

Copyright is owned by the Author of the thesis. Permission is given for a copy to be downloaded by an individual for the purpose of research and private study only. The thesis may not be reproduced elsewhere without the permission of the Author.

**CASTLE: a Computer-Assisted sentence Stress  
Teaching and Learning Environment**

A thesis presented  
in partial fulfilment of the requirements  
for the degree of  
Doctor of Philosophy  
in Computer Science  
at Massey University, Manawatu  
New Zealand

Jingli Lu

2010

# Abstract

A Computer-Assisted sentence Stress Teaching and Learning Environment (CASTLE) is proposed and developed, in order to help learners of English as a Second Language (ESL) to perceive and produce English stress correctly.

Sentence stress plays an important role in English verbal communication. Incorrect stress may confuse listeners, and even break down a conversation. Stress is also challenging for ESL learners to master. It is neither easy for them to produce nor easy to perceive stress. Learners tend to transfer the stress patterns of their first language into English, which is not always appropriate. However, stress has been overlooked in English language teaching classes, due to the time limits of a class and teachers' lack of confidence of teaching stress. Thus, CASTLE is intended to help ESL learners to use sentence stress correctly.

There are three modules in CASTLE: an individualised speech learning material providing module, a perception assistance module and a production assistance module.

Through conducting an investigation into which voice features (i.e. gender, pitch and speech rate) makes a teacher's voice preferable for learners to imitate, we find that learners' imitation preferences vary, according to many factors (e.g. English background and language proficiency). Thus, the speech material providing module of CASTLE can provide individualised speech material for different learners, based on their preferred voice features.

In the perception assistance module of CASTLE, we propose *a set of* stress exaggeration methods that can automatically enlarge the differences between stressed and unstressed syllables in teachers' voice. These stress exaggeration methods are implemented by the manipulation of different prosodic features (i.e. pitch, duration and intensity) of the teachers' voice. Our experimental results show that all our proposed exaggeration methods could help ESL learners to perceive sentence stress more accurately.

In the production assistance module of CASTLE, we propose a clapping-based sentence stress practice model that is intended to help ESL learners to be aware of the rhythm of English language. By analysing the limitation of conventional categorical representation of stress, we propose a fuzzy representation which is intended to better represent the subjective nature of stress. Based on the fuzzy representation of stress, we propose three feedback models in order to help the learners correct their stress errors.

In addition to the development of CASTLE, we also propose an enhanced fuzzy linear regression model which can overcome the spreads increasing problem encountered by previous fuzzy linear regression models.

Dedicated to my parents for their love,  
encouragement and endless support.

# Acknowledgements

I would like to take this opportunity to express my appreciation and gratitude to those people who have supported me to achieve this qualification.

My first sincere thanks go to my supervisor, Dr. Ruili Wang, for his invaluable guidance and tremendous support throughout this research. Without his tireless directions and continuing encouragement, it would have been unfeasible for me to achieve my PhD degree. His intellectual rigour and logical way of thinking have had a remarkable influence on my academic career.

I would like to express my deep gratitude to my co-supervisors Dr. Liyanage C. De Silva and Dr. Helen Zhou, for the time and effort they have spent with me, during my PhD study. I appreciate their valuable suggestions and constructive comments.

I am also grateful to Dr. Shichao Zhang, my previous supervisor for my Master's degree, who introduced me to the field of Computer Science and put my footsteps onto the research path.

I thank Claire, Rosalind and all the participants for their help in the system evaluation. Thanks to Jason, Frank, Yan and June, and other friends at Massey University, for their support and friendship.

I gratefully acknowledge the funding from the *Foundation for Research, Science and Technology* towards my study and research.

Lastly, my special thanks go to my parents for their support, understanding and encouragement.

# Contents

<b>Chapter 1. Introduction and Scope .....</b>	<b>1</b>
1.1 Introduction .....	1
1.2 Scope of this thesis .....	3
<b>Chapter 2. Motivation and Research Objectives .....</b>	<b>5</b>
2.1 Motivation .....	5
2.1.1 Importance of English stress .....	5
2.1.2 Difficulties in learning English stress faced by ESL learners .....	7
2.1.3 Current computer-assisted pronunciation teaching .....	8
2.2 Computer-Assisted sentence Stress Teaching and Learning Environment (CASTLE) .....	10
2.2.1 Research issues and proposed solutions .....	10
2.2.2 A framework for sentence stress teaching systems .....	12
2.2.3 Flowchart of CASTLE system .....	13
2.3 Summary .....	15
<b>Chapter 3. Speech Processing Techniques for CASTLE .....</b>	<b>16</b>
3.1 Literature review of automatic phoneme alignment .....	16
3.1.1 Previous work on automatic phoneme alignment .....	17
3.1.2 Performance comparison .....	20
3.2 Automatic phoneme alignment in CASTLE .....	22
3.2.1 Deficiency of previous phoneme alignment algorithms .....	22
3.2.2 Linear-regression-based flexible boundary phoneme alignment .....	24
3.2.2 TIMIT speech corpus .....	27
3.2.3 Experiments .....	28
3.3 Literature review of automatic stress detection .....	30
3.3.1 Previous work on automatic stress detection .....	30
3.3.2 Performance comparison .....	32
3.4 Automatic stress detection in CASTLE .....	33
3.4.1 Boston University Radio News speech corpus .....	33
3.4.2 Feature extraction .....	35
3.4.3 Experiments .....	37
3.5 Summary .....	38
<b>Chapter 4. Individualised Speech Material Module .....</b>	<b>40</b>
4.1 Previous research on voices suitable for learners to imitate .....	40
4.1.1 The learner's own voice .....	41
4.1.2 Voices of multiple speakers .....	42
4.2 In search of golden speaker from imitation preference perspective .....	44
4.3 Prosody modification techniques .....	46
4.3.1 Duration modification .....	46
4.3.2 Pitch modification .....	47
4.4 Experimental setup .....	49
4.4.1 Speech material .....	49
4.4.2 Participants .....	50
4.4.3 Procedures .....	50

4.5 Experimental results and discussions.....	52
4.6 Conclusions .....	57
4.7 Summary .....	58
<b>Chapter 5. Exaggeration-based Perception Assistance Module.....</b>	<b>61</b>
5.1. Hyper-pronunciation training.....	61
5.2. Pronunciation training based on prosody modification .....	63
5.3. Automatic stress exaggeration .....	65
5.3.1 Pitch-based stress exaggeration .....	66
5.3.2 Duration-based stress exaggeration .....	69
5.3.3 Intensity-based stress exaggeration.....	70
5.3.4 Combined stress exaggeration .....	71
5.4 Perceptual experiments .....	72
5.4.1 Participants.....	72
5.4.2 Speech material .....	72
5.4.3 Results and discussion .....	74
5.5 Summary .....	76
<b>Chapter 6. Production Assistance Module .....</b>	<b>78</b>
6.1 Clapping-based pronunciation practice assistance model.....	78
6.1.1 Clapping in pronunciation learning.....	78
6.1.2 Description of the CPPA model.....	79
6.2 Representation of stress.....	81
6.2.1 A limitation of the categorical representation of stress.....	81
6.2.2 A fuzzy representation of stress .....	82
6.3 Fuzzy representation based stress-error feedback models .....	83
6.3.1 Model Feedback <sub>PC</sub> .....	85
6.3.2 Model Feedback <sub>MC</sub> .....	87
6.3.3 Model Feedback <sub>DI</sub> .....	89
6.4 Flowchart of the production assistance module .....	89
6.5 Summary .....	91
<b>Chapter 7. An Enhanced Fuzzy Linear Regression Model.....</b>	<b>92</b>
7.1 Fuzzy linear regression .....	92
7.2 Fuzzy number and the spreads increasing problem .....	96
7.2.1 Fuzzy number.....	96
7.2.2 Arithmetic operations on fuzzy numbers .....	97
7.2.3 Spreads increasing problem .....	98
7.3 Review on related literature .....	99
7.3.1 Model FLR <sub>KC02</sub> and model FLR <sub>KC03</sub> .....	99
7.3.2 Model FLR <sub>NN04</sub> .....	101
7.3.3 Model FLR <sub>D'Urso03</sub> and model FLR <sub>Coppi06</sub> .....	101
7.3.4 Model FLR <sub>CD08</sub> .....	104
7.4 Flexible spreads FLR model FLR <sub>FS</sub> .....	105
7.4.1 Description of model FLR <sub>FS</sub> .....	105
7.4.2 Property of model FLR <sub>FS</sub> .....	109
7.4.3 Parameters estimation .....	110
7.5. Numerical examples.....	113
7.5.1 Initial values setting .....	113
7.5.2 Examples.....	114

7.6 Summary .....	121
<b>Chapter 8. Conclusions and Future Work .....</b>	<b>123</b>
8.1 Summary of main findings and contributions .....	123
8.1.1 Individualised speech learning material .....	123
8.1.2 Stress-exaggeration-based perception assistance .....	125
8.1.3 Production assistance .....	125
8.1.4 Linear-Regression-based flexible boundary phoneme aligner .....	127
8.1.5 An enhanced fuzzy linear regression model .....	127
8.2 Further research .....	128
<b>Appendix Questionnaire .....</b>	<b>129</b>
<b>References .....</b>	<b>130</b>
<b>Publications Related to This Research .....</b>	<b>140</b>
Published papers .....	140
Submitted papers .....	140

# List of figures

Figure 2.1 Flowchart of CASTLE system .....	13
Figure 3.1 Viterbi-based forced alignment .....	19
Figure 3.2 Comparison between estimated duration and its reference counterpart .....	23
Figure 3.3 Relationships between estimated and reference syllable durations.....	24
Figure 3.4 Possible boundary relationships of two conjunctive phonemes .....	25
Figure 3.5 Overview of the LR-FB phoneme aligner .....	26
Figure 4.1 Screenshot of CASTLE system. ....	51
Figure 4.2 Distributions of the most and the least wanted to be imitated speech.....	53
Figure 5.1 Pitch contour comparison. ....	69
Figure 5.2 Duration-based stress exaggeration .....	70
Figure 5.3 Spectrum comparison. ....	71
Figure 5.4 Boxplot of the <i>F-Measures</i> of listeners' stress pattern labeling .....	75
Figure 6.1 Illustration of the utterance.....	80
Figure 6.2 Resynthesis of clapping-based teacher's utterance.....	81
Figure 6.3 Stress difference between a teacher's syllable and a learner's imitation.....	84
Figure 6.4 Flowchart of the production assistance module .....	90
Figure 7.1 Membership functions of the estimated and observed fuzzy numbers.....	111

# List of tables

Table 3.1	Accuracies of different phoneme aligners on the TIMIT corpus.	22
Table 3.2	Parameters used to train the LR-FB phoneme aligner in CASTLE	28
Table 3.3	Performances of base phoneme aligners and the LR-FB phoneme aligner	29
Table 3.4	Performance comparison of previous stress detectors.	33
Table 3.5	ToBI labels associated with stressed syllables.	34
Table 3.6	Input features of the stress detector(s) in CASTLE	35
Table 3.7	Performances of different stress detectors	38
Table 4.1	The average of the absolute deviations from the mean	56
Table 5.1	ToBI labels and their corresponding exaggeration operations.	67
Table 5.2	Distribution of syllables and stressed syllables in sentences and clusters.	73
Table 5.3	Distribution of utterance clusters in each type of listening material.	74
Table 5.4	Comparison of listeners' stress pattern labeling accuracy	74
Table 6.1	Inputs and output of the prototype of stress-error feedback model	85
Table 7.1	Dataset1	95
Table 7.2	Dataset2	115
Table 7.3	Fuzzy regression models of dataset2	116
Table 7.4	Comparison of the performance of difference methods on Dataset2	117
Table 7.5	Dataset3	118
Table 7.6	Fuzzy regression models of dataset3	119
Table 7.7	Comparison of the performance of difference methods on Dataset3	119
Table 7.8	Dataset4: Restaurants data	120
Table 7.9	Comparison of the performance of difference methods on Dataset4	121