ECOLE NATIONALE SUPÉRIEURE
D'INFORMATIQUE ET D'ANALYSE DES SYSTÈMES

End of Second Year Project

Arabic Speech To Moroccan Sign Language Web Application

Defended on 05/06/2023
by :
AIT DAOUD KHADIJA
SEDJARI YASSINE

Supervised by :
Pr. LAZAAR MOHAMED
Jury :
Pr. NAOUM MOHAMED

Academic Year:2022-2023

Appreciation and Gratitude

In the name of Allah, the Most Gracious, the Most Merciful. All praise and thanks are due to Allah alone, the Lord of all worlds. We express our utmost gratitude and appreciation to Him for His blessings, guidance, and unwavering support throughout the completion of this project.

Acknowledgments

We would like to express our sincere gratitude and appreciation to the following individuals and organizations for their invaluable support and contributions in the completion of this project. First and foremost, we begin by thanking Professor Mohamed NAOUM for giving us the project idea. Their vision, expertise, and guidance have been instrumental in shaping the direction and scope of our work. We also extend our heartfelt thanks to our supervisor, Pr. LAZAAR Mohamed, for their guidance, expertise, and constant encouragement throughout the entire duration of this project. Their insights and feedback have been instrumental in shaping the direction and quality of our work. We would also like to thank the members of our project team for their dedication and collaborative efforts. Their commitment to excellence, teamwork, and effective communication played a crucial role in achieving our project objectives. Furthermore, we want to express our heartfelt appreciation to our families and friends for their unwavering support, understanding, and encouragement throughout this journey. Their belief in us has been a constant source of motivation. Finally, we would like to thank all the participants who volunteered their time and contributed to our research, especially our friend and brother NECHBA Mohammed. Without their involvement and cooperation, this project would not have been possible. In conclusion, we are indebted to all those who have played a part, no matter how big or small, in making this project a success. Thank you for your guidance, support, and belief in us.

Abstract

Moroccan Sign Language (MSL) plays a vital role in communication for individuals with speaking and hearing impairments in the Arabic-speaking community. While efforts have been made to create systems that convert sign language symbols to text, the development of systems that translate spoken Arabic into Moroccan Sign Language remains limited. This project focuses on the development of a comprehensive translating system for Arabic Speech to Moroccan Sign Language. The system consists of various modules designed to convert spoken Arabic audio into Arabic text. The parsed text is then transformed into a structured representation in Moroccan Sign Language. Additionally, the system removes stop words from the reordered sentence. Since Moroccan Sign Language does not support word conjugation. The system then checks each individual word against a dictionary containing videos of MSL signs. In case a word is not found in the dictionary, the system replaces it with a concatenation of its individual letters. By developing this Arabic Speech to Moroccan Sign Language System, we aim to bridge the communication gap between individuals who use spoken Arabic and those who rely on Moroccan Sign Language for effective communication.

Keywords :

Speech to text, **Sign Language Translation**, **NLP**.

Résumé

La Langue des Signes Marocaine (LSM) joue un rôle vital dans la communication des personnes ayant des troubles de la parole et de l'ouïe au sein de la communauté arabophone. Bien que des efforts aient été déployés pour créer des systèmes de conversion des symboles de la langue des signes en texte, le développement de systèmes de traduction de l'arabe parlé en Langue des Signes Marocaine reste limité. Ce projet se concentre sur le développement d'un système de traduction complet de la parole arabe en Langue des Signes Marocaine. Le système est composé de différents modules conçus pour convertir l'audio de l'arabe parlé en texte arabe. Le texte analysé est ensuite transformé en une représentation structurée en Langue des Signes Marocaine. De plus, le système supprime les mots vides de la phrase réorganisée. Étant donné que la Langue des Signes Marocaine ne prend pas en charge la conjugaison des mots, le système vérifie ensuite chaque mot individuel par rapport à un dictionnaire contenant des vidéos de signes en LSM. Si un mot n'est pas trouvé dans le dictionnaire, le système le remplace par une concaténation de ses lettres individuelles. En développant ce système de traduction de la parole arabe en Langue des Signes Marocaine, notre objectif est de combler le fossé de communication entre les personnes qui utilisent l'arabe parlé et celles qui dépendent de la Langue des Signes Marocaine pour une communication efficace.

Mots Clés :

Speech to text, **Sign Language Translation**, **NLP**.

Contents

List of Figures

General Introduction

Sign language plays a crucial role in bridging the communication gap for individuals with hearing or speech impairments who rely on sign language for everyday communication. However, the availability of efficient models that can convert text to Moroccan Sign Language (MSL) is currently limited. Adequate audio-visual support for oral communication is still lacking, posing challenges for effective communication. While significant progress has been made in the detection of sign languages in various countries using computer-based methods, research on the detection and translation of MSL remains scarce. Most existing work in this field has primarily focused on sign languages such as American Sign Language (ASL) or British Sign Language, with minimal development specifically tailored to Indian Sign Language. In the context of our project, we aim to develop a system that focuses on the translation of Arabic speech to Moroccan Sign Language. The system will gather relevant videos for each word in MSL and provide substitutes for words not present in the dictionary, considering semantic similarity and maintaining the appropriate parts of speech. Translation between spoken languages is a complex task due to variations in grammar and language rules, and this complexity is further amplified when one language is spoken and the other is a sign language. The main users of this system are individuals with hearing impairments who face communication challenges in everyday situations, as not everyone understands sign language. This poses significant obstacles to communication between individuals who can speak and those with hearing impairments. Therefore, our system aims to address these challenges by providing a tool for learning sign language and reducing communication gaps. By developing an **Arabic Speech to Moroccan Sign Language system**, we aim to empower individuals with hearing impairments to communicate effectively and facilitate interaction with the wider community that may not be familiar with sign language.

Chapter 1

Analysis and Study of Specifications

1.1 Introduction

This chapter introduces the general context of the project. We will present the project, the problem statement, and the project planning.

1.2 Study of Specifications

1.2.1 Problem Statement

Deaf individuals face significant communication challenges when interacting with non-deaf individuals, especially in critical situations such as emergencies involving police, firefighters, or ambulance services. These situations can be particularly dangerous and potentially life-threatening for deaf individuals who struggle to effectively communicate their needs and understand instructions.

In Morocco, the communication barriers experienced by deaf individuals further limit their freedom and independence. Simple tasks like driving a car become daunting challenges due to the inability to easily communicate with traffic police or other drivers. This lack of communication not only hinders their personal freedom but also poses risks to their safety and the safety of others on the road.

Currently, there is a pressing need for a solution that empowers deaf individuals in Morocco to overcome communication barriers, enabling them to interact more effectively with hearing individuals, especially during critical situations and everyday activities. By providing a reliable and efficient Arabic Speech to Moroccan Sign Language translation system, this project aims to bridge the communication gap between deaf and hearing individuals, granting deaf individuals greater freedom, independence, and the ability to fully participate in various aspects of life, including driving cars.

The development of such a system will not only enhance the safety and well-being of deaf individuals but also contribute to creating a more inclusive and accessible society in Morocco.

1.2.2 Goals

The goal of this project is to develop a web application that performs the following tasks:

- Takes Arabic Speech audio as input.

- Converts the Arabic speech into Arabic text using speech recognition techniques.
- Constructs a video representation of Moroccan Sign Language (MSL) based on the Arabic text.
- Displays the MSL video in a conversational format, allowing users to easily understand and communicate using sign language.

By achieving these objectives, the web application aims to facilitate communication between individuals who use spoken Arabic and those who rely on Moroccan Sign Language. It provides a convenient and accessible platform for users to express themselves through sign language, breaking down communication barriers and promoting inclusivity.

1.2.3 Feasibility Study

Based on the analysis of the specifications and various meetings, we have identified two requirements to accomplish this project: one is hardware-related, and the other is software-related.

On the hardware side, the application requires a server that houses a database containing all the necessary information, as well as a computer. On the development side, we need software tools to develop the dynamic web application and scripts for communication with the database server and the web client.

Regarding the data sources, we perform manual data collection of Moroccan Sign Language YouTube videos, which are then cleaned and structured for storage in our database. This data retrieval process needs to be automated to ensure that the data is up-to-date whenever a user makes a request.

In conclusion, the allocated 2-month period for this project is reasonable and sufficient to complete the project successfully.

1.2.4 Team Formation

Le projet a été divisé en trois parties :

- Design and Development : Implemented by **AIT DAOUD KHADIJA**.
- Arabic Automatic Speech Recognition : Implemented by **AIT DAOUD KHADIJA**.
- Text processing : Implemented by **SEDJARI YASSINE**
- File Management System Construction : Implemented by **SEDJARI YASSINE**
- Back-end and Front-end Development : Implemented by **SEDJARI YASSINE**

1.2.5 Proposed Solution

Our solution to successfully complete this project within the given timeframe was to divide the tasks based on functionality, namely:

Arabic Automatic Speech Recognition: This involves the use of a machine learning model

to extract Arabic text from an audio signal corresponding to the user's speech in the Arabic language.

Text Processing: This involves all the steps of transformation that the extracted text undergoes before being used to search for corresponding sign language videos. These steps include removing the definite article from the beginning of each word, removing possessive suffixes, removing stop words, and so on..

Back-end Development: The back-end development process includes designing and building the application's server infrastructure, defining the data models and database schemas, implementing server-side functionalities, and testing and debugging the code for proper functionality.

Front-end Development: Front-end development involves creating the user-facing components and interfaces of a software application or website. It focuses on the design, layout, and interactivity that users directly interact with.

1.2.6 Advantages of our Solution

Our solution perfectly meets the requirements of the specifications and brings additional improvements to the project. It offers several advantages, including:

Improved Communication: Our application addresses the significant communication challenges faced by deaf individuals when interacting with non-deaf individuals, particularly in critical situations. By providing Arabic Speech to Moroccan Sign Language translation, it enables effective communication between deaf and hearing individuals, allowing them to understand and convey their needs, especially during emergencies involving police, firefighters, or ambulance services.

Enhanced Safety: The ability to communicate effectively during critical situations is crucial for the safety and well-being of deaf individuals. Our application empowers deaf individuals in Morocco to overcome communication barriers, ensuring they can understand instructions and convey vital information, ultimately improving their safety and reducing potential risks.

Increased Independence: Our application eliminates communication barriers that limit the freedom and independence of deaf individuals in Morocco. By providing reliable translation between spoken Arabic and Moroccan Sign Language, it enables deaf individuals to engage in various daily activities with confidence, including tasks that were previously challenging, such as driving a car. This increased independence enhances their overall quality of life.

Inclusive Society: The development of our application contributes to creating a more inclusive and accessible society in Morocco. By bridging the communication gap between deaf and hearing individuals, it promotes equal participation and opportunities for deaf individuals in various aspects of life, fostering a more inclusive and diverse community.

Personal Empowerment: Our application empowers deaf individuals by providing them with the tools to effectively communicate and express themselves. It gives them the ability to be actively involved in conversations, make informed decisions, and advocate for their rights and needs, leading to greater personal empowerment.

Overall, our application offers a comprehensive solution that not only addresses the communication challenges faced by deaf individuals but also enhances their safety, independence, and overall well-being.

1.3 Specification of requirements

The specification of requirements is the initial phase of any application development process, during which we identify the needs of our application. We distinguish between functional requirements, which outline the expected functionalities of our application, and non-functional requirements, which ensure the development of a satisfactory application.

1.3.1 Functional requirements specification

In this section, we outline the various services offered by our application.

Back-end:

- Reception of audio signals from users
- Conversion of the signal into text
- Text processing
- Searching and concatenating videos from the database
- Sending the concatenated video
- Compliance with user data usage rules
- Restricted access to data
- Data security
- Updates
- Communication with the database

Interface web :

- No authentication needed.
- Direct audio recording in the browser.
- Communication in the form of a chatbot conversation.
- Interaction via audio and text.
- Reception of a video in Arabic Sign Language translating the input audio or text.

1.3.2 Non-functional requirements specification

Non-functional requirements describe all the technical, ergonomic, and aesthetic constraints that the system must meet for its development and proper functioning. As for our application, we have identified the following requirements:

- **Availability:** The application should be accessible and available for use by any user.
- **Security of access to critical information:** We must ensure the confidentiality of client data, especially in terms of authentication.
- **User-friendly interface:** The application should provide a user-friendly and intuitive interface for all types of users.
- **Scalability and flexibility:** The application should be designed in a way that allows for future enhancements and the addition of new functionalities to ensure scalability and adaptability of the solution.

1.4 Project management

1.4.1 Project Life Cycle

We have chosen the V-model because this model is characterized by parallelism.

In this model, we vertically find the development stages and horizontally the verification stages

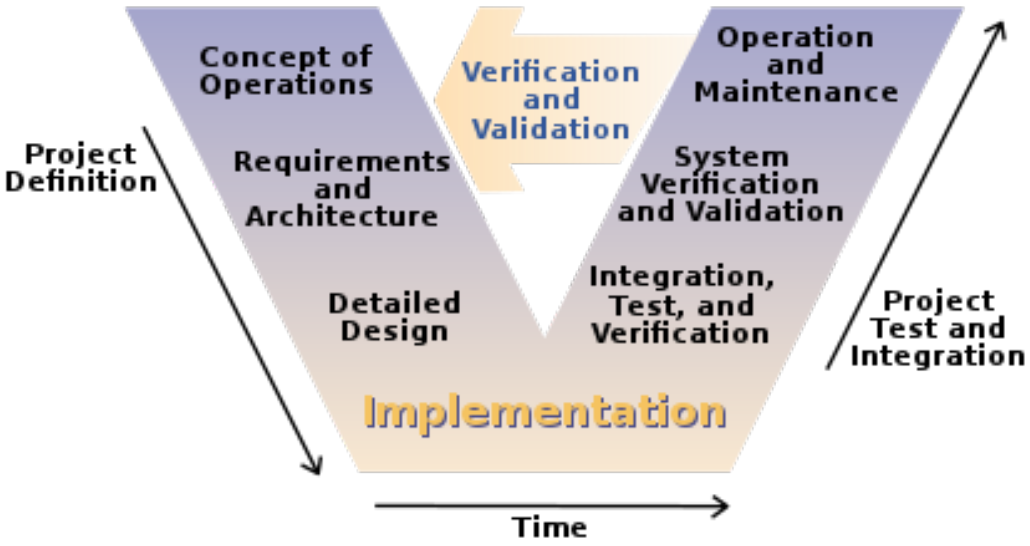


Figure 1.1: V project life cycle

1.5 Conclusion

In conclusion, our website addresses a major problem faced by our society. It serves as an intermediary between companies and job seekers. Our solution must rigorously meet the pre-established requirements.

Chapter 2

Design and Modeling

2.1 Introduction

In this section, we present some diagrams that illustrate the functionalities offered by our solution and their workflows. This representation, separate from the code, serves as a means of communication between the project owner and the developer.

2.2 UML Presentation

UML (Unified Modeling Language) is a standard language used to specify, visualize, design, and document all aspects of an information system. It provides a communication support: a graphical language consisting of 13 standard diagrams (for UML 2.0) representing different 'views' of an information system. It also allows for expressing and developing object models independently of any programming language. As an example of this modeling, we will present a set of use case diagrams and class diagrams.

Use Case Diagram: A use case diagram captures the behavior of a system and expresses the needs of system users. They provide a user-oriented view of these needs rather than a computer-oriented view.

Class Diagram: A class diagram provides an overall view of a system by presenting its classes, interfaces, collaborations, and the relationships between them. Class diagrams are static: they show what interacts but not what happens during the interaction.

Sequence Diagram: A graphical representation of interactions between actors and the system in a chronological order, as formulated in the Unified Modeling Language.

2.3 Actors

An actor is the archetype of the user (person, external process, etc.) who interacts with the system.

- **User** : can interact with the conversational agent eaither by audio recordings or by text and get videos of Moroccan Sign Language.

2.4 Use case Diagram

2.4.1 Use Case Diagram

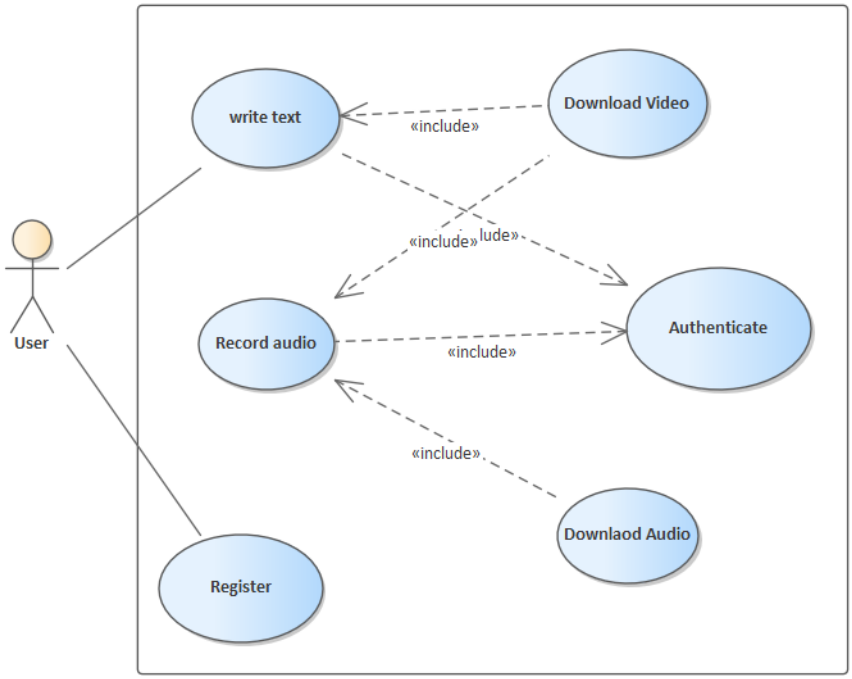


Figure 2.1: Use Case Diagram

2.4.2 Use Case Description

Account Creation

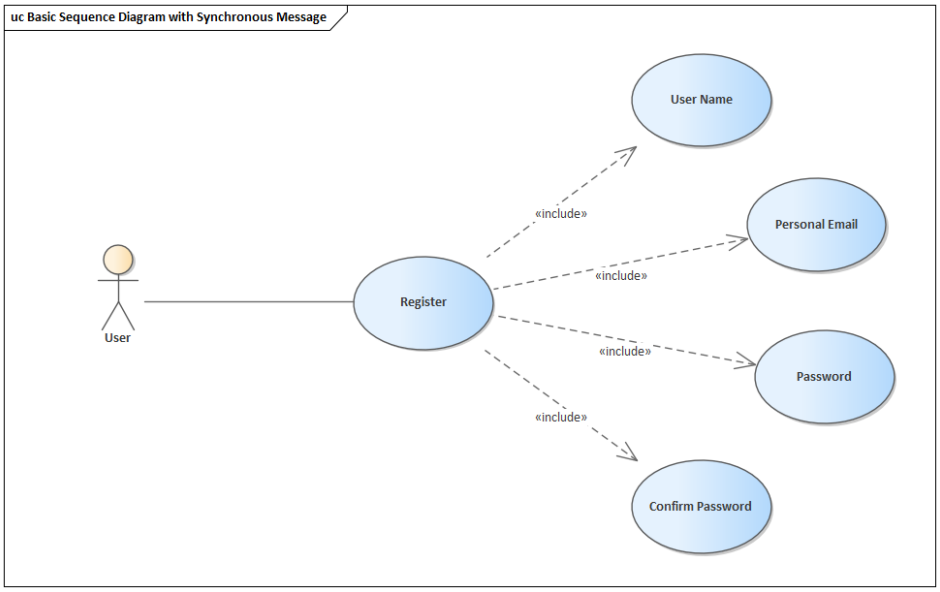


Figure 2.2: Use Case « Account Creation »

Authentication

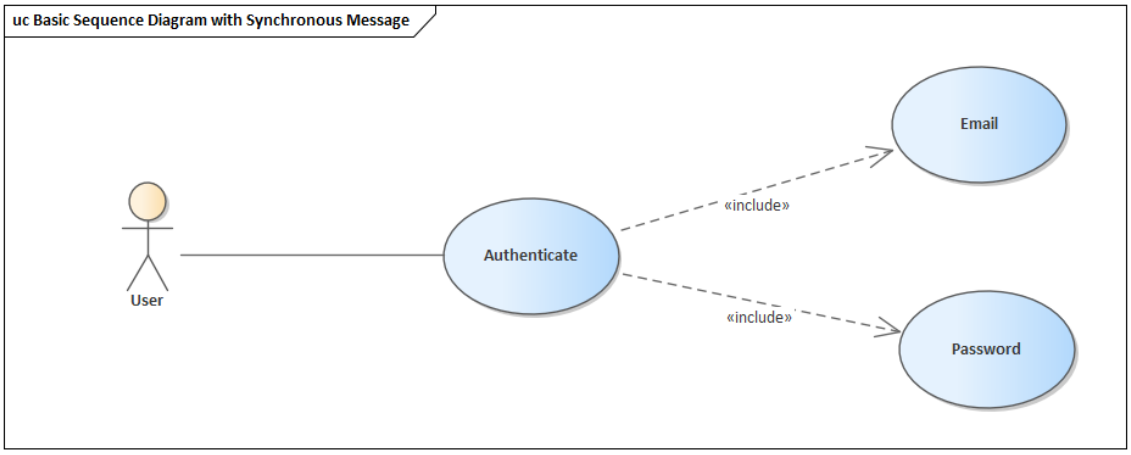


Figure 2.3: Use Case « Authentication »

2.5 Class Diagram

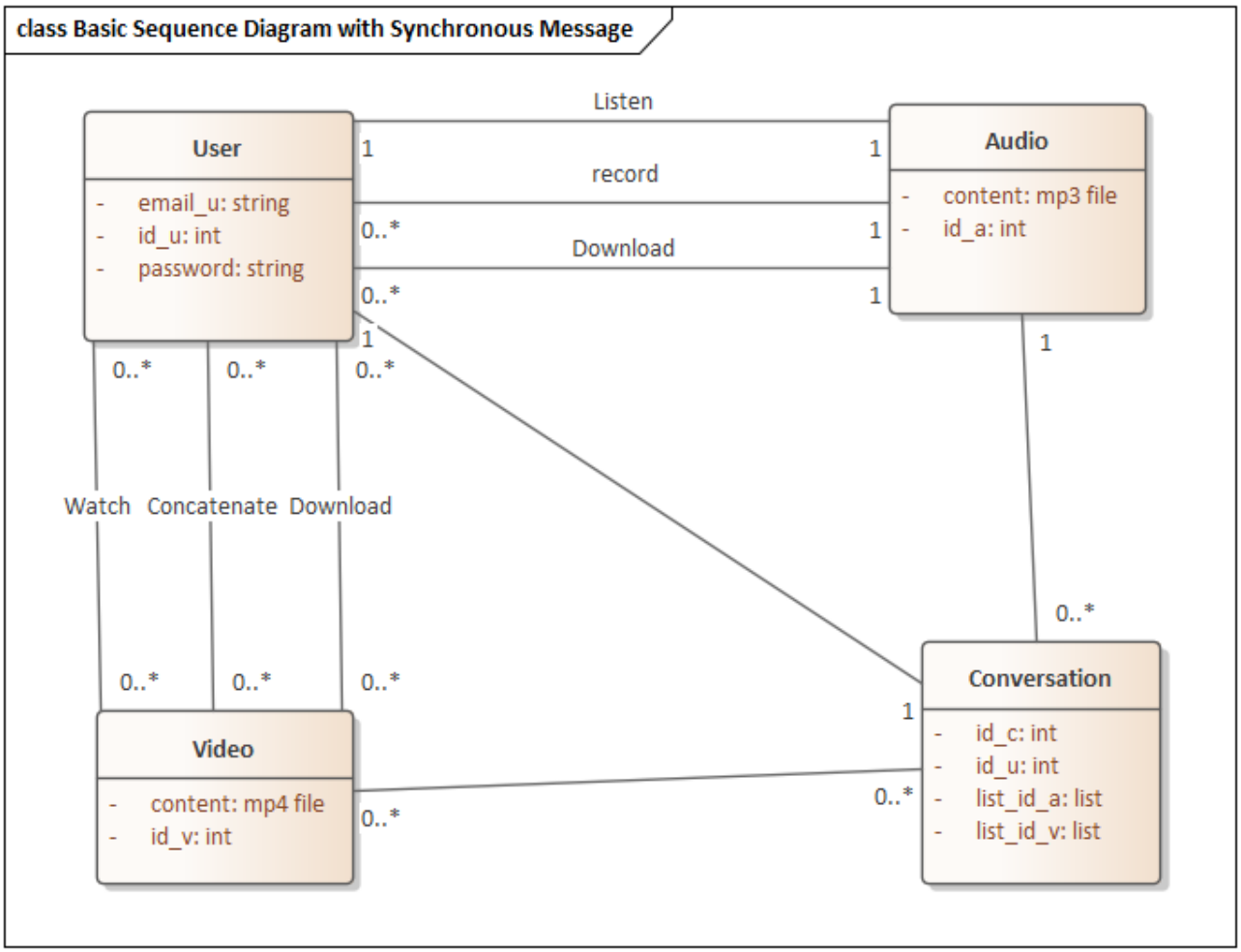


Figure 2.4: Class Diagram of the System

2.6 Sequence Diagram

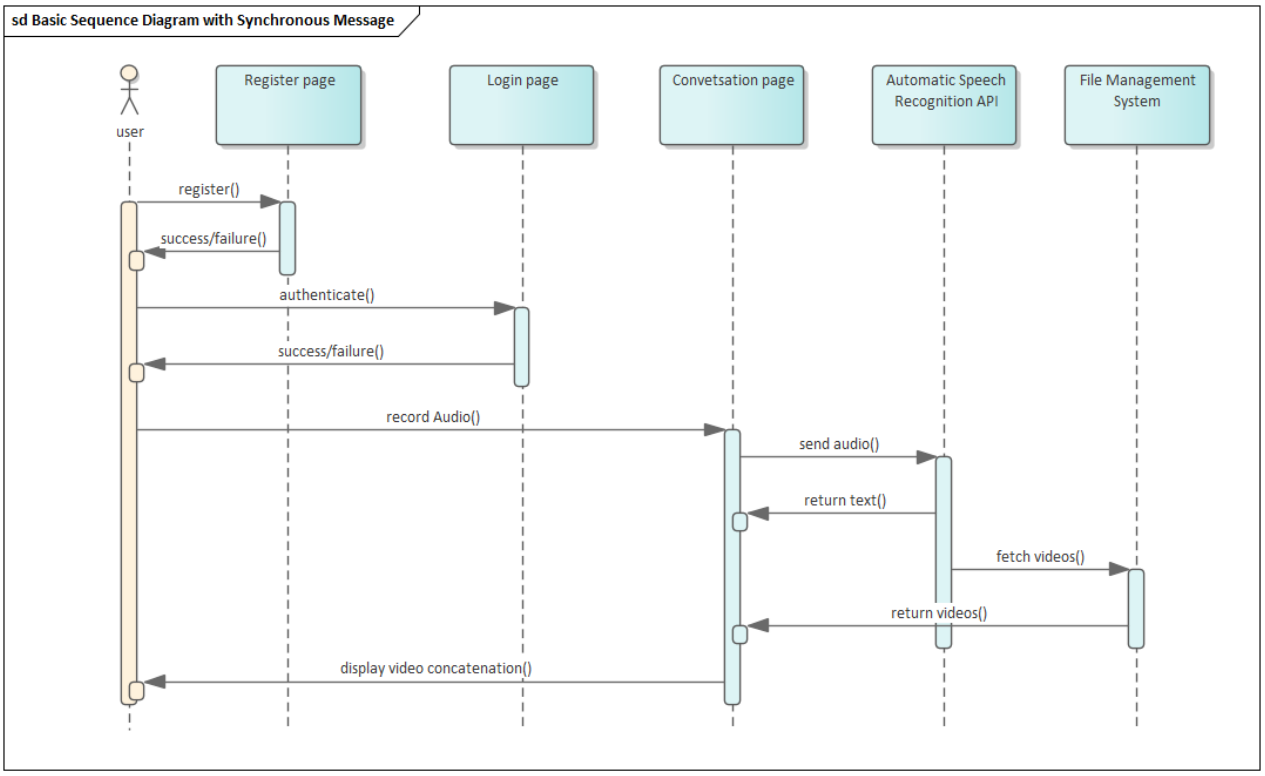


Figure 2.5: Sequence Diagram of User Interaction

2.7 Conclusion

During this chapter, we have designed the various components of our system. Now, our applications are ready to be implemented. The next chapter will focus on the deployment of our applications.

Chapter 3

Methodology

In this chapter we provide a detailed overview of the approach and techniques employed in the development of the Arabic Speech to Moroccan Sign Language translation system. This chapter outlines the step-by-step process followed to achieve the project's objectives. It begins by explaining the text-to-video architecture, which forms the foundation of the system. Subsequently, it delves into the key components of the architecture, including Automatic Speech Recognition, Text Processing, and Storage and Access. The methodology chapter serves as a guide for understanding the methods and strategies used in the implementation of the project, setting the stage for the subsequent sections of the report.

3.1 Audio-to-Video Architecture

Algorithmic Design:

1. wav2vec2 large xlsr 53 arabic is used to obtain user speech and convert it to text.

2. The sentence is restructured using stop words, definite article and possessive suffixes removal

3. Fetching for each word's (index) corresponding video in Moroccan Sign Language in the File Management System.

4. Concatenate all videos in the same order as the original input and display them to the user.

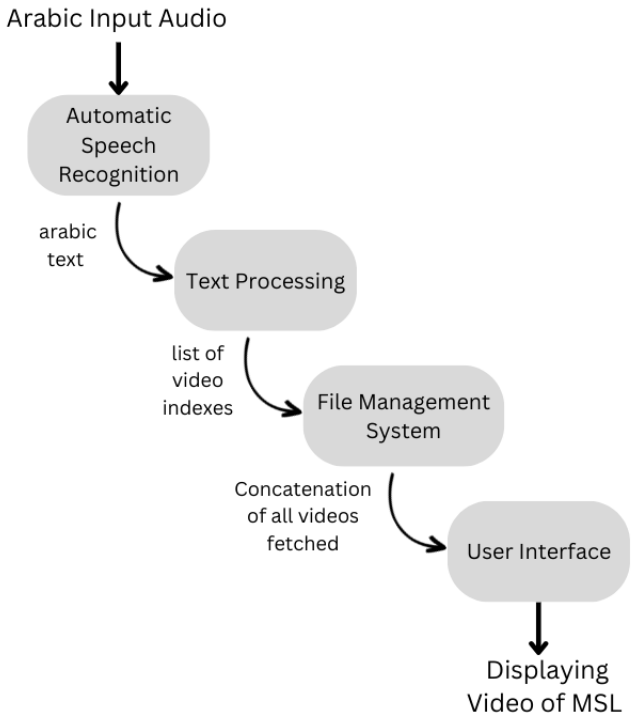


Figure 3.1: Speech-to-video low level design

Automatic Speech Recognition: This Module takes as an input an acoustic signal X of arabic language, the objective then is to determine the most probable sequence of words W that produced X. The output is an arabic text in string format.

Text Processing: This module aims to convert the input text with the grammar of Arabic into a normalized text that can be used to fetch the word indexed videos of Moroccan Sign Language in our File Management System. various methods are used such as stop words removal. The output is a list of word indices.

File Management System: In this module, the tree directory structure where we store our videos get fetched using their indicies as an input, the vedios retrieved get concatenated and delivered to the UI as an Output.

UI: The User Interface displays the video gotten from the pevious module and give the possibility to download it or change the reading speed.

3.2 Automatic Speech Recognition

3.2.1 Introduction

Automatic Speech Recognition (ASR) is a technology that aims to model the relationship between speech signals and their corresponding linguistic units, such as phonemes, words, or sentences, using mathematical and statistical models. The process of ASR can be mathematically represented as follows:

Given an acoustic signal X, ASR aims to find the most likely sequence of words W that generated that signal. This can be expressed as:

$$W^* = \arg\max P(W|X) \tag{3.1}$$

where W* represents the optimal sequence of words, and P(W|X) is the probability of word sequence W given the acoustic signal X.

ASR systems typically consist of several components:

Acoustic Model: This component models the relationship between the acoustic features of the speech signal and the corresponding phonetic units. It uses techniques such as Hidden Markov Models (HMMs) or Deep Neural Networks (DNNs) to estimate the likelihood of observing acoustic features given a particular phonetic unit.

Language Model: The language model captures the statistical properties of the language being spoken. It assigns probabilities to word sequences based on their occurrence in a given language. N-gram models, Finite State Transducers (FSTs), or Recurrent Neural Networks (RNNs) are commonly used to represent language models.

Pronunciation Model: This component provides the mapping between words and their corresponding phonetic representations. It ensures that the acoustic model and language model are compatible by aligning the phonetic units with the words in the language.

Decoder: The decoder combines the output from the acoustic model, language model, and pronunciation model to generate the most likely word sequence given the acoustic signal. It uses algorithms such as the Viterbi algorithm or the beam search algorithm to efficiently search through the space of possible word sequences.

Overall, ASR involves complex mathematical modeling and statistical estimation to transform acoustic signals into textual representations. The accuracy and performance of ASR systems heavily rely on the quality of these models and the availability of large amounts of training data.

In order to perform the automatic speech recognition, we will utilize the XLSR-53 model, which is a customized version of the Cross-lingual Speech Recognition (XLSR) system designed explicitly for accurate Arabic speech recognition. This model is developed based on the advanced XLSR-Wav2Vec2.0 architecture.

3.2.2 Wave2vec

Wav2Vec 2.0 is a speech model for self-supervised learning of speech representations that masks the speech input in the latent space and solves a contrastive task defined over a quantization of the jointly learned latent representations. It takes a float array corresponding to the raw waveform of the speech signal.

Here is an illustration of Wav2Vec 2.0 framework that learns both contextualized speech representations and a set of discretized speech representations.

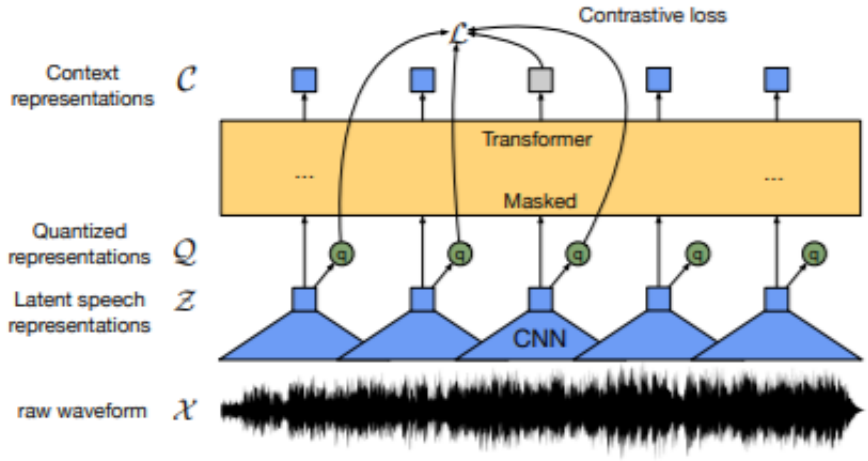


Figure 3.2: Wav2Vec2 framework

The Wav2Vec 2.0 model consists of two essential components:

Multi-layer convolutional feature encoder, denoted as f: X → Z, where X represents the raw audio input and Z represents the resulting latent speech representations (z_1, . . . , z_T) obtained over T time-steps. These latent representations capture important features from the audio input.

Transformer labeled as g: Z → C, where Z represents the latent speech representations and C represents the output representations. The Transformer takes in the latent representations and processes them to generate representations (c_1, . . . , c_T) that encapsulate information from the entire sequence. This allows the model to capture contextual dependencies and relationships between the latent speech representations.

In summary, Wav2Vec 2.0 utilizes a convolutional feature encoder to extract latent speech representations from the raw audio input, which are then further processed by a Transformer to generate comprehensive representations that capture the contextual information of the entire sequence.

3.2.3 Comparaison to BERT

Similarly to the Bidirectional Encoder Representations from Transformers (BERT), Wav2Vec2 employs a similar concept but with speech signals, but instead of masking words in text, Wav2Vec2 randomly masks short segments of the input audio waveforms. By training the model to predict the masked sections of the audio, it learns to capture contextualized representations of the speech.

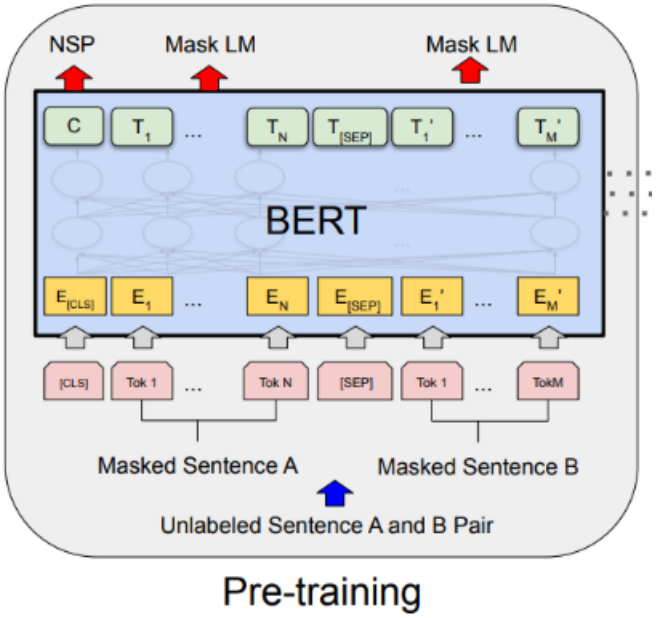


Figure 3.3: Wav2Vec2 framework

3.2.4 Dataset

The dataset used to train Wav2Vec 2.0 is a speech corpus that was recorded in south Levantine Arabic (Damascian accent) using a professional studio. Synthesized speech as an output using this corpus has produced a high quality, natural voice.

train (1.81k rows)

file (string)	text (string)	audio (audio)	phonetic (string)	orthographic (string)
"zip://arabic-speech-corpus/wav/ARA NORM 0002.wav::http://en.arabicspeechcorpus.com/arabic…	"waraj~aHa Alt~aqoriyru Al~aTHiy >aEad~ahu maEohad…	0:15 / 0:15	"sil w a r a' jj A H a tt A …	"waraj~aHa Alt~aqoriyru Al~a*iy >aEad~ahu maEohadu…
"zip://arabic-speech-corpus/wav/ARA NORM 0003.wav::http://en.arabicspeechcorpus.com/arabic…	"mim~aA qado yu&ad~iy <ilaY taraAjuEi masaAHaAti…	0:01 / 0:06	"sil m i0' mm aa q A' d y u0…	"mim~aA qado yu&ad~iy <ilaY taraAjuEi masaAHaAti…
"zip://arabic-speech-corpus/wav/ARA NORM 0004.wav::http://en.arabicspeechcorpus.com/arabic…	"wa*akara Alt~aqoriyru >ana taraAjuEa masaAHapi…	0:01 / 0:13	"sil w a * a' k a r a tt A q r…	"wa*akara Alt~aqoriyru >ana taraAjuEa masaAHapi…

Figure 3.4: Dataset Card for Arabic Speech Corpus

3.2.5 Fine-tuning

Fine-tuning in the context of Wav2Vec 2.0 involves adapting the pretrained model to a specific downstream task that allows to the model to specialize and improve its performance on a specific task.

To fine-tune Wav2Vec 2.0, a classifier is added on top of the pretrained model. This classifier represents the output vocabulary or labels of the target task, such as phonemes or words. The model is then trained on labeled data specific to the task using techniques like Connectionist Temporal Classification (CTC) loss.

3.3 Text Processing

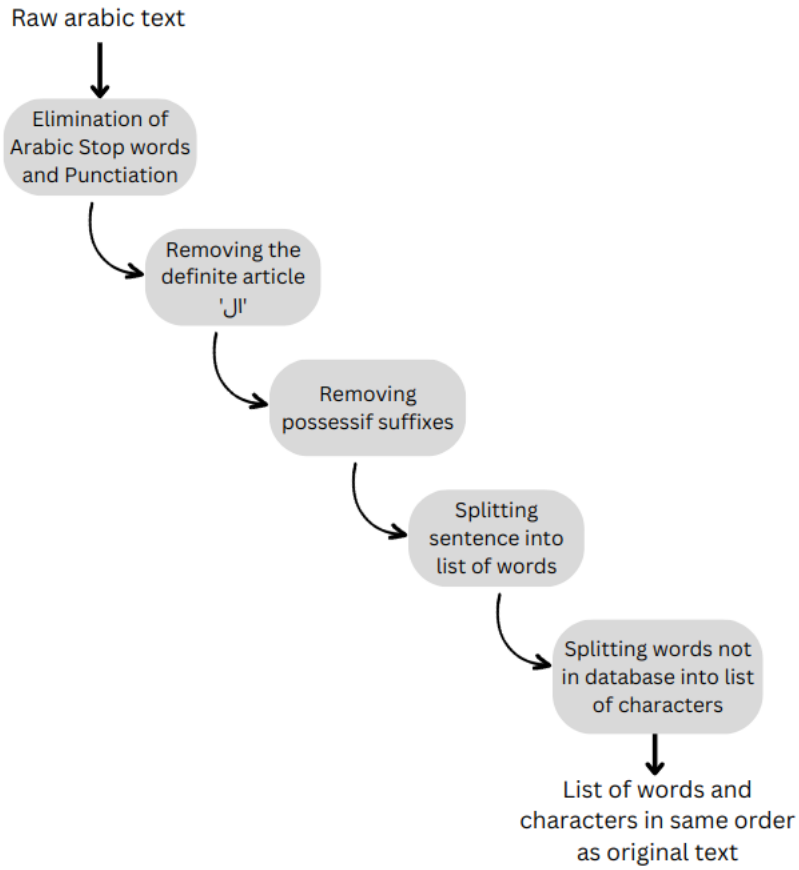


Figure 3.5: Text Processing low level design

As show in the Figure 3.5 the arabic text is processed by first removing stop words and punctuation marks, removes the definite article from words, removes possessive suffixes from words, and optionally splits words into individual letters if they are not found in a provided name directory.

3.4 Storage and Access

In our project, the choice of using a file system management approach for storage has been made based on the following considerations.

1. Directory Structure: The project utilizes a directory-based structure for storing the videos. A parent directory named "Database" contains 42 subdirectories, each named from 0 to 41. Within each subdirectory, a single video file is stored. This structured organization allows for easy indexing and retrieval of videos based on their corresponding index.

2. Video Characteristics: The videos in the project are relatively small in size, with a maximum duration of 10 seconds and a quality of 480p. These video attributes contribute to their lightweight nature, making them suitable for storage within a file system. The file system can efficiently handle the storage and retrieval of these videos without significant performance

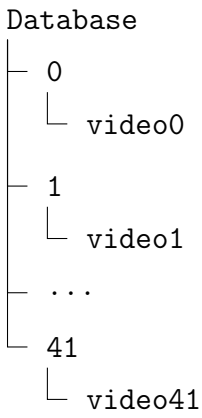


Figure 3.6: Tree Structure Directory

impact.

3. Simplified Access: By adopting a file system management approach, accessing the videos becomes straightforward. To fetch a specific video, only the corresponding index within the range of 0 to 41 needs to be provided. This simplicity allows for efficient and direct access to the desired video without the need for complex querying or data retrieval mechanisms.

4. Compatibility and Portability: File systems are widely compatible across different platforms and operating systems. The videos stored in the file system can be easily shared, accessed, and moved between systems, enhancing portability and facilitating collaboration among project stakeholders.

5. Scalability Considerations: Considering the current project scope, the chosen file system management approach is deemed appropriate. As the videos are limited in number and relatively small in size, a file system can adequately handle the storage requirements. However, it's important to note that as the project scales and the volume of videos increases significantly, alternative storage solutions such as databases may need to be considered to ensure efficient data management and scalability.

By leveraging a file system management approach, the Arabic Speech to Moroccan Sign Language project benefits from a simple, lightweight, and easily accessible storage solution. The structured directory organization and the specific video characteristics align well with the capabilities of a file system, allowing for efficient video storage and retrieval.

3.5 Conclusion

In conclusion, the methodology chapter has presented the text-to-video architecture, including Automatic Speech Recognition, Text Processing, and Storage and Access. This architecture aims to bridge the communication gap between Arabic speakers and Moroccan Sign Language users. By converting audio signals to text, processing the text, and efficiently storing and accessing videos, our project enables effective translation and enhances communication for deaf individuals. The methodology chapter lays the groundwork for the subsequent stages of implementation and evaluation, ensuring the system meets the specific requirements of Arabic Speech to Moroccan Sign Language translation.

Chapter 4

Technologies used and implementation

4.1 Introduction

The implementation stage is a crucial step in the lifecycle of our applications. This stage allows us to materialize our project by developing interfaces and implementing concrete functionalities of the system. To develop these applications, we have used several development tools.

4.2 Work environment

4.2.1 Work environment

Marque	Hp
Processeur	7th Gen Intel® Core(TM) i5-7200U CPU @ 2.50GHz
RAM	20Go
Disque dur	512Go SSD
Système d'exploitation	Linux Ubuntu 22.04.1 LTS

4.2.2 Software environment: Database management system

SQLAlchemy is an Object-Relational Mapping (ORM) tool that allows you to synchronize your classes with tables in a relational database (based on SQL). The most important step when using an ORM is to perform the "Mapping," which involves associating the elements of your classes with those of your database tables.

4.3 Technical choices

4.3.1 Programming languages

Python is a powerful and easy-to-learn programming language. It provides high-level data structures and allows for a simple yet effective approach to object-oriented programming. With its elegant syntax, dynamic typing, and interpreted nature, Python is an ideal language for writing scripts and rapidly developing applications in various domains and on most platforms.

Flask is an open-source micro web framework for Python development. It is classified as a microframework because it is lightweight. Flask aims to keep its core simple yet extensible. It does not include an authentication system, a database abstraction layer, or a form validation tool. However, many extensions are available to easily add additional functionalities.

JavaScript is a scripting language primarily used in interactive web pages and is an essential part of web applications. Alongside HTML and CSS, JavaScript is at the core of the languages used by web developers. The majority of websites use it, and most web browsers have a JavaScript engine to interpret it.

JavaScript is a prototype-based object-oriented language: the language's fundamentals and its main interfaces are provided by objects. However, unlike a class-based object-oriented language, basic objects are not instances of classes. Additionally, functions are first-class objects. The language supports the object-oriented, imperative, and functional programming paradigms.

Bootstrap is a collection of tools that are useful for creating the design (graphics, animations, browser page interactions, etc.) of websites and web applications. It is a framework that includes HTML and CSS code, forms, buttons, navigation tools, and other interactive elements, as well as optional JavaScript extensions. It is one of the most popular projects on the GitHub development platform. Before the advent of Bootstrap, there were several libraries available, which led to inconsistencies and high maintenance costs.

4.3.2 Description Languages

HTML (HyperText Markup Language) refers to a type of descriptive computer language. It is specifically a data format used in the realm of the Internet for formatting web pages. It allows for writing hypertext and introducing multimedia resources into content, among other things.

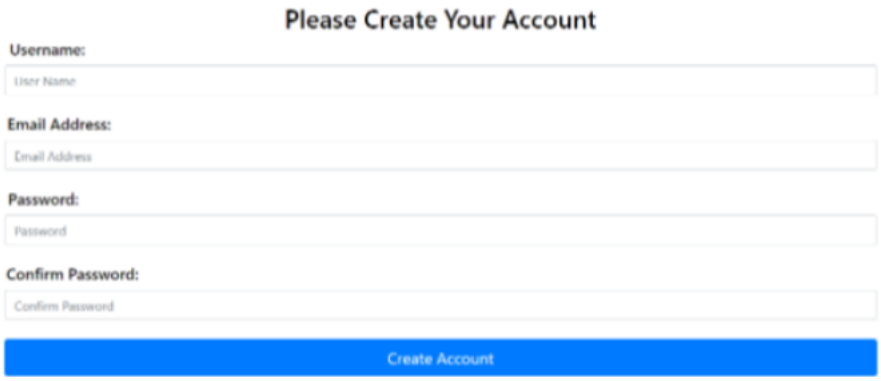
CSS (Cascading Style Sheets) is a web-based markup language used to describe the appearance and formatting of a website in a browser. It is commonly used in conjunction with HTML and is responsible for defining the visual styles and layout of web pages.

4.4 User Interfaces

The graphical interface is a crucial aspect in the development of a user-friendly and visually appealing application that provides a pleasant user experience during navigation. This factor can make a difference between two applications, even if they offer the same functionalities. Here are a series of screenshots from our applications.

4.4.1 Register Page

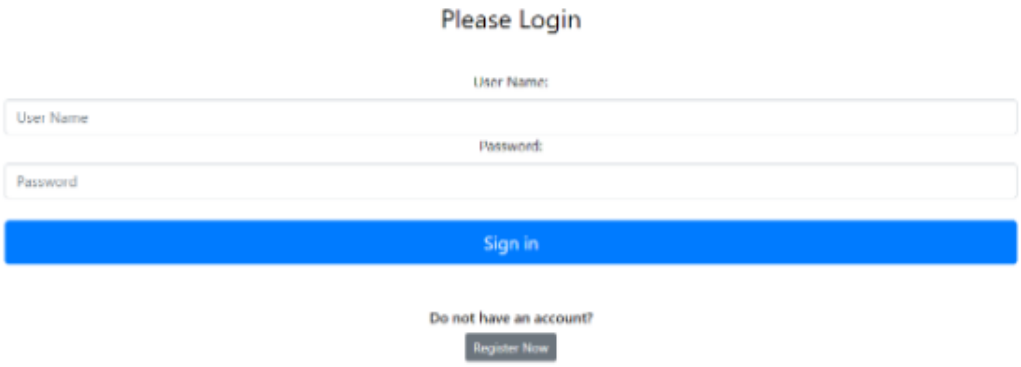
To register, the new user provides the following information: their name, email, and password. Finally, if the email is valid and has not been used before, and if the password exceeds 6 characters, by pressing "Create Account," the new user is immediately added to the SQL database.



4.4.2 Login Page

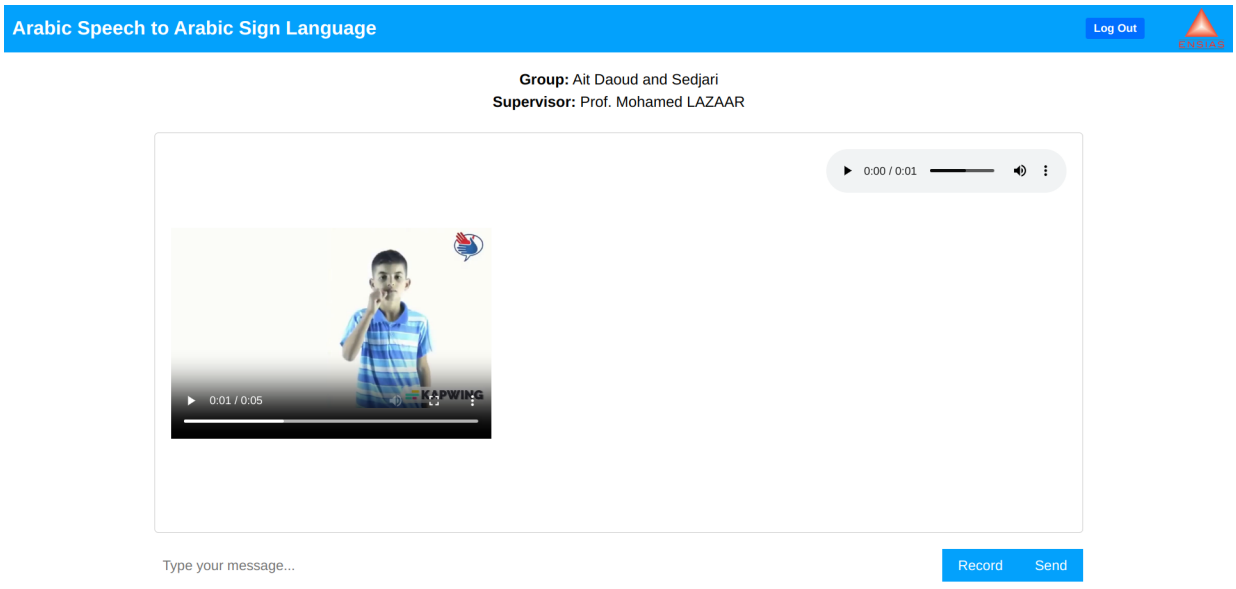
In this page, the user can authenticate to their account by providing their email and password, under the necessary condition of having an account already created. For security reasons, the

password must contain at least 6 characters. If the two entries match, the user is directly directed to their account, specifically to the conversation page.



4.4.3 Conversation Page

In the conversation page, the user can record audio by clicking on the record button. Clicking on the record button again stops the recording. Immediately and automatically, the audio is sent to the conversational agent. The user can click on the play button to listen to the audio and also download it. After sending the audio, within a few seconds, the system generates a video translation of the audio into Moroccan sign language. Similarly, the user can watch and download the video. Additionally, the user can directly write Arabic text and send it to be translated into video sign language. It's worth noting that there is also a logout button available for logging out of the system.



4.5 Conclusion

In this chapter, we have presented in detail the development of our system. We started by introducing the hardware and software environment, the technical choices, and the working tools on which our application is based. Finally, we concluded with the testing and validation scenarios of the application.

General Conclusion

In conclusion, this project has successfully addressed the challenge of developing a comprehensive system that enables Arabic speech to be translated into Moroccan Sign Language. Through a thorough analysis of requirements and the utilization of appropriate technologies, we have designed and implemented a robust architecture that encompasses various components such as Automatic Speech Recognition, Text Processing, and Video Translation. The project has showcased the effectiveness of combining machine learning algorithms, linguistic analysis, and multimedia processing to bridge the communication gap between spoken Arabic and sign language. Overall, this project represents a significant contribution towards enhancing accessibility and inclusivity for individuals with hearing impairments, and it paves the way for further advancements in the field of multimodal communication systems.

References

- Hemang Monga, Jatin Bhutani, Muskan Ahuja, Nikita Maida and Himangi Pande "Speech to Indian Sign Language Translator", a School of Computer Engineering an Technology, MIT World Peace University, Pune, India
- Alexis Conneau, Alexei Baevski, Ronan Collobert, Abdelrahman Mohamed, Michael Auli, "Unsupervised Cross-lingual Representation Learning for Speech Recognition" , arXiv:2006.13979v2 [cs.CL] 15 Dec 2020