

**ACOUSTIC PROFILES BASED ON QURANIC MAQAMAT
AUDIO FEATURES**

Farah Hanim bt Seman @ Abd Jabar

Thesis submitted in partial fulfilment for the degree of
DOCTOR OF PHILOSOPHY IN
SCIENCE AND TECHNOLOGY

UNIVERSITI SAINS ISLAM MALAYSIA

November 2022

AUTHOR DECLARATION

I hereby declare that the work in this thesis is my own unless specified and duly acknowledged by quotation.

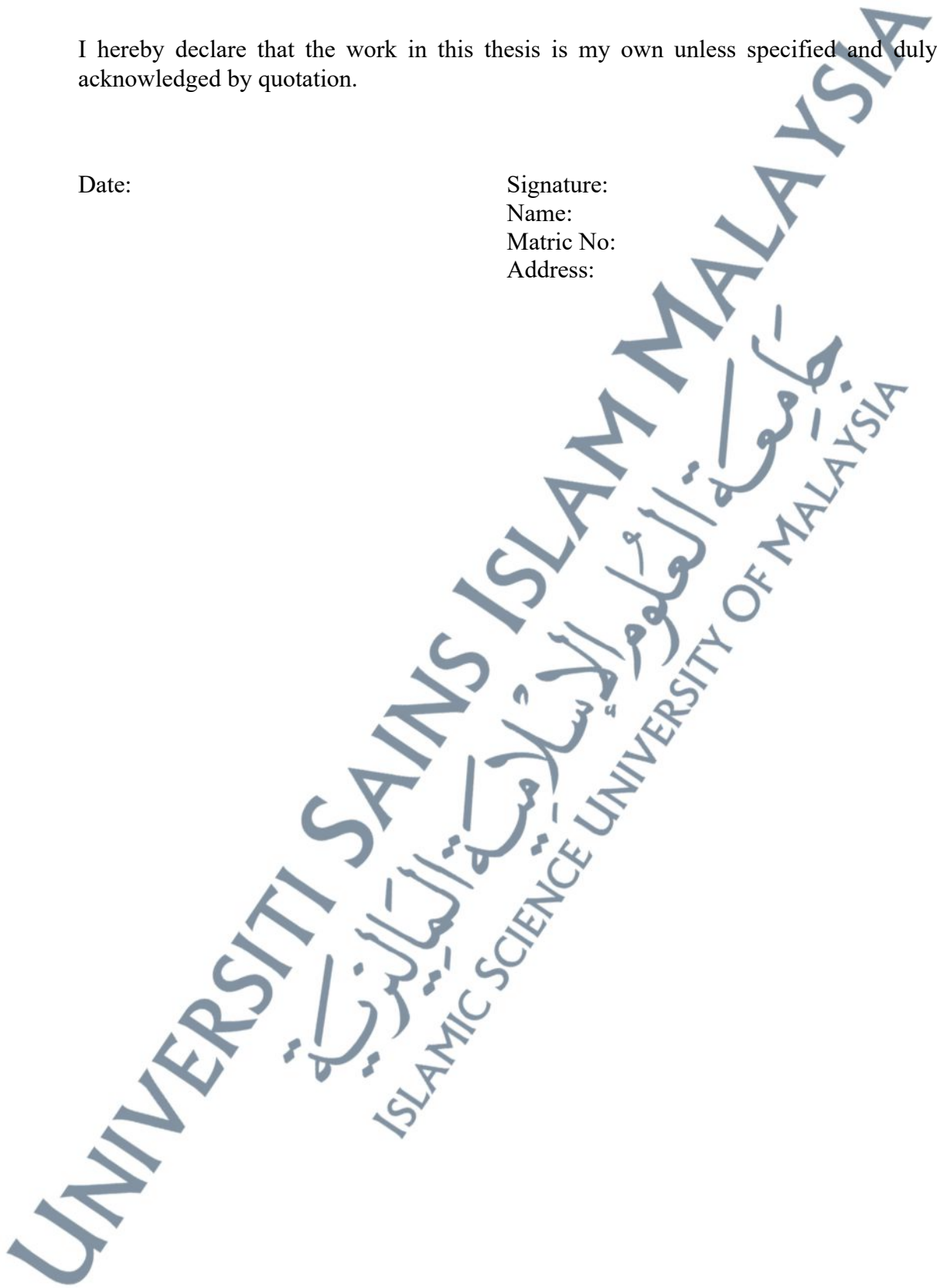
Date:

Signature:

Name:

Matric No:

Address:



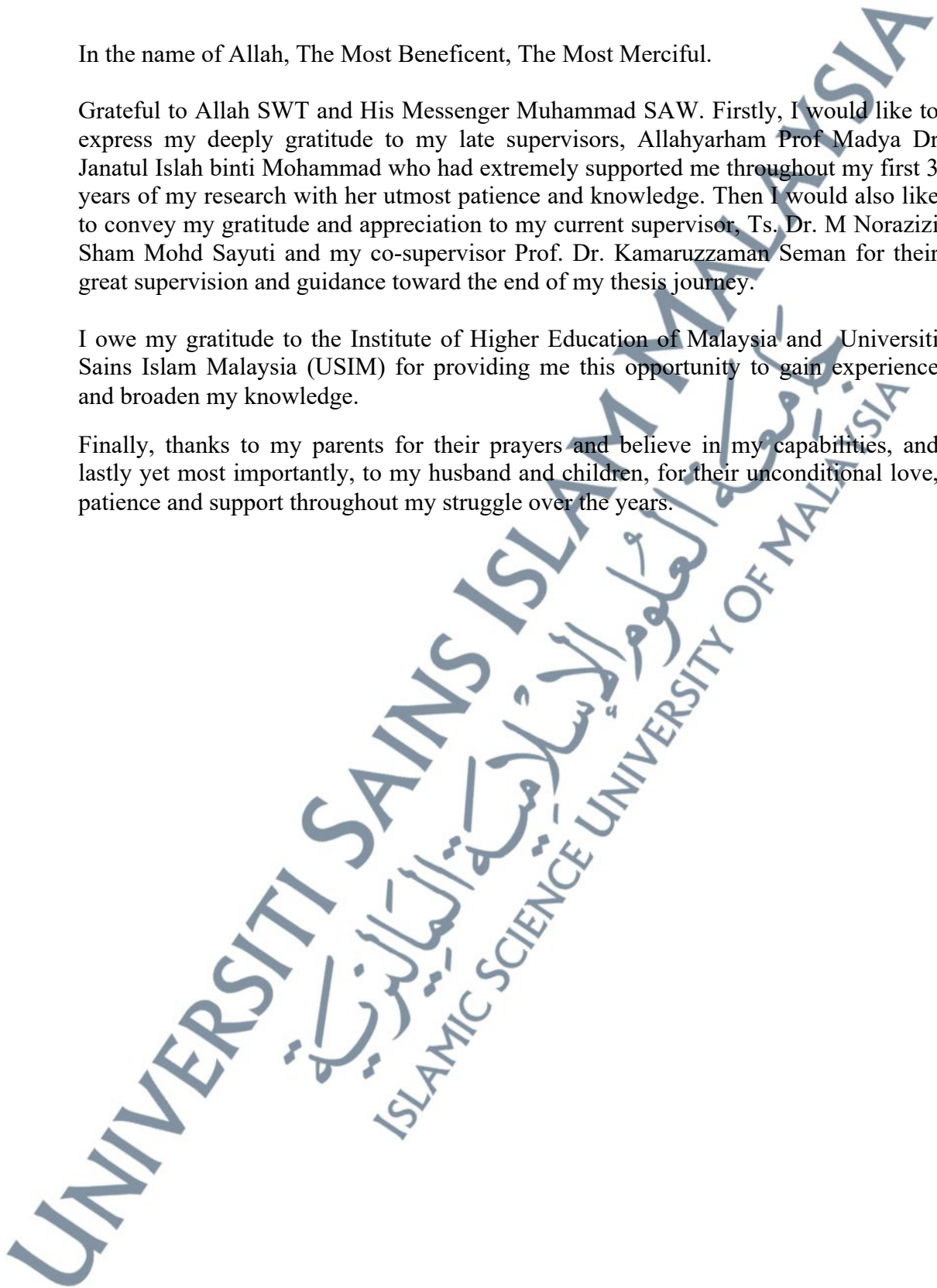
ACKNOWLEDGEMENTS

In the name of Allah, The Most Beneficent, The Most Merciful.

Grateful to Allah SWT and His Messenger Muhammad SAW. Firstly, I would like to express my deeply gratitude to my late supervisors, Allahyarham Prof Madya Dr Janatul Islah binti Mohammad who had extremely supported me throughout my first 3 years of my research with her utmost patience and knowledge. Then I would also like to convey my gratitude and appreciation to my current supervisor, Ts. Dr. M Norazizi Sham Mohd Sayuti and my co-supervisor Prof. Dr. Kamaruzzaman Seman for their great supervision and guidance toward the end of my thesis journey.

I owe my gratitude to the Institute of Higher Education of Malaysia and Universiti Sains Islam Malaysia (USIM) for providing me this opportunity to gain experience and broaden my knowledge.

Finally, thanks to my parents for their prayers and believe in my capabilities, and lastly yet most importantly, to my husband and children, for their unconditional love, patience and support throughout my struggle over the years.



ABSTRAK

Pengekstrakan Sifat Audio memberikan peranan besar dalam pemrosesan pertuturan terutama dalam analisis audio bersemantik yang memberikan maklumat sifat audio berasaskan intonasi, emosi dan rentak. Ramai penyelidik telah mengutarakan cabaran yang dihadapi oleh penutur Bahasa Arab terutama di dalam Al-Quran, disebabkan oleh perbezaan dalam teknik penulisan dan bacaan. Intonasi perkataan di dalam Al-Quran akan memberikan maksud terjemahan atau persepsi kepada objektif surah secara amnya atau ayat secara khususnya. Antara kaedah untuk mengekstrak sifat jelas yang mencirikan keaslian signal pertuturan bermelodi yang kompleks ini adalah analisis *kepstral*. Kini, kajian lepas telah menunjukkan kebanyakan penyelidikan tentang ciri-ciri akustik dan retorik di dalam Al-Quran memfokuskan analisis teks berbanding bunyi bacaan. Dalam kajian ini, analisis spektral dipertingkatkan telah dicadangkan bagi mengekstrak sifat audio yang berasaskan ontologi yang terkandung dalam rakaman audio Bacaan Al-Quran Bermaqamat. Sistem yang dicadangkan ini dimulakan dengan mengekstrak sifat audio yang terkandung dalam 242 buah rakaman audio menggunakan analisis berspektral sedia ada dan yang dipertingkatkan berasaskan fungsi meledingkan. Sebanyak sembilan set sifat audio diekstrak dari algoritma deskriptor berspektral dan dua set adalah dari algoritma pengekstrakan sifat audio yang dipertingkatkan dengan fungsi yang diledingkan. Kerangka ini berupaya mengesan sampel mean spektral dan deskriptor spektral sebagai sifat audio yang jelas yang digunakan sebagai entiti bagi penandaan atribut dan padanan peraturan. Peringkat akhir adalah analisis semantik berasaskan ontologi Maqamat dengan penandaan atribut dan padanan peraturan bagi binaan pengkalan data. Hasil penyelidikan dan prestasi dianalisis dengan semantik audio berdasarkan bentuk frekuensi dari sifat spektral. Kajian ini diharapkan dapat membantu dalam pemahaman kandungan maqamat dan menyumbang kepada meningkatkan inisiatif dalam memprofil sifat akustik dalam analisis penuturan kompleks bagi pemrosesan penuturan masa hadapan.

ABSTRACT

Audio feature extraction underpins a massive proportion of speech processing, mainly in semantic audio analysis which retrieves sound features information based on intonation, emotion and rhythm. Many researchers have addressed various challenges that most speakers faced when dealing with Arabic language especially in the Quran, due to its differences in written and recital technique. The intonation of words in the Quran will bring different translation or perception to the objective of the chapter generally or to the exact verse particularly. Among other methods to extract strong features that characterise the complex nature of complex and melodious speech signals is cepstral analysis. Nowadays, existing literature have shown that most of the study on acoustical and rhetorical element of the Quran is focusing on the textual analysis rather than recitation of the Quran. In this research an enhanced audio feature extraction has been presented to extract significant ontological audio features contained in Quranic Maqamat Recitation audio recording. The proposed system is initiated by extracting the maqamat features from 242 sets of audio files using existing and enhanced cepstral analysis based on warping function. Nine sets are extracted from spectral descriptors algorithm and the other 2 sets are extracted from the audio feature extraction techniques enhanced with warping function. The framework managed to detect the mean of spectral envelope and spectral descriptors as significant audio features which used as entities for attributes tagging and rule matching. The final stage is performing the semantic analysis based on the proposed ontology with attribute tagging and rule matching for the insight knowledge base construction. The results is analysed in semantic audio analysis and their performance based on the formants frequencies extracted from the spectral properties. This study will initiate an understanding on the characteristics of the maqamat content and hopefully contribute to raising initiatives in profiling the acoustical elements of complex speech analysis for further speech processing analysis.

الملخص

يرتكز استخراج الميزة الصوتية بشكل كبير على معالجة الكلام خاصة حين يتعلق الأمر ب تحليل الصوت وقد تناول مجموعة .الدلالي الذي يستخلص معطيات الميزات الصوتية بناء على التنغيم والعاطفة والإيقاع من الباحثين بالدراسة التحديات التي يواجهها غالبية المتحدثين مع اللغة العربية وخاصة القرآن لاختلاف ومن ثم كان لتنغيم أو ترتيب الكلمات في القرآن أن يضيف للسورة بصفة عامة .أسلوب الكتابة والتلاوة ومن بين طرق استخراج الميزات القوية الأخرى التي تميز طبيعة .والآية بصفة خاصة ترجمة أو تصورا مختلفا Cepstral إشارات الكلام المعقد والمنغوم المعقدة التحليل الطيفي الذي ساعتمد عليه في هذا البحث . وتكشف لنا المؤلفات اليوم أن معظم الدراسات حول عنصري الصوت والبلاغة تهتم بالتحليل *Analysis* لهذا أقدم في هذا البحث خاصية لاستخراج الميزات الصوتية المحسنة من .النصي بدلا من تلاوة القرآن أجل استخراج الميزات الصوتية الوجودية المهمة التي يتضمنها التسجيل الصوتي لتلاوة المقامات القرآنية. وقد خلصت إلى هذا النظام المقترح باستخراج ميزات المقامات من 242 مجموعة من الملفات الصوتية باستعمال التحليل الطيفي الراهن والمحسن المبني على وظيفة التزيف و 9 مجموعات من الميزات الصوتية التزيف. من خوارزمية الواصفات الطيفية ومجموعتين من تقنيات استخراج الميزات الصوتية المحسنة بوظيفة وقد نجح هذا الإطار في رصد متوسط الغلاف الطيفي والواصفات الطيفية باعتبارها مميزات صوتية هامة ومطابقة القواعد. وتمثل المرحلة النهائية في *Attributes Tagging* تستخدم ككيانات لوسم الصفات التحليل الدلالي بناء على الأنطولوجيا المقترحة مع وسم الصفات ومطابقة القواعد لبناء قاعدة معرفة عميقة. أما النتائج فقد تم تحليلها في التحليل الصوتي الدلالي بناء على ترددات صفات أصوات الكلام المستخرجة من الخصائص الطيفية. إن هذه الدراسة تقدم فهما جديدا لخصائص محتوى المقامات على أمل أن تساهم في رفع مبادرات لتحديد خصائص عناصر التحليل الكلامي المعقد الصوتية لأجل المزيد من تحاليل معالجة الكلام.

TABLE OF CONTENTS

CONTENT	PAGE
AUTHOR DECLARATION	ii
ACKNOWLEDGEMENTS	iii
ABSTRAK	iv
ABSTRACT	v
AL-MULAKHKHAS	vi
CONTENT PAGE	vii
LIST OF TABLES	x
LIST OF FIGURES	xi
LIST OF EQUATIONS	xiv
LIST OF APPENDICES	xvi
LIST OF ABBREVIATIONS	xvii
LIST OF ACRONYMS	xviii

CHAPTER 1 : INTRODUCTION

1.1 Introduction	1
1.2 Problem Statement	4
1.3 Objectives	5
1.4 Motivations	6
1.5 Research Scope	7
1.6 Benefits	7
1.7 Structure of thesis	8
1.8 Summary	10

CHAPTER 2 : AUDIO FEATURE EXTRACTION IN SPEECH PROFILING FOR QURANIC SEMANTIC AUDIO

2.1 Introduction	11
2.2. Quranic Maqamat Recitation	12
2.2.1. Overview of Quranic Maqamat	12
2.2.2. Types of maqamat	13
2.2.3. Characteristics and Sound element in QMR	19
2.3. Audio Feature Extraction	20
2.3.1. AFE for Complex Spectrum	22
2.3.2. Cepstral Analysis	23
2.3.3. Mel-Frequency Cepstral Coefficient (MFCC)	25
2.3.4. Frequency warping	31
2.3.5. Spectral Descriptors	32
2.4. Automatic Speech recognition for Speech Profiling	34
2.4.1. An overview of speech recognition	35
2.4.2. ASR for Quranic Semantic Search	37
2.4.3. Ontologies for Semantic Audio Analysis	38
2.4.4. Ontology design principles	38

2.5. Semantic Audio Analysis (SAA) for Knowledge base construction	39
2.5.1. Utilities and applications	39
2.5.2. Semantic Audio Tool	42
2.6 Summary	45

CHAPTER 3 : RESEARCH METHODOLOGY

3.1 Introduction	46
3.2 Research Design Framework	46
3.3 Research method	47
3.4 Data collection.....	50
3.4.1. Primary Datasets	51
3.4.2. Secondary Datasets	52
3.5 Audio Feature Extraction	53
3.5.1. Spectral Descriptors	53
3.5.2. Cepstral coefficient analysis (CCA)	53
3.5.3. MFCC Analysis.....	54
3.5.4. W-DFT via MFCC	56
3.6 QMR Knowledgebase construction	56
3.7 Overall Proposed Framework with outcome.....	57
3.7.1. Problem awareness	59
3.7.2. Proposing the system architecture of MPS	59
3.7.3. Implementing the System Development.....	59
3.7.4. Evaluation of MPS	60
3.7.5. Analysis of findings	61
3.8 Summary	61

CHAPTER 4 : AUDIO FEATURE EXTRACTION FOR QMR

4.1 Introduction	62
4.2 Audio Feature Extraction	63
4.2.1. Pre-processed data	63
4.2.2. Spectral Descriptors	64
4.2.3. Cepstral extraction with CCA	65
4.2.4. MFCC with W-DFT	65
4.3 Result & Discussions	70
4.3.1. Spectral Descriptors	70
4.3.2. CCA, MFCC and W-DFT	72
4.3.3. Spectral envelope of W-DFT	74
4.4 Summary	77

CHAPTER 5 : MPS CLASSIFICATION FOR QMR KNOWLEDGEBASE CONSTRUCTIONS

5.1	Introduction	78
5.2	Work process	79
5.3	Features Classification and Recognition	80
	5.3.1. Peak-magnitude-to-RMS-ratio	81
	5.3.2. Classes	81
5.4	The Maqamat Ontology	82
	5.4.1. Motivations.....	83
	5.4.2. Selecting of Ontologies	84
	5.4.3. Entities and Annotations	84
	5.4.4. Attribute Tagging for Rule Matching.....	85
	5.4.5. Workflows, Events and Timelines	86
5.5	Design Framework for QMR Knowledge Base Construction	87
	5.5.1. Pre-processing	88
	5.5.2. Profiling.....	88
	5.5.3. Post-processing.....	89
5.6	Experimental result and discussion	90
5.7	Summary	94

CHAPTER 6 : PERFORMANCE EVALUATION WITH CROSS VALIDATION TECHNIQUES

6.1	Introduction	96
6.2	Cross Validation technique	97
	6.2.1. Hold out.....	97
	6.2.2. k-fold	99
6.3	Result and discussion	99
6.4	Summary	104

CHAPTER 7 : CONCLUSIONS AND FUTURE WORKS

7.1.	Conclusions	105
7.2.	Achievements	108
7.3.	Contributions to Knowledge	109
7.4.	Limitations	109
7.5.	Future Works.....	110

REFERENCES	112
------------------	-----

APPENDIX A	118
------------------	-----

APPENDIX B	126
------------------	-----

APPENDIX C	133
------------------	-----

APPENDIX D	141
------------------	-----

LIST OF TABLES

Tables		Page
Table 2.1	Summary of Maqamat types with its features and themes.	19
Table 2.2	Principal dimensions of elementary audio sound	41
Table 3.1	Primary datasets from selected surahs	51
Table 3.2	Secondary datasets from selected surahs	52
Table 5.1	Annotation label for maqamat type	82
Table 5.2	Entities and its annotation in MO	85
Table 5.3	Some examples of rules applied in rule matching for attribute tagging	86
Table 5.4	Example of Numerical data attributes of 3 spectral descriptors for peak-magnitude-to-RMS-ratio, mean and standard deviation respectively	90
Table 5.5	Performance evaluation of proposed technique from primary datasets	94
Table 5.6	Example of result from selected real-world datasets after implementing the framework	94
Table 6.1	Accuracies of k-fold and holdout validations with various classifiers used	103

LIST OF FIGURES

Figures		Page
Figure 1.1	The foundations of digital signal processing in sound	3
Figure 2.1	Basic concept of Audio Feature Extraction	21
Figure 2.2	Block diagram representing computation of spectrum	24
Figure 2.3	Mel-filter bank	28
Figure 2.4	MFCC feature extraction	30
Figure 2.5	Example of spectral centroid from human speech	33
Figure 2.6	Four basic stages of Speech recognition	36
Figure 2.7	Knowledge representation model between analysis and information layers	44
Figure 2.8	Knowledge representation model between information and application layers	45
Figure 3.1	The fundamental research design and flow process	48
Figure 3.2	Flow diagram describing the work process of proposed framework for MPS development	50
Figure 3.3	Conventional MFCC using HMM model	55
Figure 3.4	QMR knowledge base construction model	57
Figure 3.5	Overall proposed system framework	58
Figure 4.1	Basic Audio Feature Extraction	63

Figure 4.2	Software for audio editing tool	64
Figure 4.3	Screenshot of MATLAB command lines for HTK routines	67
Figure 4.4	Feature extraction of MFCC in HMM model	69
Figure 4.5	Flowchart of MFCC with W-DFT	69
Figure 4.6	Example of plotted graph for Spectral Descriptors or Maqamat Bayyati of Surah Al-Ikhlās	71
Figure 4.7	Example of plotted graph for the three Spectral Descriptors in same time-domain	71
Figure 4.8	Results for Quranic maqamat audio signals using cepstral analysis and MFCC techniques	73
Figure 4.9	Example of frequency spectrum for 3 surahs using W-DFT for maqamat	73
Figure 4.10	Spectral envelope of WDFT with high-low and extracted mean value	74
Figure 4.11	Spectral envelope of WDFT with (a) $T_s = 25\text{ms}$	75
Figure 4.12	Spectral envelope of WDFT with (a) $T_s = 250\text{ms}$	75
Figure 4.13	Filterbank energy, spectrum and spectral envelope of WDFT	76
Figure 4.14	The overall W-DFT features for each maqamat	77
Figure 5.1	Work process for classification model	80
Figure 5.2	Screenshot of command line used in MATLAB as preliminaries rules	85
Figure 5.3	Overview of Maqamat Ontology	87
Figure 5.4	Design framework in constructing QMR Knowledge base	88

Figure 5.5	Example of output when tested with Surah Abasa from secondary datasets	91
Figure 5.6	Bar chart showing percentage of intelligibility accuracy	91
Figure 5.7	Gaana.com and Qurandb.com database	93
Figure 6.1	Hold-out validation distribution ratio	98
Figure 6.2	The process of k-fold validations with 5-Fold validations (k=5)	99
Figure 6.3	Scatter Plot with prediction model	101
Figure 6.4	Confusion matrix for TPR and FNR	102
Figure 6.5	Confusion matrix for PPV and FDR	102
Figure 6.6	Parallel coordinates plot with model predictions	103
Figure 7.1	The overall process of MPS system	107

LIST OF EQUATIONS

Equations		Pages
Eq. (2.1)	Complex spectrum	22
Eq. (2.2)	Complex spectrum in real and imaginary part	23
Eq. (2.3)	Speech sequence in frequency domain	23
Eq. (2.4)	Magnitude spectrum	23
Eq. (2.5)	Logarithmic magnitude spectrum	24
Eq. (2.6)	Logarithmic IDFT	24
Eq. (2.7)	First-order FIR high-pass filter	26
Eq. (2.8)	FIR filter in z-domain	26
Eq. (2.9)	Discrete Fourier Transform	27
Eq. (2.10)	Mel frequency	27
Eq. (2.11)	Mel filterbanks	28
Eq. (2.12)	Discrete Cosine Transform	29
Eq. (2.13)	Dynamic features of n_{th} time frame	30
Eq. (2.14)	Spectral centroid	32

Eq. (2.15)	Spectral spread	33
Eq. (2.16)	Spectral Roll-off point	34
Eq. (3.1)	Warped DFT	57
Eq. (3.2)	Elements of Warped DFT	57
Eq. (5.1)	Peak-magnitude-to-RMS-ratio	82

UNIVERSITI SAINS ISLAM MALAYSIA
جامعة العلوم الإسلامية
ISLAMIC SCIENCE UNIVERSITY OF MALAYSIA

LIST OF APPENDICES

Appendices		Pages
Appendix A	Speech signal, spectral envelope and 3D spectrum	114
Appendix B	Mean spectral envelope	127
Appendix C	Numerical data of RMS, mean and standard deviation	135
Appendix D	Signal pattern extracted from primary dataset using W-DFT for each maqamat based on attributes	141

UNIVERSITI SAINS ISLAM MALAYSIA
جامعة العلوم الإسلامية
ISLAMIC SCIENCE UNIVERSITY OF MALAYSIA

LIST OF ABBREVIATIONS

Notation	Descriptions
c_r	Real spectrum
c_j	Imaginary part
n	Signal sequence
e	Excitation sequence
h	Vocal tract filter sequence
s	Speech sequence
E	Excitation sequence in frequency domain
H	Vocal tract filter sequence in frequency domain
S	Speech sequence in frequency domain
ω	Spectrum sequence
α	pre-emphasis parameter
a	coefficient
H	Transfer function
Z	z-domain
f_{mel}	Mel-frequency
H_m	Weight of given spectrum relative to energy spectrum bin
M	Total number of triangular mel weighting filters
c	Cepstral coefficients
C	Total number of MFCCs
T	Total number of successive frames
N	Number of points used to compute the DFT
T_w	Time frame
T_s	Shift frame

LIST OF ACRONYMS

DSP	Digital Signal Processing
QMR	Quranic Maqamat Recitation
CCA	Cepstral Coefficient Analysis
IDFT	Inverse Discrete Fourier Transform
DFT	Discrete Fourier Transform
IDCT	Inverse Discrete Cosine Transform
DCT	Discrete Cosine Transform
FIR	Finite Impulse Response
SD	Spectral Descriptors
ASR	Automatic Speech Recognition
HMM	Hidden Markov Models
HTK	HMM Toolkit
SAA	Semantic Audio Analysis
W-DFT	Warped DFT
MFCC	Mel-Frequency Cepstral Coefficient