

**Democratic and People's Republic of Algeria**  
**Ministry of Higher Education and Scientific Research**  
**University: Saad Dahleb, Blida 1**  
**Faculty: Science**  
**Department: Computer Science**



## **MASTER THESIS**

Option : Systèmes Informatiques et Réseaux

---

*Developing an application that detects vocal errors in  
the recitation of the holy Quran*

---

*By: :REKAB Romaissa*

*President: ZAHRA Fatima zohra*

*Supervisor: Ms. Malek ARKAM*

*Supervisor: YKHLEF Hadjer*

**November 2020**

---

## *Abstract*

---

The goal of this thesis is to facilitate the process of receiving the Noble Qur'an by creating a program that detects voice errors for the Qur'an reciter, this was done by means of a "web speech recognizer" which recognizes the reciter's reading as the first step and then converts it into a visible text for the user via the Speech-To-Text process. This text is compared with the original text of the Noble Qur'an which has been downloaded in the form of a database, after full compatibility between the two texts, the user is notified of his correct reading. In the event of a mistake in user recitation, the user's text differs from the original text, then he is notified of his incorrect reading.

After creating a web-based application, we searched for similar works. We found three works: "Tamkeen", "Read Along" Mobile Applications and "Web captioner". Each of them has its own goal, but they meet our application in at least one of its goals.

Finally, in this thesis, we experimented the effectiveness of our application through its 20 experiments, saving the results of these experiments, discussing them, and giving possible reasons for these results with proposing future solutions.

**Key words:** The Holy Quran, Speech-To-Text process, Arabic language, web-based application.

## *Résumé*

---

Le but de cette thèse est de faciliter le processus de réception du Noble Coran en créant une application qui détecte les erreurs de voix pour le récitant du Coran, cela a été fait au moyen d'un « logiciel de reconnaissance vocale » qui reconnaît la lecture du récitant comme la première étape puis le convertit en texte visible pour l'utilisateur via le processus Speech\_To\_Text. Ce texte est comparé au texte original du Noble Coran qui a été téléchargé sous forme de base de données, après compatibilité totale entre les deux textes, l'utilisateur est averti de sa lecture correcte. En cas d'erreur de récitation de l'utilisateur, le texte de l'utilisateur diffère du texte original, puis il est informé de sa lecture incorrecte.

Après avoir créé une application Web, nous avons recherché des travaux connexes. Nous avons trouvé trois travaux : « Tamkeen », « Read Along » Applications mobiles et « Web captioner ». Chacun d'eux a son propre objectif, mais ils répondent à notre application dans au moins un de ses objectifs.

Enfin, dans cette thèse, nous avons expérimenté l'efficacité de notre application à travers ses 20 expériences, en sauvegardant les résultats de ces expériences, en les discutant, et en donnant les raisons possibles de ces résultats en proposant des solutions futures.

**Mots clés :** Le Saint Coran, processus Speech-To-Text, langue arabe, application Web.

## ملخص

الهدف من هذه الأطروحة هو تسهيل عملية تلقي القرآن الكريم من خلال اختراع برنامج يكتشف الأخطاء الصوتية لقارئ القرآن الكريم، تم ذلك بواسطة "متعرف الكلام الخاص بالويب" والذي يقوم بالتعرف على قراءة القارئ كأول خطوة ثم يحولها إلى نص ظاهر للقارئ عبر عملية Speech\_To\_Text، يتم مقارنة هذا النص مع النص الأصلي للقرآن الكريم والذي تم تحميله على شكل قاعدة بيانات، بعد التوافق التام بين النصين يتم إخبار المستعمل بصحة قراءته. في حال خطأ التلاوة، يختلف نص المستعمل على النص الأصلي حينها يتم إخباره بخطأ في قراءته. بعد إنشاء تطبيق على شبكة الإنترنت، بحثنا عن أعمال مشابهة له. وجدنا ثلاثة أعمال: "تمكين" و "اقرأ على طول" تطبيقات الهاتف المحمول و "Web captioner". لكل واحد منهم هدف خاص به لكنهم يجتمعون مع تطبيقنا على الأقل في واحدة من أهدافه. وأخيرا قمنا في هذه الأطروحة بتجربة فعالية تطبيقنا من خلال تجربته ٢٠ تجربة وتسجيل نتائج هذه التجارب ومناقشتها وإعطاء أسباب محتملة لهذه النتائج مع اقتراح حلول مستقبلية.

**الكلمات المفتاحية:** القرآن الكريم، عملية تحويل الكلام إلى نص، اللغة العربية، التطبيق المستند إلى الويب.

---

## *Acknowledgement*

---

"At first, I thank Allah who helped me to realize this work, and who has been with me at every moment of our study path. To my dear relatives my mother and father for their Patience, their love, their support and their encouragement. Our sincere thanks to all the members of the jury who have done us the honor of reviewing this work.

A special thank you to M. ARKAM Malek for her patience, helps me and her advice to develop this work.

One Person I must mention for his help and patience in the development of this work is my husband HASSNAOUI Abd Allah, a big thanks for his support."

---

# Contents

---

<b>1</b>	<b>Introduction .....</b>	<b>1</b>
1.1	The holy Quran .....	2
1.2	Voice recognition.....	2
1.3	Objective .....	2
1.4	Thesis structure.....	3
<b>2</b>	<b>Chapter 1: background .....</b>	<b>4</b>
2.1	Introduction .....	5
2.2	Speech recognition process .....	5
2.2.1	Signals .....	5
2.2.2	Speech signal .....	5
2.2.3	Speech Pre-processing.....	6
2.2.4	Feature Extraction.....	6
2.2.4.1	MFCC .....	7
2.2.4.2	LPC.....	7
2.2.5	Speech Classification .....	7
2.2.5.1	HMM .....	7
2.2.5.2	DTW .....	8
2.2.5.3	VQ .....	8
2.2.6	ARTIFICIAL NEURAL NETWORK FROM THE VIEWPOINT OF SPEECH RECOGNITION.....	8
2.2.6.1	ANNs .....	8
2.2.6.2	Artificial Neuron .....	8
2.2.7	TYPES OF ARTIFICIAL NEURAL NETWORK.....	9
2.2.7.1	Feedforward Network.....	9
2.2.7.2	RNN.....	10
2.2.7.3	MNN.....	11
2.2.7.4	Kohonen Self Organizing Maps.....	11
2.2.8	ADVANTAGES OF ARTIFICIAL NEURAL NETWORK .....	11
2.3	ASR errors detection.....	12
2.3.1	Decoder based features.....	12
2.3.2	Non-decoder based features.....	13
2.3.3	The DSR errors choosing in the project: .....	13

2.4	Related work.....	16
2.4.1	Tamkeen mobile application .....	16
2.4.1.1	Comparison .....	16
2.4.1.2	Analysis .....	16
2.4.2	Read along mobile application .....	17
2.4.2.1	Comparison .....	17
2.4.2.2	Analysis .....	17
2.4.3	Web captioner.....	17
2.4.3.1	Comparison:.....	18
2.4.3.2	Analysis .....	18
2.5	Conclusion.....	18
<b>3</b>	<b>Chapter2: Conception .....</b>	<b>19</b>
3.1	Introduction .....	20
3.2	Define actors and jobs.....	20
3.3	Use case diagram .....	20
3.4	Sequence diagram.....	21
3.4.1	Principal page sequence diagram.....	21
3.4.2	Evaluation page sequence diagram .....	22
3.5	Class diagram .....	22
3.5.1	Shift to relational.....	23
3.6	Conclusion.....	24
<b>4</b>	<b>Chapter 3: Experiments and Results .....</b>	<b>25</b>
4.1	Introduction .....	26
4.2	Web-based application.....	26
4.3	Project building methodology .....	26
4.3.1	Choose an SR.....	26
4.3.2	Download Quran database.....	27
4.3.3	Errors detection.....	27
4.4	Tools .....	27
4.5	Evaluation of memorization and recitation of the holy Quran.....	29
4.5.1	Evaluation of memorization .....	29
4.5.1.1	Case 1: Recitation is true .....	30
4.5.1.2	Case 2: Recitation is false .....	32
4.5.1.3	Case 3: No recitation to evaluate it.....	33
4.5.2	Evaluation of recitation .....	34
4.6	Experiments.....	35
4.7	Performance.....	37
4.8	Discussion .....	39

4.9	Conclusion.....	39
<b>5</b>	<b>Conclusion.....</b>	<b>40</b>
<b>6</b>	<b>References.....</b>	<b>42</b>

## List of Figures

FIGURE 2.1 - EXAMPLE OF A RECORDING OF SPEECH. THE TIME WAVEFORM SHOWS THE VARIATION OF ACOUSTIC PRESSURE AS A FUNCTION $S(T)$ OF TIME FOR THE WORD “SIGNAL” [12] .....	6
FIGURE 2.2 - PROCESS OF SPEECH RECOGNITION.....	6
FIGURE 2.3 - BASIC ARTIFICIAL NEURON [2].....	9
FIGURE 2.4 - A FULLY CONNECTED FEEDFORWARD WITH ONE HIDDEN LAYER AND ONE OUTPUT LAYER. [2] .....	10
FIGURE 2.5 - STRUCTURE OF A RECURRENT NEURAL NETWORK.....	10
FIGURE 2.6 - MODULE NEURAL NETWORK ARCHITECTURE. [2] .....	11
FIGURE 3.1 - USE CASE DIAGRAM.....	20
FIGURE 3.2 - PRINCIPAL PAGE SEQUENCE DIAGRAM.....	21
FIGURE 3.3 - EVALUATION PAGE SEQUENCE DIAGRAM .....	22
FIGURE 3.4 - CLASS DIAGRAM .....	23
FIGURE 4.1 - PRINCIPAL PAGE SCREENSHOT WITH PORT NUMBER ILLUSTRATION .....	28
FIGURE 4.2 - NEW USER PAGE.....	28
FIGURE 4.3 - PRINCIPAL PAGE SCREENSHOT WITH USER DATA.....	29
FIGURE 4.4 - EVALUATION PAGE SCREENSHOT WITH USER DATA.....	30
FIGURE 4.5 - EVALUATION PAGE SCREENSHOT FOR MIC REQUEST .....	30
FIGURE 4.6 - SCREENSHOT FOR EVALUATION PAGE – TRANSCRIPTION CASE TRUE & VERSE 1 RESULT - .....	30
FIGURE 4.7 - SCREENSHOT FOR EVALUATION PAGE - TRANSCRIPTION CASE FALSE - .....	32
FIGURE 4.8 - SCREENSHOT FOR EVALUATION PAGE -VERSE 3 FALSE- .....	33
FIGURE 4.9 - SCREENSHOT FOR EVALUATION PAGE -VERSES AND USER RECITATION- .....	33
FIGURE 4.10 - SCREENSHOT FOR EVALUATION PAGE – USER ALERT FOR LACK OF RECITATION TO EVALUATE- .....	33
FIGURE 4.11 – SYSTEM OVERVIEW.....	34
FIGURE 4.12 - EVALUATE RECITATION BUTTON.....	35



FIGURE 4.13 - THE RELATIONSHIP BETWEEN THE NUMBER OF VERSES AND THE ACCEPTANCE RATION  
OF OUR APPLICATION DURING PREVIOUS EXPERIENCES .....38

## List of Tables

TABLE 2.2 - COMPARISON BETWEEN TAMKEEN AND OUR APPLICATION .....16  
TABLE 2.3 - COMPARISON BETWEEN READ ALONG AND OUR APPLICATION .....17  
TABLE 2.4 - COMPARISON BETWEEN WEB CAPTIONER AND OUR APPLICATION .....18  
TABLE 3.1 - USER'S JOBS .....20  
TABLE 4.1 - LIST OF VERSES INCLUDED IN THE TEST .....36  
TABLE 4.2 - PERFORMANCE RESULT .....37

---

## *Notation and Acronyms*

---

**ANNs:** Artificial Neural Networks.

**ASR:** Automatus Speech Recognition

**DSR:** Dictation Speech Recognition.

**GUI:** Graphic User Interface.

**HMM:** Hidden Markov Model.

**IA:** Intelligence Artificial.

**NB:** Naïve Bayes.

**NN:** Neural Network.

**SR:** Speech recognition

**STT:** Speech To Text.

**SVM:** Support Vector Machine.

**VQ:** Vector Quantization

**DTW:** Dynamic Time Warping

**MFCC:** Mel Frequency Cepstrum Coefficients

**RNN:** Recurrent Neural Network A Recurrent Neural Network

**LPC:** Linear Predictive Coding

**MNN:** Modular Neural Network A Modular Neural Network

---

## *Introduction*

---

### **1.1 The holy Quran**

The Holy Quran is the word of God that descended on the heart of the Prophet Muhammad (pbuh), Who taught his companions the Qur'an as he received it from his Lord. In turn, the companions (Sahaba) taught the followers, who in turn taught their followers...

This precise method, called receiving and sharing (التلقي والمشافهة), has brought us the Holy Quran without misrepresentation or misinformation, this precise method occurs in only one language, Arabic. This language chosen by God to send his message to the world is one of the 10 most widespread languages in the world. Despite this, it has a lot of technological deficiencies such as (voice recognition, artificial intelligence...) which forces us to develop and improve new applications that interact with this rich language. Therefore, it has become important to design applications capable of analyzing, detecting and extracting errors and malfunctions from the Arabic speech that are entered by the user.

### **1.2 Voice recognition**

As speech signals are non-stationary in nature, speech recognition is a complex task due to the differences in gender, emotional state, accent, pronunciation, articulation, nasality, pitch, volume, and speed variability in people speak. Presence of background noise and other types of disturbances also makes a speech processing system complex and difficult. The performance of a speech processing system is usually measured in terms of recognition accuracy. Speech processing is useful for various applications such as like mobile applications, weather forecasting, agriculture, healthcare, automatic translation, robotics, video games, transcription, audio and video database search, household applications and language learning applications etc. [19]

### **1.3 Objective**

In this project, the objective is to facilitate the process of memorizing the Holy Qur'an by creating an application that detect errors in user recitation of the Noble Qur'an using the speech recognition.

## **1.4 Thesis structure**

This thesis consists of **two primary parts**. The first part covers the state-of-the-art notions that are necessary for understanding the ideas developed in this thesis. In the **Chapter 1** we present the basic knowledge in Arabic speech recognition and related works, While **The second chapter** will be a Conception and models used in this project. The second half of this thesis describes the methodology that we have followed for created this application. We give in the **Chapter 3** a description of our application, tests due in this project and results obtained with its discussion, we present the obtained results through performance table.

Finally, we conclude by summarizing the contributions of this thesis, the lines of limitations and future work.

---

*Chapter 1: background*

---

## 2.1 Introduction

This chapter will present a fundamentals knowledge that let us understand more this project. Such as our project includes two parts, so we will address each part separately, starting with the speech recognition process included signals, speech signal, speech pre-processing, Feature extraction and software methods to realize speech recognition, passing by ASR errors detection, and ending with related work.

## 2.2 Speech recognition process

Speech recognition, also known as automatic speech recognition (ASR), computer speech recognition, or speech-to-text, is a capability which enables a program to process human speech into a written format. While it's commonly confused with voice recognition, speech recognition focuses on the translation of speech from a verbal format to a text one whereas voice recognition just seeks to identify an individual user's voice. [14]

To understand more ASR, it will be necessary to study the basis of this software, we will see the meaning of the word "signal" by the following

### 2.2.1 Signals

**Signal** is defined as any physical quantity that varies as a function of time, space, or any other variable or variables. Signals convey information in their patterns of variation. The manipulation of this information involves the acquisition, storage, transmission, and transformation of signals. [12]

In the nature there are many signals can be studies, in our case we are interested to speech signal.

### 2.2.2 Speech signal

The speech signal, shown as a *time waveform* in Figure 1.1, represents the variations of acoustic pressure converted into an electric signal by a microphone. We note that different sounds correspond to different patterns of temporal pressure variation. [12]

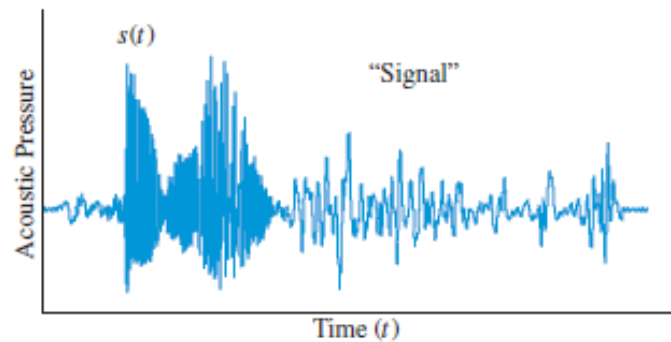


Figure 2.1 - Example of a recording of speech. The time waveform shows the variation of acoustic pressure as a function  $s(t)$  of time for the word “signal” [12]

### 2.2.3 Speech Pre-processing

Speech pre-processing is intended to solve such problems. This plays an important role in eliminating the irrelevant sources of variation. It ultimately improves the accuracy of speech recognition. The speech pre-processing generally involves noise filtering, smoothing, end point detection, framing, windowing, reverberation cancelling and echo removing. [2]

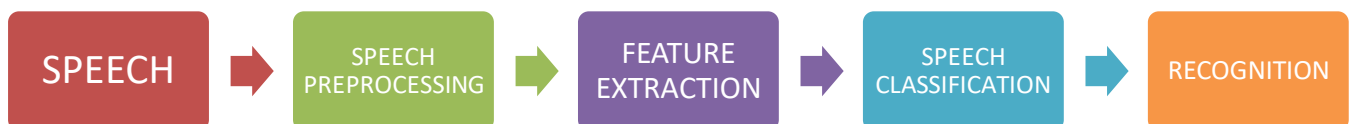


Figure 2.2 - Process of Speech Recognition

### 2.2.4 Feature Extraction

The speech varies from person-to-person. This is due to the fact that every person has different characteristics embedded in utterance. Theoretically, possibility should be there to recognize speech from the digitized waveform. But due to the large variation in speech signal, there arise a need to perform some feature extraction to reduce that variations. The following section



summarizes some of the feature extraction technologies that are in use nowadays. These techniques are also useful in other areas of speech processing. [2]

### **2.2.4.1 MFCC**

is the most prominent method used in the process of feature extraction in speech recognition. It is based on the frequency domain which is based on Mel scale based on human ear scale. MFCCs, being frequency domain features, are more accurate than time domain features [8]. MFCC represents the real cepstral of windowed short time signal which is derived from Fast Fourier Transform (FFT). These coefficients are robust and reliable for variations of speaker and operation environment. [2]

### **2.2.4.2 LPC**

is a tool most widely used for medium or low bit rate coder. Digital signal is compressed for efficient transmission and storage. Computation of parametric model based on least mean squared error theory is known as linear prediction (LP). The signal is expressed as a linear combination of previous samples. Formant frequencies are the frequencies where resonance peak occurs. [2]

## **2.2.5 Speech Classification**

The most common techniques used for speech classification are discussed in short. These systems involve complex mathematical functions and they take out hidden information from the input processed signal. [2]

### **2.2.5.1 HMM**

is the most successfully used pattern recognition technique for speech recognition. It is a mathematical model signalized on the Markov Model and a set of output distribution. This technique is more general and has a secure mathematical foundation as compared to knowledge based approach and template based approach. In this method, speech is split into smaller audible entities and these entities represent a state in the Markov Model. According to the probabilities of transition, there exists a transition from one state to another. [2]

### **2.2.5.2 DTW**

Is a technique compares words with reference words. It is an algorithm to measure the similarity between two sequences that can vary in time or speed [11]. In this technique, the time dimensions of the unknown words are changed until they match with that of the reference word. [2]

### **2.2.5.3 VQ**

Is a technique in which the mapping of vector is performed from a large vector space to a finite number of region in that space. This technique is based on block coding principle. Each region is called as cluster and can be represented by its centre known as a code-word. Code book is the collection of all code-words. [2]

## **2.2.6 ARTIFICIAL NEURAL NETWORK FROM THE VIEWPOINT OF SPEECH RECOGNITION**

### **2.2.6.1 ANNs**

Are nothing but the crude electronic models based on neural structure of brain. The human brain basically learns from the experiences. It is a fact that some problems which are beyond the scope of current computers can be are easily solvable by energy efficient packages. Such type if brain modelling also provides a less technical path for the development of machine solution. ANN are computer having their architecture modelled after the brain. They mainly involve hundreds of simple processing units wired together in complex communication network. Each simple processing unit represents a real neuron which sends off a new signal or fires if it receives a strong signal from the other connected unit. [2]

### **2.2.6.2 Artificial Neuron**

Artificial Neurons are the basic unit of Artificial Neural Network which simulates the four basic function of biological neuron. It is a mathematical function conceived as a model of natural neuron. The following figure shows the basic artificial neuron. [2]

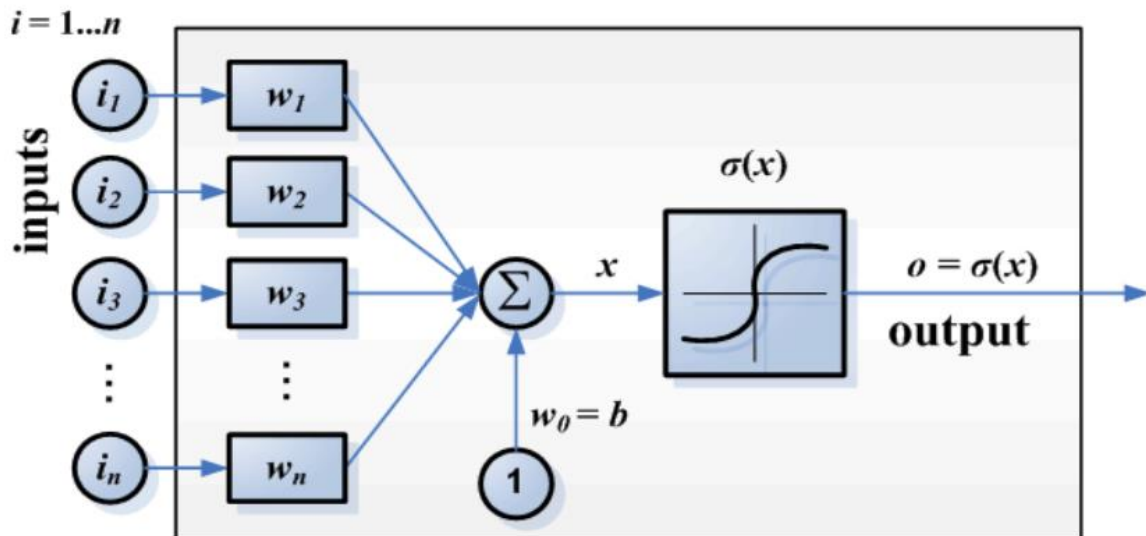


Figure 2.3 - Basic Artificial Neuron [2]

In this figure, various inputs are shown by the mathematical symbol,  $i(n)$ . Each of these inputs are multiplied by connecting weights  $w(n)$ . Generally, these products are simply summed and fed to the transfer function to generate the output results. The applications like text recognition and speech recognition are required to turn these real world inputs into discrete values. These applications don't always utilize networks composed of neurons that simply sum, and thereby smooth, inputs. In the software packages, these neurons are called as processing elements and have many more capabilities than the basic artificial neuron described above. [2]

## 2.2.7 TYPES OF ARTIFICIAL NEURAL NETWORK

Researchers from the world have found out countless different structures of Artificial Neural Network. Short descriptions of each are given below. [2]

### 2.2.7.1 Feedforward Network

Is the first and the simplest form of ANN. In this network, the information flows only in one i.e. forward direction from input node via hidden nodes to the output node. This network contains no loops or cycles. A neuron in layer 'a' can only send data to a neuron in layer 'b' if  $b > a$ . Learning is the adaptation of free parameters of neural network through a continuous process of stimulation by the embedded environment. Learning with a teacher is called as (a) supervised training; and learning without a teacher is called as (b) unsupervised training. The back-propagation algorithm has emerged to design the new class of layered feedforward network called as Multi-Layer Perceptrons

(MLP). It generally contains at least two layers of perceptrons. It has one input layer, one or more hidden layers and output layers. The hidden layer plays very important role and acts as a feature extractor. It uses a nonlinear function such as sigmoid or a radial-basis to generate complex

functions of input. To minimize classification error, the output layer acts as a logical net which chooses an index to send to the output on the basis of input it receives from the hidden layer.

[2]

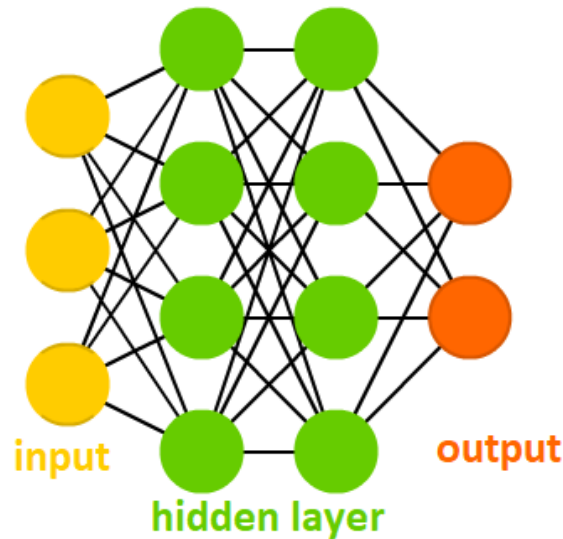


Figure 2.4 - A Fully Connected Feedforward with Two Hidden Layer and One Output Layer. [3]

### 2.2.7.2 RNN

Is a neural network that operates in time. RNN accepts an input vector, updates its hidden state via non-linear activation function and uses it to make prediction on output. In this network, the output of the neuron is multiplied by a weight and fed back to the inputs of neuron with delay. RNN have achieved better speech recognition rates than MLP, but the training algorithm is again more complex and dynamically sensitive, which can cause problems [15].

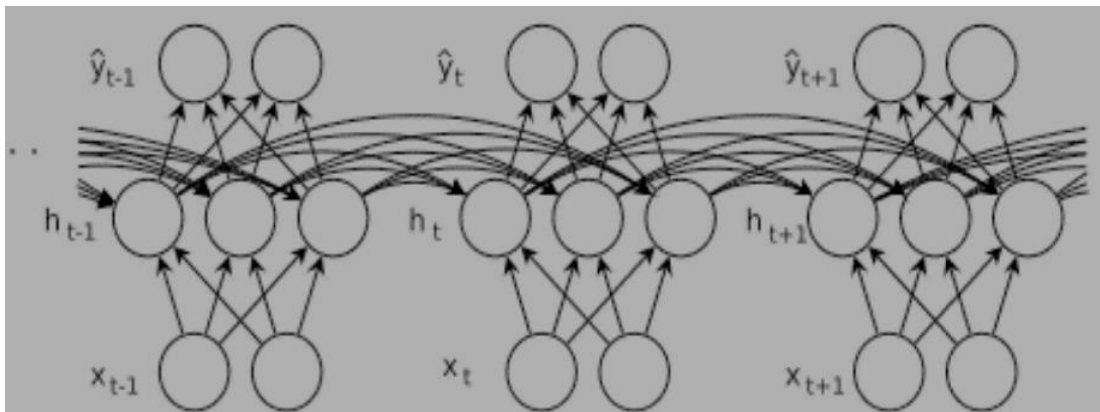


Figure 2.5 - Structure of A Recurrent Neural Network.

### 2.2.7.3 MNN

Consist of several modules, each module carrying out one sub task of the neural network's global task, and all module functionally embedded. The global task can be any NN application, e.g., mapping, clustering, function approximation or associative memory application [16].

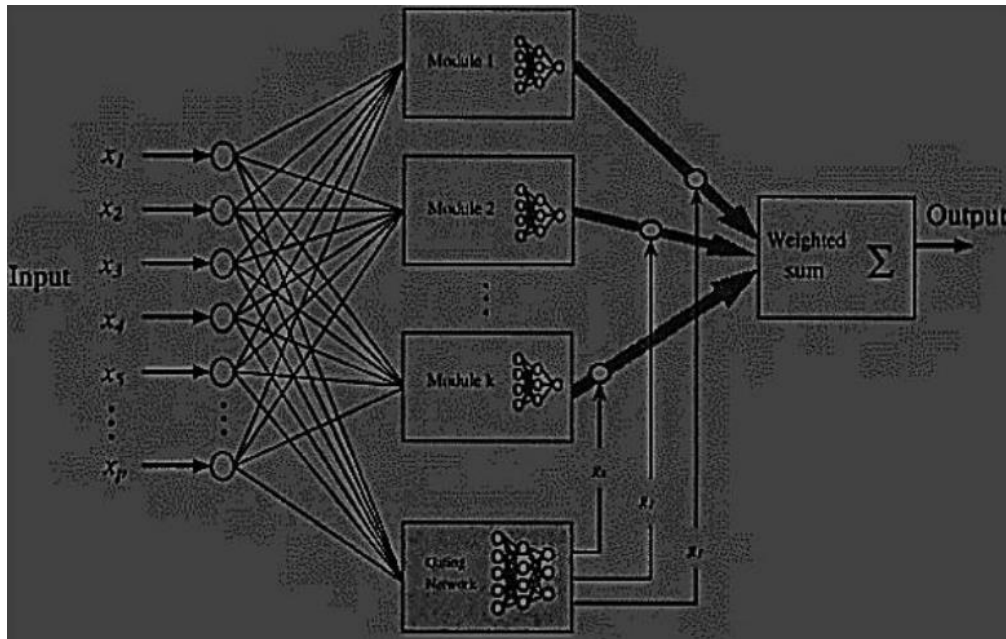


Figure 2.6 - Module Neural Network Architecture. [2]

### 2.2.7.4 Kohonen Self Organizing Maps

Kohonen self-organizing maps are a type of neural network. They require no supervision and hence called as 'Selforganizing'. They learn on their own unsupervised competitive learning. They are called as 'Maps' because they attempt to map their weight to conform to the given input data [17].

## 2.2.8 ADVANTAGES OF ARTIFICIAL NEURAL NETWORK

- ANN have the ability to learn how to do task based on the data given for training, learning and initial experience.
- ANN can create their own organisation and require no supervision as they can learn on their own unsupervised competitive learning.
- Computations of ANN can be carried out in parallel.

- ANN can be used in pattern recognition which is a powerful technique for harnessing the data and generalizing about it.

## 2.3 ASR errors detection

The goal of errors detection is to determine whether an error has occurred in the transcription using features generated from the ASR system, such as confidence scores, language model, and confusion network density. Those features will be used later to classify a hypothesis word to two possible classes, either a correct word or an error. There are two categories of research that addressed the subject of errors detection in ASR systems: the first one focused on features generated from the ASR decoder, such as confidence scores, linguistic information, and confusion networks, and the second one used additional features generated from hypothesized word sequence, such as n-grams, parts of speech, syntactic features, and semantic features.[15]

### 2.3.1 Decoder based features

In this category of research there are four study are mentioned:

- First study addressed the issue of errors detection in ASR, especially in DSR, by using data-mining techniques. This study consists of using three different data-mining classifiers, including NB, NN, and SVM for detecting errors in DSR. The three models were trained to identify errors using features extracted from DSR output, including confidence scores and linguistic information. Results of this study have shown that those systems could **identify until 50% of output errors**.[15]
- Another study proposed the use of additional features extracted from the confusion networks, and estimated a correctness probability using logistic regression based on those features. The proposed system achieved a **classification error rate of 12.3%** on a French broadcast news corpus. [15]
- The third study investigated the use of a Markov Chains (MC) classifier with two states: error state and correct state, to model errors using a set of 15 common features in errors detection. The resulted system was tested on American English broadcast news speech

NIST corpus, and has achieved 860 errors correctly detected with only **16.7% classification error rate**. [15]

- Fourth study proposed a system for errors detection in conversational spoken languages translation, in addition to traditional features obtained from ASR outputs. This system used additional features provided as the feedback of Statistical Machine Translation (SMT), including SMT confidence estimates and posteriors from named entity detection (NED). Furthermore, this system used an automated word boundary detector based on acoustic-prosodic features to verify the existence of ASR-hypothesized word boundaries, in order to improve the ASR errors detection. This system provided **2.8% absolute improvement in error detection** over a simple error detector based on features traditionally employed in the literature. [15]

### 2.3.2 Non-decoder based features

In this category of research, the use of non-decoder based features are suggested, extracted from other sources different than the ASR decoder, in addition to the traditional decoder-based features. A binary word match feature that presents a binary comparison between two different ASR systems, bigram hit feature measuring the number of hits found by querying a very popular Web search engine, and a topic feature to identify if a word is out of the global topic of the hypothesized sentence. The introduction of this non-decoder-based features led to significant improvements, from **13.87% to 12.16% classification error rate** with a maximum entropy model, and from **14.01% to 12.39% classification error rate** with linear-chain conditional random fields, comparing to a baseline using only decoder-based features. [15]

### 2.3.3 The DSR errors choosing in the project:

We choose for our application the second category of research “Non-decoder based features” which use a binary word match feature.

**Justification:** In view of the development that Non-decoder based features has achieved compared to decoder based features we decided to choose it to be the errors detector for our project.

And this is the pseudo-algorithm that we follow to establish this step:

```
transcript = Array of user recitation
Ayah = Array of verses choosing by user from Quran db
j= the length of transcript array
normalize_text(String) :function that eliminate special
characters in the Ayah Array
trim(): function that eliminate white space in the Ayah Array
```

```
function evaluation() {

if(j==0) alert("لا يوجد شيء لتقييمه")

else{
    for (var i = 0; i < Ayah.length; i++) {

        var bar=normalize_text(Ayah[i].textContent).trim()
        var myr=transcript[i].toString()
        var bol =bar==String(myr).trim();

        if (bol) {

            alert("الآية["+i+1+"]: صحيحة")

        }

        else {

            alert("الآية["+i+1+"]: خاطئة ")

        }

    }

}
```



```
    }  
  }  }}
```

During our work on this project, we faced several challenges, including tashkeel in the Noble Qur'an, which is considered as a character, while that SR system does not recognize this tashkeel from user recitation, which we suggest to be a topic for future research. This challenge led us to learn about an `normalize_text(Ayah[i].textContent)` which is a function that remove all unwanted characters:

```
normalize_text = function(text) {  
  
  //remove special characters  
  text = text.replace(/([\u0621-\u063A\u0641-\u064A\u0660-\u0669a-zA-Z 0-9])/g, '');  
  
  //normalize Arabic  
  text = text.replace(/(|!|!)/g, '');  
  text = text.replace(/(°)/g, '°');  
  
  //convert arabic numerals to english counterparts.  
  var starter = 0x660;  
  for (var i = 0; i < 10; i++) {  
    text.replace(String.fromCharCode(starter + i), String.fromCharCode(48 + i));  
  }  
  
  return text;  
}
```

## 2.4 Related work

We will pass now to works related to our application, we have searched a lot for works corresponding to one of our application goals, we have found three works: “Tamkeen” and “Read along” mobile applications and web captioneer, the first one makes memorizing the Holy Quran to the user easily using speech recognition while the second one teaches the children how to pronounce their language using the speech recognition, the last one makes the event, speech, classroom lecture, or church service accessible with real-time captioning, all of them include Arabic language.

### 2.4.1 Tamkeen mobile application

Tamkeen is a volunteer youth initiative created to help anyone who wants to memorize the Noble Qur'an and review its memorization. "Tamkeen" is a mobile application under construction that provides a memorization schedule and a personal electronic audio that listens to the reader and alerts him to his mistakes, draws results and plans the student's level and progress.[4]

#### 2.4.1.1 Comparison

Table below presents a comparison between Tamkeen and our application:

<b>Criteria</b>	<b>Tamkeen</b>	<b>Our application</b>
<b>Application type</b>	mobile application	Web based
<b>Processed data</b>	Quran text	Quran text
<b>Data processed by</b>	Word	Verses
<b>Recognition used by</b>	Google API	Web speech API

Table 2.1 - comparison between Tamkeen and our application

#### 2.4.1.2 Analysis

This application has a lot of advantages, one of which is that it shows the Qur'anic words that the reader reads correctly gradually, while if you read them wrongly does not show any Qur'anic text and leaves the Qur'anic page white.

One of the drawbacks of this application is that it considers the word correct even though its tashkeel is wrong.

## 2.4.2 Read along mobile application

Read Along is a free reading app for Android that helps children have fun while they learn to read, this app has an in-app reading buddy that listens to your young learner read aloud, offers assistance when they struggle and rewards them with stars when they do well – guiding them along as they progress. It works best for children who already have some basic knowledge of the alphabet.

After the initial download, the app works securely offline. Inspire a love of reading in young learners. Languages available: With Read Along, children can read a variety of fun and engaging stories in different 9 languages including [5] + Arabic is recently added.

### 2.4.2.1 Comparison

Table below present a comparison between read along and our application:

<b>Criteria</b>	<b>Read along</b>	<b>Our application</b>
<b>Application type</b>	mobile application	Web based
<b>Processed data</b>	Kids stories	Quran text
<b>Data processed by</b>	Word	Verses
<b>Recognition used by</b>	Google API	Web speech API

Table 2.2 - comparison between Read along and our application

### 2.4.2.2 Analysis

This application has many good features including that it teaches the child to read correctly through play, another feature is that it contains many languages which makes it a global application. The negativity of the tashkeel is present strongly in this application with Arabic language, which leads to the wrong education of the Arab child.

## 2.4.3 Web captioner

Web Captioner is a web-based speech recognition tool that displays a real-time transcription in your browser. It's free and works with a microphone and a PC, Mac or Chromebook running Google Chrome. It relies on Google's implementation of the Web Speech API, which provides fast and accurate speech recognition. For many, it works with equipment you already have. [17]

### 2.4.3.1 Comparison:

Table below present a comparison between web captioner and our application:

<b>Criteria</b>	<b>Web captioner</b>	<b>Our application</b>
<b>Application type</b>	Web based	Web based
<b>Processed data</b>	transcript	Quran text
<b>Data processed by</b>	Word	Verses
<b>Recognition used by</b>	Web speech API	Web speech API

Table 2.3 - comparison between Web captioner and our application

### 2.4.3.2 Analysis

This app is considered the closest to our application in terms of work because it web based application and use a web speech API but it is far the most in terms of the goal because it does not correct any error said it only displays what is said

## 2.5 Conclusion

In this chapter the fundamentals are discussed, such as signal, speech pre-processing, feature extraction, classification, ANNs mentioning their types and cite some of its advantages.

Passing to ASR errors detection with decoder based features or non-decoder based features, given the several approaches that included above this two category with its results, and we explained our category chosen with a justification.

Finally, we present related works that we found in our research which are Tamkeen and Read along mobile applications and web captioner, these three applications are compared with our applications in some criteria and analysis each one of them.

---

## *Chapter 2: Conception*

---

### 3.1 Introduction

In our curriculum we learned that a good application means a good conception, there are three important diagrams that allow us to build a consistent application, which are: use case diagrams, class diagrams, sequence diagrams.

### 3.2 Define actors and jobs

Actor is a person who have a precise job in our application, more explanation in this table below:

Actor	Jobs
User	<ul style="list-style-type: none"> <li>• Identification</li> <li>• Recitation recording</li> <li>• Evaluate recitation</li> </ul>

Table 3.1 - User's jobs

### 3.3 Use case diagram

Figure 3.1 explain more user's job:

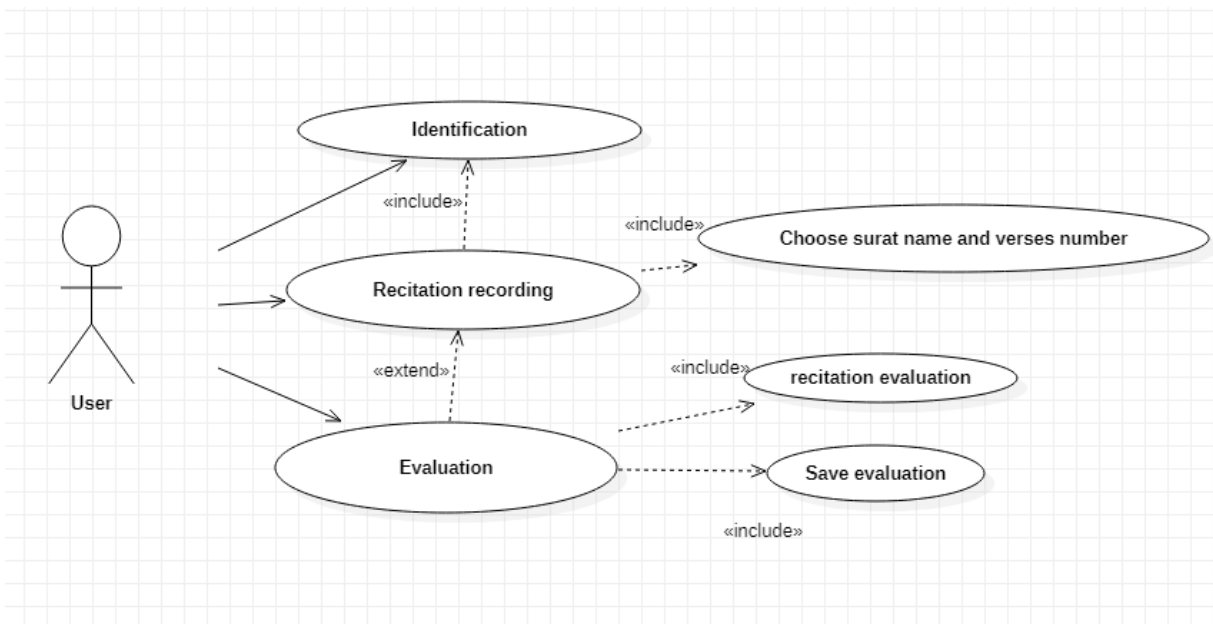


Figure 3.1 - Use case Diagram

### 3.4 Sequence diagram

Sequence diagram represent our application scenario, we divide our application into two scenes: principal page and evaluation page.

#### 3.4.1 Principal page sequence diagram

First, the principal page sequence diagram is shown in the following Figure :

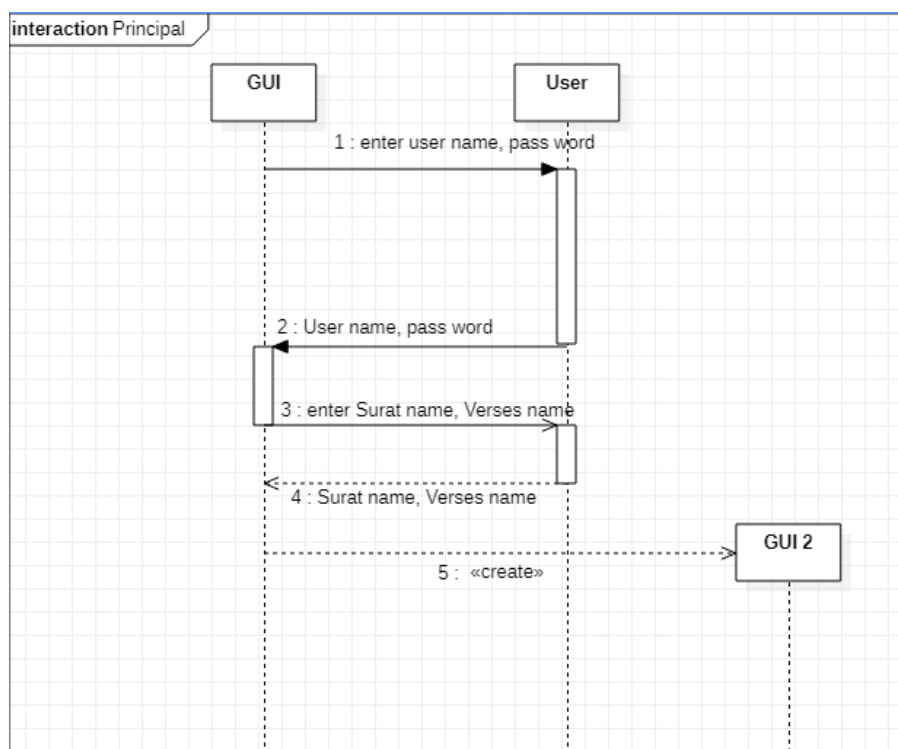


Figure 3.2 - Principal page Sequence diagram

### 3.4.2 Evaluation page sequence diagram

Secondly, the evaluation page sequence diagram is shown in the following Figure :

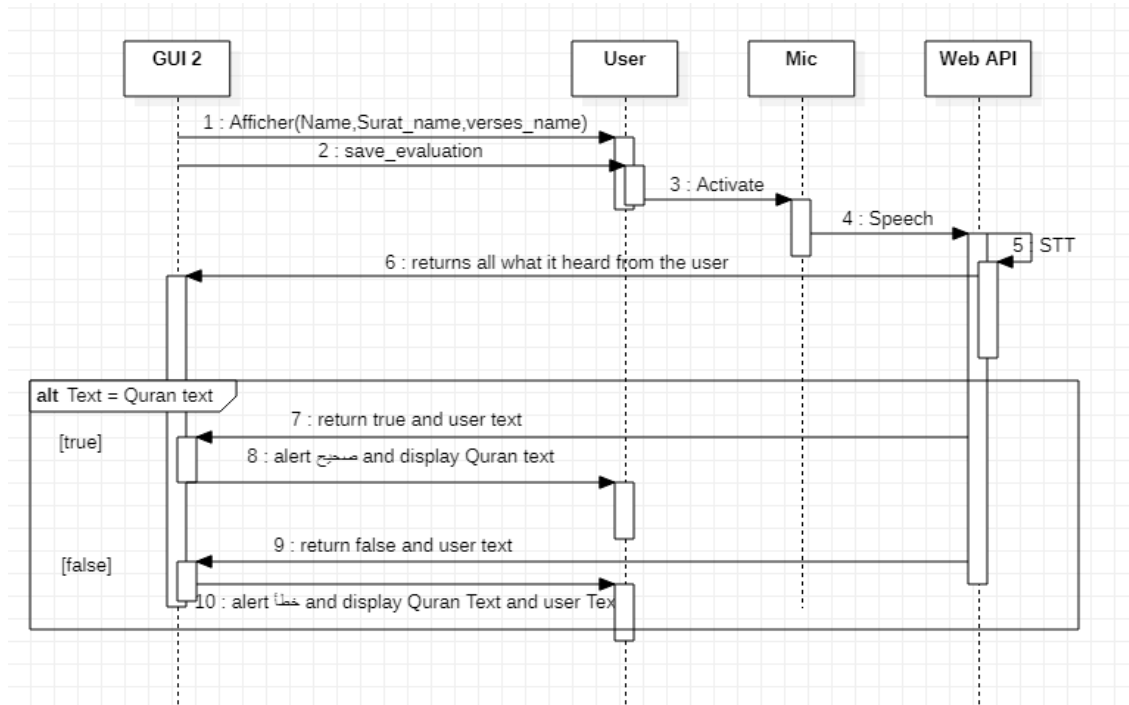


Figure 3.3 - Evaluation page Sequence diagram

### 3.5 Class diagram

Class diagram is a static diagram. It represents the static view of our application, more details in Figure:



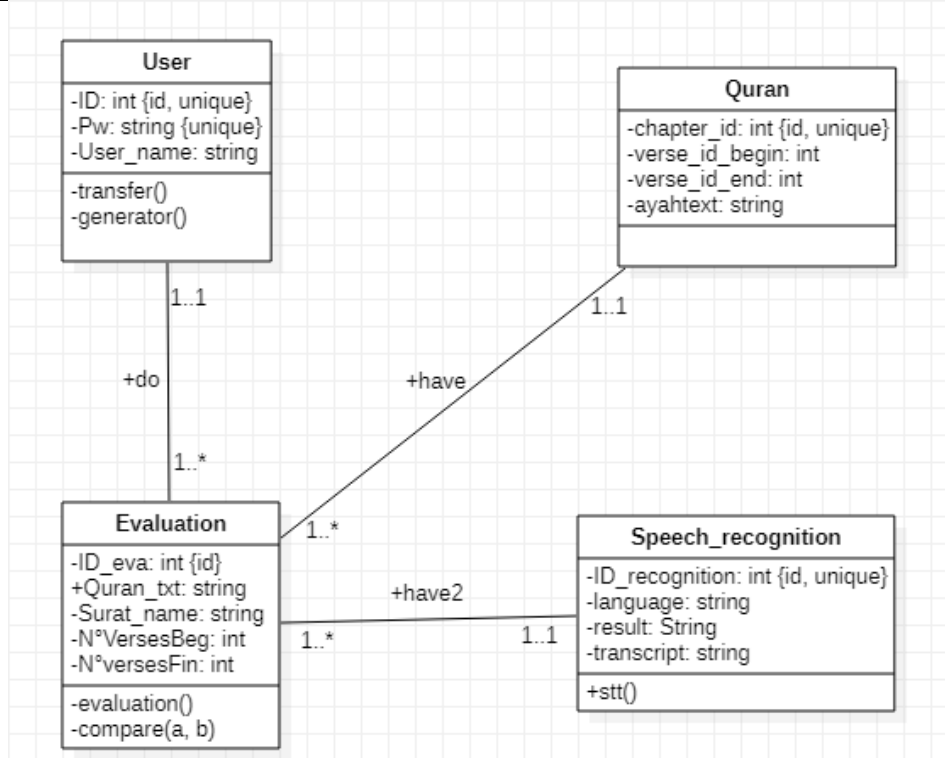


Figure 3.4 - Class diagram

This diagram contains 4 classes:

- User class represents the identification part.
- Evaluation class that owns an evaluate () function that evaluate user recitation (transcript) according to the holy Quran (quran\_txt) with using a compare (a, b) function that comparing two text.
- Speech recognition class that listens to user speech and transform it to text
- Quran class that contains all the holy Quran text (database)

### 3.5.1 Shift to relational

User (ID: int, Pw: string, User\_name: string, # ID\_eva: int)

Evaluation (ID\_eva: int, Quran\_txt: string, Surat\_name: string, N°VersesBeg: int, N°versesFin: int)

Quran (chapter\_id: int, verse\_id\_begin: int, verse\_id\_end: int, ayahtext: string, #ID\_eva: int)

Speech\_recognition (ID\_recognition: int, language: string, result: String, transcript: string, #ID\_eva: int)

### **3.6 Conclusion**

This chapter present the different interaction between user and system showing the job of each one, we use three several diagrams that permit the building of this project which are: use case diagram, and sequence diagram, and class diagram.

---

## *Chapter 3: Experiments & Results*

---

## **4.1 Introduction**

In this chapter we will describe the methodology that we have followed to create our application, which is composed from two parts. In the first part we present the experimental setup defined to evaluation of user memorization and his recitation of the holy Quran, whereas in the second part, we present the results of our experiments and we discussed them.

## **4.2 Web-based application**

We have chosen web-based application because it has many features unlike desktop app and mobile app, and there are some features that we found it in [13]:

- Web applications are published centralized in one place and the whole world can see them.
- Simultaneous work of a large number of users.
- Ability to work on multiple platforms.
- Low cost compared to the average number of users.
- Ease of use by end users.
- Centralized upgrade.

## **4.3 Project building methodology**

To build our project we follow three main steps, we will mention them in detail in the following:

### **4.3.1 Choose an SR**

For our application we choose as an ASR the web speech API which is according to [16] supported by five popular browsers Google Chrome, Mozilla Firefox, Microsoft Internet Explorer, Safari, and Opera in use.

### 4.3.2 Download Quran database

For a Quran text version, we need to download its database in order to have a reliable reference.

The user has the freedom to choose whatever verses he wants to recite and then the program will extract the selected verses from the database using a select query with a precise condition.

### 4.3.3 Errors detection

This step is the results of the two previous steps, we need the transcript result in 2.4.1 and the specific verses chosen by the user in 2.4.2 to compare between them and finally print the result.

## 4.4 Tools

We have created our web application using NodeJS 14.8.0 which is a JavaScript runtime built on Chrome's V8 JavaScript engine, and MySQL Database Service is a fully managed database service [11], we add to MySQL the holy Quran data base that contain all his verses [10], as Figure 4.1 presents the storage of Quran database in a localhost address and which be connected to our app when we start the server which has **5002** as a port number, we manage Quran database by using phpMyAdmin which is a free software, intended to handle the administration of MySQL over the Web [12].

First, as the Figure 4.2 shows we allowed the user to enter his name, his password and the name of the surah with the specification of the verses he wanted to test his preservation for it.

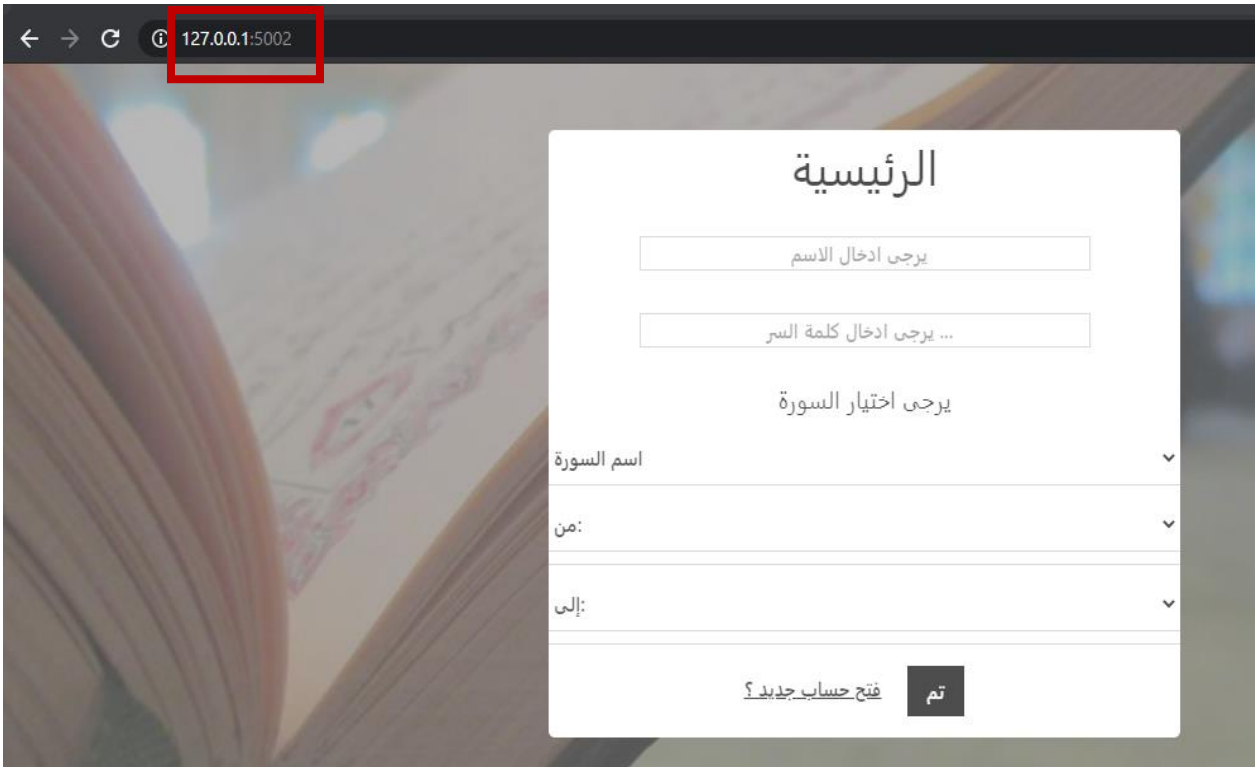


Figure 4.1 - Principal page Screenshot with port number illustration

When the link “فتح حساب جديد؟” is clicked, the server transfers the user to “*new user*” page, this page permits the user to subscribe in our website by entered his name, password and password confirm

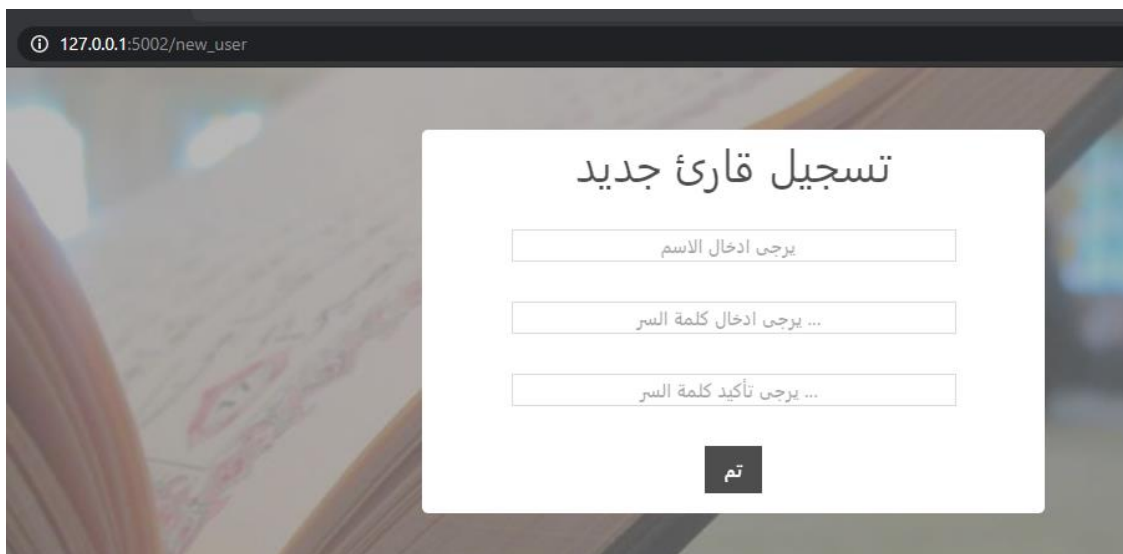


Figure 4.2 - New user page

When the user press "تم" he will be finished the subscribing step and he will be referred to "principal" page and his username and password are added to our database.

## 4.5 Evaluation of memorization and recitation of the holy Quran

After the subscribing to our application user can be access to his profile to evaluate his memorisation and recitation of the holy Quran.

### 4.5.1 Evaluation of memorization

When the button "تم" is clicked, the server transfer user data to "evaluation" page.

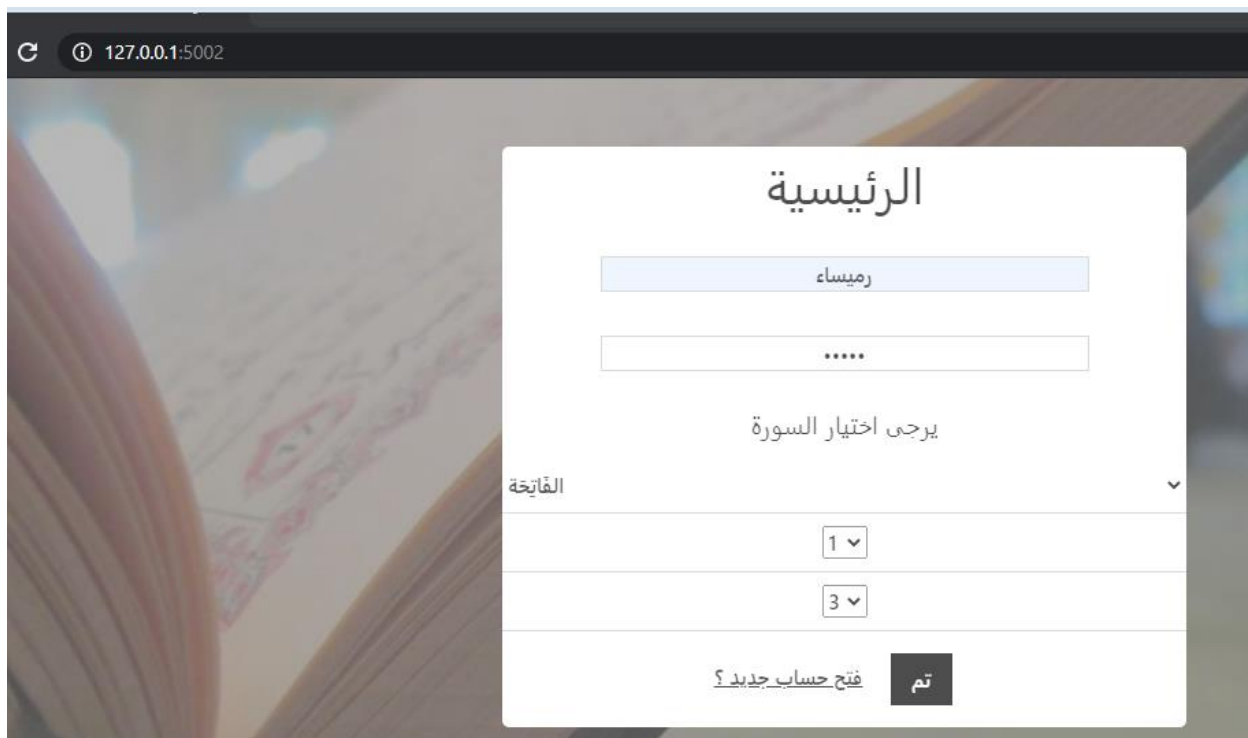


Figure 4.3 - Principal page Screenshot with user data

At this stage, As Figure 4.4 shows the user presses the "اختبار الحفظ" button and begins reciting the verses he has identified in phase 4.4 to begin the process of voice recognition that was explained in phase 1.2. In the coming we will take some screenshots to Quran user recitation (specifically Sura al-Fatiha 1-3) correctly and wrongly and the results of our application.

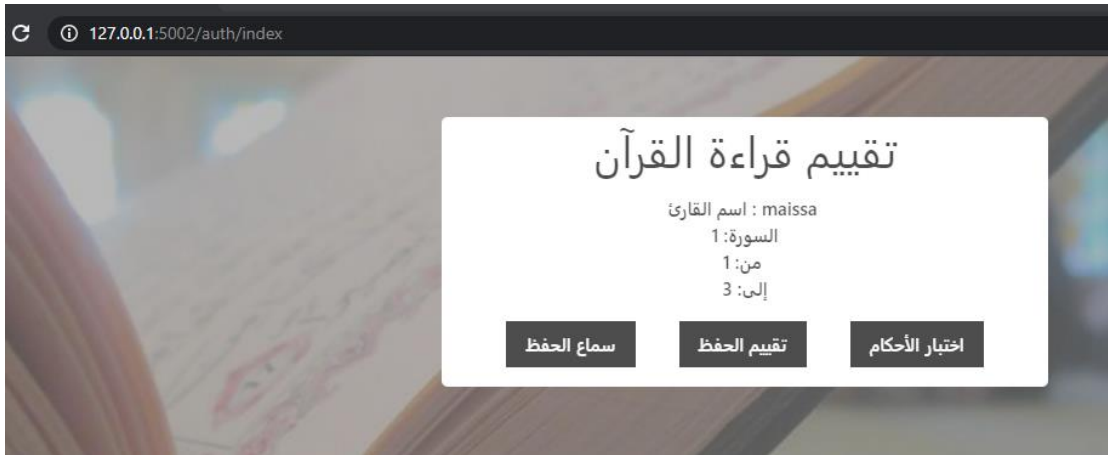


Figure 4.4 - Evaluation page Screenshot with user data

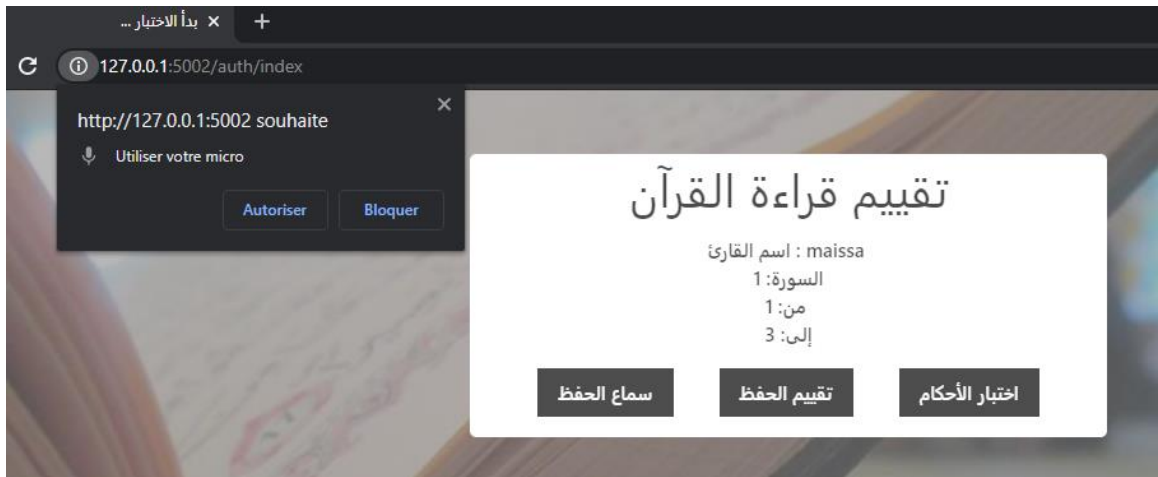


Figure 4.5 - Evaluation page Screenshot for mic request

#### 4.5.1.1 Case 1: Recitation is true



Figure 4.6 - Screenshot for evaluation page – transcription case true & verse 1 result -





Figure 4.7 - Screenshot for evaluation page -verse 2 result-



Figure 4.8 - Screenshot for evaluation page -verse 3 result-

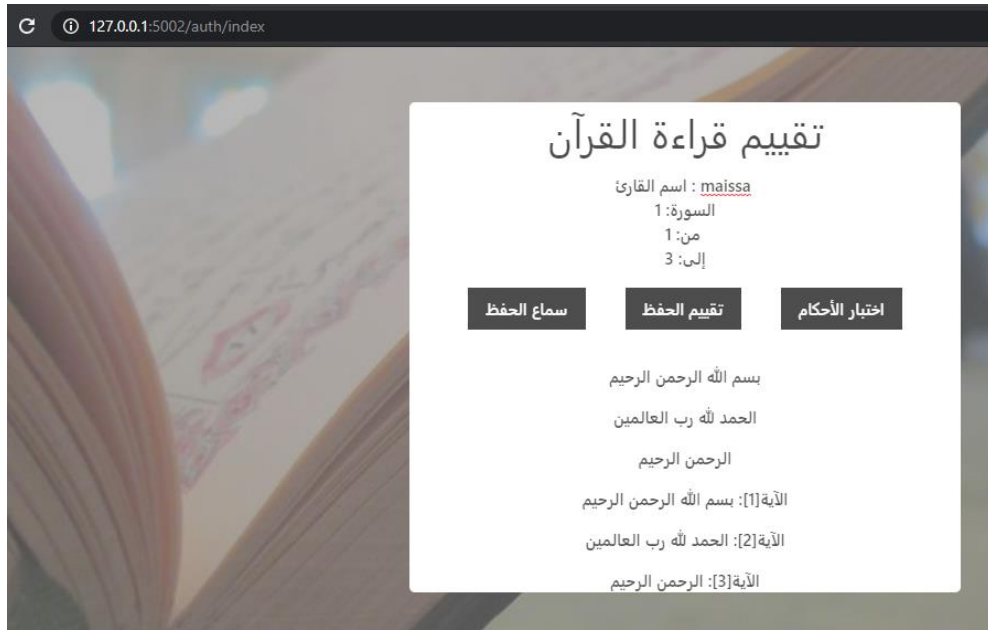


Figure 4.9 - Screenshot for evaluation page -verses-

#### 4.5.1.2 Case 2: Recitation is false

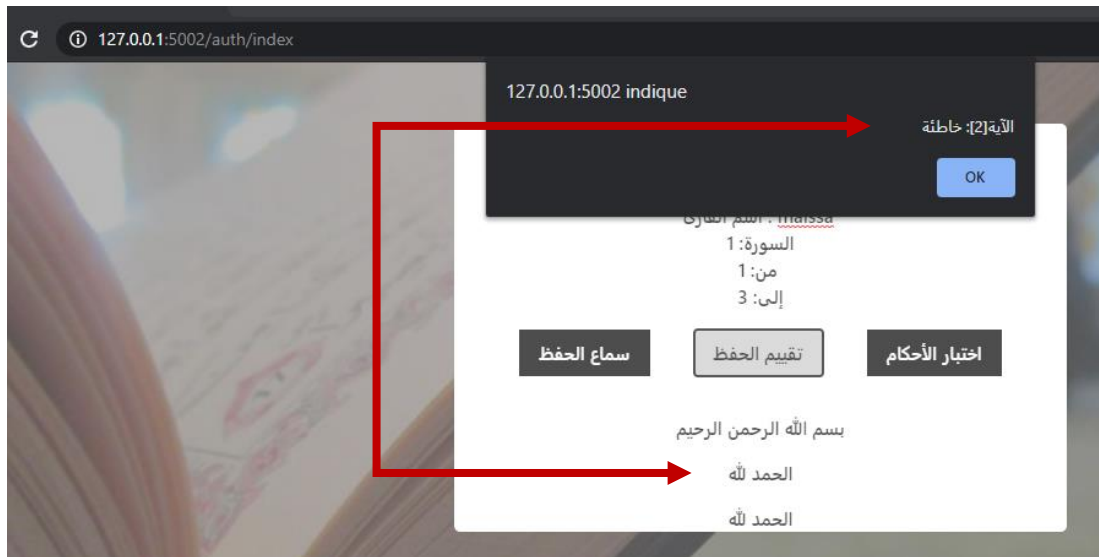


Figure 4.7 - Screenshot for evaluation page - transcription case false -



Figure 4.8 - Screenshot for evaluation page -verse 3 false-

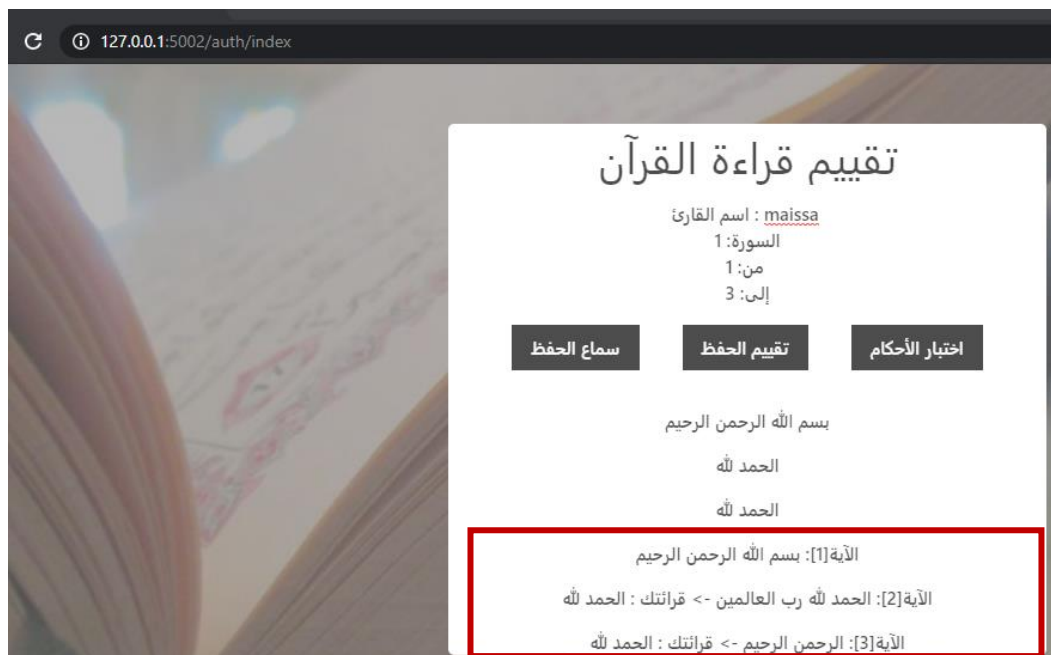


Figure 4.9 - Screenshot for evaluation page -verses and user recitation-

#### 4.5.1.3 Case 3: No recitation to evaluate it

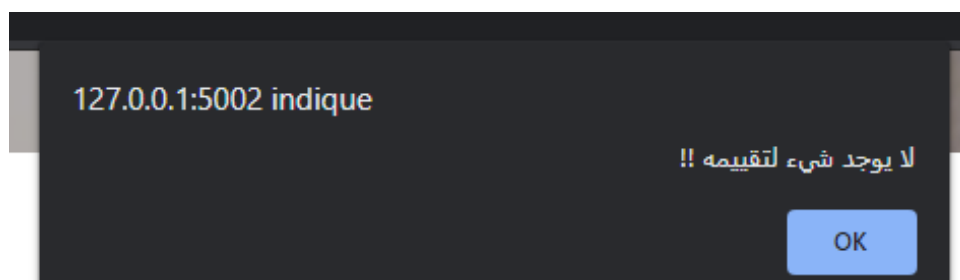


Figure 4.10 - Screenshot for evaluation page – User alert for lack of recitation to evaluate-

## 4.5.2 Evaluation of recitation

This step can be established when we complete the previous step (4.5.1) because it's combined prediction from verbal content (textual transcript) as well as the non-verbal content (by direct signal analysis).

For realize this step we propose the using of audio fingerprint which is: extracted from spectrograph by a process called feature extraction. Track identification using audio fingerprinting is considered to be one of the fastest and accurate forms of music recognition. [20]

To extract these fingerprints, we first transform the audio signal into a spectrogram. We convert the resulting spectrogram into a set of 2-D binary images. Finally, we extract the top-d salient regions from each binary image.

Figure 4.13 represent a system overview and continuation of the previous steps in the form of diagram.

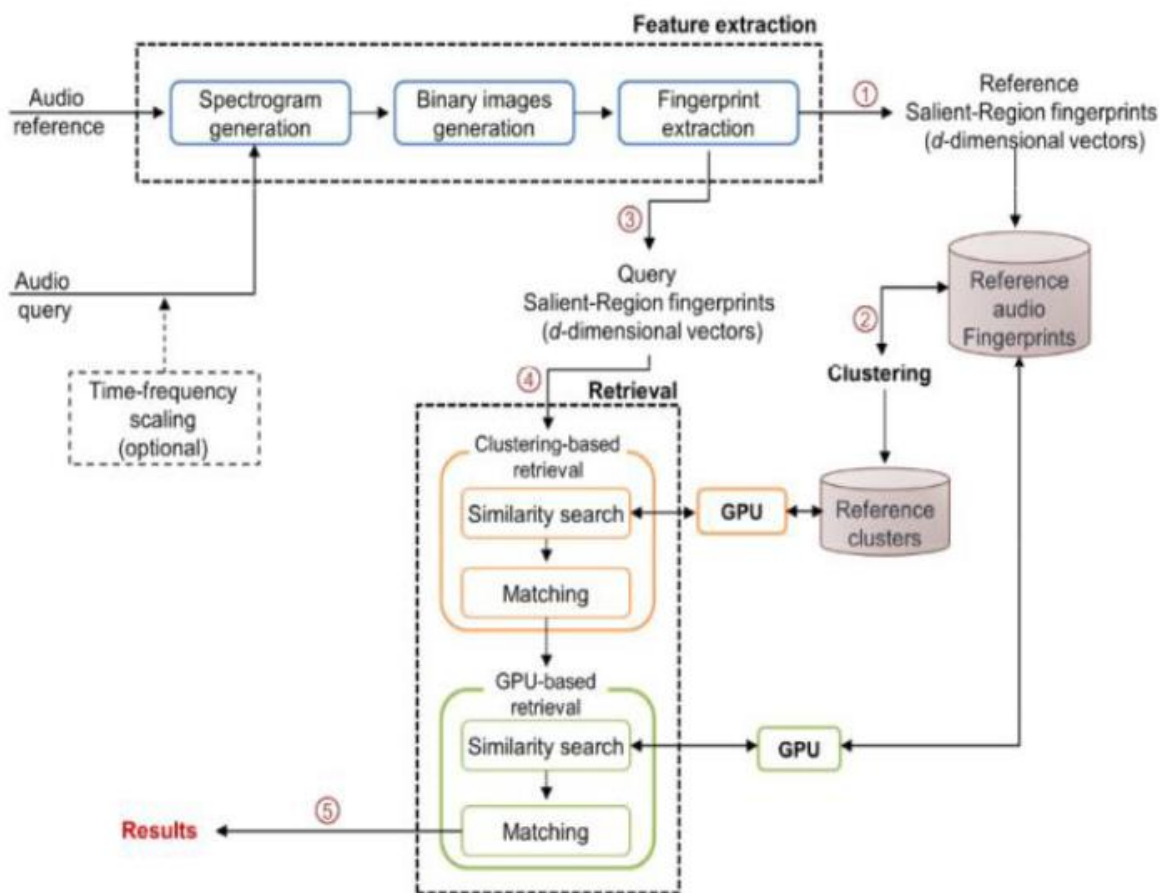


Figure 4.11 – System overview

Unfortunately, we were not able to reach this part of the project because of the confinement which disrupted our working session with Ms. Arkam who guide us for the best way to accomplish this step which is represented by the second button:

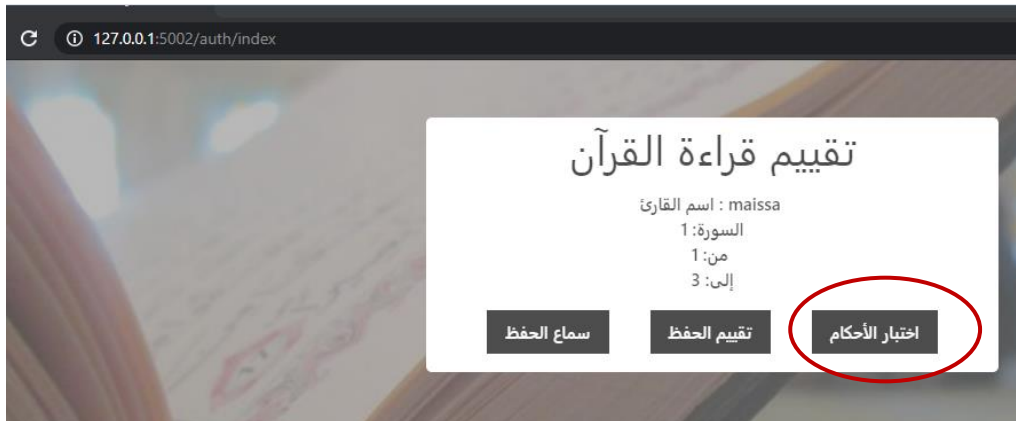


Figure 4.12 - evaluate recitation button

## 4.6 Experiments

Starting with testing our application 20 times in different cases, Table 1.4 show our tests specifications. We began our experiments according to the order of the Surahs of the Qur'an with a random selection of the verses in terms of their location and quantity.

Test #	Verses as said is:		Verses specifications			
	True	False	Surat number	Number of verses	From	To
01	*		1	3	1	3
02	*		7	6	5	10
03	*		15	6	41	46
04	*		17	2	25	26
05	*		20	4	25	28
06	*		26	5	77	81
07	*		37	5	135	139
08	*		44	4	51	54
09	*		86	4	1	4
10	*		114	3	1	3
11		*	2	2	42	43
12		*	6	1	91	91
13		*	8	2	22	23
14		*	14	2	28	29
15		*	20	6	25	30
16		*	38	3	37	39
17		*	51	5	2	6
18		*	60	1	3	3
19		*	70	5	21	25
20		*	112	4	1	4

Table 4.1 - List of verses included in the test

## 4.7 Performance

After the experiments whose characteristics are described in 1.2, we obtained the results shown in Table 4.2:

<b>Test #</b>	<b>#Accepted test / all tests</b>	<b>Acceptance ratio</b>
<b>01</b>	3/3	<b>100%</b>
<b>02</b>	0/6	<b>0%</b>
<b>03</b>	3/6	<b>50%</b>
<b>04</b>	0/2	<b>0%</b>
<b>05</b>	$\frac{3}{4}$	<b>75%</b>
<b>06</b>	3/5	<b>60%</b>
<b>07</b>	4/5	<b>80%</b>
<b>08</b>	2/4	<b>50%</b>
<b>09</b>	4/4	<b>100%</b>
<b>10</b>	3/3	<b>100%</b>
<b>11</b>	2/2	<b>100%</b>
<b>12</b>	0/1	<b>0%</b>
<b>13</b>	$\frac{1}{2}$	<b>50%</b>
<b>14</b>	$\frac{1}{2}$	<b>100%</b>
<b>15</b>	4/6	<b>66%</b>
<b>16</b>	2/3	<b>66%</b>
<b>17</b>	4/5	<b>80%</b>
<b>18</b>	1/1	<b>100%</b>
<b>19</b>	5/5	<b>100%</b>
<b>20</b>	4/4	<b>100%</b>
<b>average acceptance ratio</b>		<b>68.85%</b>

Table 4.2 - performance result

**#Accepted test:** Describe the number of correct interactions of our application, i.e., in reading the verse correctly, if the application returns "true", this means adding one to the correct result, while if the verse is read incorrectly if the application returns "false" we will add one to the correct result.

The Figure below represents the relationship between the number of verses and the acceptance ration of our application during previous experiences.

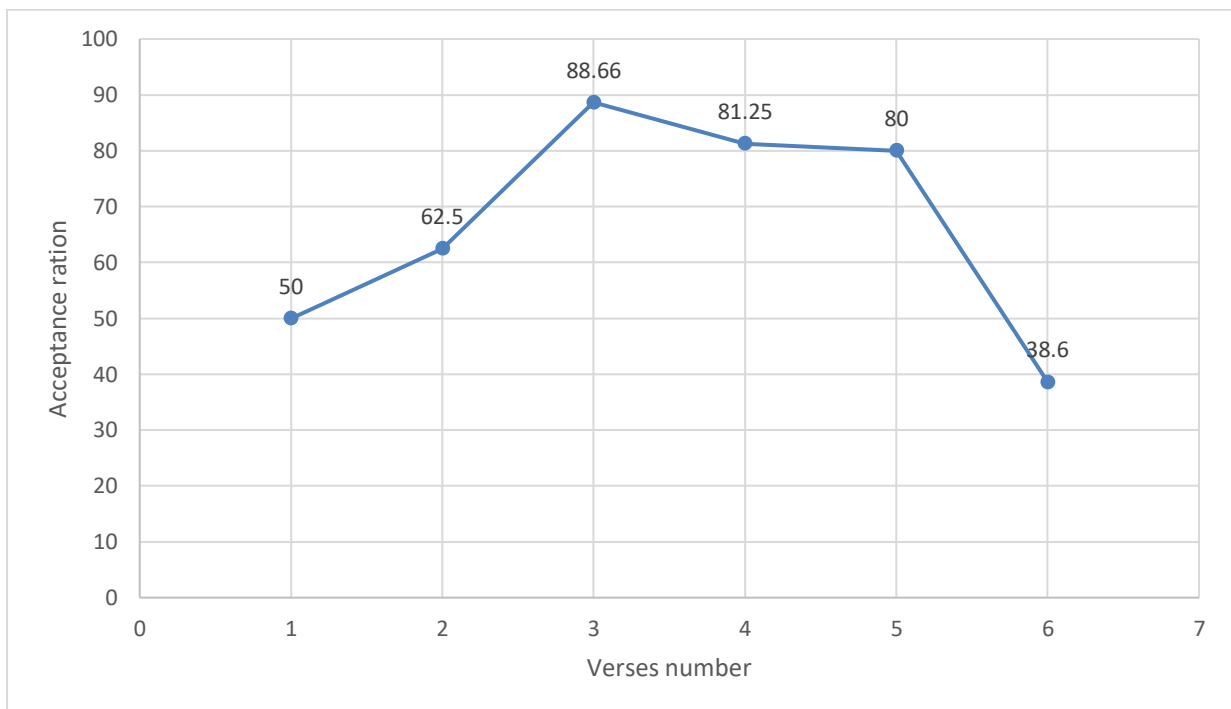


Figure 4.13 - The relationship between the number of verses and the acceptance ration of our application during previous experiences



## 4.8 Discussion

From the results shown in Table 4.2 and Figure 4.14 we observe that when user recite one verse the acceptance ratio be 50% while in two verses the curve creases to 62.5%. In three verses the curve reaches 88.66% which is the high point in these tests, after that we observe a continued decrease for recite 4,5,6 verses we got the order 81.25%,80%,38.6%.

From the previous data analyze, we note that the larger number of verses, the lower the response rate of ASR, we due that to:

1. noise that is picked up by the microphone
2. the inability of the device to receive all the verses because of the weakness of speech recognition in Arabic.

## 4.9 Conclusion

In conclusion we presented in this chapter two parts, first one present our application and the tools used to create it and the method of use by the user, in addition we presented two cases for correct and wrong reading and the desired results of the application.

Second part, shows 20 experiments that we have conducted and analysed the results obtained by these experiments and the conclusion is that the possibility of recognizing Arabic speech must be developed in order to be able to develop effective applications in this field.

---

## *Conclusion*

---

The primary goal of this thesis was to conduct an **application that detect the voice errors of user recitation of the holy Quran**. To this end, we have created a web-based application that used Arabic speech recognition software using web speech API and we carried out 20 experiments to analyze its accuracy using a combination of verses that had choosing randomly.

From this experimental study, we can derive the following conclusion:

- ❖ The larger number of verses, the lower the response rate of ASR

This thesis has revealed several interesting areas for improvement. Based on the insights gained from the experimental findings, we have concluded that the Arabic language still have a weakness in speech recognition comparing with other language.

During this work we have faced several difficulties, most notably, web speech API was unable to recognize tashkeel of word, which is an important thing in the Arabic language.

As a project continuation we have proposed a future research: evaluation of recitation of the holy Quran, given a proposal method that present in fingerprint and its algorithms.

## 6 References

- [1] Prerana Das, K. A., Pranab Das and Vijay Prasad (2015). "VOICE RECOGNITION SYSTEM: SPEECH-TO-TEXT." *Journal of Applied and Fundamental Sciences*.
- [2] Kamble1, B. C. (2016). "Speech Recognition Using Artificial Neural Network." *Journal of Computing, Communications & Instrumentation Engg (IJCCIE) Vol. 3* (Issue 1).
- [3] INSTITUTE, T. A. (2016,SEPTEMBER 14). "THE NEURAL NETWORK ZOO." from <https://www.asimovinstitute.org/neural-network-zoo/>.
- [4] Tamkeen (August 16, 2020). Tamkeen. V. 1.0.7. <https://play.google.com/store/apps/details?id=com.mighty.tamkeen&hl=en>.
- [5] LLC, G. (2020,November 10). Read Along. [https://play.google.com/store/apps/details?id=com.google.android.apps.seekh&hl=en\\_US&gl=US](https://play.google.com/store/apps/details?id=com.google.android.apps.seekh&hl=en_US&gl=US).
- [6] [https://www.investopedia.com/terms/a/artificial-neural-networks-ann.asp#:~:text=An%20artificial%20neural%20network%20\(ANN\)%20is%20the%20component%20of%20artificial,to%20produce%20the%20desired%20output.](https://www.investopedia.com/terms/a/artificial-neural-networks-ann.asp#:~:text=An%20artificial%20neural%20network%20(ANN)%20is%20the%20component%20of%20artificial,to%20produce%20the%20desired%20output.)
- [7] <https://psns.org/spectrograms/what-is-a-spectrogram>
- [8] <https://lipn.univ-paris13.fr/~gerard/uml-s2/uml-cours05.html>
- [9] Manolakis, D. G., & Ingle, V. K. (2011). *Applied digital signal processing: theory and practice*. Cambridge University Press.
- [10] [www.tanzil.info](http://www.tanzil.info)
- [11] <https://www.mysql.com/cloud/>
- [12] <https://www.phpmyadmin.net>
- [13] Djordjevic, N. (2017). "Evaluation of the usability of web-based applications." *Vojnotehnicki glasnik* **65**(3): 785-802.
- [14] [https://www.ibm.com/cloud/learn/speech-recognition?mhsrc=ibmsearch\\_a&mhq=compare%20two%20audio](https://www.ibm.com/cloud/learn/speech-recognition?mhsrc=ibmsearch_a&mhq=compare%20two%20audio)
- [15] Rahhal Errattahia, A. E. H., Hassan Ouahmane (2018). "Automatic Speech Recognition Errors Detection and Correction." *Procedia Computer Science*.
- [16] Adorf, J. (2013, May 27). "Web Speech API." KTH Royal Institute of Technology, Stockholm.

- [17] <https://webcaptioner.com/help/getting-started/what-is-web-captioner>
- [18] <https://www.research.ibm.com/haifa/dept/imt/mst.shtml>
- [19] Vibhute, A. and R. Hibare (2014). "Feature Extraction Techniques in Speech Processing: A Survey." International Journal of Computer Applications **107**(5): 1-8.
- [20] Patil, S., et al. (2016). "Fast Audio Fingerprinting System using GPU and a Clustering-Based Technique." Ijarcce **5**(12): 300-304.