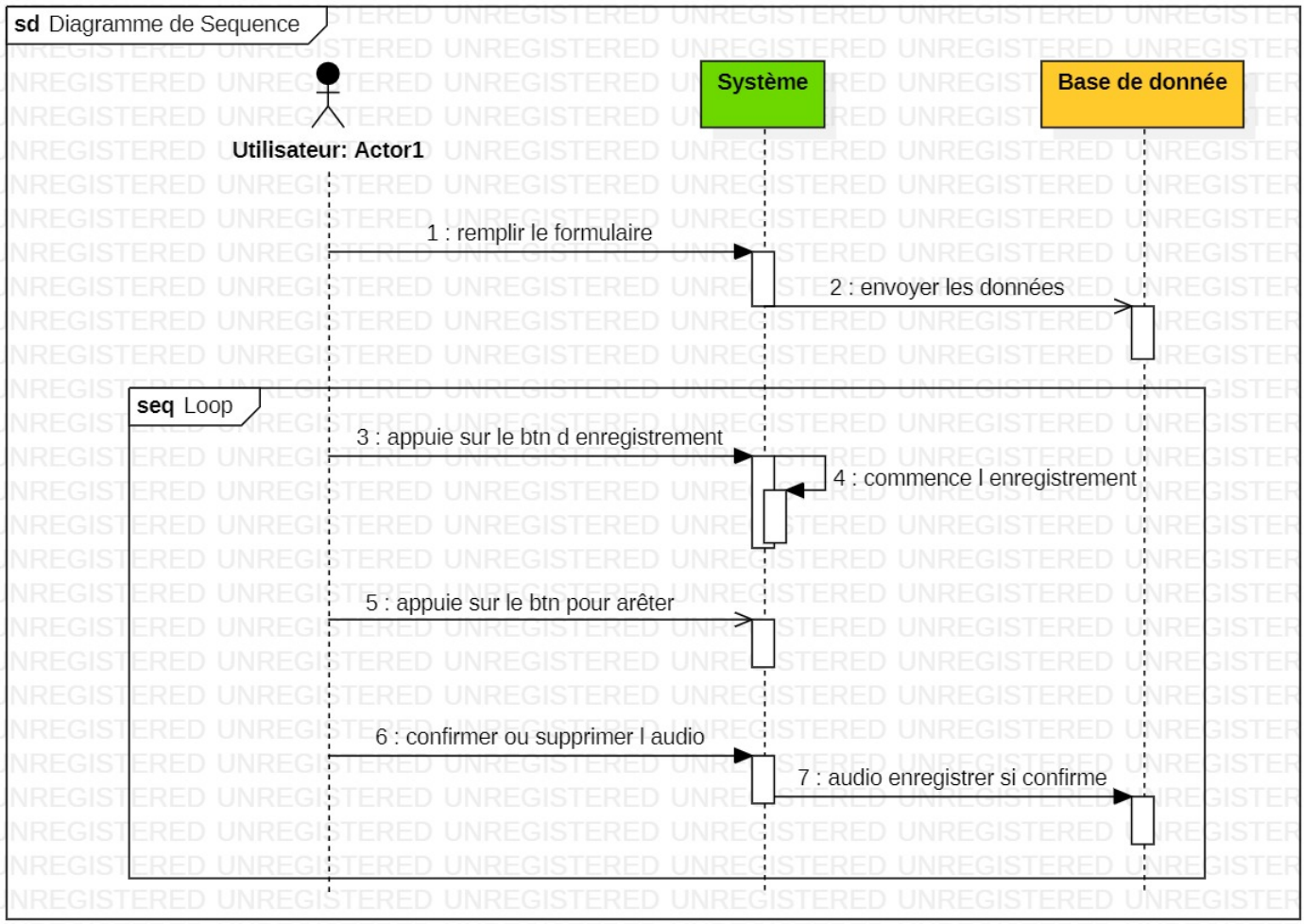

FIGURE 3.4 – Diagramme de séquence

3.4 Langage de programmation :

- **Java :** est un langage de programmation orienté objet.Java a été officiellement présentée le 23 mai 1995 au SunWorld. La particularité et l'intérêt de Java réside dans sa portabilité entre les différents systèmes d'exploitation tels que Unix, Windows, ou MacOS.Un programme développé en langage Java, peut ainsi s'exécuter sur toutes les plateformes grâce à ses frameworks associés visant à garantir cette portabilité[21].

- **Kotlin :** est un projet Open Source disponible sans frais sous la licence Apache 2.0.[8], Il réduit le temps passé à écrire et à maintenir le même code pour différentes plates-formes tout en conservant la flexibilité et les avantages de la programmation native. Les applications Kotlin fonctionneront sur différents systèmes d'exploitation, tels que iOS, Android macOS, Windows, Linux, watchOS et autres.Kotlin bénéficie d'un grand soutien et de nombreux contributeurs dans sa communauté mondiale en pleine croissance.[9]

- **Dart :** est un langage open source développé dans Google dans le but de permettre aux développeurs d'utiliser un langage orienté objet avec analyse de type statique. Depuis la première version stable en 2011, Dart a beaucoup changé, à la fois dans le langage lui-même et dans ses objectifs principaux. Avec la version 2.0, le système de type de Dart est passé de l'optionnel au statique, et depuis son arrivée, Flutter est devenu la cible principale du langage.[12]

3.5 Frameworks

- **Flutter :** est un framework Dart open source de Google permettant de créer des applications multiplateformes compilées nativement à partir d'une base de code unique.Le code Flutter se compile en code machine ARM ou Intel ainsi qu'en JavaScript, pour des performances rapides sur n'importe quel appareil. [13]

3.6 Environment

- **Visual Studio Code :** est un environnement de développement intégré open source développé par Microsot. Il reconnaît beaucoup de langages de programmation, supporte plusieurs technologies du web et offre la possibilité d'installer des extensions pour avoir plus de fonctionnalités et interagir avec des technologie comme Docker, Git, etc.[31]

3.7 Prétraitement des données

Le prétraitement du signal de parole est la première et la plus importante étape du processus de reconnaissance automatique de la parole. Il englobe un ensemble de techniques et de méthodes visant à préparer les données avant de les utiliser pour l'entraînement et l'évaluation de modèles de reconnaissance de la parole.Parmi les techniques couramment utilisées, on trouve :

- **Validation des données :** est le processus visant à garantir que les données vocales sont exactes et complètes. Cela implique de vérifier les transcriptions des enregistrements audio pour s'assurer qu'elles sont correctes et de supprimer tout enregistrement corrompu ou non pertinent.

- **Augmentation des données de parole :** Nous reconnaissions l'importance de la diversité des conditions réelles dans lesquelles la reconnaissance de la parole sera effectuée. Pour simuler ces variations, nous appliquerons des transformations à nos enregistrements de parole existants. Cela comprendra des techniques telles que la modification de la hauteur du signal vocal pour représenter différentes caractéristiques vocales et l'ajout de bruit de fond ou de sons environnementaux pour reproduire des conditions acoustiques variées. Cette augmentation de données renforcera la capacité de modèle à traiter des environnements bruyants et divers.
- **Normalisation :** La normalisation des caractéristiques consiste à les convertir dans une plage standard, généralement comprise entre 0 et 1. Cela permet de donner à toutes les caractéristiques un poids égal lors du processus d'apprentissage, ce qui améliore la précision du modèle global.

3.8 Conclusion

Dans ce chapitre, nous avons présenté CollectVoice, une application qui permet de collecter des données vocales arabes de locuteurs natifs de tamazight. L'application dispose d'une interface utilisateur intuitive qui permet aux utilisateurs de s'enregistrer en prononçant des phrases spécifiques. Nous avons également présenté les diagrammes UML que nous avons utilisés pour concevoir l'application.

Chapitre 4

Analyse et Statistiques de Dataset

Contents

4.1 Introduction

Dans ce chapitre, nous procéderons à une analyse approfondie de notre dataset de reconnaissance de la parole arabe pour les locuteurs Tamazigh. Nous explorerons les caractéristiques clés de l'ensemble de données, y compris le nombre d'audios, les informations sur l'âge et le sexe des locuteurs, ainsi que la structure du fichier TSV du dataset. Cette analyse est essentielle pour comprendre la composition et la distribution des données, ce qui peut avoir un impact significatif sur la conception et la performance de nos systèmes de reconnaissance automatique de la parole.

4.2 Description du Dataset

4.2.1 Nombre d'échantillons vocaux

Le dataset que nous avons rassemblé comprend un total de **400** enregistrements audio. Chaque enregistrement audio est une instance d'un discours parlé par un locuteur Tamazigh. Cette quantité de données représente la richesse de notre dataset en termes de données vocales disponibles pour la formation, la validation et les tests de système de reconnaissance d e la parole.

La taille du dataset est un facteur important, car elle peut influencer la capacité de système à généraliser et à traiter efficacement une variété de locuteurs et de conditions d'enregistrement. Elle détermine également la diversité linguistique et sociodémographique de notre échantillon ce qui est particulièrement important pour assurer la représentativité de notre modèle de reconnaissance de la parole.
Le tableau 4.1 contient toutes les informations concernant le dataset.

N°	Désignation	Valeur
1	Taille compressée	65 Ko
2	Taille décompressée	135 Mo
3	Total d'heures	1.18
4	Nombre de voix	400
5	Format audio	wav
6	Nombre de locuteurs	70

TABLE 4.1 – Détails du dataset

4.2.2 Répartition par âge

Une analyse détaillée de la répartition par âge des locuteurs dans notre dataset de reconnaissance de la parole arabe pour les locuteurs Tamazigh révèle une diversité significative au sein de notre échantillon. Cette diversité est essentielle pour garantir que le modèle de reconnaissance de la parole soit robuste et capable de traiter un large éventail de groupes d'âge.

Le graphique suivant illustre cette répartition par âge :

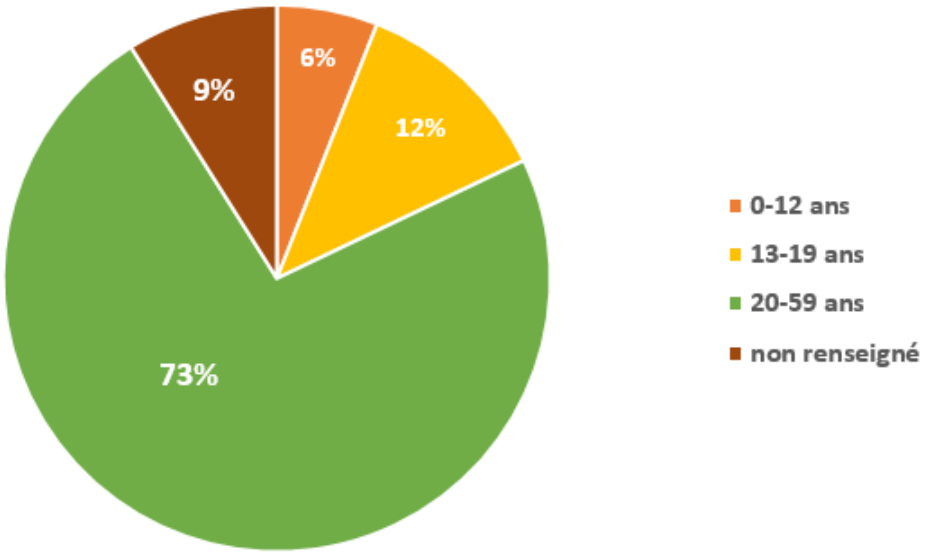

FIGURE 4.1 – Répartition par âge

La répartition par âge des locuteurs dans un dataset de reconnaissance de la parole est un facteur important à prendre en compte pour garantir la robustesse et l'équité du système. En effet, les locuteurs de différents âges utilisent le langage de manière différente, et une répartition équilibrée permet de capturer cette diversité.

4.2.3 Impact de la répartition par âge sur le système de reconnaissance de la parole

Un système de reconnaissance de la parole qui est formé sur un dataset qui ne comprend pas une répartition équilibrée des locuteurs de différents âges peut être biaisé vers un groupe d'âge

particulier. Cela peut entraîner des résultats inexacts ou discriminatoires.

Par exemple, un système de reconnaissance de la parole qui est formé sur un dataset qui comprend principalement des locuteurs adultes peut avoir du mal à reconnaître le langage utilisé par les enfants. Cela peut conduire à des erreurs d'identification ou à des résultats inappropriés.

À l'inverse, un système de reconnaissance de la parole qui est formé sur un dataset qui comprend principalement des locuteurs enfants peut avoir du mal à reconnaître le langage utilisé par les adultes. Cela peut également conduire à des erreurs d'identification ou à des résultats inappropriés.

4.2.4 Répartition par sexe

Une analyse de la répartition par sexe des locuteurs montre une répartition équilibrée. Les hommes représentent 60% des échantillons vocaux, tandis que les femmes représentent 40%.

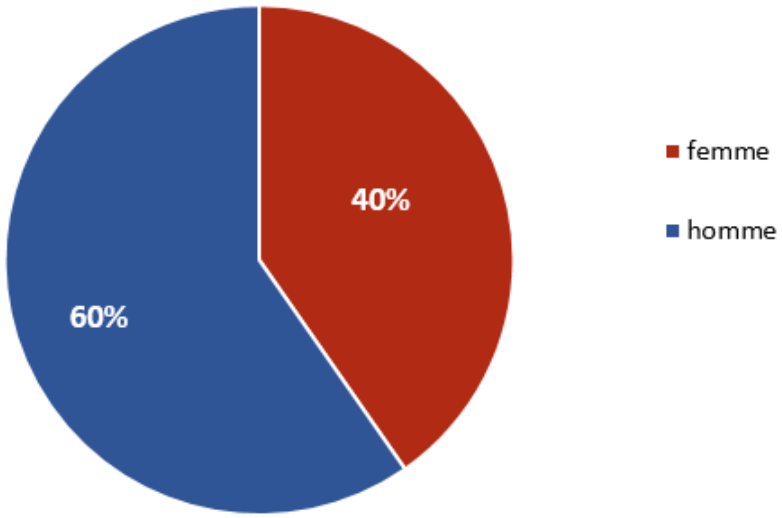

FIGURE 4.2 – Répartition par sexe

Cette équité entre les sexes dans le dataset est essentielle pour garantir une représentativité adéquate. En effet, les hommes et les femmes utilisent le langage de manière différente, et une répartition équilibrée permet de capturer cette diversité.

En outre, une répartition équilibrée entre les sexes est importante pour garantir que le modèle de langage ne soit pas biaisé. Si la répartition est déséquilibrée, le modèle peut être biaisé vers un sexe ou un autre, ce qui peut entraîner des résultats inexacts ou discriminatoires.

4.2.5 Structuration de l'Ensemble de Données

ous avons opté pour l'utilisation du format TSV pour organiser notre ensemble de données. Dans ce format, Chaque ligne d'un fichier TSV contient des informations essentielles sur les locuteurs, notamment leur adresse, âge, sexe, langue maternelle, ainsi que la phrase transcrite et le chemin relatif vers le fichier audio correspondant. Cette structuration détaillée des données nous a permis de créer un ensemble de données complet et bien organisé, facilitant ainsi l'analyse et l'entraînement de modèle de reconnaissance vocale. . La figure 4.3 montre le format des fichier TSV inclus dans le dataset.

__path__	Adresse	`Native language`	audio_url	user_age	user_gender	user_id	sentence
Ø§ÙØ¯Ù'Ø¯ÙŽØ£Ù' ÙfÙÙ„Ù'Ù'ÙŽ ÙŠÙŽÙ^Ù'Ù		Kabyle	https://firebasestorage.g		Male	ZHS3GDUCYshapcnXJTE	إِبْدَأْ كُلّ يَوْم بِفِكْرٍ إِيجَابِيٍّ وَقَلْبٍ مُمْتَنّ
Ø§ÙØ¯Ù'Ø¯ÙŽØ£Ù' ÙfÙÙ„Ù'Ù'ÙŽ ÙŠÙŽÙ^Ù'Ù		Kabyle	https://firebasestorage.g		Female	jImKFrYnFHWkt9QpyfE3	إِبْدَأْ كُلّ يَوْم بِفِكْرٍ إِيجَابِيٍّ وَقَلْبٍ مُمْتَنّ
Ø§ÙØ¯Ù'Ø¯ÙŽØ£	bouira	Kabyle	https://firebase	27	Female	lQnB3l8k8QOtlkKPU1W(	إِبْدَأْ كُلّ يَوْم بِفِكْرٍ إِيجَابِيٍّ وَقَلْبٍ مُمْتَنّ
ØªÙŽØªÙŽØ·ÙŽÙ	Ath-Ouacif Tizi Ouzou	Kabyle	https://firebase	26	Female	3dBOx1vKlqhzw8rHs0R(	أَتَطَلَّعُ إِلَى مُسْتَقْبَلٍ مُشْرِقٍ وَمَلِيءٍ بِالْفُرَصِ وَالنَّجَدَيَات
ØªÙŽØªÙŽØ·ÙŽÙ	Tamanrasset	Kabyle	https://firebase	39	Male	hRaF3qkbeagYH6KW96	أَتَطَلَّعُ إِلَى مُسْتَقْبَلٍ مُشْرِقٍ وَمَلِيءٍ بِالْفُرَصِ وَالنَّجَدَيَات
ØªÙŽØªÙŽØ·ÙŽÙ	taghzout Bouira	Kabyle	https://firebase	33	Male	i3dEwEnbxraNJyERlnstu	أَتَطَلَّعُ إِلَى مُسْتَقْبَلٍ مُشْرِقٍ وَمَلِيءٍ بِالْفُرَصِ وَالنَّجَدَيَات
ØªÙŽØªÙŽÙ...ÙŽÙ	bouira	Kabyle	https://firebase	32	Female	lQnB3l8k8QOtlkKPU1W(	أَتَمَنَّى لَكَ يَوْمًا سَعِيدًا وَمَلِيئًا بِالنَّجَاحِ وَالسَّعَادَة
ØªÙŽØ-Ù'ØªÙŽØ	haizr	Kabyle	https://firebase	11	Male	biZF2xWP2YhnMIkgsscu	أَحْتَاجُ إِلَى مَزِيدٍ مِن الوَقْت
ØªÙŽØ-Ù'ØªÙŽØ	bouira	Kabyle	https://firebase	32	Female	lQnB3l8k8QOtlkKPU1W(	أَحْتَاجُ إِلَى مَزِيدٍ مِن الوَقْت

FIGURE 4.3 – Ficheier tsv de dataset

Les fichiers audio, constituants essentiels de notre ensemble de données, ont été organisés et stockés dans un sous-répertoire spécifique que nous avons nommé **"clips"**. Cette stratégie de stockage a simplifié considérablement la gestion de nos données en les regroupant de manière cohérente et en facilitant leur accès.

4.3 Traitement des données audio

Les données audio sont omniprésentes dans notre vie quotidienne, que ce soit sous forme de musique, de discours ou de sons environnementaux. Le traitement de ces données peut permettre de réaliser une grande variété de tâches, telles que la reconnaissance vocale, la classification d'instruments de musique, l'analyse des émotions dans la voix, et bien plus encore.

Python [1] offre des outils puissants pour travailler avec ces données de manière efficace.

4.3.1 Bibliothèques Python pour le Traitement Audio

Avant d'aller plus loin dans notre exploration des sons et de leur analyse, nous devons préparer notre boîte à outils. Pour cela, nous utilisons plusieurs bibliothèques Python qui nous aident à comprendre et à manipuler les sons.

- **pandas :** Pandas est la bibliothèque python standard pour travailler avec des dataframes. Contrairement à R, cela ne fait pas partie de la base python et doit être importé séparément. Il est généralement importé sous forme de : pd.[29]

- **NumPy et SciPy :** sont des modules complémentaires open-source pour Python qui fournissent des routines mathématiques et numériques courantes sous forme de fonctions précompilées et rapides. Ce sont des packages très matures qui offrent des fonctionnalités numériques équivalentes, voire supérieures, à celles des logiciels commerciaux tels que MatLab. Le package NumPy (Numeric Python) fournit des routines de base pour manipuler de grandes matrices et tableaux de données numériques. Le package SciPy (Scientific Python) étend les fonctionnalités de NumPy avec une collection substantielle d'algorithmes utiles tels que la minimisation, la transformation de Fourier, la régression et d'autres techniques mathématiques appliquées.[28]

- **Matplotlib :** est probablement le paquet Python le plus utilisé pour les graphiques 2D. Il fournit à la fois un moyen très rapide de visualiser les données de Python et des chiffres de qualité publication dans de nombreux formats.[30]

- **Librosa 0.10.1 :** est un paquet Python pour la musique et l'analyse audio. Il fournit le bâtiment Blocs nécessaires à la création de systèmes de récupération d'informations musicales.[26]

  - **librosa.load :** [27] Chargez un fichier audio en tant que série chronologique en virgule flottante.

1. https ://www.python.org/

L'audio sera automatiquement rééchantillonné à la fréquence donnée (par défaut).sr=22050

Pour conserver la fréquence d'échantillonnage native du fichier, utilisez .sr=None

- **Seaborn 0.12.2 :** Seaborn est une bibliothèque de visualisation de données Python basée sur matplotlib. Il fournit une interface de haut niveau pour le dessin Graphiques statistiques attrayants et informatifs. [22]

4.3.2 Lecture des Fichiers Audio

Nous commencerons par identifier et lister les fichiers audio disponibles dans un répertoire donné à l'aide de la bibliothèque**"glob"**. Nous pouvons ensuite charger et écouter un fichier audio spécifique à l'aide de **librosa** et **IPython.display**.

Reading in Audio Files

There are many types of audio files: mp3, wav, m4a, flac, ogg

```
Entrée [229]: glob('D:\Dataset\clips\*.wav')
```

```
Out[229]: ['D:\\Dataset\\clips\\Test Test.wav',
 'D:\\Dataset\\clips\\العدل هو اساس الدولة الديموقراطية.wav',
 'D:\\Dataset\\clips\\بَقِيَّةِ-التَّعَرُّفِ-اللَّقَائِي-عَلَى-الكَلَامِ-تَتَطَلَّبُ-قَاعِدَةَ-بَيَانَاتِ-صَوْتِيَّةً-كَبِيرَ.wav',
 'D:\\Dataset\\clips\\خطأ بجعلك متواضع خير من انجاز يصيبك بالغرور.wav',
 'D:\\Dataset\\clips\\ربي اغفر لي و لوالدي يوم يقوم الحساب.wav',
 'D:\\Dataset\\clips\\ لا تؤمنو بالفشل لان الاستمتاع بالتجربة لا يعد فشلا.wav',
 'D:\\Dataset\\clips\\يمكن للأطباء علاجك ولاكن الله وحده قادر على شفائك.wav',
 'D:\\Dataset\\clips\\يَا-مُقَلِّبَ-القُلُوبِ-ثَبِّتْ-قَلْبِي-عَلَى-دِينِكْ.wav',
 'D:\\Dataset\\clips\\2_يَجِب-عَلَيْنَا-أَنْ-نَتَعَلَّم-مِنْ-أَخْطَائِنَا-وَنَعْمَل-عَلَى-تَجَنُّبِهَا-فِي-المُسْتَقْبَل.wav',
 'D:\\Dataset\\clips\\3_يَجِب-عَلَيْنَا-أَنْ-نَتَعَلَّم-مِنْ-أَخْطَائِنَا-وَنَعْمَل-عَلَى-تَجَنُّبِهَا-فِي-المُسْتَقْبَل.wav',
 'D:\\Dataset\\clips\\يَجِب-عَلَيْنَا-أَنْ-نَتَعَلَّم-مِنْ-أَخْطَائِنَا-وَنَعْمَل-عَلَى-تَجَنُّبِهَا-فِي-المُسْتَقْبَل.wav',
 'D:\\Dataset\\clips\\2_يَجِب-عَلَيْنَا-أَنْ-نَحْتَرِم-حُقُوق-الآخَرِين-وَأَنْ-نُعَامِلَهُمْ-بِاللُّطْفِ-وَالاحْتِرَام_.wav',
 'D:\\Dataset\\clips\\يَجِب-عَلَيْنَا-أَنْ-نَحْتَرِم-حُقُوق-الآخَرِين-وَأَنْ-نُعَامِلَهُمْ-بِاللُّطْفِ-وَالاحْتِرَام.wav',
 'D:\\Dataset\\clips\\يُؤْتِكُم-خَيْرًا-مِمَّا-أَخَذَ-مِنْكُم.wav',
 'D:\\Dataset\\clips\\يُمْكِن-لِلْأَطِبَّاءِ-عِلَاجُك-وَلَكِنَّ-اللَّهَ-وَحْدَهُ-قَادِرٌ-عَلَى-شِفَائِك.wav']
```

```
Entrée [230]: audio_files = glob('D:\Dataset\clips\*.wav')
```

```
Entrée [231]: # Play audio file
ipd.Audio(audio_files[2])
```

```
Out[231]:
```

FIGURE 4.4 – Lecture des Fichiers Audio

4.3.3 Analyse du Fichier Audio

Nous extrairons des informations essentielles sur l'audio, telles que la forme du signal audio (y) et le taux d'échantillonnage (sr).

```
In [13]: print(f'y: {y[:10]}')
         print(f'shape y: {y.shape}')
         print(f'sr: {sr}')

         y: [0. 0. 0. 0. 0. 0. 0. 0. 0. 0.]
         shape y: (194746,)
         sr: 22050
```

FIGURE 4.5 – Analyse du Fichier Audio

4.3.4 Affichage du Signal Audio Brut

Pour visualiser à quoi ressemble le signal audio brut, nous utiliserons **matplotlib**.

```
In [16]: pd.Series(y).plot(figsize=(10, 5),
                           lw=1,
                           title='Raw Audio Example',
                           color=color_pal[2])
         plt.show()
```

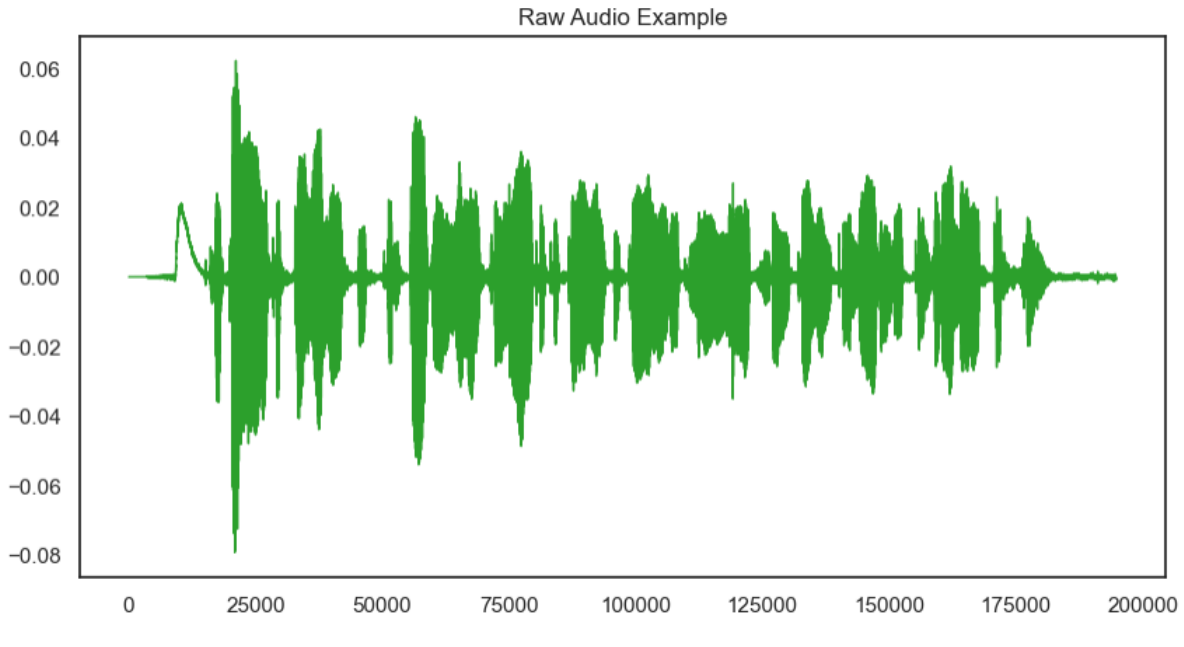

FIGURE 4.6 – Affichage du Signal Audio Brut

4.3.5 Élimination des Silences Inutiles

Nous pouvons améliorer notre analyse en supprimant les parties de silence inutiles du signal audio.

```
In [17]: # Trimming leading/lagging silence
         y_trimmed, _ = librosa.effects.trim(y, top_db=15)
         pd.Series(y_trimmed).plot(figsize=(10, 5),
                                   lw=1,
                                   title='Raw Audio Trimmed Example',
                                  color=color_pal[8])
         plt.show()
```

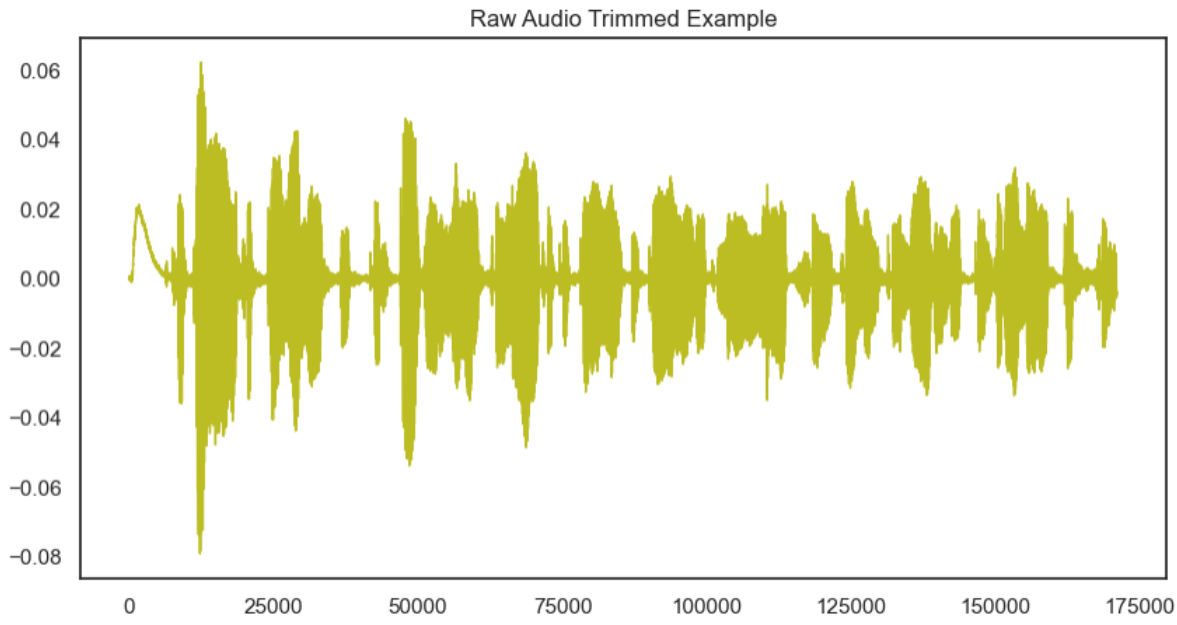

FIGURE 4.7 – Élimination des Silences Inutiles

4.3.6 Création du Spectrogramme

Pour mieux comprendre la composition fréquentielle de l'audio, nous calculerons et afficherons son spectrogramme.

Spectrogram

```
In [22]: D = librosa.stft(y)
         S_db = librosa.amplitude_to_db(np.abs(D), ref=np.max)
         S_db.shape

Out[22]: (1025, 381)

In [23]: # Plot the transformed audio data
         fig, ax = plt.subplots(figsize=(10, 5))
         img = librosa.display.specshow(S_db,
                                        x_axis='time',
                                        y_axis='log',
                                        ax=ax)
         ax.set_title('Spectogram Example', fontsize=20)
         fig.colorbar(img, ax=ax, format=f'%0.2f')
         plt.show()
```

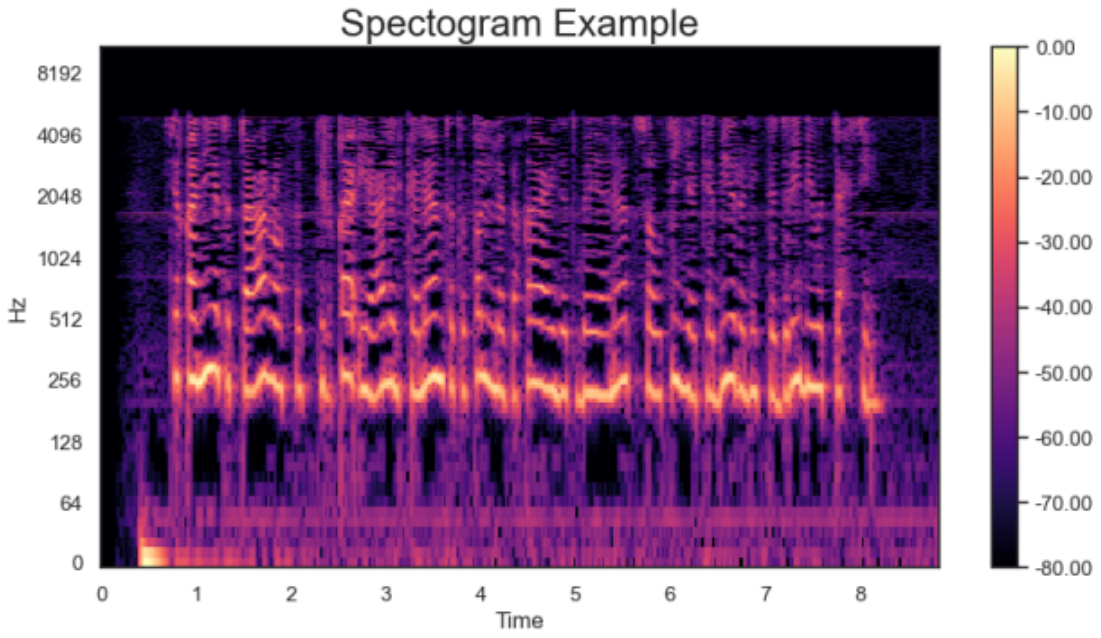

FIGURE 4.8 – Création du Spectrogramme

4.4 Conclusion

En conclusion, ce chapitre a permis une analyse approfondie de notre ensemble de données de reconnaissance de la parole en arabe pour les locuteurs Tamazigh. Nous avons examiné le nombre total d'échantillons vocaux, la répartition par âge et par sexe des locuteurs, ainsi que la structure du fichier TSV du dataset. Ces données sont essentielles pour guider la conception de nos systèmes de reconnaissance vocale, en garantissant leur robustesse et leur équité face à une diversité de locuteurs.

Conclusion générale

La reconnaissance de la parole est une technologie complexe qui nécessite une compréhension approfondie de la parole humaine et de la façon dont elle est produite. Elle est également confrontée à un certain nombre de défis, tels que la diversité des accents et des dialectes, la présence de bruits de fond et la complexité de la langue humaine.

Ce projet visait à développer un ensemble de données pour la reconnaissance de la parole arabe parlée par les locuteurs tamazight. Bien que nous ayons rencontré des défis pour collecter suffisamment de données, nous avons réussi à créer un ensemble de données de taille respectable contenant des enregistrements de locuteurs tamazight de différentes régions et ayant des accents différents.

L'un des principaux défis auxquels nous avons été confrontés était la difficulté de convaincre un nombre significatif de personnes et de volontaires de télécharger et d'utiliser notre application pour nous aider à collecter et à améliorer les données de reconnaissance automatique de la parole arabe. Cette expérience nous a enseigné des leçons précieuses sur les complexités de la collecte de données et l'importance de l'engagement des utilisateurs.

Ce projet présente un aperçu des défis et des opportunités de la reconnaissance vocale arabe. Il souligne l'importance de la collecte et de l'utilisation de jeux de données exhaustifs pour développer des systèmes plus précis et fiables. En surmontant ces limitations, les chercheurs peuvent ouvrir la voie à des applications de la reconnaissance vocale arabe dans des domaines tels que les appareils intelligents et les machines automatiques.‘

Bibliographie

[1] Sherif Mahdy Abdou et Abdullah M Moussa. "Arabic speech recognition : Challenges and state of the art". In : *Computational linguistics, speech and image processing for arabic language* (2019), p. 1-27.

[2] Ahmed Ali, Stephan Vogel et Steve Renals. "Speech recognition challenge in the wild : Arabic MGB-3". In : *2017 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)*. IEEE. 2017, p. 316-322.

[3] Ahmed Ali et al. "The MGB-2 challenge : Arabic multi-dialect broadcast media recognition". In : *2016 IEEE Spoken Language Technology Workshop (SLT)*. IEEE. 2016, p. 279-284.

[4] Shipra J Arora et Rishi Pal Singh. "Automatic speech recognition : a review". In : *International Journal of Computer Applications* 60.9 (2012).

[5] SOUSSI Haithem BESSAAD Oussama. "Conception et réalisation d'une application Web". UNIVERSITE DE SOUSSE INSTITUT SUPERIEUR DE GESTION DE SOUSSE, 2011/2012.

[6] Shammur Absar Chowdhury et al. "Towards one model to rule all : Multilingual strategy for dialectal code-switching Arabic ASR". In : *arXiv preprint arXiv :2105.14779* (2021).

[7] *Common Voice - Languages*. Rapp. tech. consulter 17/06/2023. URL : https://commonvoice.mozilla.org/fr/languages.

[8] JetBrains Open-source Contributors. *Kotlin : Concise. Cross-platform. Fun.* Rapp. tech. Consulté le 10/07/2023. 2023. URL : https://kotlinlang.org/?fromMenu.

[9] Android Developer. *Kotlin sur Android Developer*. Rapp. tech. Consulté le 10/07/2023. 2023. URL : https://developer.android.com/kotlin?hl=fr.

[10] Mourad DJELLAB et al. "Algerian Modern Colloquial Arabic Speech Corpus (AMCASC) : regional accents recognition within complex socio-linguistic environments". In : *Language Resources and Evaluation* 51 (2017), p. 613-641.

[11] Mourad DJELLAB et al. "Algerian Modern Colloquial Arabic Speech Corpus (AMCASC) : regional accents recognition within complex socio-linguistic environments". In : *Language Resources and Evaluation* 51 (2017), p. 613-641.

[12] IONOS Digital GUIDE. *Dart : présentation du langage de programmation*. Rapp. tech. Consulté le 10/07/2023. 2020. URL : https://www.ionos.fr/digitalguide/sites-internet/developpement-web/le-langage-de-programmation-dart/.

[13] IONOS Digital GUIDE. *Dart : présentation du langage de programmation*. Rapp. tech. Consulté le le 19/07/2023. 2020. URL : https://www.ionos.fr/digitalguide/sitesinternet/developpement-web/dart-le-langage-de-programmation/.

[14] L. HAMILTON. *Legal Requirements for Collecting Personal Data*. Rapp. tech. 2023. URL : https://www.termsfeed.com/blog/legal-requirements-collect-personal-data/.

[15] Abdelhakim HAMMADECHE et Mohamed TAKI. "Reconnaissance automatique de la parole arabe continu". Université Saad Dahleb, Blida 1, 2018/2019.

[16] MED LGHANDOR. *Analyse Orientée Objet UML*. Rapp. tech. 3/08/2023. URL : https://www.academia.edu/7054722/AnalyseOrienteObjetUML.

[17] LINGUA.EDU. *Les langues les plus parlées dans le monde*. Rapp. tech. Consulté le 02/05/2023. URL : https://lingua.edu/the-most-spoken-languages-in-the-world/.

[18] Hamdy MUBARAK et al. "QASR : QCRI Aljazeera Speech Resource–A Large Scale Annotated Arabic Speech Corpus". In : *arXiv preprint arXiv :2106.13000* (2021).

[19] Elmoukhtar NOUREDDINE. *Diagramme de cas d'utilisation*. Rapp. tech. 2/08/2023. URL : https://www.academia.edu/29941682/Diagrammedecasdutilisation.

[20] Zealouk OUISSAM. "Système de reconnaissance automatique de l'Amazighe et analyse des formants pour le diagnostic vocal". In : (2020).

[21] PIERREB. *Le langage Java : histoire, caractéristiques popularité*. Rapp. tech. Consulté le 10/07/2023. 2018. URL : https://www.silkhom.com/langage-java-histoire-caracteristiques-popularite/.

[22] PYDATA DEVELOPMENT TEAM. *Seaborn*. Rapp. tech. Consulté le 06/08/2023. URL : https://seaborn.pydata.org/.

[23] Hassan SATORI et al. "Investigation Arabic Speech Recognition Using CMU Sphinx System." In : *International Arab Journal of Information Technology (IAJIT)* 6.2 (2009).

[24] Manoj Kumar SHARMA et O KUMAR. "Speech recognition : A review". In : *International Journal of Advanced Networking and Applications (IJANA)* (2014), p. 62-71.

[25] Suwon SHON et al. "ADI17 : A fine-grained Arabic dialect identification dataset". In : *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE. 2020, p. 8244-8248.

[26] librosa development TEAM. *librosa*. Rapp. tech. Consulté le 06/08/2023. URL : https://librosa.org/.

[27] librosa development TEAM. *librosa Documentation*. Rapp. tech. Consulté le 06/08/2023. URL : https://librosa.org/doc/latest/generated/librosa.load.html.

[28] UCSB COLLEGE OF ENGINEERING. *An Introduction to NumPy and SciPy - UCSB College of Engineering*. Rapp. tech. Consulté le 06/08/2023. 2022. URL : https://sites.engineering.ucsb.edu/~shell/che210d/numpy.pdf.

[29] UDEMY. *Analyse de Données avec Python : Numpy, Pandas et Matplotlib*. Rapp. tech. Consulté le 06/08/2023. 2022. URL : https://www.udemy.com/course/manipulation-de-donnees-en-python-mai.

[30] UNIVERSITY OF CALIFORNIA, BERKELEY. *Matplotlib Tutorial - University of California, Berkeley*. Rapp. tech. Consulté le 06/08/2023. URL : https://www.stat.berkeley.edu/~nelle/teaching/2017-visualization/README.html.

[31] VISUAL STUDIO CODE. *Visual Studio Code*. Rapp. tech. Consulté le 19/07/2023. URL : https://code.visualstudio.com/.