

UNIVERSITI TEKNOLOGI MARA

**GRAMMAR-BASED PROSODY
MODIFICATION FOR EXPLICIT
CONTROL MALAY LANGUAGE
STORYTELLING SPEECH
SYNTHESIS**

MUHAMMAD IZZAD BIN RAMLI

Thesis submitted in fulfillment
of the requirements for the degree of
Doctor of Philosophy

Faculty of Computer and Mathematical Sciences

May 2018

CONFIRMATION BY PANEL OF EXAMINERS

I certify that a panel of examiners has met on 23th January 2018 to conduct the final examination of Muhammad Izzad bin Ramli on his Doctor of Philosophy thesis entitled “Grammar-Based Prosody Modification for Explicit Control Malay Language Storytelling Speech Synthesis” in accordance with Universiti Teknologi MARA Act 1976 (Akta 173). The Panel of Examiners recommends that the student is awarded the relevant degree. The panel of Examiners was as follows:

Puzziawati Ab. Ghani, PhD
Associate Professor Datin
Faculty of Computer & Mathematical Sciences
Universiti Teknologi MARA (UiTM)
(Chairman)

Norhaslinda Kamaruddin, PhD
Senior Lecturer
Faculty of Computer & Mathematical Sciences
Universiti Teknologi MARA (UiTM)
(Internal Examiner)

Uma Shanker, PhD
Professor
Indian Institute of Information Technology, India
(External Examiner)

Shyamala Doraisamy, PhD
Associate Professor
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(External Examiner)

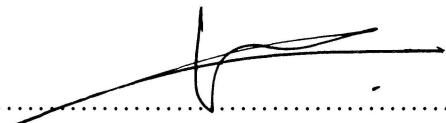
PROF SR DR HJ ABDUL HADI
HJ NAWAWI
Dean
Institute of Graduates Studies
Universiti Teknologi MARA
Date: 16th May 2018

AUTHOR'S DECLARATION

I declare that the work in this thesis was carried out in accordance with the regulations of Universiti Teknologi MARA. It is original and is the result of my own work unless otherwise indicated or acknowledged as referenced work. This thesis has not been submitted to any other academic institution or non-academic institution for any degree or qualification.

I, hereby, acknowledge that I have been supplied with the Academic Rules and Regulations for Post Graduate, Universiti Teknologi MARA, regulating the conduct of my study and research.

Name of Student : Muhammad Izzad bin Ramli
Student I.D. No. : 2014859498
Programme : Doctor of Philosophy of Science - CS950
Faculty : Computer and Mathematical Sciences
Thesis Title : Grammar-Based Prosody Modification for Explicit
Control Malay Language Storytelling Speech
Synthesis

Signature of Student : 

Date : May 2018

ABSTRACT

Storytelling speech synthesis is a process of converting written text to the spoken speech in storytelling speaking style. It has gained much interest in the area of digital storytelling and storytelling humanoid robot for children in learning environment. Reviews have shown that storytelling speech synthesis can be developed using implicit control, explicit control or playback approach. The literatures stated that each approach has its own drawbacks and needs to be tackled for a better quality synthesized speech. In this thesis, explicit control is selected because it is commonly used in the storytelling speech synthesis and has shown to produce good intelligibility and reasonably natural speech. However, modification of prosody in explicit control approach remains a problem as it may lead to speech quality degeneration due to extreme over-exaggeration of speech. Furthermore, perception evaluation showed that the similarity score between the natural and synthesized speech can also be improved for a more satisfactory result. Therefore, this research aims to introduce a new prosody modification technique to reduce over-exaggeration and simultaneously improve the similarity between the natural and synthesized speech. Three narrative children short stories in neutral and storytelling styles are recorded by nine storytellers. A total of 522 speech sentences, 5,238 words and 12,294 syllables are collected to be utilized as experimental datasets and prosody analysis. Based on the prosody analysis, a grammar-based prosody modification rules are proposed by integrating grammatical structure. Consequently, new rules and algorithm that is *MustFront* rule, *limitation* rule, and *two-steps pitch contour* algorithm are introduced to increase the synthesized speech quality. Using Harmonic Noise Model (HNM) as the synthesizer, the grammar-based prosody modification rules are used to produce the synthesized storytelling speech. The synthesized storytelling speech is then compared to baseline methods of synthesized storytelling speech that are global and local prosody modification rules. The evaluation of the synthesized storytelling speech was conducted using objective test (Perceptual Evaluation of Speech Quality (PESQ) test, and aspects or components test) and perceptive test (naturalness, intelligibility, similarity test, and recognition test). The result of PESQ test showed that grammar-based prosody modification with limitation rule produced the highest Mean Opinion Score (MOS) of 3.35 based on five-point scale. The prosody parameters test also demonstrated that the synthesized storytelling speech using grammar-based with limitation rule is much closer to the natural storytelling speech. As for the perception test evaluated by nine native speakers, results showed that grammar-based rule with limitation rule is able to outperform local and global rules by achieving the naturalness, intelligibility and similarity Mean Opinion Score (MOS) of 4.11, 4.47 and 4.06, respectively using a five-point scale. The proposed rule also managed a high accuracy rate of 92% for the recognition test. As conclusion, the performance of the synthesized storytelling speech using grammar-based with limitation rule is better than local and global rules.

TABLE OF CONTENTS

	Page
CONFIRMATION BY PANEL OF EXAMINERS	ii
AUTHOR'S DECLARATION	iii
ABSTRACT	iv
ACKNOWLEDGEMENTS	v
TABLE OF CONTENTS	vi
LIST OF TABLES	xi
LIST OF FIGURES	xiii
LIST OF ABBREVIATIONS	xv
CHAPTER ONE: INTRODUCTION	1
1.1 Background	1
1.2 Problem Statements	4
1.3 Research Questions	6
1.4 Objectives	6
1.5 Research Scope	6
1.6 Significance of Study	8
1.7 Organization of The Thesis	9
1.8 Summary	10
CHAPTER TWO: LITERATURE REVIEW	11
2.1 Introduction	11
2.2 Speech Synthesis	11
2.3 Expressive Speech Synthesis	15
2.3.1 Explicit Control	16
2.3.2 Playback Approach	18
2.3.3 Implicit Control	18
2.3.4 Discussion on Expressive Speech Synthesis Approach	19