DEVELOPMENT OF AN ARABIC TEXT-TO-SPEECH SYSTEM

BY

MUSTAFA ZEKI OBAID

A dissertation submitted in fulfilment of the requirement for the degree of Master of Science in Computer and Information Engineering

Kulliyyah of Engineering

International Islamic University
Malaysia

JUNE 2011

ABSTRACT

The field of speech synthesis or Text-To-Speech (TTS) has rapidly expanded during the last few years due to the wide range of applications that require human-machine interaction. Arabic language, being the fourth most spoken language on the globe, has received the attention of the researchers in development of an intelligible and close to natural Text-To-Speech system. Most of the available Arabic TTS Systems have drawbacks such as discontinuity, limited vocabulary and poor naturalness. A rule-based Arabic Text-To-Speech system using hybrid synthesis is presented in this dissertation. Sinusoidal model and concatenation using phonemes as the unit of speech were used to implement the hybrid synthesis system. A set of rules was constructed to achieve letter-to-sound mapping and an exception lexicon to cover the words that does not follow a certain pronunciation rule. Using rule-based synthesis makes the system vocabulary independent. A special phonemes inventory was constructed to be used for the Arabic TTS. Applying the accurate stress pattern for Arabic words was successfully achieved which contributed to the overall quality of the system. The evaluation results shows high level of intelligibility with acceptable naturalness with an overall rate of categorical estimation is 3.458 out of 5.

خلاصة البحث

أن مجال تركيب الكلام الآلي أو تحويل النصوص الى كلام (تي تي أس) قد اتسع مؤخرا نظرا للتطبيقات الكثيرة التي تتطلب التفاعل الصوتي بين الانسان والآلة. لقد اكتسبت اللغة العربية والتي تعتبر رابع لغة من حيث عدد الناطقين في العالم ولخصوصيتها الدينية اهتمام الباحثين في هذا المجال لتطوير نظام نطق آلي عربي واضح ومفهوم وقريب من النطق البشري. أن اغلب أنظمة النطق الآلي العربي المتوفرة تعاني من عيوب مثل عدم الأستمرارية أثناء نطق المقاطع الصوتية، تغطيتها لمفردات محدودة، أو جودة الصوت بعيدة عن الصوت البشري. في هذا البحث، نقدم نظام نطقٍ آلى عربي معتمد على مجموعة من القواعد اللفظية يستخدم تقنية تركيب الكلام الهجينة. تم استخدام نموذج الموجات الجيبية إضافة الى تجميع الاصوات المسجلة باعتماد الفونيم كوحدة صوت أساسية لغرض تطبيق النظام الهجين المقترح. أن استخدام مركِّب الكلام المعتمد على قواعد اللفظ يجعل النظام الناتج غير معتمد على نوع واحد من الالفاظ ويغطي جميع المفردات العربية. ولغرض الحصول على نطق عالي الجودة تم بناء مستودع خاص للأصوات العربية. لقد تم بنجاح تطبيق أضافة النّبر على الكلمات العربية بناءاً على قواعد النّبر الخاصة والذي ساهم بشكل واضح في تحسين جودة الصوت الناتج. أظهرت نتائج الاختبار والتقويم مستوى عالي من الوضوح مع مستوى مقبول ومقارب لجودة الصوت البشري وكان المعدل الكلي للتقدير المطلق هو 3,458 من اصل 5.

APPROVAL PAGE

I certify that I have supervised and read this study and that in my opinion, it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a thesis for the degree of Master of Science in Computer and Information Engineering.

……
Othman O. Khalifa
Supervisor

……
Ahmed W. Naji
Co-Supervisor

I certify that I have read this study and that in my opinion, it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a thesis for the degree of Master of Science in Computer and Information Engineering.

……
Teddy Surya Gunwan
Internal Examiner

……
Nooritawati Md Tahir
External Examiner

This dissertation was submitted to the Department of Electrical and Computer Engineering and is accepted a fulfilment of requirement for the degree of Master of Science in Computer and Information Engineering.

……
Othman O. Khalifa
Head, Department of Electrical and Computer Engineering

This dissertation was submitted to the Kulliyyah of Engineering and is accepted a fulfilment of requirement for the degree of Master of Science in Computer and Information Engineering.

……
Amir Akramin Shafie
Dean, Faculty of Engineering

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.

Mustafa Zeki Obaid

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

INTERNATIONAL ISLAMIC UNIVERSITY MALAYSIA

DECLARATION OF COPYRIGHT AND AFFIRMATION OF FAIR USE OF UNPUBLISHED RESEARCH

DEVELOPMENT OF AN ARABIC TEXT-TO-SPEECH SYSTEM

I hereby affirm that the International Islamic University Malaysia (IIUM) holds all rights in the copyright of this work and henceforth any reproduction or use in any form or by means whatsoever is prohibited without the written consent of IIUM. No part of this unpublished research may be reproduced, stored in a retrieval system or transmitted, in any form or by means, electronic, mechanical, photocopying, recording or otherwise without prior written permission of the copyright holder.

Affirmed by Mustafa Zeki Obaid

....

Signature Date

This humble effort is fully dedicated to the Islamic nation where Arabic Language is the essence of their faith. For the language of Qur'an I present this work.

ACKNOWLEDGEMENTS

All praise to Almighty Allah (SWT) and salutation upon the Prophet Muhammad (PBUH). All creations are blind without His guidance (وَمَنْ يُضْلِلِ اللَّهُ فَمَا لَهُ مِنْ هَادٍ).

I wish to thank my parents for them I will remain beholden for good. My father as an example to learn from, I have always admired his invaluable and wise advice and teachings; my mother for her extraordinary care, kindness, and love. There are no words that can exhibit my gratitude to them. I thank my wife for her patience, encouragement, and support, to all my family thank you so much.

I would like to show my sincere gratitude to a special person. This thesis would never have been possible without his insightful observations and advice. It has been a privilege and pleasure to have worked with my supervisor Prof. Dr. Othman O. Khalifa. A special thank to my co-supervisor senior assistant Prof. Ahmed W. Naji for his exceptional effort and patience while helping me during research journey. I never forget to thank Prof. Momoh Jimoh E. Salami for the continuous supervision and informative recommendations during the progress of this research.

All my friends at the faculty of engineering for your help and support, all the friends inside and outside the International Islamic University Malaysia, everyone how contributed directly or indirectly to this work, thank for everything.

TABLE OF CONTENTS

Abstract ii
Abstract in Arabic iii
Approval Page iv
Declaration v
Copyright Page vi
Dedication vii
Acknowledgements viii

CHAPTER ONE: INTRODUCTION 1
1.1 Introduction 1
1.2 Problem Statement and Its Significance 4
1.3 Research Objectives 5
1.4 Methodology and Tools 5
1.5 Text-to-Speech Technology Commercial Future 7
1.6 Research Scope 7
1.7 Conclusion 8
1.8 Thesis Outline 9

CHAPTER TWO: LITERATURE REVIEW 11
2.1 Introduction 11
2.2 Current Techniques For Speech Synthesis 14
2.2.1 Articulatory Synthesis 14
2.2.2 Concatenative Synthesis 15
2.2.3 Formant Synthesis 16
2.3 Sinusoidal Models 17
2.4 Hybrid System 18
2.4.1 Harmonic and Noise Models (HNM) 19
2.5 Arabic Language 21
2.5.1 Historical Review of Arabic Language 21
2.5.2 Arabic Alphabet, Morphology, and Prosody 24
2.6 Challenges of The Arabic Language for TTS System 24
2.6.1 Diacritization Issue 24
2.6.2 Differences in Gender 25
2.7 History of Arabic Speech Synthesis 26
2.8 Summary 35

CHAPTER THREE: PROPOSED SYSTEM 36
3.1 Introduction 36
3.2 Arabic Phonological System 37
3.2.1 Consonants 37
3.2.1.1 Place of Articulation 38

3.2.1.2 Degree of Aperture 39
3.2.1.3 Manner of Articulation 39
3.2.1.4 Special Attributes for Arabic Consonants 40
3.2.2 Vowels 42
3.2.3 Diphthongs 43
3.3 Arabic Syllables Types 44
3.3.1 Proposed Arabic Words syllabification 47
3.4 Normalization 47
3.5 The Letter-To-Sound (LTS) Rules 48
3.6 Overall System Description 50
3.7 Summary 53

CHAPTER FOUR: SYSTEM DESIGN AND IMPLEMENTATION 54
4.1 Introduction 54
4.2 Proposed Arabic Text-to-speech design 54
4.2.1 Espeak As A Speech Engine 56
4.2.2 Proposed Exceptions Lexicon 56
4.2.3 Proposed Letter- to- Sound Rules 58
4.2.4 Numbers Handling 61
4.2.5 Prosody and Stress Shift 62
4.2.6 Phoneme Database Construction 63
4.2.6.1 Vowels and Diphthongs 65
4.2.6.2 Consonants 67
4.2.7 List of Phonemes Used in Arabic TTS 69
4.4 Prototype of the proposed system 71
4.5 Summary 75

CHAPTER FIVE: SYSTEM IMPLEMENTATION AND EVALUATION 76
5.1 Introduction 76
5.2 Test group 77
5.3 Evaluation Methodology 77
5.4 Test and Evaluation Result 79
5.4.1 Words Perception Test 80
5.4.2 Sentence Perception Test 80
5.4.3 Numbers Perception Test 82
5.4.4 Categorical Estimation (CE) Evaluation 83
5.4.5 Comparative Test 85
5.4.5.1 Categorical Estimation 85
5.4.5.2 System Spesifications 86
5.5 Objective Evaluation Test 86
5.7 Summary 9**0**

CHAPTER SIX: CONCLUSION AND RECOMMENDATIONS 91
6.1 Conclusion 91
6.2 Recommendations 92
6.2.1 NLP Module 92

6.2.2 DSP Module 93

BIBLIOGRAPHY 94

APPENDIX A: Glossary 98
APPENDIX B: Arabic Letter-To-Sound Rules 99
APPENDIX C: Arabic Exception Lexicon 105
APPENDIX D: Arabic Phonemes Database 108
APPENDIX E: Evaluation Questionnaire 117

LIST OF TABLES

Table No.		Page No.
2.1	Milestones of speech synthesis.	13
2.2	Current Arabic TTS systems and their features.	34
3.1	Arabic consonants and their place and manner of articulation.	41
3.2	Arabic vowels.	42
3.3	Arabic semi vowels.	43
3.4	Arabic diphthongs in different positions.	44
3.5	Arabic syllable types.	46
4.1	Flags used in exception lexicon and the definitions.	58
4.2	Translation of the word "المَكتَبة".	60
4.3	Special characters in both positions.	61
4.4	Special characters only in following positions.	61
4.5	Number- To- Word (NTW) translation.	62
4.6	Symbol for places of articulations used in phonemes definition.	68
4.7	Type and properties of a phoneme definition.	68
4.8	List of vowels and semi vowels used in Arabic TTS.	69
4.9	List of consonants used in Arabic TTS.	70
4.10	List of special graphemes used in Arabic TTS.	71
5.1	Age and gender distribution of the listeners.	77
5.2	Attributes used in Categorical Estimation (CE) evaluation.	79
5.3	Result of Categorical Estimation (CE) test.	84
5.4	Performance Metrics.	86

LIST OF FIGURES

Figure No.		Page No.
1.1	Simple text-to-speech synthesis procedure.	2
1.2	Research Methodology	7
2.1	Kratzenstein's resonators.	17
2.2	Wheatstone's reconstruction of Von Kempelen's speaking machine.	12
2.3	Sinusoidal analysis/ synthesis system.	18
2.4	Basic idea of hybrid synthesis system.	19
3.1	Places of articulation and vocal organs.	37
3.2	Arabic vowels chart.	43
3.3	General form of an Arabic syllable.	45
3.4	Proposed stress pattern selection scheme for Arabic words.	50
3.5	Flowchart of the proposed Arabic TTS system.	52
4.1	A general architecture of the Arabic TTS.	55
4.2	Screenshot of Arabic LTS rules.	60
4.3	Code snippets for stress and pause marker for Arabic.	63
4.4	Example of phonemes construction: vowel <ي>.	66
4.5	Example of phonemes construction: diphthong < ـَيـ >.	67
4.6	Example of phonemes construction: consonant <ظ>.	69
4.8	Screenshot of the Arabic TTS system prototype.	72
4.9	Screenshot of the Arabic TTS system prototype.	73
4.10	Rules Snapshot from Arabic TTS system.	74
5.1	Word perception test result.	80

5.2	Sentence perception result (section 3)	81
5.3	Sentence perception result (section 4).	82
5.4	Result of Categorical Estimation (CE) test.	85
5.5	Visual perceptual test for the phrase “ بِسمِ اللهِ الرَّحمن الرَّحيم “.	88
5.6	Spectrograph for natural and synthesized “بِسمِ اللهِ الرَّحمن الرَّحيم”	89

LIST OF ABBREVIATIONS

CE	Categorical Estimation
DRT	Diagnostic Rhyme Test
DSP	Digital Signal Processing
GUI	Graphical User Interface
HMM	Hidden Markov Model
HNM	Harmonic Noise Model
IPA	International Phonetic Alphabet
KACST	King Abdulaziz City for Science and Technology
KATTS	King Abdulaziz Text-To-Speech
LPC	Linear Predictive Coding
LTS	Letter-To-Sound
MBROLA	Multi Band Resynthesis OverLap Add
MFCCs	Mel-frequency cepstral coefficients
MRT	Modified Rhyme Test
MSA	Modern Standard Arabic
NLP	Natural Language Processing
PSOLA	Pitch Synchronous Overlaps Add
TD-PSOLA	Time Domain Pitch Synchronous Overlap-add
TTS	Text-To-Speech
VODER	Voice Operating Demonstrator

CHAPTER ONE

INTRODUCTION

1.1 INTRODUCTION

Speech is considered the primary means of communication and interaction between people. The dream of machine-human communication or producing a talking machine started during the 18th century with the use of mechanical systems and continues until current time with the use of computer (Schroeder, 1993). Producing a machine-speech is called speech synthesis. Today, the most common interfaces for human-machine interaction are still keyboards, keypads, and mice. However, an increasing necessity to interface with machines in mobile environments is leading to speech becoming a required means to interface with machines and automated information services. For this reason the automatic generation of speech from text, referred to as text-to-speech (TTS) synthesis, which has been extensively researched and improved upon over the last two decades, has been gaining significant interest in commercial applications over the last few years (Shukla, 2007).

Recent progress in speech synthesis has produced synthesizers with high intelligibility for some major languages like English, but the sound quality and naturalness remain a major problem. However, speech synthesis for Arabic language is still in its early steps. This fact makes speech synthesis an important field for investigation and improvement for the major languages including Arabic, the forth most spoken language on the globe (Lewis, 2009).

Speech synthesis or text-to-speech (TTS) procedure consists of two main phases. The first one is text analysis, where the input text is transcribed into a phonetic

or some other linguistic representation, and the second one is the generation of speech waveforms, where the acoustic output is produced from phonetic and prosodic information.

These two phases are usually called as high- and low-level synthesis. A simplified version of the procedure is shown in Figure 1.1. The input text might be for example data from a word processor, standard ASCII from e-mail, a text-message, or scanned text from a book. The character string is then pre-processed and analyzed into phonetic representation which is usually a string of phonemes with some additional information for correct intonation, duration, and stress. Speech sound is finally generated with the low-level synthesizer by the information from high-level one.

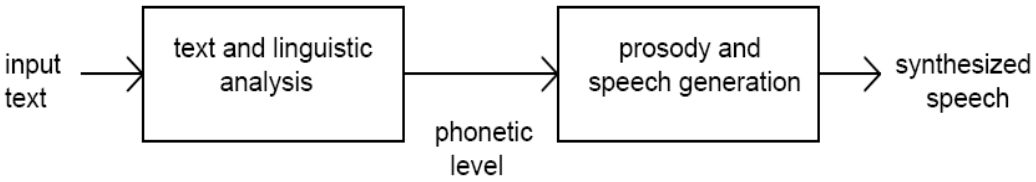

Figure 1.1: Simple text-to-speech synthesis procedure.

The main techniques used in speech synthesis design are Articulator synthesis, Formant synthesis, and Concatenative synthesis. Articulatory synthesis attempts to model the human speech production system directly. Formant synthesis, which models the pole frequencies of speech signal or transfer function of vocal tract based on source-filter-model. Concatenative synthesis, which uses different length pre-recorded samples derived from natural speech.

In theory, the most accurate method is articulatory synthesis which models the human speech production system directly, but it is also the most difficult approach, because it synthesizes speech by controlling the speech articulators and determines the

characteristics of the vocal tract filter by means of a description of the vocal tract geometry and places the potential sound sources within this geometry (Assaf, 2005). For example, the vocal cord model is a two-mass model with two vertically moving masses, due to such complexity articulatory synthesis has not been realized yet.

Therefore, the available TTS systems mostly use either concatenative or formant synthesis technique. Each technique has its own points of strength and weakness and suits a specific language while doesn't for others. An interesting approach is to use a hybrid system where the formant and concatenative methods have been applied in parallel to phonemes where they are the most suitable (Fries, 1994). In general, combining the best parts of the basic methods is a good idea, but in practice, controlling of synthesizer may become difficult (Lemmetty, 1999).

Since the quality of synthetic speech is improving steadily, the application field is also expanding rapidly. Synthetic speech may be used to read e-mail and mobile messages, in multimedia applications, or in any kind of human-machine interaction. In-vehicle environment also benefit from TTS technology since reading text on a display is dangerous while driving. Other hands-busy, eyes-busy applications such as industrial applications or mobile network environment also are ideal for TTS technology.

TTS is essential in assistive applications for blind or illiterate users and can be used as a substitute voice for users who are able to type or otherwise create text but are not able to speak. It can be used also in many educational tasks like spelling and pronunciation teaching aid for Arabic language.

The evaluation of synthetic speech is also an important issue, but it is challenging because the speech quality is a multidimensional term. This leads to the large number of different tests and methods to evaluate different features in speech.

Today, speech synthesizers of various qualities are available as several different products for most common languages; however it is not there for Arabic.

Arabic language has the privilege that, unlike English or French, the correspondence between how words are written and how they are spoken is always according to specific pronunciation rules with some minor exceptions. That characteristic encourages development of Arabic Text-To-Speech system using rule-based synthesis which is implemented in this research. There are other challenges beside utterance during the development of Arabic TTS like dependency between adjacent words and lack of vowelization in most Arabic texts. Compared to other major languages, the field of Arabic TTS sill needs more research. This fact explains why some of the available Arabic TTS systems mostly lack naturalness, cover limited range of Arabic vocabulary, or not flexible to be used in more than one environment.

1.2 PROBLEM STATEMENT AND ITS SIGNIFICANCE

There is no language aid available with the current technologies for the visually handicapped Arabic native speakers despite the fact that the number of blinds in the Arab World is around 5 millions living in a population around 340 million people according to the World Health Organization (Regional Office of The Middle East). It is a must to build a reliable, intelligible, and user friendly Arabic TTS system to give those people a chance to use the technologies like text messages, emails, and other web services.

There are still limitations and room for improvement for Arabic TTS even though there are number of researches. These limitations are the lack of fluency, large system size, close source software, and intonation and stress handling. On top of that there are the language dependent challenges which are the lack of diacritization in

most Arabic texts and feminine and masculine words related to the Arabic numbering system.

Arabic is the fourth most spoken language in our world with more than 442 million speaker spread in 23 countries as an official language (Bateson, Mary Catherine, 2003). Furthermore it carries a religious value for more than 1.6 billion Muslim (Pew Research Center, 2009). Moreover, most of the current technologies were first developed based on English language as a standard in the market, nowadays other languages like Arabic, have heavily involved in these technologies, such as operating systems in personal computer or cell phone, websites, e-books, and other telecommunication applications.

1.3 RESEARCH OBJECTIVES

The objectives of this study are as follows:

a. To design an Arabic TTS system using rule-based hybrid synthesis technique.
b. To develop the proposed system as a set of rules and constructing its Arabic phonemes database.
c. To evaluate the output speech of the developed system checking the level of intelligibility and naturalness.
d. To validate the performance of the developed system and benchmark it with the available Arabic TTS systems.

1.4 METHODOLOGY AND TOOLS

1) Study the background and theory of TTS technology.

2) Critical review of the relevant work and up-to-date research works done in this field.
3) Specify the criteria of the Arabic language, its phonological system, and the specification of implementation of Arabic language in TTS synthesis.
4) Propose a rule-based hybrid synthesis Arabic TTS system.
5) Construct the proposed Arabic TTS system using eSpeak speech engine (Duddington, 2007), and C++ programming language.
6) Develop a stress and intonation pattern for the pre-processing stage using a special stress pattern for Arabic language.
7) Develop NLP (Natural Language Processing) module which mainly contains Arabic rules and exceptions dictionary data base, where NLP is considered as the processing stage. Special phonemes database for Arabic language is to be constructed (based on the conducted study) using inherited phonemes from other languages with modification, or pre-recorded WAV phonemes modified using PRAAT sound analyser (Weenink, 2009).
8) Validate the developed system using both objective and subjective criteria by applying listening test on phoneme, word, and sentence level.
9) Compare the overall performance of the designed system with the current Arabic TTS systems using both subjective and objective criteria.

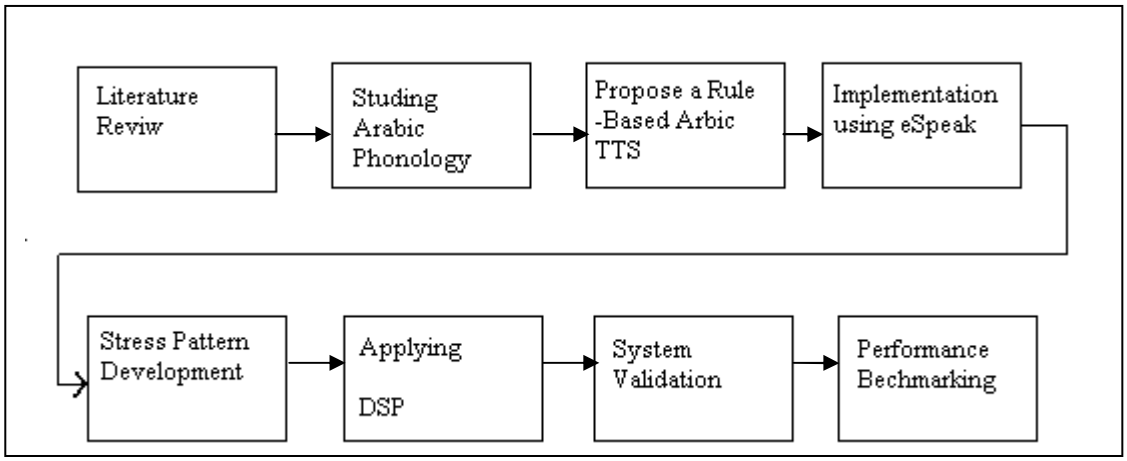

Figure 1.2: Research Methodology

1.5 TEXT-TO-SPEECH TECHNOLOGY COMMERCIAL FUTURE

With the tremendous expanding of the use of e-readers like Kindle, iPad, and other commercial products, imbedding Text-To-Speech is a must in such devises. All kind of machinery one day would be provided with a talking ability to interact with human users, altogether with the speech recognition making it possible to have a simple chat with a microwave, car, ATM, and other daily life machines.

1.6 RESEARCH SCOPE

In this thesis, the focus is on the design of Arabic Text-To-Speech synthesizer system using rule-based Hybrid synthesis technique. This task include four main steps, the first one is to build Arabic phoneme data base that contains the parameters of some phonemes that would be generated using Formant synthesis and wave files for other phonemes that are synthesized using concatenation technique. The second step is to establish a set of pronunciation rules that change each text character to its equivalent phoneme representative. Exception dictionary data base contains a list of all the words

with special pronunciation that don't follow a certain rule is also built. The input text is supposed to be partially vowelized which is considered one of the limitation of the proposed system.

1.7 CONCLUSION

1) A hybrid Text-To-Speech synthesis system has been developed for Arabic language, where pharyngeal, fricative, and stop consonants are generated using concatenation synthesis while formant synthesis is applied to generate vowels and some other consonants from summation of multiple sine waveforms with time varying amplitudes and frequencies.
2) The proposed Arabic TTS system is vocabulary independent, it can handle all types of input text. Furthermore it is a small size application that can work in many platforms like Microsoft Windows, Linux, and Widows Mobile.
3) The proposed system has the flexibility of changing the speaker from male to female and other sound variants like whispering as it is an open source system that is inviting any modification or enhancement.
4) A new approach to overcome the absence of vowelization diacritics in Arabic text is proposed. Our Arabic TTS system offers the ability to enrich the exceptions dictionary by listing the exact pronunciation of the common words, set of words, or even sentences. Therefore, unvowelized (without Diacritization marks like *fatha* and *thamma*) or partially vowelized Arabic text, is vowelized using the exception dictionary in order for TTS to translate it correctly.

5) A new Arabic inventory phoneme database was constructed to suit the developed TTS system.
6) Arabic letter-to-sound (LTS) rules were built and a morphophonemic model has been structured.
7) An Arabic dictionary was constructed containing the correct utterance of exception words or phrases that don't follow rules, some abbreviations, stress pattern for other words, and other flags that decide the naturalness of the generated speech.

1.8 THESIS OUTLINE

This thesis will be presented in five chapters. Chapter Two describes the theoretical and background of speech synthesis with brief listing of the most commonly used technologies and algorithms. Some comprehensive explanations on the available Arabic TTS synthesis systems including the technology used will be discussed in this chapter.

Chapter Three will discuss the NLP module of the proposed Arabic TTS synthesizer system that deal with the preprocessing a raw text, started with recognizing a grapheme, and assigning the phonetical entities before converting it into sound, basically the formulation of the phonemic and phonetic rules into algorithms that are applicable to the computer-based processing of input Arabic text in NLP module. Stress pattern for each word will be explained in Chapter Three too.

Chapter Four will focus on design and implementation process including phoneme database building, recording special consonants, setting pronunciation rules, and produce the output speech.