بسم الله الرحمن الرحيم

Sudan University of Science and Technology

College of Graduate Studies

Designing a Language Model for some Arabic Words and Letters Recognition based on the Concept of Al-Qaidah ALnoraniah

تصميم نموذج لغوي للتعرف الصوتي على بعض الكلمات والحروف العربية بناءً على مفهوم القاعدة النورانية

A dissertation submitted in partial fulfilment of the requirements for the award of the degree of Master of Computer Science

Prepared by/

Abdelrahman Ali Salih Ahmed

Supervisor/

Dr. Omer Balola Ali

February 2022

APPROVAL

We hereby declare that we have read this thesis or dissertation and in our opinion this thesis is sufficient in terms of scope and quality for the award of the degree of Master of computer sciences.

Signature ________________________
Name of Supervisor ________________________
Date ________________________

Signature ________________________
Name of External Examiner ________________________
Date ________________________

Signature ________________________
Name of Internal Examiner ________________________
Date ________________________

DECLARATION

I hereby declare that this thesis or dissertation “Designing Qaidah Noraniah model for Arabic Speech Recognition” is the result of my own research except the citations in the references. The thesis has not been accepted for any degree nor concurrently submitted in candidature for any other degree.

Signature : ________________________________
Name of Student : ________________________________
Date : ________________________________

قال تعالى:

أعوذ بالله من الشيطان الرجيم

{وَقُل رَّبِّ زِدۡنِی عِلۡمًا}

[سُورَةُ طه: ١١٤]

DEDICATE

I would like to dedicate my thesis
To my parents, they always ask Allah to make me successful in my life.

To my wife, she stands with me and helped me in the rough times.

To my brothers and sisters, they encourage me and helped me.

To my friends and fellow members

To all my doctors and teachers, they advised me and guide me
without whom it was almost impossible for me
to complete my thesis work

ABSTRACT

The Arabic language is one of the most popular languages due to the number of Arab people and due to its religious status, which makes non-Arabs interested in learning it. However, there are not many speech recognition systems for the Arabic language, or there is no integrated system for it. Because of its extensive vocabulary and morphology, it is more difficult to develop systems to recognize it. For some of these reasons, it uses agglutinative letters and diacritics (الحركات). The Qaidah Noraniah language model was designed to help solve some of these problems. The model was trained with a half hour of data using the CMU Sphinx program, and the model was tested by dependent and independent speakers. The model achieved good results in recognizing single letters, recognizing letters with diacritics and the ability to differentiate between different diacritics of a single letter, as well as recognizing words not used in training where they consisted of letters used in training. The model achieved a letter recognition rate of 67.17%.

Keywords: speech recognition, Qaidah Noraniah, language model.

المستخلص

تعد اللغة العربية من أكثر اللغات شيوعا لعدد العرب المتحدثين بها ونظرا لمكانتها الدينية مما يجعل غير العرب مهتمين بتعلمها، بالرغم من ذلك لا توجد العديد من أنظمة التعرف على الصوت للغة العربية أو لا يوجد نظام متكامل لها، وذلك لكثرة مفرداتها وتركيبتها اللغوية، مما زاد من صعوبة بناء أنظمة للتعرف عليها، من بعض هذه الأسباب وجود الحروف الملتصقة مع بعضها والحركات المرتبطة بالحروف. تم إنشاء نموذج القاعدة النورانية اللغوي ليساعد في حل بعض هذه المشاكل وتم تدريب النموذج بنصف ساعة من البيانات باستخدام برنامج CMU Sphinx وتم اختبار النموذج من قبل متحدثين من داخل وخارج النظام، أحرز النموذج نتائج جيدة في التعرف على الأحرف المنفردة والتعرف على الحركات والقدرة على التفريق بين الحركات المختلفة للحرف الواحد وأيضا التعرف على كلمات لم تستخدم في التدريب حيث تتكون من حروف مستخدمة في التدريب، وحقق النموذج معدل تعرف قدره 67.17%.

TABLE OF CONTENTS

TITLE **PAGE**

LIST OF TABLES

TABLE NO. **TITLE** **PAGE**

LIST OF FIGURES

FIGURE NO. **TITLE** **PAGE**

LIST OF ABBREVIATIONS

ASR	**-**	**Automatic Speech Recognition**
DCT	**-**	**Discrete Cosine Transformation**
MFCC	**-**	**Mel Frequency Cepstral Coefficients**
HMM	**-**	**Hidden Markov Models**
CMU	**-**	**Carnegie Mellon University**
NN	**-**	**Neural Networks**
KNN	**-**	**K-Nearest Neighbor**
DTW	**-**	**Dynamic Time Warping**
VoIP	**-**	**Voice over IP**
EFCC	**-**	**Ear Frequency Cepstral Coefficient**
LPC	**-**	**Linear Predictive Coding**
LSP	**-**	**line spectral frequency**
RMS	**-**	**Root Mean Square**
STE	**-**	**Short Time Energy**
MDVP	**-**	**Multi-Dimensional Voice Program**
SDC	**-**	**Shifted Delta Cepstral Coefficient**
PSD	**-**	**Power Spectral Density**
SVM	**-**	**Support Vector Machine**
GMM	**-**	**Gaussian Mixture Model**
Pdf	**-**	**probability distribution functions**
HSMM	**-**	**Hidden Semi-Markov Model**
DNN	**-**	**Deep Neural Network**
LSTM	**-**	**Long Short-Term Memory**
MLP	**-**	**Multilayer Perceptron**
RBF	**-**	**Radial Basis Function**

GA	**-**	**Genetic Algorithm**
DSR	**-**	**Distributed Speech Recognition**
NSR	**-**	**Network Speech Recognition**
WER	**-**	**Word Error Rate**

LIST OF APPENDICES

CHAPTER 1

INTRODUCTION

1.1.1 Background:

1.1.1 Automatic Speech Recognition (ASR):

Automatic Speech Recognition (ASR) is a technology that has been developed to allow the computer to understand spoken speech and convert it to text(Saini and Kaur, 2013), in order to make the computer behave like a human, which provides an environment for the user to use the computer in a simpler and easier way. Many ASR integrated systems have been developed for some languages, such as English, French, and Chinese, but some languages, despite their popularity and spread, have no integrated systems or are still in early stages(Ahmed and Ghabayen, 2017). This is due to several reasons, including language morphology and rich vocabulary. If we take the Arabic language as an example, it is one of the most widely spoken languages in the world. It represents the mother tongue for all Arabs and also for its religious status as the language of the Holy Quran, which makes non-Arab Muslims have an interest in learning the language. The agglutinative letters, letters with diacritics (الحركات), rich vocabulary and multitude of dialects are some of the reasons that make it difficult to create an Arabic ASR integrated system(Absa *et al.*, 2018a; Al-Sabri, Adam and Rosdi, 2018; Saeed, Salman and Ali, 2019).

1.**1.2 Qaidah Noraniah:**

The Qaidah Noraniah method is one of the greatest sciences related to the Holy Quran. This method is used to teach Arabic by teaching the pronunciation of letters, then how to connect letters with each other, and then learning to link diacritics (الحركات) with letters. It is also used for teaching other things like the prolongation (المد) and Shadda (الشدة) using a scientific method of gradual and by the teaching of voice(Mohammed, 2018).

The name of this rule comes from the person who created it, Noor Mohammed Haquani, who wrote a book called Al-Qaidah Al-Noriah (القاعدة النورية), and there is another name for the developed version of it, AL-Qaidah Al-Noraniah Al-Fathia Al-Motawra (القاعدة النورانية الفتحية المطورة), which was created by Osama Quari.

The objectives of Qaidah Noraniah are:

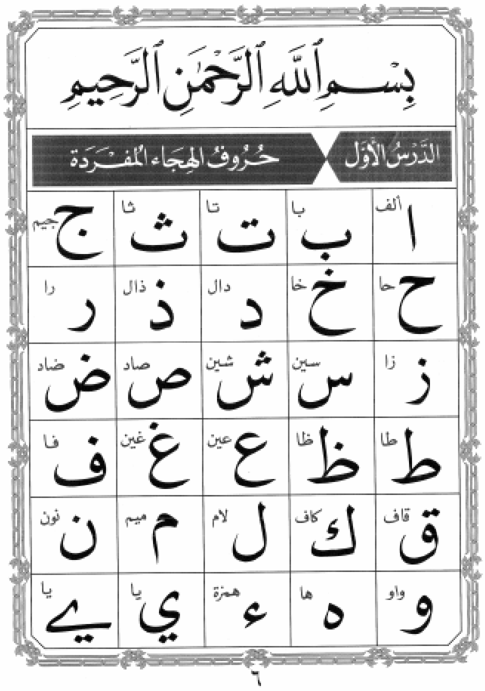
- Learn the perfect way to pronounce the alphabet in the correct way.
- Facilitating the learning of the Quran for all ages.
- Developing the understanding, awareness and perception.
- To consolidate all the rules of Tajweed and learn to apply them.

There are lots of benefits to Qaidah Noraniah, some as in the following points:

- Identify each letter of the Holy Quran, as well as the letters of the Arabic language.
- It helps to learn the Arabic letters of Quran to develop the exits of letters, and to give each letter what it deserves.

- The Qaidah Noraniah contains most of the important rulings of Tajweed.
- Service and support of the Islamic religion through the mastery of the Arabic language.
- Qaidah Noraniah is the easiest way to teach beginners how to read and memorize the Quran in shortest time, and as little effort as possible.
- Anyone who masters this rule, even if he is young, can read the Holy Quran with no difficulties.

The Qaidah Noraniah has a unique learning method that must be followed exactly as described by the author, with each stage preparing you for the next. some of these stages are: Single Alphabet letters: The letters of the Arabic alphabet (أ، ب، ت، ...) are pronounced many times by the teacher at this stage, and the students repeat after him. Composite Alphabet letters: When two or more letters are combined (i.e.,بلب، من، تح), the first letter is pronounced separately, followed by the second letter, and so on (i.e., بلب: بَاء، لام، با). Separated letters: at this stage, the teacher focuses on pronouncing the prolongation (المد) for fourteen letters that contain four groups. Each group has some properties that differ from the other groups. The three diacritics (الحركات): The Arabic language have diacritics that when added to letter gave different pronunciation, these diacritics are: Fatha (فتحة ـَ), kasra (كسرة ـِ) and Dhama (ضمة ـُ), each letter is pronounced using these diacritics. Tanween: its letter (ن) added to the end of the noun as a pronunciation, and it is written as (ـً, ـٍ or ـٌ) depending on the letter's diacritic. The vowel letters: The Arabic language has three vowel letters (ا، ب، ي)(Haqqani, 1998).



بِسْمِ ٱللَّهِ ٱلرَّحْمَٰنِ ٱلرَّحِيمِ

الدَّرْسُ الأَوَّل حُرُوفُ الهِجَاء المُفْرَدَة

ألف ا	با ب	تا ت	ثا ث	جيم ج
حا ح	خا خ	دال د	ذال ذ	را ر
زا ز	سين س	شين ش	صاد ص	ضاد ض
طا ط	ظا ظ	عين ع	غين غ	فا ف
قاف ق	كاف ك	لام ل	ميم م	نون ن
واو و	ها ه	همزة ء	يا ي	يا ے

٦

Figure 1-1: Page of the first lesson of Qaidah Noraniah

1.2 Problem Statement:

This research focused on the problem of creating a dataset that contains all Arabic words due to the lack of audio files for the Arabic language and because most Arabs speak a dialect. And there's the issue of distinguishing between the diacritics of words with similar sets of letters.

1.3 Objectives:

The object is to develop a system for Qaidah Noraniah that can perform the following tasks:

- Recognize the correct pronunciation of Arabic letters.

- Recognize Arabic words generated from a dataset of letters.
- The ability to distinguish between the diacritics of Arabic letter.

1.4 Scope of the Research:

The model design is based on 9 characters of the Arabic language. (ا, ب, خ,س , ع, ل, م, ن, ي) that represent the letters of Arabic sentence: (بُنِيَ الْإِسْلَامُ عَلَى خَمْسٍ) which used to train the model.

1.5 Expected Outcome:

At the end of this research, the developed system will have the ability to recognize words generated from letters and to recognize the difference between diacritics for a letter.

1.6 Thesis Structure:

This thesis is organized as follows:

- Chapter 1 contains an introduction of the thesis, an introduction of Qaidah Noraniah, a problem statement, an objective of the research, a scop of the research, and an expected outcome.
- Chapter 2 contains the literature review, a detailed description of the automatic speech recognition system and related work.
- Chapter 3 describes the methodology of the research starting from preparing the data until creating the model.
- Chapter 4 contains the result of training and testing the proposed model and discusses the results and compares it to other research in the same field.
- Chapter 5 contains the conclusion of the research.

CHAPTER 2

THEORETICAL BACKGROUND

2.1 Background:

Automatic Speech Recognition (ASR) is a technology developed to allows the computer to understand spoken speech, and converting it to text, in order to make the computer behave like a human, which provides an environment for the user to deal with the computer in a simpler and easier way. ASR system consist of three main things: Feature extraction, Acoustic and Language models and classification There are two main important methods on ASR, first is feature extraction and second is classification, **Figure *2-1*** shows the main processes of ASR system(Saini and Kaur, 2013).

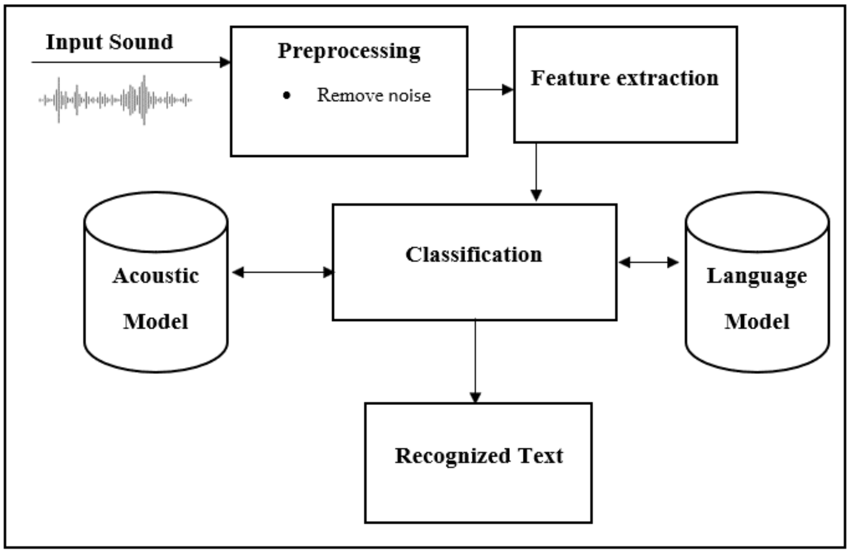


Figure 2-1: Automatic speech recognition (ASR) system

2.1.1 Feature extraction:

The process starts with converting speech signals from analog to digital, then an enchantment is made to the samples, such as removing background noise. Then comes the step of extracting the features. The feature extraction minimizes the size of data without affecting the characteristics of voice and transforms it into a parameterized representation. There are a lot of techniques to extract features. These techniques differ in their methods and types.

2.1.1.1 Mel Frequency Cepstral Coefficients (MFCC):

FCC is the most used among other techniques for its high performance in extracting features (Sakka, Techini and Bouhlel, 2017; Pradana, Adiwijaya and Wisesty, 2018) mentioned that MFCC is the dominant and most used approach for feature extraction. It uses the knowledge of the human auditory system, which can't perceive frequencies over 1 KHz. **Figure 2-2** presents MFCC steps(Adiwijaya *et al.*, 2017; Suresh and Thorat, 2018).

- **Framing:** Framing is the process of dividing a spoken stream into little chunks of 20 or 30 milliseconds called frames(Suresh and Thorat, 2018).
- **Windowing:** The hamming window procedure is used to reduce and eliminate discontinuity between the start and finish of each frame of the signal.
- **Fast Fourier Transform (FFT):** For each frame sample, this technique converts the time domain to the frequency domain.
- **Mel Filter Bank:** In comparison to high frequency, the low frequency component of speech includes useful information. A number of triangular filter banks are needed to execute Mel-scaling; therefore, a bank of triangular filters is generated during the MFCC calculation.
- **Discrete Cosine Transformation (DCT):** The inverse the process of FFT is the final stage in the MFCC feature extraction process(Adiwijaya *et al.*, 2017).

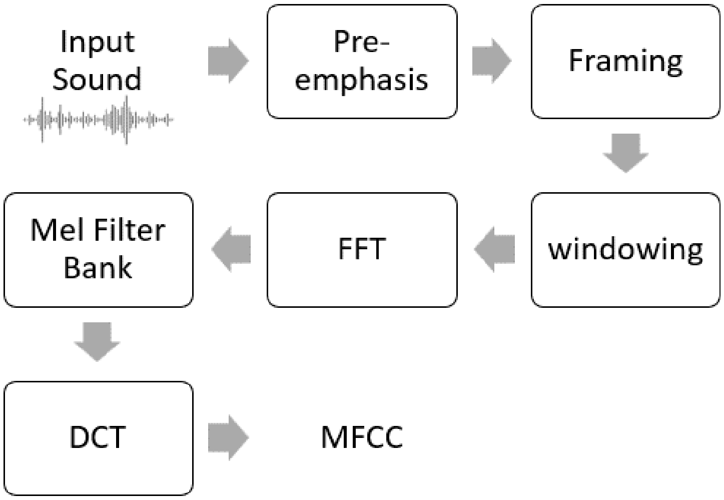


Figure 2-2: MFCC steps.

2.1.2 Acoustic and language models:

Acoustic and language models are important components of ASR because sounds carry a lot of information that can be a conversation, birds singing, animals sounding, and so on.(Suresh and Thorat, 2018). To make the ASR system for special tasks, we must make acoustic and language models for specific purposes.

2.1.2.1 Acoustic model:

An acoustic model consists of phonemes. Phonemes represent letters with probability values to guide the ASR system to search for correctly pronounced letters based on analyzed text databases. Each language has its own set of phonemes. The English language has about 40 phonemes (AL Bizzocchi, 2017). Increasing the data leads to an increase in the accuracy of the model. These probabilities are counted by training algorithms such as Hidden Markov Models (HMM), which give a probability for each phoneme according to word grammar. An acoustic model contains word and phoneme models(Zarrouk and Benayed, 2016).

On **Table 2-1** the words and their pronunciation are generated using the Carnegie Mellon University (CMU) Lexicon Tool (Carnegie Mellon, no date). Some words have one form of pronunciation, as in word (World), and others have more than one form of pronunciation, as in word (Hello). The training algorithm matches phoneme sequences with the correct word pronunciation.

Table 2-1: An example of words and their pronunciation

word	pronunciation	
Hello	HELLO	HH AH L OW
	HELLO (2)	HH EH L OW

World	WORLD	W ER L D
Speech	SPEECH	S P IY CH
Language	LANGUAGE	L AE NG G W AH JH
	LANGUAGE (2)	L AE NG G W IH JH
Jump	JUMP	JH AH M P
Run	RUN	R AH N
Read	READ	R EH D
	READ (2)	R IY D

2.1.2.2 Language model:

The language model has a similar method of work as the acoustic model, but it looks for the best sequence of words by using grammar or analyzing a large amount of data to determine the correct sequence of words(Ahmed and Ghabayen, 2017).

The type of system and data determine the language model design. As a result, there are different types of language models(Saini and Kaur, 2013). The N-gram language model and the Neural language model are the most well-known.

2.1.2.2.1 N-gram language model:

The N-gram is a sequence of text or sound samples collected from a database. N-grams have different sizes based on system needs: unigram, bigram, trigram and n-gram. The size for unigram is 1 gram, bigram is 2 grams, and trigram is 3 grams.(Ahmed and Ghabayen, 2017).

On Table ***2-2*** three types of n-gram size; unigram, take one word at a time; each word represents one gram; bigram, take two words at a time, the first with second, then second with third to the end of the text; trigram, take 3 words at a time. For more than that, we call it an n-gram. n is the size of grams that can be 4 or more than that(Ahmed and Ghabayen, 2017).

Table 2-2: N-gram size examples

N-gram size	example
unigram	There is always a butter place
bigram	There is is always always a a butter butter place
trigram	There is always is always a always a butter a butter place

2.1.2.2.2 Neural Network language model:

The neural network language model represents words with vectors and uses a neural network to predict continuous speech from these vectors(Bengio, 2008). To produce the best results, the neural network requires a large amount of data. The data is received by the input layer, which gives it weight before passing it to the hidden layers, which process it and then pass it to the output layer, these layers are presented on **Figure *2-3***(Zerari *et al.*, 2018).

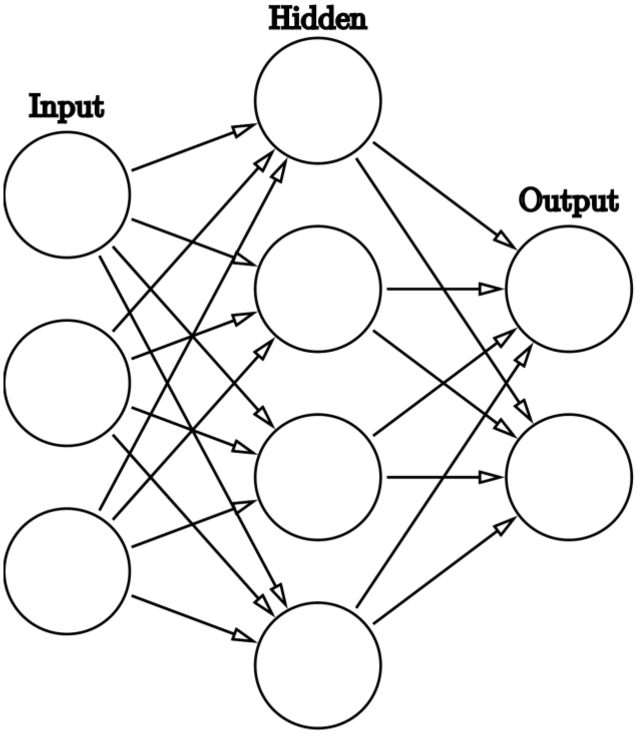


Figure 2-3: Neural Network layers

2.1.3 Classification:

Classification is the most important step in speech recognition. It comes after the feature extraction step. The features are passed to a classifier to find the best match between features, and to make a decision for speech utterance depending on acoustic and language models. There are a lot of techniques used for classification. The most used are Neural Networks (NN), Hidden Markov Models (HMM), K-Nearest Neighbor (KNN) and Dynamic Time Warping (DTW)(Saini and Kaur, 2013).

2.2 Related works:

Much research has been done to improve speech recognition or to provide new methods. Some research focused on using suitable feature extraction as in (Klaylat *et al.*, 2018; Yasmine *et al.*, 2019), and other research on classification techniques as in (Morsy *et al.*, 2018; Saeed, Salman and Ali, 2019) , or improving feature extraction or classification by combining several techniques together as in (Mnassri, Bennasr and Cherif, 2017; Bouchakour and Debyeche, 2018), while others focused on building special acoustic and language models as in (Ahmed and Ghabayen, 2017; Smit *et al.*, 2018).

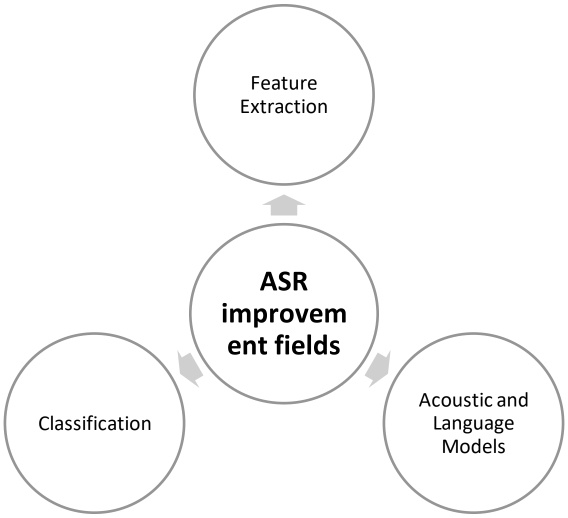


Figure 2-4: ASR improvement fields

2.2.1 Improve ASR using Feature Extraction and classification techniques:

2.2.1.1 Feature Extraction:

Some of the research used single feature extraction and others used more than one. Table ***2-3*** show feature extraction techniques that use single feature extraction.

Some researches make improvement to techniques by making change in the structure of technique, In (Yasmine *et al.*, 2019) to improve speech recognition over Voice over IP (VoIP) the system used modified MFCC called Ear Frequency Cepstral Coefficient (EFCC), by incorporating two models simulates mid-external and inner ear into MFCC, to make it more similar to human ear, the result showed that EFCC make improvement by 2% using 43 EFCC coefficient compared to 60 from MFCC.

Table 2-3: Using single feature extraction

NO	Feature Extraction	Reference	Purpose	Method	Pros	Cons
1	MFCC	(Mohammed, Sunar and Salamb, 2015)	Building verification system for Quranic verses	Train and test acoustic and language models	-	-
		(Afrillia *et al.*, 2017)	Recognize the Nagham of the Quran	Detect the type of Nagham by classification	-	Cannot be applied on any Quran recitation
		(Alkhatib *et al.*, 2017)	Correct mispronounced word	Comparing sound signal	Classify MFCC vector	the accuracy will be low on noisy environment
		(Sakka, Techini and	Improve recognition performance of	Enhance the spectral subtraction	Dealing with problem of noise	-

		Bouhlel, 2017)	Arabic language	by geometric approach		
		(Menacer, Mella, Fohr, Jouvet, Langlois and Smaïli, 2017)	Developing speech recognition system for Arabic dialect	Building new model by combining two models	Using more than one model	the dialect model audio data was recorded by 3 male persons
		(Mnassri, Bennasr and Cherif, 2017)	Propose new recognition technique	Optimizing classification technique	Improve classification ability	-
		(Yousaf *et al.*, 2018)	Design mobile application to for Deaf-Mute people	Convert speech to sign for Deaf-Mute	-	-
		(Yousfi, Zeki and Haji, 2018)	Speech recognition to distinguish between prolongation and recitation types	Match the speech with recorded samples	-	Doesn’t build custom language model
		(Zerari *et al.*, 2018)	Speech recognition system for Arabic digit	recurrent neural networks (RNN) for	efficiency	-

				processing features		
		(Bouchakour and Debyeche, 2018)	Improve continues Arabic speech recognition	Hybrid classification technique	Take advantage of MFCC sup types.	-
		(N. M. Arafa *et al.*, 2018)	Automatic pronunciation error detection	Create a dataset	-	-
		(Pradana, Adiwijaya and Wisesty, 2018)	Develop Arabic speech recognition system	Using multiple layers of classifier	-	-
2	EFCC	(Yasmine *et al.*, 2019)	Improve speech recognition over VoIP	Modify MFCC technique	-	-

Some research uses more than feature extraction techniques. These techniques are normally used to increase the accuracy of speech recognition systems. The most commonly used are statistical features, which are: MFCC, zero crossing rate, signal energy, temporal centroid, energy entropy, spectral flux, spectral energy, and Root Mean Square (RMS). The statistical features represent the core of the signals; each one extracts specific values to be classified. In (Suresh and Thorat, 2018) a language identification system used MFCC combined with Shifted Delta Cepstral Coefficient (SDC). MFCC can extract vocal tract shape, which gives temporal and acceleration information, and SDC gives information about the phonemes. MFCC features are extracted and then passed to SDC as 4 parameters

(N-d-p-k). N represents the number of coefficients; d determines the spread of calculated delta; p determines the gaps between delta; k represents the number of blocks. The combination of MFCC with SDC gave high results on language identification. **Table *2-4*** contains additional researches.

Table 2-4: Using more than feature extraction

NO	Feature Extraction	Reference	Purpose	Method	Pros	Cons
1	MFCC, Linear Predictive Coding (LPC)	(Adiwijaya *et al.*, 2017)	A system to recognize Arabic letters	Compare the result of two technique	-	-
2	MFCC, intensity, Fundamental frequency (F0), F0 envelope, zero crossing rates, line spectral frequency (LSP)	(Klaylat *et al.*, 2018)	Enhance emotion recognition system	Using multiple feature extraction techniques to get most of audio features	-	-

3	MFCC, SDC	(Suresh and Thorat, 2018)	Language identification system	Combining the advantages of techniques	Increase in accuracy	Time and computational work have increased.
4	MFCC, zero crossing rate, signal energy, temporal centroid, energy entropy, spectral flux, Spectral energy, RMS	(Salman, Saeed and Ali, 2018; Saeed, Salman and Ali, 2019)	Improve speech recognition for Arabic letters	Using statistical features	Increase recognition rate	the experiment is for 5 letters
5	energy, zero crossing rate, entropy	(Absa *et al.*, 2018b)	Algorithm for Arabic speech segmentation	Using three techniques for segmentation	-	-
6	Short Time Energy (STE), Fundamental frequency	(Al-Sabri, Adam and Rosdi, 2018)	A system to automatically detect Shadda in	By combining techniques that depend on the	Combining STE with Intensity give	The pattern of Shadda is different for Arabic letters

	(F0), Intensity		Arabic speech	loudness and pitch of sound	accurate result	
7	spectrogram features, Power Spectral Density (PSD)	(Khairuddin *et al.*, 2017)	Classification system for Quranic letters	Analyze the classification result of each feature technique	PSD has good accuracy	Combination has no effect
8	MFCC, LPC, LPCC, PLP, RASTA-PLP, Multi-Dimensional Voice Program (MDVP)	(Mesallam *et al.*, 2017)	Developing Arabic pathology database	Benefit from the properties of feature techniques	-	-
9	MFCC, LFCC, Timber features	(Sardar and Shirbahadurkar, 2018)	Identify speaker whisper sound	Comparing the result of feature extraction techniques	High accuracy of whisper training-whisper testing	low accuracy of whisper training-neutral testing

Some research compares results of feature extraction techniques to get the most accurate one of these techniques, as in (Adiwijaya *et al.*, 2017), while others compare techniques separately, and then these results are compared with combined values. In

(Khairuddin *et al.*, 2017) a system was made to classify the correct pronunciation of Quran for both males and females. The system extracts formants from speech signals using the spectrogram feature and also PSD extracted using MATLAP software. The two features have been classified separately and then combined together, but no effect has been gained from this combination.

2.2.1.2 Classification:

The classification technique is also used, a single technique or more than one, for the purpose of achieving accuracy and a high recognition rate.

In (Pradana, Adiwijaya and Wisesty, 2018) the system used Support Vector Machine (SVM) as a classifier. SVM divides the data into two classes and maximizes the margins between hyperplanes for the closest data class. The most optimal hyperplane is called the support vector. The classification was made for dependent and independent speakers. The result showed the dependency affected the classification, which gave high accuracy for dependent speakers.

The Dynamic Time Warping (DTW), which measures the similarity between two sequences of time, is used as a classifier for (Alkhatib *et al.*, 2017), a modified DTW used to calculate cosine similarity between vectors. The experiment led to good results.

In (Mohammed, Sunar and Salamb, 2015; Yousaf *et al.*, 2018) HMM was used as a classifier. In (Khelifa *et al.*, 2017) instead of using HMM, the system used modified HMM, which is a Hidden Semi-Markov Model (HSMM), with probability distribution functions (pdf) incorporated into HMM to deal with the problem when duration probabilities decrease over time. The results showed a little enhancement.

In (Suresh and Thorat, 2018) the system used Gaussian Mixture Model (GMM) as a classifier. Like HMM, GMM is based on probability estimation and likelihood. The system used a combination of MFCC and SDC for feature extraction. The result using GMM showed high accuracy.

There are a lot of neural networks classification techniques, each one with special properties. In (Morsy *et al.*, 2018) a system for detection of speech attributes for the Arabic language used Deep Neural Network (DNN) for classification with filter bank features. DNN classifies the frame to determine if it belongs to a specific attribute or not. The classification results have a high average accuracy. In (Zerari *et al.*, 2018) the system used neural networks to classify spoken Arabic digits. The used technique is Long Short-Term Memory (LSTM) to deal with sequence-to-sequence problem. Features extracted by MFCC are then processed by the LSTM layer to encode the sequence to a fixed vector and then passed to Multilayer Perceptron (MLP) to classify it. The results showed the effectiveness of this method, In (Saeed, Salman and Ali, 2019) a type of neural networks called Radial Basis Function (RBF) is used. It is optimized for noisy environments and has faster convergence, smaller extrapolation errors, and higher reliability. **Table *2*-*5*** contains researches used single classifier.

Table 2-5: Using single classifier

NO	Classifier	Reference	Purpose	Method	Pros	Cons
1	SVM	(Pradana, Adiwijaya and Wisesty, 2018)	Develop Arabic speech recognition system	Compare result of dependent and	-	-

				independent speakers		
2	DTW	(Alkhatib *et al.*, 2017)	Correct mispronounced word	Modify DTW	Butter result than normal DTW	Low accuracy
3	KNN	(Adiwijaya *et al.*, 2017)	A system to recognize Arabic letters	classify data of two extraction techniques	High accuracy by using small number of KNN	Not adjustable
		(Sardar and Shirbahadurkar, 2018)	Identify speaker whisper sound	classify data of extraction techniques	-	-
4	HMM	(Mohammed, Sunar and Salamb, 2015)	Building verification system for Quranic verses	Train and test data	-	-
		(Khelifa *et al.*, 2017)	System to recognize Arabic phoneme	Modify HMM	Increase in accuracy	There isn't much improvement

		(Yousaf *et al.*, 2018)	Design mobile application to for Deaf-Mute people	Train and test data	-	-
5	GMM	(Suresh and Thorat, 2018)	Language identification system	Train and test data	-	-
6	Neural Networks (NN)	(Morsy *et al.*, 2018)	Detect speech attribute in Arabic language	Train DNN to recognize attribute	-	-
		(Saeed, Salman and Ali, 2019)	Improve speech recognition for Arabic letters	RBF as neural network classifier	Higher reliability	-

The purpose of conducting some research using more than feature extraction can be to increase the accuracy of the system or to compare the feature extraction techniques. Using more than one classifier is mostly used for optimization.

In (Bouchakour and Debyeche, 2018) a system for Arabic Continuous Speech Recitation for Distributed Speech Recognition (DSR) and Network Speech Recognition (NSR), the system uses a hybrid DNN-HMM technique. It is known that HMM is based on probability estimation and maximization of likelihood, but DNN is based on maximum posterior criterion. The results showed the success of DNN-HMM and its ability to improve the performance of ASR.

In (Zarrouk and Benayed, 2016), SVM-HMM and MLP-HMM were compared, whereas in (Menacer, Mella, Fohr, Jouvet, Langlois and Smaili, 2017), HMM-GMM and HMM-DNN were used to develop six different acoustic models, with the combination of techniques producing a better result than using a single technique. **Table *2-6*** shows researches used more than one classifier.

Table 2-6: Using more than classifier

NO	Classifier	Reference	Purpose	Method	Pros	Cons
1	SVM, KNN, Neural Network MLP, Random Forest, Naive Bayes	(N. M. Arafa *et al.*, 2018)	Automatic pronunciation error detection	Appling more than classifier to get best result	-	-
2	Genetic Algorithm (GA)-SVM	(Mnassri, Bennasr and Cherif, 2017)	Propose new recognition technique	Optimize SVM by GA	High accuracy	-
3	Hybrid DNN-HMM	(Bouchakour and Debyeche, 2018)	Improve continues Arabic speech recognition	Combining DNN with HMM	Improve performance	-
4	SVM-HMM, MLP-HMM	(Zarrouk and Benayed, 2016)	A system for Arabic phoneme recognition	Comparing the result of hybrid methods	-	-

5	HMM-GMM, HMM-DNN	(Menacer, Mella, Fohr, Jouvet, Langlois and Smaïli, 2017)	Enhance Arabic recognition system	Training and testing with two hybrid techniques	-	-

2.2.2 Improve ASR by Acoustic and language models:

Some researchers used various methods to improve ASR systems, such as devolving and enhancing acoustic and language models. In (Smit *et al.*, 2018), a character-based language model was created for continuous speech recognition, with the goal of predicting words that were not seen in training data and producing a large vocabulary of words. The corpus of the language model is split into characters rather than words to allow reconstructing word boundaries from character data. In (Ahmed and Ghabayen, 2017) three approaches were used to enhance the Arabic ASR system: the first approach to deal with the problem of Tashkeel of words by using a Decision Tree to generate pronunciation variants; the second approach by using hybrid acoustic models from two models; and the third approach by using processed text to improve the language model because of the lack of Arabic resources. The results of these improvements reduced the Word Error Rate (WER).

Table ***2-7*** shows examples of improvement made on ASR systems based on the fields of ASR, by displaying the type of ASR field, the purpose of the improvement, the method used and the results.

Table 2-7: Examples of improvement of ASR

ASR improving Method	Type	Reference	Purpose	Method	Result
Features extraction	Single feature extractor	(Yasmine *et al.*, 2019)	Improve speech recognition over VOIP	Using EFCC (modified version of MFCC)	Improvement using EFCC by 2% compared to MFCC
	Multi feature extractor	(Suresh and Thorat, 2018)	Butter identification of spoken languages	Combine the properties of MFCC and SDC	Combining techniques gives high accuracy
Classification	Single classifier	(Alkhatib *et al.*, 2017)	Pronunciation correction	Modify DTW	Good result
		(Khelifa *et al.*, 2017)	Accurate Arabic recognizer	Using HSMM (modified version of HMM)	Increase in accuracy
	Multi classifier	(Bouchakour and Debyeche, 2018)	Improve client-server communication and increase ASR performance	Hybrid classifier out of two DNN and HMM	Improve performance
Acoustic and language models	Acoustic model	(Smit *et al.*, 2018)	Produce large vocabulary and predict word	Split corpus of language	Good result in predicting sup-words

			not in training data	model into characters rather than words	not in training data
		(Ahmed and Ghabayen, 2017)	Enhance Arabic speech recognition	Hybrid acoustic model	Reduce WER by 1.2%
	Language model	(Ahmed and Ghabayen, 2017)	Enhance Arabic speech recognition	Using processed text	Reduce WER by 1.9%

2.3 Summary:

This chapter presents how the automatic speech recognition systems work, discusses some of the studies that used these methods, and presents some of the research's pros and cons.

CHAPTER 3

METHODOLOGY

3.1 Introduction:

This chapter outlines the steps for designing this research model. The first step is gathering and preparing audio and text data. After that, a preprocess will be made for audio and text data. Then, the language model will be created, and finally the data training.

3.2 Preparation of data:

This step contains two important steps: preparing text data and collecting and preprocessing audio data. Each step stage will be described.

3.2.1 Text data:

The words are divided into letters. The first stage contains only the first letter. Then at each stage, the next character is added with the previous separately and then integrated into the group of two letters. Because the pronunciation of the word is determined by pronouncing a single letter or two letters together. The words are divided into letters. The first stage contains only the first letter. Then at each stage, the next character is added with the previous separately and then integrated into the group of two letters. Because the pronunciation of the word is determined by pronouncing a single letter or two letters together. **Figure *3-1*** depicts word dividing stages for the Arabic word Marhaba (مرحبا). The first stage begins with the first letter, then adding the second letter to the first letter. At stage 2, the letter can be separated or compounded with the previous letter. Some letters must be compounded to appear on pronunciation, such as the letter which has Sokoon (سكون ـْ) must be compounded with the previous letter, as in this case, the letter RA (رْ). Stage 5 represents all the letters of the word but with different forms, and one of these forms can be used to correctly pronounce the word.

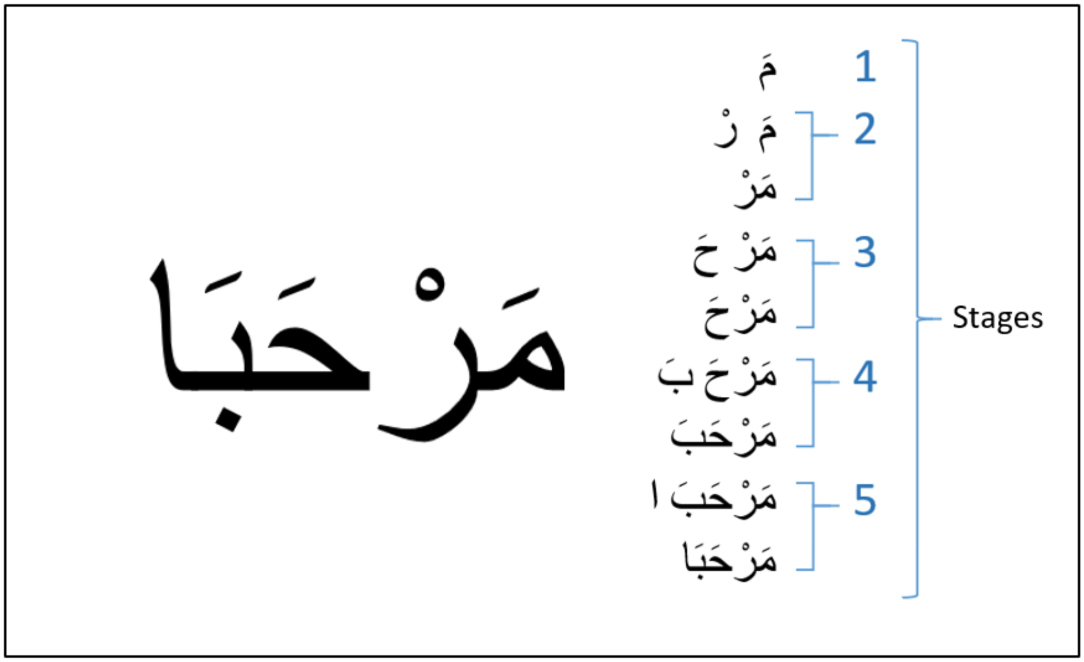


Figure 3-1: Word divide stages

In this study, the letters of the Arabic sentence (بُنِيَ الْإِسْلَامُ عَلَى خَمْسٍ), which contains 9 Arabic letters on **Table *3-1*** from 28 Arabic letters are used for training the model.

Table 3-1: Arabic letters used for training

letter
ا
ب
خ
س
ع
ل
م
ن
ي

The Arabic sentence is sliced into 17 parts. Every part contains a single letter or composed letters. These represent the pronunciation of a word or part of the word from the sentence.

Some Arabic letters have multiple forms; for example, the phoneme Alif (ا) in some words has multiple forms depending on the word (ا, ى). One form will be used for butter recognizing. **Table *3-2*** shows the words and the letters divided from them.

Table 3-2 Letters of Arabic sentence (بُنِيَ الْإِسْلَامُ عَلَى خَمْسٍ)

word	letter
بُنِيَ	بُ
	بُنِ
	نِ
	نِيَ
	يَ
الْإِسْلَامُ	ا
	الْ
	إِ

	إِسْ
	لَ
	لاَ
	مُ
عَلَى	عَ
	عَلَ
	لَ
	لاَ (لَى)
	ا (ى)
خَمْسٍ	خَ
	خَمْ
	سٍ

An N-gram is generated for the Arabic sentence based on the letters in **Table *3-2***. Every letter has more than one pronunciation based on diacritics, so we used only one form of letter pronunciation on the table.

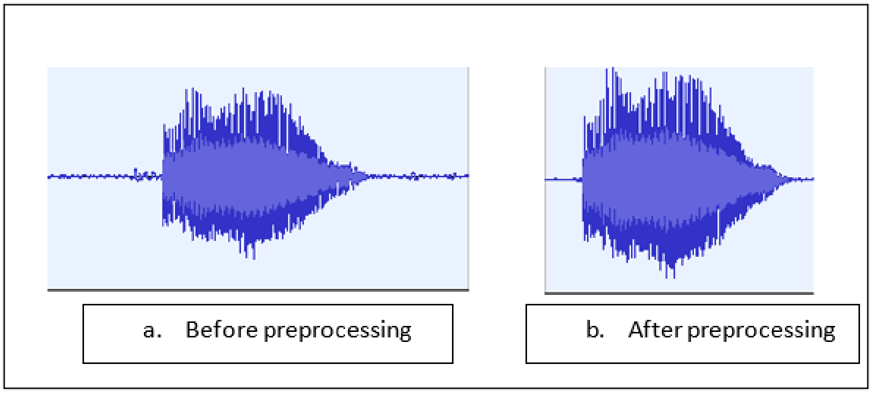
3.2.2 Audio data:

The audio data will be recorded for text data. The data will be recorded for both genders male and female, at different ages. Each person will be asked to record 10 audio samples. Then preprocessing will be done for these samples, such as removing noise, removing silence, and enhancing audio quality, and then the audio will be divided into letters and sub-words from **Table *3-2***. Four other words will be used in **Table *3-3***. These words are constructed from the letters of **Table *3-2***.

Table 3-3 Words constructed from letters of Arabic sentence (بُنِيَ الْإِسْلَامُ عَلَى خَمْسٍ)

Word
عَمُلَ
إِسْمُ
عَلَمُ
خَلَعَ

The audio preprocessing will be done by software to remove noise and silence, enhance by applying normalize to adjust the peak of amplitude, and then divide the audio into samples. **Figure** ***3-2*** presents a phoneme before and after preprocessing. The audio samples will be converted to wav files with a 16 KHz sample rate and a mono channel in order to be used for training. The CMU Sphinx toolkit (Carnegie Mellon, 2015) will be used for training and testing the model. To give the best results, the application will be configured based on the amount of data that will be trained.


a. Before preprocessing

b. After preprocessing

Figure 3-2 Phoneme Alif (أ) before and after preprocessing

To get the best results for this experiment, the sound samples will be evaluated by experts to determine the mispronounced letters and words. Any mistake in pronouncing will lead to bad results in training and testing for the model. These steps are shown in **Figure** ***3-3***.

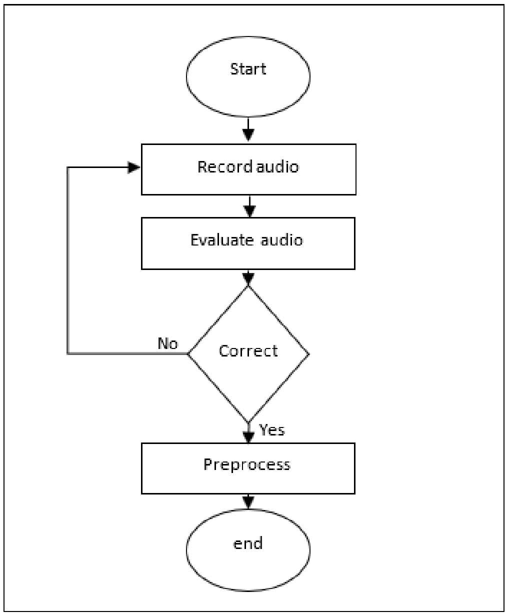
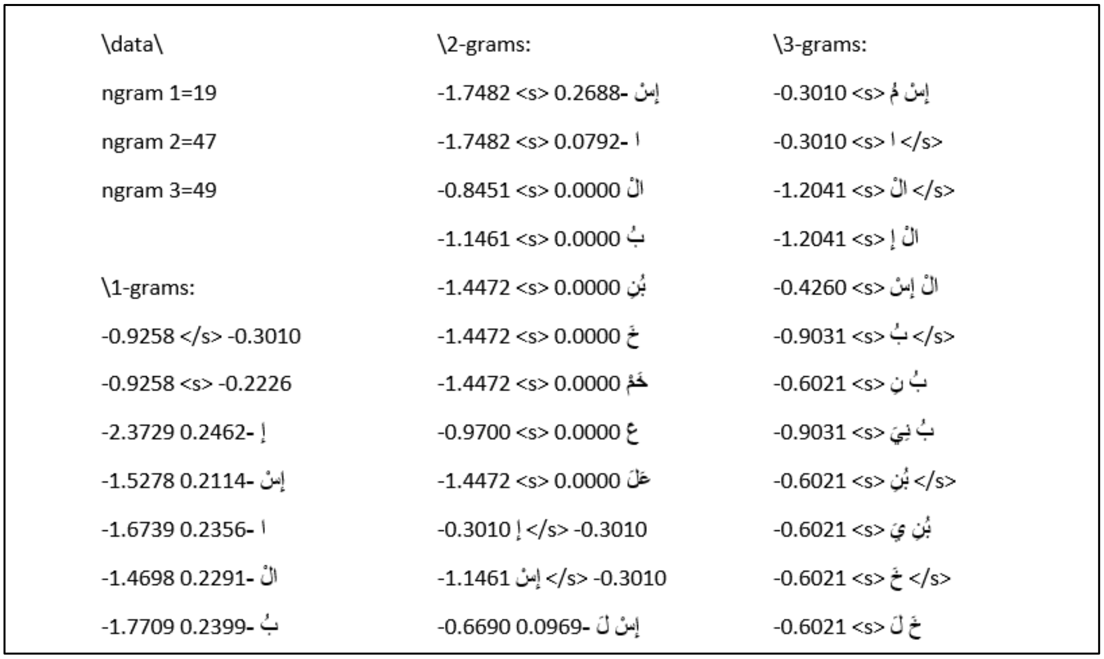


Figure 3-3 Audio sample evaluation process

3.3 Designing Acoustic model and Language model:

The acoustic model presents the phoneme of the language. To build the acoustic model, the phoneme of each letter with different diacritics and two combined letters must be generated using an algorithm.

The language model of this model is based on N-gram. N-gram is used to predict the probability of the next item in a sequence. In this case, n-gram will be used to predict the next letter or compound letters that represent the full word, Sphinx Knowledge Base Tool VERSION 3 (Alexander and Carnegie Mellon, 2010) used for generating n-gram and giving probability based on text data.



```
\data\
ngram 1=19
ngram 2=47
ngram 3=49

\1-grams:
-0.9258 </s> -0.3010
-0.9258 <s> -0.2226
-2.3729 0.2462- إ
-1.5278 0.2114- إسْ
-1.6739 0.2356- ا
-1.4698 0.2291- الْ
-1.7709 0.2399- بُ

\2-grams:
-1.7482 <s> 0.2688- إسْ
-1.7482 <s> 0.0792- ا
-0.8451 <s> 0.0000 الْ
-1.1461 <s> 0.0000 بُ
-1.4472 <s> 0.0000 بُنِ
-1.4472 <s> 0.0000 خَ
-1.4472 <s> 0.0000 خَمْ
-0.9700 <s> 0.0000 ع
-1.4472 <s> 0.0000 عَلَ
-0.3010 إ </s> -0.3010
-1.1461 إسْ </s> -0.3010
-0.6690 0.0969- إسْ لَ

\3-grams:
-0.3010 <s> إسْ مُ
-0.3010 <s> ا </s>
-1.2041 <s> الْ </s>
-1.2041 <s> الْ إ
-0.4260 <s> الْ إسْ
-0.9031 <s> بُ </s>
-0.6021 <s> بُ نِ
-0.9031 <s> بُ نِيَ
-0.6021 <s> بُنِ </s>
-0.6021 <s> بُنِ يِ
-0.6021 <s> خَ </s>
-0.6021 <s> خَ لَ
```

Figure 3-4: N-gram with probability

Figure ***3-4*** presents the n-gram for the Arabic letters in **Table *3-2*** by one, two and three grams; the symbol (<s>) present the silent before the phoneme, and the symbol (</s>) present the silent after the phoneme. Numbers such as (0.3010) represent the probability of the letter or word. If the number is small, it has a high probability; if the number is large, it has a low probability.

3.4 Training the model:

After the preprocessing step comes the model training step. This step has two processes: feature extraction and classification. The feature will be extracted from audio samples and passed to the classifier to match these samples with audio transcription.

3.4.1 Feature extraction:

The feature will be extracted using MFCC. MFCC is the most widely used extraction technique. MFCC will be used to extract features from audio samples for each person.

3.4.2 Classification:

The feature will be passed to the classifier to match the feature with the data from the model. HMM is used for classification. It is based on probabilities; it is used to find the next sequence based on probability(Eddy, 2004). The steps for training the model are explained in **Figure 3-5**.

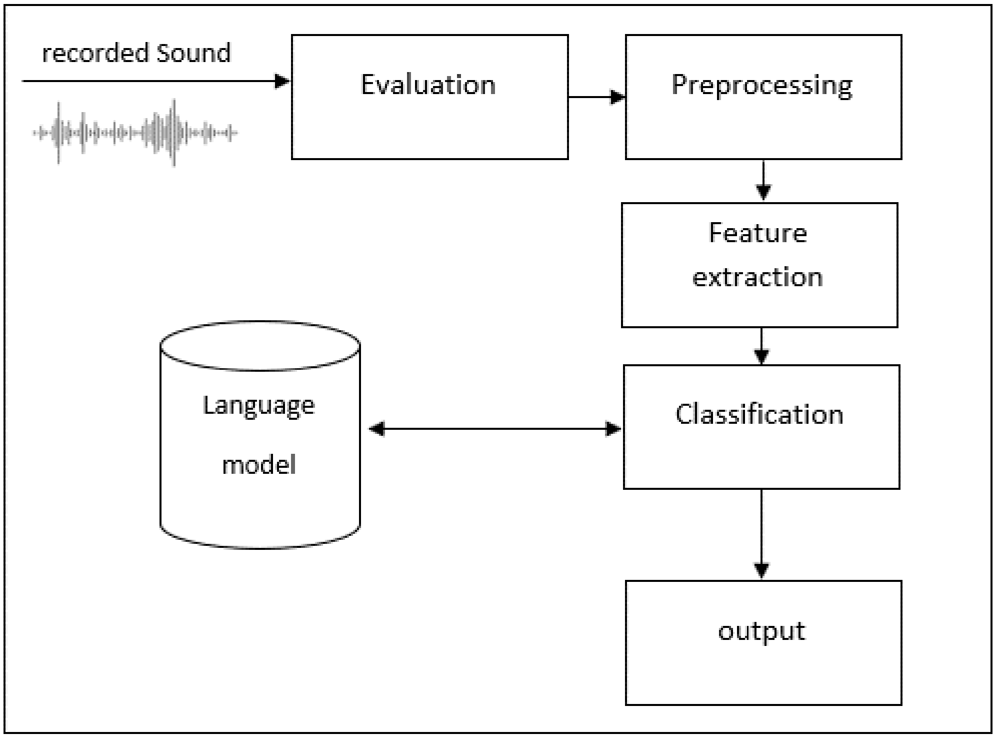


Figure 3-5: training process

The steps of the Qaidah Noraniah model are represented in **Figure 3-6**. The overall process of the training is represented in **Figure 3-7**. The sound sample will be evaluated by an expert to determine if the pronunciation is correct, and then the preprocessing step will be performed to remove noise and silence, enhance the audio sample, and slice the audio into samples. Then the features will be extracted for these samples, and they will be matched with the language model and give the output.

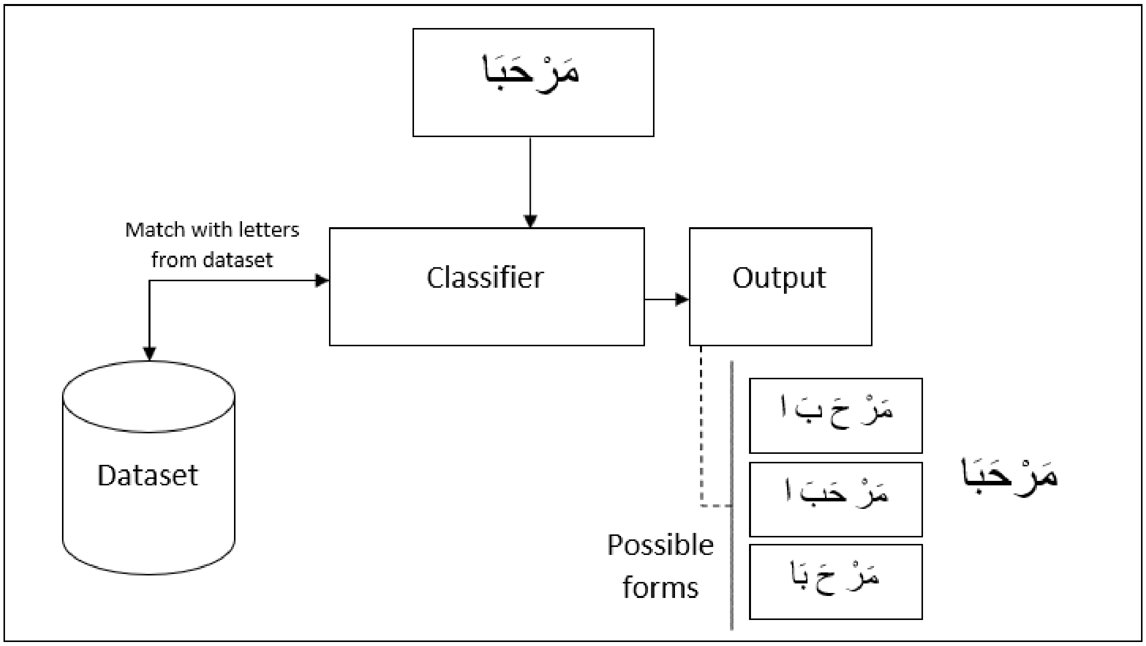


Figure 3-6 Qaidah Noraniah model steps

In **Figure 3-7** by using the letters of the Arabic word (بُنِيَ), instead of only recognizing this word, the system is able to recognize every letter or word composed from these letters. **Table 3-4** contains letters, composed letters and words that can be recognized from this Arabic word. The number of these words can be more based on the numbers of letters in the Arabic word and the correct Arabic words that can be constructed from these letters.

Table 3-4: list of letters and word can be recognized from the Arabic word (بُنِيَ)

Letter	Two letters	Word
بُ	بُنِ	بُنِيَ
نِ	نِيَ	
يَ		

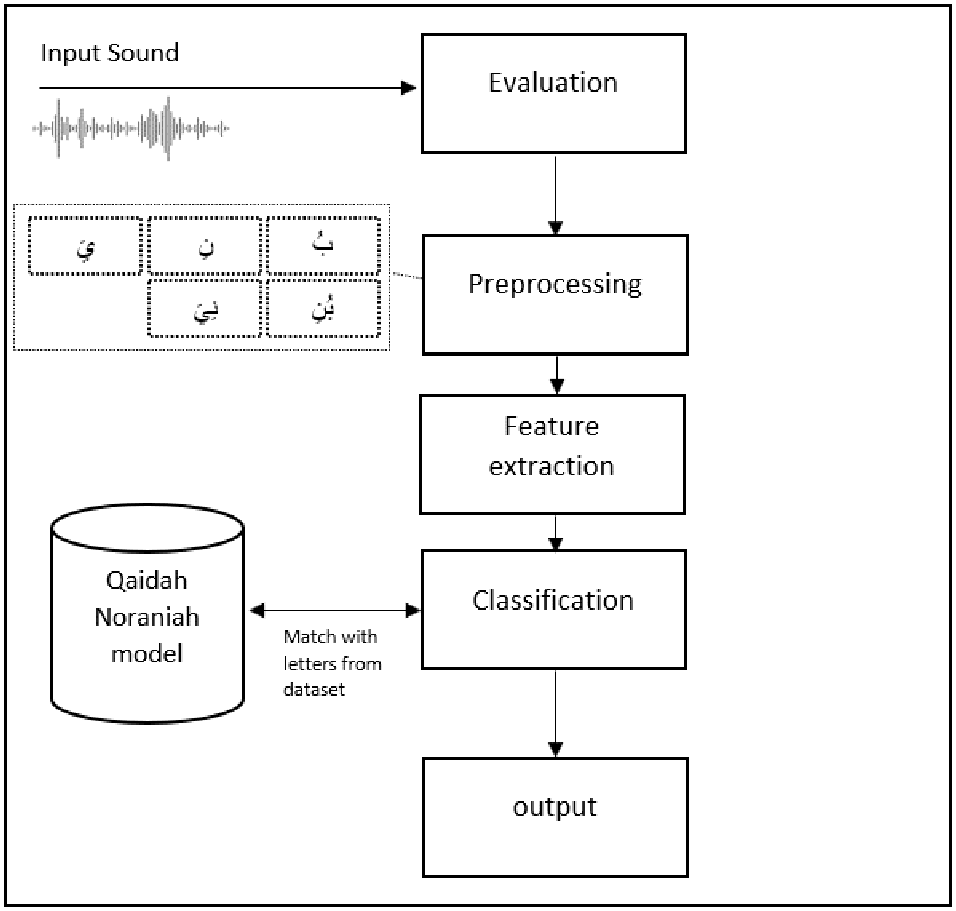


Figure 3-7: Proposed model process

3.4 Summary:

The proposed method is described in this chapter, starting by preparing text and audio data and the step of preprocessing these data, then designing the language model, starting by extracting features from audio files, then classification, and finally showing the illustrations of the proposed method.

CHAPTER 4

RESULT

In this chapter, the results of the model testing will be presented, then discussed, and the study will be compared to other studies in the same field.

4.1 Training the model:

4.1.1 Audio data:

A total of 30 people, 21 male and 9 female, recorded audio data for text data. There ages ranged from 5 to 60 years old. Each person has 10 audio samples. Then preprocessing was done for the samples. The audio preprocessing was done by open source software called Audacity (Audacity, 2017). The audio samples are converted to wav files with a 16 KHz sample rate and mono channel in order to be used for training. The total amount of data is more than half an hour.

4.1.2 Feature Extraction:

For training, the feature will be extracted using MFCC. MFCC will be used to extract features from 4,250 audio samples that have been recorded by 25 people.

For both male and female, each person has 10 recorded audio files. Every file contains 17 phonemes.

4.2 Testing the model:

The model is tested by 10 speakers, 5 dependent speakers and 5 independent speakers (not used to train the model) from both genders. The test is done by recording audio samples for the words, sup-words, and letters in **Table *3-2*** and **Table *3-3***. Each person is asked to record 5 audio samples. The total of audio samples is 1250.

4.3 Results:

The test results of the model for dependent speakers and independent speakers are described in **Table *4-1***. Letters and words are tested separately. The results are measured by WER in the following equation.

Equation 4-1 Word error rate

$$WER = \frac{Number\ of\ errors}{Total\ number\ of\ data}$$

Table 4-1 Test results of dependent and independent speaker

Type	Amount	Word Error Rate (WER)
Letter	17	32.9%
Word	8	68.75%

The results also showed letter recognition. Some letters are hard to recognize because the pronunciation of the letter makes it sound like another letter. In this test, the letter Alif (ا) is recognized in some incorrect letter recognition as the letter Aa (ع). Table ***4-2*** show the error rate percentages of letters pronunciation. Some of the error rates for dependent speakers for letters (عَلَ, يَ) have been observed to be high, owing to mispronunciations and unclear pronunciation.

Table 4-2: Error rate of letters

NO	Letter	Error rate	
		Dependent Speakers	Independent Speakers
1	ا	12%	26%
2	الْ	8%	16%
3	إِ	32%	46%
4	إِسْ	6%	20%
5	بُ	8%	8%
6	بُنِ	2%	2%
7	خَ	12%	18%
8	خَمْ	10%	20%
9	سِ	4%	12%
10	عَ	6%	32%
11	عَلَ	16%	10%
12	لَ	12%	14%
13	لَا	30%	38%
14	مُ	12%	28%
15	نِ	34%	40%

16	نِيَ	14%	0%
17	يَ	8%	2%

On Table ***4-2*** the 1st letter (ا) Alif with diacritic Fatha (ـَ فتحة) and 3rd letter (إِ) Alif with diacritic kasra (ـِ كسرة) are recognized as different letters.

The results for training the words and constructed words are shown in **Table *4-3*** and **Table *4-4***. The error rate percentage is presented for words from both dependent and independent speakers. The error rate of words (بُنِيَ) is higher for dependent speakers due to mispronunciations of some speakers.

Table 4-3: Error rate of words of the Arabic sentence

NO	Word	Error rate	
		Dependent Speakers	Independent Speakers
1	بُنِيَ	6%	4%
2	الْإِسْلَامُ	42%	50%
3	عَلَى	2%	16%
4	خَمْسٍ	4%	10%

Table 4-4: Error rate of constructed words

NO	Word	Error rate	
		Dependent Speakers	Independent Speakers
1	عَمُلَ	40%	50%
2	إِسْمُ	38%	50%

3	عَلَمُ	28%	46%
4	خَلَعَ	30%	50%

On **Table *4-4*** the error rate of words is not high for words not in training data. The error rate of the 4th word is high because the system recognizes the first and second letters as (عَلَ) and not (خَلَ). This is because the pronunciation sounds similar.

Figure *4-1* shows the total error rate of letters and words for dependent and independent speakers.

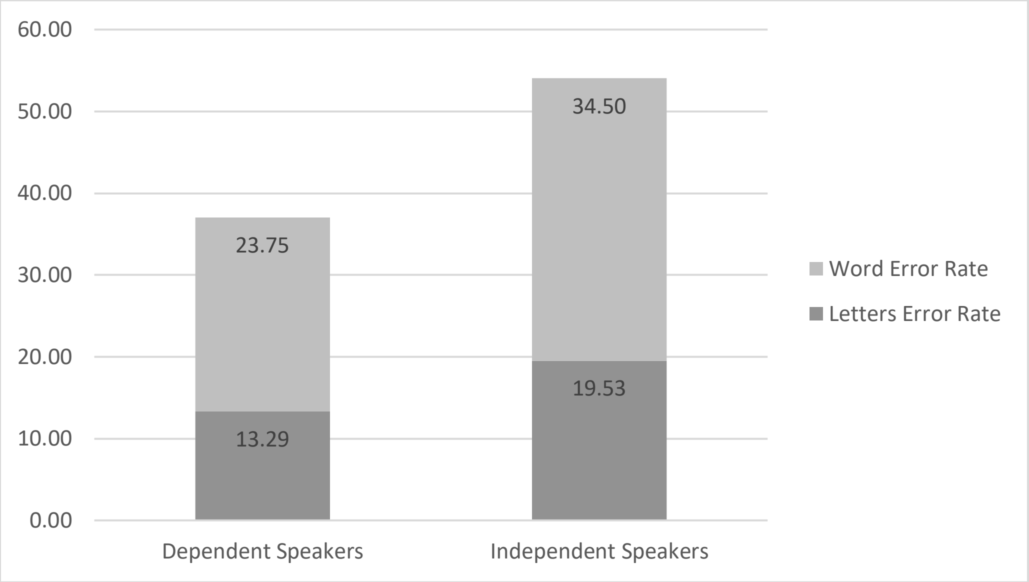


Figure 4-1: Total error rate for dependent and independent speakers

Table *4-5* shows the total number of letters and words used for testing and also shows the error rate for both dependent and independent speakers.

Table 4-5: Total data amount and error rate

Name	Description
Total letters	850
Total words	400
Total data amount	**1250**
Dependent error rate of letters	13.29%
Dependent error rate of words	23.75%
Dependent error rate	**16.64%**
Independent error rate of letters	19.52%
Independent error rate of words	34.5%
Independent error rate	**24.32%**
Total error rate for letters	**32.82%**
Total error rate for words	**58.25%**
Total error rate	**40.96%**

The total error rate is 40.96%. This rate is achieved by re-evaluating the test results. A letter or word can have more than one form. See **Figure *3-1***. The error rate is reduced by giving all types of forms of composed letters or words. This helps in making the recognition process fixable, which reduces the error rate.

Table 4-6: Example of possible forms of word

Word	Possible forms
بُنِيَ	بُ نِ يَ
	بُ نِيَ
	بُنِ يَ

4.3 Discussion:

There are some obstacles to creating the system. These are: some of the audio samples were unclear and the pronunciation was not correct or clear; some people have a problem when they pronounce single or composed letters but they don't have a problem when they pronounce the word. These reasons have affected the recognition rate for the system.

The total recognition rate for the proposed method is 59.04%, but the recognition rate that is achieved for letter recognition is 86.7% for dependent speakers and 80.47% for independent speakers. The total recognition rate for both is 67.17%. In comparison to a system made for recognizing single Arabic letters (Adiwijaya et al., 2017), the system used MFCC and LPC for extracting features for comparison study. The recognition rate achieved by MFCC is 59,87% and by LPC is 78,92%. Although the system is intended to recognize Arabic single letters with diacritics, it has difficulty distinguishing between diacritics for single letters. Another research(Khairuddin *et al.*, 2017) this research is classify the correct pronunciation for Arabic letter with Sukoon (ْ) for both male and female. The system extracted the formant of audio signal and used PSD as a classifier. The total recognition rate is 62%. these comparisons are presented in **Table 4-7**.

Table 4-7: A comparison of the proposed method with other research

Reference	**Feature extraction**	**Classification**	**Amount of data used for training**	**Recognize word from letters**	**Recognition rate**
(Adiwijaya *et al.*, 2017)	MFCC, LPC	KNN	4032 samples from 6 people	no	59,87% by MFCC, 78,92% by LPC
(Khairuddin *et al.*, 2017)	formant	PSD	22 samples from 22 people	no	62%
This research	MFCC	HMM	4250 samples from 35 people	yes	67.17%

This proposed method has the ability to recognize words that are constructed from the letters used for the training, the recognition rate is 76.25% for dependent speakers and 65.5% for independent speakers, the total recognition rate for both is 41.75%, the system achieved this rate although it was not trained by these words. The system also has the ability to distinguish between the pronunciation of diacritics of single letter.

4.3 Summary:

The steps made to train and test the model are mentioned at the start of this chapter. The result of testing the model showed a good result. The recognition rate achieved for letters is 67.17% and 41.75% for words. A comparison was made between this research and other research in the same fields. This research achieved good results and has the ability to recognize words from the letters used for training.

CHAPTER 5

5.1 Conclusion:

Arabic speech recognition systems are facing many problems. There are a lot of factors that affect recognition due to the Arabic language morphology. Some of these factors were explained in the first chapter, and one of them is letter diacritics (الحركات).

By developing the Qaidah Noraniah model, which focuses on the pronunciation of letters with diacritics to help in differentiating between words with the same letters and different diacritics, and by offering a solution by utilizing letter recognition to recognize words, the proposed method seeks to solve some of these issues.

Using the Qaidah Noraniah model for speech recognition proves the ability to recognize words constructed from letters used in training and recognize the difference between diacritics of one letter. However, the total recognition rate of both dependent and independent speakers for letters was 67.17%, which is not great for some reasons; first the amount of audio data used for training, second the number of letters, and finally the speaker's pronunciation of letters. However, the system is able to recognize words constructed from letters used for training. The recognition rate for words was 41.75%.

The working method of the Qaidah Noraniah model for dividing words into letters and using many forms for words helps reduce error rates and gives the ability to recognize single letters, composed letters, words, and words generated from sets of letters.

5.2 Future Work:

A lot of work can be done to reduce the error rate by improving the model, such as increasing the amount of text and audio data and using different techniques for feature extraction or classification.

REFERENCES

Absa, A. H. A. *et al.* (2018a) 'A hybrid unsupervised segmentation algorithm for arabic speech using feature fusion and a genetic algorithm (July 2018)', *IEEE Access*, 6(July), pp. 43157–43169. doi: 10.1109/ACCESS.2018.2859631.

Absa, A. H. A. *et al.* (2018b) 'A hybrid unsupervised segmentation algorithm for arabic speech using feature fusion and a genetic algorithm (July 2018)', *IEEE Access*, 6(July), pp. 43157–43169. doi: 10.1109/ACCESS.2018.2859631.

Adiwijaya *et al.* (2017) 'A comparative study of MFCC-KNN and LPC-KNN for hijaiyyah letters pronounciation classification system', *2017 5th International Conference on Information and Communication Technology, ICoIC7 2017*, 0(c), pp. 2–6. doi: 10.1109/ICoICT.2017.8074689.

Afrillia, Y. *et al.* (2017) 'Performance Measurement of Mel Frequency Ceptral Coefficient (MFCC) Method in Learning System of Al-Qur'an Based in Nagham Pattern Recognition', *Journal of Physics: Conference Series*, 930(1), pp. 0–6. doi: 10.1088/1742-6596/930/1/012036.

Ahmed, B. H. A. and Ghabayen, A. S. (2017) 'Arabic Automatic Speech Recognition Enhancement', *Proceedings - 2017 Palestinian International Conference on Information and Communication Technology, PICICT 2017*, (May), pp. 98–102. doi: 10.1109/PICICT.2017.12.

Al-Sabri, A., Adam, A. and Rosdi, F. (2018) 'Automatic detection of Shadda in modern standard Arabic continuous speech', *International Journal on Advanced Science, Engineering and Information Technology*, 8(4–2), pp. 1810–1819. doi: 10.18517/ijaseit.8.4-2.6813.

Alexander, R. and Carnegie Mellon, U. (2010) *Sphinx Knowledge Base Tool VERSION 3*. Available at: http://www.speech.cs.cmu.edu/tools/lmtool-new.html (Accessed: 15 August 2021).

Alkhatib, B. *et al.* (2017) 'BUILDING AN ASSISTANT MOBILE APPLICATION FOR TEACHING ARABIC PRONUNCIATION USING A NEW APPROACH FOR ARABIC SPEECH', *Journal of Theoretical*, 95(3), p. 8645. Available at: http://www.jatit.org/volumes/Vol95No3/3Vol95No3.pdf.

Audacity, T. (2017) 'Audacity', *The Name Audacity (R) Is a Registered Trademark of Dominic Mazzoni Retrieved from http://audacity. sourceforge. net*. Available at: http://thurs3.pbworks.com/f/audacity.pdf (Accessed: 23 June 2022).

Bengio, Y. (2008) 'Neural net language models', *Scholarpedia*, 3(1), p. 3881. doi: 10.4249/scholarpedia.3881.

AL Bizzocchi (2017) 'How many phonemes does the English language have?', *International Journal on Studies in English Language and Literature (IJSELL) 5*, pp. 36–46. Available at: https://www.joseheras.com/www/pdfs/ijsell/v5-i10/6.pdf (Accessed: 17 June 2022).

Bouchakour, L. and Debyeche, M. (2018) 'Improving continuous arabic speech recognition over mobile networks DSR and NSR using MFCCs features transformed', *International Journal of Circuits, Systems and Signal*

Processing, 12, pp. 379–386.

Carnegie Mellon, U. (2015) ‘CMUSphinx Open Source Speech Recognition’. Available at: https://cmusphinx.github.io/ (Accessed: 23 June 2022).

Carnegie Mellon, U. (no date) *CMU Lexicon Tool*. Available at: http://www.speech.cs.cmu.edu/tools/lextool.html (Accessed: 28 April 2021).

Eddy, S. R. (2004) ‘What is a hidden Markov model?’, *Nature Biotechnology 2004 22:10*, 22(10), pp. 1315–1316. doi: 10.1038/nbt1004-1315.

Haqqani, N. M. (1998) *القاعدة النورانية*. Jeddah: Al Madinah Al Munawarah Printing & Publishing Co.

Khairuddin, S. *et al.* (2017) ‘Classification of the Correct Quranic Letters Pronunciation of Male and Female Reciters’, *IOP Conference Series: Materials Science and Engineering*, 260(1), pp. 0–12. doi: 10.1088/1757-899X/260/1/012004.

Khelifa, M. O. M. *et al.* (2017) ‘An accurate HSMM-based system for Arabic phonemes recognition’, *9th International Conference on Advanced Computational Intelligence, ICACI 2017*, pp. 211–216. doi: 10.1109/ICACI.2017.7974511.

Klaylat, S. *et al.* (2018) ‘Enhancement of an Arabic Speech Emotion Recognition System’, *International Journal of Applied Engineering Research*, 13(5), pp. 2380–2389. doi: 10.1007/s10470-018-1142-4.

Menacer, M. A., Mella, O., Fohr, D., Jouvet, D., Langlois, D. and Smaili, K. (2017) ‘An enhanced automatic speech recognition system for Arabic’, in

Proceedings of the Third Arabic Natural Language Processing Workshop, pp. 157–165.

Menacer, M. A., Mella, O., Fohr, D., Jouvet, D., Langlois, D. and Smaïli, K. (2017) ‘Development of the Arabic Loria Automatic Speech Recognition system (ALASR) and its evaluation for Algerian dialect’, *Procedia Computer Science*, 117, pp. 81–88. doi: 10.1016/j.procs.2017.10.096.

Mesallam, T. A. *et al.* (2017) ‘Development of the Arabic Voice Pathology Database and Its Evaluation by Using Speech Features and Machine Learning Algorithms’, *Journal of Healthcare Engineering*, 2017. doi: 10.1155/2017/8783751.

Mnassri, A., Bennasr, M. and Cherif, A. (2017) ‘GA Algorithm Optimizing SVM Multi-Class Kernel Parameters Applied in Arabic Speech Recognition’, *Indian Journal of Science and Technology*, 10(27), pp. 1–9. doi: 10.17485/ijst/2017/v10i27/114943.

Mohammed, A. (2018) *Noorani Qaida*. Available at: https://mawdoo3.com/ما_هي_القاعدة_النورانية.

Mohammed, A., Sunar, M. S. and Salamb, M. S. H. (2015) ‘Quranic verses verification using speech recognition techniques’, *Jurnal Teknologi*, 73(2), pp. 99–106. doi: 10.11113/jt.v73.4200.

Morsy, H. *et al.* (2018) ‘Automatic Speech Attribute Detection of Arabic Language’, *International Journal of Applied Engineering Research*, 13(8), pp. 5633–5639. Available at: http://www.ripublication.com.

N. M. Arafa, M. *et al.* (2018) ‘A Dataset for Speech Recognition to Support Arabic Phoneme Pronunciation’, *International Journal of Image, Graphics* *and Signal Processing*, 10(4), pp. 31–38. doi: 10.5815/ijigsp.2018.04.04.

Pradana, W. A., Adiwijaya and Wisesty, U. N. (2018) 'Implementation of support vector machine for classification of speech marked hijaiyah letters based on Mel frequency cepstrum coefficient feature extraction', *Journal of Physics: Conference Series*, 971(1). doi: 10.1088/1742-6596/971/1/012050.

Saeed, T. R., Salman, J. and Ali, A. H. (2019) 'Classification improvement of spoken arabic language based on radial basis function', *International Journal of Electrical and Computer Engineering*, 9(1), pp. 402–408. doi: 10.11591/ijece.v9i1.pp402-408.

Saini, P. and Kaur, P. (2013) 'Automatic Speech Recognition : A Review', *International Journal of Engineering Trends and Technology*, 4(iii), pp. 132–136.

Sakka, Z., Techini, E. and Bouhlel, M. S. (2017) 'Using geometric spectral subtraction approach for feature extraction for DSR front-end Arabic system', *International Journal of Speech Technology*, 20(3), pp. 645–650. doi: 10.1007/s10772-017-9433-1.

Salman, J., Saeed, T. R. and Ali, A. H. (2018) 'Improve the Recognition of Spoken Arabic Letter Based on Statistical Features', *Iraqi Journal of Computer, Communication, Control and System Engineering*, 18(3), pp. 26–32. doi: 10.33103/uot.ijccce.18.3.3.

Sardar, V. M. and Shirbahadurkar, S. D. (2018) 'Speaker Identification of Whispering Sound using Selected Audio Descriptors', *International Journal of Applied Engineering Research ISSN*, 13(9), pp. 6660–6666.

Smit, P. *et al.* (2018) 'Character-based units for unlimited vocabulary continuous speech recognition', *2017 IEEE Automatic Speech Recognition and Understanding Workshop, ASRU 2017 - Proceedings*, 2018-Janua, pp. 149–156. doi: 10.1109/ASRU.2017.8268929.

Suresh, M. and Thorat, S. . . (2018) 'Language identification system using MFCC and SDC feature', *media.neliti.com*, (2581), pp. 113–119. Available at: https://media.neliti.com/media/publications/342680-language-identification-system-using-mfc-45bf091f.pdf.

Yasmine, Z. K. *et al.* (2019) 'Improved Speaker Recognition over VoIP using Auditory Features', *Proceedings of IEEE/ACS International Conference on Computer Systems and Applications, AICCSA*, 2018-Novem. doi: 10.1109/AICCSA.2018.8612835.

Yousaf, K. *et al.* (2018) 'A Novel Technique for Speech Recognition and Visualization Based Mobile Application to Support Two-Way Communication between Deaf-Mute and Normal Peoples', *Wireless Communications and Mobile Computing*, 2018, pp. 1–12. doi: 10.1155/2018/1013234.

Yousfi, B., Zeki, A. M. and Haji, A. (2018) 'Holy Qur ' an Speech Recognition System Distinguishing the Type of prolongation', 2(1).

Zarrouk, E. and Benayed, Y. (2016) 'Hybrid SVM/HMM model for the arab phonemes recognition', *International Arab Journal of Information Technology*, 13(5), pp. 574–582.

Zerari, N. *et al.* (2018) 'Bi-directional recurrent end-to-end neural network classifier for spoken Arab digit recognition', *2nd International Conference on Natural Language and Speech Processing, ICNLSP 2018*, (April), pp. 1–6. doi: 10.1109/ICNLSP.2018.8374374.

APPENDICES

APPENDICES A: LIST OF ARABIC TERMS

Arabic term	description
الحركات	the diacritics of letters
المد	the prolongation of pronunciation
الشدة	emphasis in the pronunciation
القاعدة النورية	A way of teaching Arabic language
القاعدة النورانية الفتحية المطورة	A way of teaching Arabic language
فتحة ـَ	One of Arabic diacritics
كسرة ـِ	One of Arabic diacritics
ضمة ـُ	One of Arabic diacritics
سكون ـْ	One of Arabic diacritics
RA (ر)	Arabic letter
Alif (ا)	Arabic letter