REAL-TIME AUDIO-BASED TRAINING SYSTEM FOR PROPER QUR'ANIC LETTER PRONUNCIATION

BY

TAREQ M. K. ALTALMAS

A thesis submitted in fulfilment of the requirement for the degree of Doctor of Philosophy (Engineering)

Kulliyyah of Engineering
International Islamic University Malaysia

JULY 2021

ABSTRACT

Recitation of the ***Quran*** is an essential activity for every Muslim to understand the message from Allah to His servant. ***Al-Quran*** is written in the Arabic language, and it is important to recite it as it was written, based on what has been practiced by Prophet Muhammad *s.a.w*. However, this task is a great challenge, especially for those of non-Arab descent. The face-to-face traditional and prevalent method of teaching and learning the ***Quran*** with ***Tajweed*** rules starts at early ages and is time-consuming as it requires extensive practice sessions with a qualified teacher. The teacher can only see the recitation or the pronunciation of the student by looking at the face correctly and by listening, then making the corrections based on his experiences immediately. In learning the ***Quran***, knowing the unique articulation point (***Makhraj***) and special characteristics (***Sifaat***) are the basic but essential things emphasised significantly. Although the traditional method in Quranic teaching and learning is accepted worldwide, particularly in Muslim populations, this demand for qualified teachers may not be fulfilled in many places. This issue can be overcome with an efficient learning platform to complement the existing conventional technique employing the computer and technology. Previous literature shows no similar approach highlighting the efficient Quranic learning system focusing on the basic ***Makhraj*** and ***Sifaat*** but instead concentrating on the accuracy of the Quranic verses as a whole that leads to the unsolvable problem of ***Tajweed***. Therefore, this research embarks on developing real-time Quranic teaching and learning interactive platform, known as Computer-assisted pronunciation training (CAPT) systems, to serve as a complementary tool to systematically help Muslims recite the ***Quran,*** as a significant solution for ***Tajweed*** teaching and learning. The research was started with modelling the correct pronunciation of each letter based on the speech acoustic recorded from the experts in the ***Quran***. Then the investigation of the combinations of the unique features of each letter concerning ***Makhraj*** and ***Sifaat*** was conducted. For ***Sifaat*** features representation, the results showed that the combinations of Mel-frequency cepstral coefficients (MFCC), and perceptual linear prediction coefficients (PLP) were the best for identifying the ***Sifaat*** of the Quranic letters. On the other hand, the combination of Mel-frequency cepstral coefficients (MFCC) and the linear prediction cepstral coefficients (LPCC) was the best way to identify the ***Makhraj*** of the Quranic letters. The process was continued with the analysis of ***Sifaat*** then ***Makhraj***, where weighted k-nearest neighbours (KNN), medium Gaussian support vector machine (SVM), and random under-sampling boosted trees (RUSBoosted) were selected for different conditions. In this research, five classification models were developed to evaluate five pairs of the ***Sifaat***. Another four classification models were developed to evaluate the main four ***Makhraj*** of the Quranic letters. Once the representation of the letter was completed and ready, Matlab Application Designer was used to design and build the real-time Quranic teaching and learning platform. It is very important for the proposed system to successfully work in real-time as it should mimic the process of the conventional Quranic teaching and learning where the learning happens in real-time, with immediate feedback is given to the student for improvement. To ensure the consistency of the system's accuracy, two levels of evaluation were conducted. The first one was to test each classifier alone with a new dataset, where the classifier

models have shown a good performance ranging from 70% to more than 90%. The second evaluation was conducted on the real-time system developed, where the results of the system were compared to the evaluation of human experts. The system's accuracy score was good, where the accuracy was about 88%. The system's ability to identify good pronunciation has also outperformed with 92%, and its ability to categorise the wrong pronunciation as "incorrect "was good, where the accuracy was about 76%. Therefore, the system developed has successfully represented the correct pronunciation of all of the Quranic letters based on ***Makhraj*** and ***Sifaat*** on an interactive computer-assisted pronunciation training, which will be a significant platform as a complementary tool for conventional ***Tajweed*** teaching and learning.

ملخص البحث

تلاوة القراءة ضرورة لكل مسلم لفهم الرسالة السماوية، فلغة القرآن هي اللغة العربية ومن الضروري أن يُتلى بالعربية كما تلاه النبي محمد صلى الله علي وسلم، إن قراءة القرآن بالطريقة المثلى لهي مهمة عظيمة خاصة لمن لغتهم الأم ليست العربية. طريقة التلقي (وجه لوجه) التقليدية والسائدة في تعليم القرآن وأحكامه تبدأ في سن مبكرة، وتستغرق وقتًا طويلاً، وتتطلب جلسات تدريب مكثفة مع معلم مؤهل، بينما يستطيع المعلم فقط رؤية التلاوة أو نطق الطالب من خلال النظر إلى الوجه بشكل صحيح والاستماع الشامل، ثم إجراء التصحيحات بناءً على تجاربه على الفور. من الأساسيات المهمة في تعلم القرآن والتي يتم التأكيد عليها بشكل كبير هي معرفة مخارج الحروف والصفات. وعلى الرغم من أن الطريقة التقليدية في تعليم القرآن وتعلمه مقبولة في جميع أنحاء العالم، لا سيما بين المسلمين ، إلا أن هذا الطلب على المعلمين المؤهلين قد لا يتم تلبيته في العديد من الأماكن، يمكن التغلب على هذه المشكلة من خلال منصة تعليمية فعالة لمساندة الطريقة التقليدية الحالية، عن طريق الكمبيوتر والتكنولوجيا. تُظهر الدراسات السابقة أنه لا يوجد نظام يُسلط الضوء على أساسيات المخارج والصفات، بل يركز على دقة الآيات القرآنية ككل مما يؤدي إلى مشكلة لم تحل في فحص التجويد. لذلك، يشرع هذا البحث في تطوير منصة تفاعلية للتعليم والتعلم القرآني في الوقت الفعلي، والمعروفة باسم أنظمة التدريب على النطق بمساعدة الكمبيوتر (CAPT) لتكون بمثابة أداة تكميلية لمساعدة المسلمين على تلاوة القرآن بشكل ممنهج كحل مهم لتعليم مهارة التجويد. بدأ البحث بنمذجة النطق الصحيح لكل حرف بناءً على النطق الصوتي المسجل من خبراء القرآن. ثم تم التحقيق في توليفات السمات المميزة لكل حرف فيما يتعلق بالمخارج والصفات. بالنسبة لتمثيل الصفات، أظهرت النتائج أن توليفات معامِلات (MFCC) و (PLP) كانت الأفضل في التعرف على صفات الحروف القرآنية. من ناحية أخرى ، كان الجمع بين معاملات (MFCC) و (LPCC) هو الأفضل للتعرف على مخارج الحروف القرآنية. استمرت العملية بتحليل الصفات ثم المخارج، حيث تم اختيار خوارزمية أقرب الجيران KNN، متجه الدعم الآلي SVM و الاشجار المعززة RUSBoosted للمصنفات المستخدمة في الحالات المختلفة. في هذا البحث تم تطوير خمسة نماذج تصنيف لتقييم خمسة أزواج من الصفات. تم تطوير أربعة نماذج تصنيف أخرى لتقييم المخرجات الأربعة الرئيسية للحروف القرآنية. بمجرد الانتهاء من نماذج التصنيف لمخارج والصفات، تم استخدام مصمم تطبيق MATLAB لتصميم وبناء منصة تعليمية للقرآن في الوقت الفعلي. من المهم جدًا أن يعمل النظام المقترح بنجاح في الوقت الفعلي لأنه يجب أن يحاكي

عملية التدريس والتعلم القرآني التقليدي حيث يحدث التعلم في الوقت الفعلي، مع تقديم ملاحظات فورية للطالب للتحسين. لضمان دقة النظام، تم إجراء مستويين من التقييم. الأول كان اختبار كل مصنف بمفرده باستخدام مجموعة بيانات جديدة، حيث أظهرت نماذج المصنف أداءً جيدًا يتراوح من 70٪ إلى أكثر من 90٪. تم إجراء التقييم الثاني على نظام الوقت الفعلي الذي تم تطويره حيث تمت مقارنة نتائج النظام بتقييم الخبراء البشريين. كانت نتيجة دقة النظام جيدة، حيث بلغت الدقة حوالي 88٪. كما تفوقت قدرة النظام على التعرف على النطق الجيد بنسبة 92٪، كما أن قدرته على تصنيف النطق الخاطئ على أنه "غير صحيح" كانت جيدة أيضاً، حيث بلغت الدقة حوالي 76٪. لذلك، فقد نجح النظام الذي تم تطويره في تمثيل النطق الصحيح لجميع الحروف القرآنية بناءً على المخارج والصفات في تدريب النطق بمساعدة الكمبيوتر، والذي سيكون منصة مهمة كأداة تكميلية لتدريس وتعلم التجويد التقليدي.

APPROVAL PAGE

The thesis of Tareq M K ALTalmas has been approved by the following:

Salmiah Bt. Ahmad
Supervisor

Wahju Sediono
Co-Supervisor

Nik Nur Wahidah Bt. Nik Hashim
Co-Supervisor

Muhammad Mahbubur Rashid
Co-Supervisor

Khairul Azami Sidek
Internal Examiner

Kamarul Hawari Ghazali
External Examiner

Siti Anom Ahmad
External Examiner

Ismaiel Hassanien Ahmed
Chairman

DECLARATION

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

Tareq M K ALTalmas

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

INTERNATIONAL ISLAMIC UNIVERSITY MALAYSIA

DECLARATION OF COPYRIGHT AND AFFIRMATION OF FAIR USE OF UNPUBLISHED RESEARCH

REAL-TIME AUDIO-BASED TRAINING SYSTEM FOR CORRECT QUR'ANIC LETTER PRONOUNCIATION

I declare that the copyright holders of this thesis are 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 Commercialisation policy.

Affirmed by Tareq M K AlTalmas

…….. ……

Signature Date

This thesis is dedicated to my dear parents, my wife and my family

ACKNOWLEDGEMENTS

Firstly, it is my utmost pleasure to dedicate this work to my dear parents, my wife, and my family, who granted me the gift of their unwavering belief in my ability to accomplish this goal: thank you for your support and patience.

A special thanks to Dr Salmiah Ahmad, Dr Wahju Sediono, Dr Nik Nur Wahidah Binti Nik Hashim, and Muhammad Mahbubur Rashid for their continuous support, encouragement, and leadership, and for that, I will be forever grateful.

I wish to express my appreciation and thanks to ***Dar Al-Quran Al-Karim wa Sunnah*** in Gaza for their contribution in collecting the data for the research.

Finally, I would like to acknowledge the Ministry of Higher Education Malaysia (MOHE) for funding this project through my study under two different grants, which were RACE14-009-0015, and RIGS15-146-0146

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ANN	Artificial Neural Networks
ASR	Automatic Speech Recognition
AVSR	Audio-Visual Speech Recognition
AUC	Area Under the Curve
CALL	Computer-aided Language Learning
CAPT	Computer-aided Pronunciation Training
CM	Confusion Matrix
CNN	Convolutional Neural Network
CV	Cross-validation
DCT	Discrete Cousin Transform
DFT	Discrete Fourier Transform
DTW	Dynamic Time Warping
FFT	Fast Fourier Transform
FN	False Negative
FP	False Positive
GMM	Gaussian Mixed Model
GLL	Global Average Log Likelihood
GOP	Goodness of Pronunciation
GUI	Graphical User Interface
HMM	Hidden Markov Model
KNN	K-nearest Neighbors
LDA	Linear Discriminant Analysis
LPC	Linear Predictive Coding
LPCC	Linear Predictive Cepstral Coefficients
MFCC	Mel Frequency Cepstral Coefficients
MLP	Multilayer perceptron
PCA	Principal component analysis
PLP	Perceptual linear predictive
PSD	Power Spectral Density
QDA	Quadratic Discriminant Analysis
RASTA-PLP	Relative spectral-perceptual linear prediction
ROC	Receiver operating characteristic
RUSBoosted	Random Undersampling Boosting
SR	Speech Recognition
SVM	Support Vector Machine
TN	True Negative
TP	True Positive
VQ	Vector Quantization

LIST OF SYMBOLS

α	FIR Filter Coefficient
α_k	LPC Predictor coefficients
$A(z)$	Inverse Filter of Speech Representation
$c(n)$	cepstral coefficients
C	order of the MFCC analysis
$d(x_i, x_l)$	Euclidean Distance
$e_n(m)$	Error the Difference Between Actual and Estimated Speech
E_n	Squared Error
Σi	Covariance Matrices
$F_{1\text{ - }score}$	F1-score for the Classifiers
f	Physical frequency
f_{Bark}	Bark-scale frequency
$f_{c(bark)}$	Centre Frequency of One Filter in Bark Scale
f_{mel}	Mel-scale frequency
G	Gain of the Excitation.
$g_i(x)$	Quadratic Discriminant Function
$J(y, \hat{y})$	Jaccard index
K	Number of the Nearest Neighbors
L	Total Number of Triangular Mel Weighing Filter
λi	Roots of the Inverse Filter
$M_l(k)$	l^{th} Filter of the Filter Bank
μ_i	Class Means
p	LPC Analysis Order
$p(X\|y = i)$	Posterior probability.
$S(z)$	Current Speech Sample
$\tilde{s}(n)$	Actual Speech Sample at Time n
$\tilde{s}(n)$	Estimated Speech Signal
σ^2	Gain of the LPC Model
$X[k]$	Frequency Domain Representation of the Speech Signal
X[n]	Discrete-Time Representation of the Speech Signal
x_n	Feature n in the Features Vector
$U(z)$	Excitation the Source Signal of the Speech Production Model
y	Actual Value
$\hat{y}$	Predicted Value

CHAPTER ONE

INTRODUCTION

1.1 RESEARCH BACKGROUND

The twentieth century witnessed considerable advancement in many domains of technologies, where significant development of machine-based applications contributed to many aspects of human life. This era shows severe machine intrusion into people's daily lives, making the machines smarter and the people more dependent on them. Human-machine interaction is a result of progress in computer power and abilities. This interaction can be as simple as pressing the buttons to a much complex form as having a conversation with the machine. The term machine here means computer or smartphone. Education is essential to humanity; people spend one-fourth of their lives acquiring new knowledge and skills, which help them to compete and succeed. Machines with the latest technologies are required nowadays as an innovative platform for education and learning. It can be a way of distance learning using websites or smart applications that can perform few tasks, and it helps expediting the learning process.

Quran is the holy book for Muslims, and it has been sent and written in the Arabic language as narrated in Surah Yusuf verse number 2 *"Indeed, We have sent it down as an Arabic Qur'an that you might understand."* The virtues of reading the ***Quran*** are ten rewards for every recited letter. The people who used to recite the ***Quran*** are the best, and last but not least Muslim position and rank in Jannah are determined based on the amount of ***Quran*** memorized in his life. In fact, when

someone makes reading and understanding the ***Quran*** as a daily activity, it will significantly help those who have passed away rest peacefully in their graves.

Recitation of the ***Quran*** with ***Tajweed*** is an essential task as a Muslim. It is necessary to fulfill other worships, such as praying, fasting, and performing Hajj, which occupy the use of ***Quranic*** reading knowledge. It is very much needed for a Muslim to recite the ***Quran*** as close as how our Prophet Muhammad (s.a.w.) did. To do it in such a way, learning ***Tajweed*** becomes necessary in Islam from an early age. The word ***Tajweed*** means to improve or to make better. It is the rule and knowledge that help people to recite the Holy ***Quran*** similar to what was recited by Prophet Mohammed (s.a.w). Moreover, an important part of ***Tajweed*** is pronouncing the letters from its correct articulations (***Makhraj***) and giving the letters their inherent characteristics (***Sifaat***) and dues in conditional attributes. This requirement is somehow challenging for non-Arab Muslims. Characteristics (***Sifaat***) of Quranic letters help differentiate the letters that have the same articulations, where they are divided into two groups: characteristics with opposites and characteristics without opposites. The pronunciation of the Quranic letters from their correct articulation points and their characteristics is not easy for people from non-Arab backgrounds, where they need much effort to learn the correct way of pronouncing the Quranic letters and therefore, should be helped.

Interestingly, most Muslims are non-native Arabic speakers and live in societies where Arabic is not the official language, as only 20 % of Muslims are those with Arabic mother tongue. Malay Muslims are distributed around Malaysia, Indonesia, and the Philippines, representing 14% of the total Muslim population worldwide (Anon 2018). Such statistics show another problem in addition to the required time to the learning process, where teachers are unavailable in the non-Arab