MODELING FOR RECOGNIZING RECITATION TYPE OF THE HOLY QUR'ĀN

BY

BILAL YOUSFI

A thesis submitted in fulfilment of the requirement for the degree of Doctor of Philosophy in Computer Science

Kulliyyah of Information and Communication Technology
International Islamic University Malaysia

SEPTEMBER 2021

ABSTRACT

Computer science and speech recognition have enjoyed a long and fruitful relationship for decades. Speech recognition has been beneficial for capturing and producing an accurate transcription of spoken words. In computer science, a prime challenge is to interpret these signals into meaningful data and to develop algorithms and applications to establish an interface between the human's voice signal and computer. Moreover, the major concerns of automatic speech recognition (ASR) are determining a set of classification features and finding a suitable recognition model for these features. Hidden Markov Models (HMMs) have been demonstrated to be powerful models for representing time-varying signals. The act of reading Qur'ān and pronouncing its sound dwells on the type of recitation. These are referring to the recitation of Warsh or the recitation of Hafss. According to the science of Qira'āt, it is essential to recognize the type of recitations, especially with the diversity and the spread of Qira'āt in the world. There are numerous efforts made by previous systems on the development of feasible guiding techniques to the act of reading the Holy Qur'ān (Tajweed rules). Unfortunately, liking the major control variables of the practices of both Usūl al Qira'āh (general principles) and Farsh al-huruf (specific variants) in those approaches were neglected. In order to fill this gap, this research thesis attempts to design and fabricate a speech recognition system that distinguishes the types of recitations (Qira'āt of Hafss An Assim and the Qira'āt of Warsh An Naafi') while reciting the Qur'ān. The proposed system is capable of recognizing, identifying, pointing out the mismatch and discriminate between two types recitations for Hafss and Warsh. An experiment among user's recitation for Hafss and Warsh with the recitation made by the expert Qur'ān reader stored in a database has been done. This thesis investigates acoustic models based on the Hidden Markov Models (HMM) classifier together with clustering algorithm for Qur'ān Speech Recognition. A significant improvement on the recognition performance was achieved when the HMM-clustering model was implemented compared to the baseline model's (single HMM model (conventional MFCC)) result. The results show that the proposed model has a faster ability for recognizing phonemes sequences than the (conventional MFCC model. Model. The adoption of the k-means algorithm for acoustic modeling is seen to be a more valid model for acoustic modeling speech recognition. However, our developed system shows a lower performance in some instances when it was compared to the other systems recently reported in the literature that used the same data. This due to the small size of training dataset used for this research, hardware availability and noise from the environment and from the speakers which can affect the improvement the results in this thesis as our aim is to investigate the proposed models for speech recognition and to make a direct comparison between these models.

خلاصة البحث

يتمتع علم الحاسوب وتقنية التعرف على الكلام بعلاقة وطيدة منذ عقود. حيث تعتبر هذه التقنية مفيدة في التقاط وإنتاج نسخة دقيقة من الكلمات والمقاطع والأصوات المنطوقة. حيث يكمن التحدي الرئيسي في علم الحاسوب في ترجمة هذه الإشارات وتحويلها إلى بيانات ذات معنى ومفهومه للحاسب لتطوير خوارزميات وتطبيقات لإنشاء حلقة وصل بين الإشارة الصوتية للإنسان ولغة الكمبيوتر. علاوة على ذلك، فإن الاهداف الرئيسية للتعرف التلقائي على الكلام (ASR) هي تحديد مجموعة من ميزات التصنيف وإيجاد نموذج للتعرف عليها، حيث أثبتت نماذج ماركوف المخفية (HMM) أنها نماذج قوية لتمثيل إشارات الصوتية وتصنيف ميزاتها. تعتمد طريقة نطق آيات القرآن الكريم وإخراج حروفه ولفظه الصوتي على نوع القراءة. القراءات القرآنية من أهم علوم القرآن، حيث يعتبر من الموضوعات الشديدة الصلة بنص القرآن الكريم، لأنه يعنى بكيفية النطق بألفاظ القرآن، وتحقيق الروايات المنقولة في ذلك عن أئمة القراءة. من الضروري التعرف على نوع القراءة المتبعة عند قراءة القران، لا سيما مع تنوع وانتشار القراءات في العالم. هناك العديد من الجهود التي بذلتها النظم السابقة في تطوير تقنيات توجيه قراءة القرآن الكريم (قواعد التجويد). لسوء الحظ، لم تعالج هذه الابحاث أحكام علم القراءات من أصول القراءات (المبادئ العامة) وفرش الحروف (المتغيرات المحددة). من أجل سد هذه الفجوة، تحاول هذه الدراسة البحثية تصميم نظام للتعرف وتمييز نوع القراءة (قراءة حفص عن عاصم وقراءة ورش عن نافع) من خلال قراءة القرآن عن طريق استعمال تقنية التعرف على الكلام. النظام المقترح قادر على التعرّف على عدم التطابق وتحديده والتمييز بين نوعين القراءة، حيث تم إجراء تجارب بين قراءات المشتركين مع اخرى لقراء مشاهير مخزنة في قاعدة البيانات. كما تبحث هذه الأطروحة عن مدى جدوى استعمال النماذج الصوتية المستندة إلى مصنف نماذج ماركوف المخفية (HMM) جنبًا إلى جنب مع خوارزمية التجميع clustering للتعرف على الكلام في القرآن. اشارت النتائج ان النظام حقق تحسن كبير في أداء التعرف باستعمال نموذج HMM-Clustering مقارنتا بنموذج MFCC - HMM التقليدي. علاوة على ذلك، اشارت النتائج الى أن النموذج المقترح لديه قدرة أسرع على التعرف على عينات الكلمات مقارنتا بنموذج MFCC التقليدي. كما بينت الدراسة ان اعتماد خوارزمية k- means للنمذجة الصوتية يُعتبر نموذجًا أكثر فعالية للتعرف على الكلام في النمذجة الصوتية. غير ان النظام المطوّر اظهر أداءً أقل جودة في بعض الحالات عند مقارنته بالأنظمة المنشورة سابقا والتي تستخدم نفس البيانات. هذا نظرًا لصغر حجم مجموعة البيانات التدريبية المستخدمة في هذا البحث ولوجود الضوضاء من البيئة ومن مكبرات الصوت والتي يمكن أن تؤثر سلبا على تحسين النتائج في هذه الرسالة .

APPROVAL PAGE

The thesis of Student's Name has been approved by the following:

_____
Akram M Z M Khedher
Supervisor

_____
Abdul Wahab Abdul Rahman
Internal Examiner

_____
Syed Ahmad Sheikh Aljunid
External Examiner

_____
Shahrul Na'im Sidek
Chairman

DECLARATION

I hereby declare that this dissertation is the result of my own investigations, except where otherwise stated. I also declare that it has not been previously or concurrently submitted as a whole for any other degrees at IIUM or other institutions.

Bilal Yousfi

Signature ..... Date .....

INTERNATIONAL ISLAMIC UNIVERSITY MALAYSIA

DECLARATION OF COPYRIGHT AND AFFIRMATION OF FAIR USE OF UNPUBLISHED RESEARCH

MODELING FOR RECOGNIZING RECITATION TYPE OF THE HOLY QUR'ĀN

I declare that the copyright holder of this thesis is jointly owned by the student and IIUM.

By signing this form, I acknowledged that I have read and understand the IIUM Intellectual Property Right and Commercialization policy.

Affirmed by Bilal Yousfi

……..…….. ……..

Signature Date

*This thesis is dedicated to my beloved parents (Abdullah and Sassia)***.**

To

My dearest brothers (Mohammed Ridha and Alaeddine), sisters (Nadjah, Amira and Chaima), nephews (Abdelbaset Abdisamad, Ayoub and Jawad), and nieces (Bayan, Maria and Rosline)

For their love, prayers, encouragement, perseverance, and faith in me.

ACKNOWLEDGEMENTS

In the name of Allah, the Most Gracious, the Most Merciful. He, who has given me the strength to continue doing this research. Alhamdulillah, All Praise to Allah that he has answered my prayers to submit this thesis with the hope to pass with excellence

All the languages of the world cannot produce words that thank my parents for their great support to my education, eventually resulting in successful completion of this work. They were the cause, and the inspiration without which this work would have never been completed. I wish a happy long life for my parents; without them I simply would not be where I am now.

I would like to express my deep gratefulness, admiration, and honor to my supervisor Dr. Akram M Z M Khedher. With such richness in experience, knowledge, and dedication, he supported and provided me with the direction to complete this work; always having time to listen, always prepared to help and always open to discussion. I must present my thanks to him for his advice, valuable feedback and suggestions and support during the research journey.

Likewise, I am greatly indebted to my brothers and sisters, Mohammed Radha, Nadjah, Asma, Chaima, for their continuous support and words of encouragement throughout the course of this Ph.D. study. This thesis would not have been completed without their help and support.

I would also like to thank my nephew and nieces, Bayan, Abdelbaset Abdessamed, Ayoub, Djouad, Maria, and Rosline.

I would also like to thank KICT Post Graduate Office Members, Kyra, Narita, Izzati, and Sarah for their support and help.

In the same vein, I am equally grateful to all my colleagues and friends at the IIUM for providing soundboards and light relief.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

List of Abbreviations

AI	Artificial intelligence
ASR	Automatic Speech Recognition
DFT	Discrete Fourier transform
DTW	Dynamic Time Warping
FIR	Finite Impulse Response
HMM	Hidden Markov model
LPC	Linear Prediction Coding
MFCC	Mel-Frequency Cepstral Coefficients
MSA	Modern Standard Arabic
WRA	Word Recognition Accuracy
ZCR	Zero Crossing Rate

CHAPTER ONE

INTRODUCTION

1.1 OVERVIEW

Allah Almighty said: "And make loose the knot (the defect) from my tongue, (i.e., Remove the incorrectness from my speech) " [20:27].

﴿وَٱحۡلُلۡ عُقۡدَةٗ مِّن لِّسَانِي٢٧﴾

Language can be described as a system of speech sounds, words, and patterns used by humans to communicate their thoughts and feelings (Goldstein, 2014). In other words, speech is a natural communication method between humans and is a meaningful, useful and the fastest form of data input between humans and technology. That is why there are a different variety of languages used in verbal communication. However, it is estimated that there are almost 16 languages which are spoken by more than 500 million people, i.e., English, Arabic, Chinese, Italian, Spanish, Russian, French, Hindi, Portuguese, Bengali, German, Japanese and Urdu.

The Arabic language is considered one of the most ancient living and Semitic languages in the world and it is also the Fifth most generally used language. The Arabic language is the mother tongue for roughly 240 million people (Garcia, 2018), and the Arabic language is linked with Islam since it is the language of the Qur'ān, and as well Arabic is the language that is used in most of the Islamic literature.

The Qur'ān is the Speech (Kalaam) of Allah Almighty, which he revealed to The Prophet (ﷺ) in wording and meaning, and which has been preserved in the Mushafs, and has reached us by Mutawātir transmissions.

As mentioned in the Holy Qur'ān itself, the Arabic language is the language of the Qur'ān, and with it, the Qur'ān was revealed to the Prophet (ﷺ). The knowledge of the Arabic language is essential for every Muslim so that they can perform their religious acts of worship and they can be proficient in the recitation of the Qur'ān. Allah Almighty says in His Book:

إِنَّا **أَنزَلْنَاهُ** قُرْآنًا عَرَبِيًّا لَّعَلَّكُمْ تَعْقِلُونَ(2)

"Verily. We have revealed this as an Arabic Qur'ān" [12:2]

وهذا لِسَانٌ عَرَبِيٌّ مُّبِينٌ (103)

.... this (the Qur'ān) is in a clear Arabic tongue" [16:103]

وَكَذَلِكَ أَوْحَيْنَا إِلَيْكَ قُرْآنًا عَرَبِيًّا... (7)

«And thus, we have inspired you with an Arabic Qur'ān… [142:7].

The practice of reciting the Qur'ān differs from the normal reading of Arabic text, and it is performed according to a set of guidelines known as Tajweed. Therefore, learning how to recite the Qur'ān correctly is an obligatory act upon every Muslims, as in Sahih al-Bukhari Hadith number 5027: The Prophet (ﷺ) said, "The best among you (Muslims) are those who learn the Qur'ān and teach it."(Al-Bukẖārī & Ḫān, 1997). Many people, especially beginners or new learners have difficulties in reading the Qur'ān base on the designed rules of recitation. The Majority of them mismatch their act of recitations, from the basic prescribed ones and this is most likely caused by their lack of oral practices and monitoring. A research finding from (Mohd, 2006) states that "most Muslims are weak in reading the Qur'ān in the field of Islamic education for some time". Furthermore, some Muslims only practices reading the Qur'ān during their lesson in school as this could be attributed to the lack of resources that can ease their recitations

skills and encourage them to learn the recitation types according to the rules governing those styles.

There are seven accepted readings in the system of Qira'āt. The number seven is based on a well-known hadith of several variants, in the Hadith that was narrated by al-Bukhari, 3047; Muslim, 819: The Prophet (ﷺ) said: "The Qur'ān has been revealed in seven different ways (modes (Ahruf)), so recite it whichever is easiest for you ". (Al-Bukẖārī & Ḫān, 1997) , which means that the Holy Qur'ān was revealed in seven variants Ahruf (dialect), the Ahruf which refers to different dialects among the Arabs at the time of the revelation of the Qur'ān and the term Qira'āt (recitations) usually denotes the accepted variant Ahruf of the Qur'ān. Moreover, the word Qira'āt is the variations in words and pronunciations of the Qur'ānic verses (Qadhi, 2003). In another definition, Qira'āt refers to the changes which occur in the words of the Qur'ān with regard to the prolongation (Muddud), shortening (Kassar), the punctuation of the written text and the pronunciation of the Qur'ānic words. (Dogan, 2014)

In the 4th century of Hijra, Abū Bakr b. Mujāhid (d. 324/936) standardized the number of recitations to seven acts of Qira'āt and these are recognised as the seven well-known reading modes of the Holy Qur'ān. The seven readings that were accepted in Ibn Mujāhid's time as the accepted Qira'āt includes: Ibn Kathīr (Mecca, d. 120/738), Nāfi` (Medina, d. 169/785), Ibn `Āmir (Damascus, d. 118/736), Abū `Amr (Baṣra, d. 154/770), `Assim (Kūfa, d. 127/745), Ḥamza (Kūfa, d. 156/773), and al-Kisā'ī (Kūfa, d. 189/804). In the Qira'āt science today, the most popular readings are those transmitted by Hafss (d. 180/796) on the authority of `Assim and Warsh (d. 197/812) on the authority of Nafi. (Dogan, 2014).

Computer science and speech recognition have enjoyed a long and fruitful relationship for decades. Automatic Speech Recognition (ASR) is a field of computer

science which deals with designing computer systems that can recognise spoken words or take dictation which involves the ability to match a voice pattern against a provider or acquired vocabulary (Abushariah & Gunawan, 2011). According to Rabiner (Rabiner & Schafer, 1978), Speech recognition is the process of converting or translating an acoustic signal, captured by the microphone to a set of textual words. A word is recognised by extracting features from the captured signal and classifies the features with the given voice sample in the database. One of the earliest researches of ASR is in Human-Computer Interaction (HCI) such as speech to speech translation (Lavie et al., 1997; Wahlster. 2000), Computer-aided language learning (Witt and Young, 1997; Xu et al., 2009; Peabody. 2011), dictation (Murveit et al., 1993; Lee, 1997; Lee et al., 2009), and voiced based information retrieval (Franz and Milch, 2002; Zue et al., 2000) etc.

Currently, there have been numerous researches related to Al-Qur'ān and Arabic speech; Such as speech recognition for Tajweed rules has been found to focus on Automated Tajweed checking rules engine for Qur'ānic learning (Raja-Jamilah Raja-Yusof and Fadila Grine, 2013), as well as Tajweed checking system to support recitation (Ahsiah, Noor, & Idris, 2013). There are other studies which also streamline Qur'ānic Verse recitation recognition module for support in J-QAF learning (abbreviation to Jawi, Qur'ān, Arabic, Fardu Ain, which is a religious learning approach in Malaysian schools), and Makhraj recognition using speech processing (Wahidah et al., 2012) in order to isolate the right sound speech for the right Qur'ān words.

In Consistent with the previous studies, the goal of this research was to investigate the relationship between reciting the Qur'ān and understanding the type of recitation using spectrogram voice analysis. To accomplish this goal, new speech recognition model for Qur'ān have been developed based on the two famous Qira'āt

which are the Qira'āt of Hafss An Assim (Hafss from Assim) and the Qira'āt of Warsh An Naafi' (Warsh from Naafi').

This study aims to deliver types of recitation identification system that uses pattern recognition and classification methods as this would assist people who wish to practice reading the Qur'ān following one of the two famous Qira'āt, while constituting a step toward developing a machine interface that can respond to the human type of recitation following during the learning of the Qur'ān.

1.2 BACKGROUND ABOUT SPEECH RECOGNITION

Speech recognition systems, which come under the natural language processing umbrella, in particular, have witnessed a significant breakthrough based on the advances in Machine Learning (ML) and it can be categorized into two modes. The first is isolated word recognition which is the simplest speech recognition mode, where each word is surrounded by silence, so that word boundaries are well known. The second is the continuous speech recognition which is more natural than the isolated mode. This mode is more difficult than the isolated word recognition because the word boundaries are difficult to recognise (detecting the beginning and the end of each word). In speech conversation, a word may be uttered in a different way due to the variations in dialects, pronunciations, accents, volumes, gender and age. (Tebelskis, 1995). Also, the speaker variability such as emotion and illness is combined with the attention to the environment. Speech recognition systems are also influenced by variations between the environments which can introduce corruption into the speech signal as a result of background noise, transmission channels, and microphone characteristics. (Jou, 2008; Yuk, 1999)

The initial speech recognition stage is the pre-processing where recording, speech signals are passed through the pre-processing block for determining the correct boundary of the isolated words and rejecting the artifacts of speech such as noise and intra-word stops. With regards to large vocabulary continuous speech boundary detection, the problem gets to be much more difficult due to the intra-word silences and other artifacts. Generally, these problems are reduced by applying speech boundary detection algorithms. The most prevalent methods used for endpoint detection are energy profile and Zero Crossing Rate (ZCR). (Chen, 1988).

The most prominent phase in speech recognition is feature extraction which is designed to extract unique, robust, discriminative, and computationally efficient characteristic from the speech signals. The most common feature extraction techniques are Mel Frequency Cepstral Coefficients (MFCC) (S. Davis & Mermelstein, 1980), perceptual linear prediction coefficients (PLR) (Hermansky, 1990), and Fast Fourier Transform (FFT) (Van Loan, 1992). Typically, the ASR system represents the speech signal with state-of-the-art Mel Frequency Cepstral Coefficients (MFCCs).

The main concerns in the automatic speech recognition (ASR) are to Figure out a set of classification features and finding a suitable and appropriate model for these features. Hidden Markov Models (Rabiner, 1989), which used a special case of regular Markov models, have been demonstrated to be a powerful model for representing time-varying signals as a parametric random process (X.D. Huang et al., 2001, Rabiner, 1989). Artificial Neural Networks (ANNs) such as Multi-Layer Perceptron (MLP) (Bengio, 2009; Glorot & Bengio, 2010), Deep Neural Networks (DNN) (Glorot & Bengio, 2010), Recurrent Neural Networks (RNN), and Echo State Network (ESN) (Jaeger, 2002), have also been widely used for representing time-varying quasi-stationary signals.

Despite the use of the above techniques, current speech extraction and recognition techniques encounter errors in recognition accuracy rates and delay in processing time. An alternative is to perform the recognition process after using a clustering algorithm to classify the feature vectors, and the HMM is trained to recognise it in the same manner as the other patterns.

1.3 PROBLEM STATEMENT

The act of reading and pronouncing the Qur'ān dwells on the type of recitation as these refers to the recitation of Warsh or Hafss. It is very important to recognise the type of recitations, especially with the diversity and the spread of Qira'āt in the world (Qadhi, 2003). People are having difficulties in reading the Qur'ān with the correct rules of recitation type where they cannot recognise which recitation type they are following when reciting the Qur'ān due to lack of oral practices and monitoring.

Topics associated with the process of learning and memorization of the Qur'ān have been a subject of interest to several researchers. Al- Qur'ān words are believed by Muslims to be the words of Almighty Allah, which have distinctive characteristics and has been revealed in its precise meaning and wording through the Angel Gabriel. Moreover, The Qur'ān provides some instruction about how to perform its recitation, as in

﴿... وَرَتِّلِ الْقُرْآنَ تَرْتِيلًا ٤﴾ [المزمل: 4]

«…, and recite the Qur'ān with measured recitation (4) » [Al-Muzzammil: 4]

In spite of the large amount of research on the Qur'ān and the act of learning and memorizing the Qur'ān in proper Qur'ānic recitation, an analysis of studies and sources conducted in literature (see section 03.06) indicate that—unfortunately—very little scientific research is available concerning the recitation types of the Qur'ān.