

TABLE OF CONTENTS

CHAPTER NO.	TITLE	PAGE NO.
	ABSTRACT	iv
	LIST OF TABLES	xiii
	LIST OF FIGURES	xiv
	LIST OF SYMBOLS AND ABBREVIATIONS	xvi
1	INTRODUCTION	1
1.1	TEXT TO SPEECH SYNTHESIS	1
1.1.1	Terminology	2
1.1.2	An overview of a Text to Speech Synthesis System	5
1.2	OBJECTIVE OF THE RESEARCH	8
1.3	SCOPE FOR A TEXT TO SPEECH SYNTHESIS SYSTEM	8
1.4	TEXT-TO-SPEECH SYNTHESIZERS FOR INDIAN LANGUAGES	10
1.5	ISSUES FOUND IN DIPHONES	10
1.6	ISSUES FOUND IN SYLLABLES	11
1.6.1	Syllabification Rules	11
1.6.2	Sample Syllabification	14
1.6	OBJECTIVE AND SCOPE OF THE THESIS	15
1.7	ORGANIZATION OF THE THESIS	17
1.8	MAJOR CONTRIBUTIONS OF THE RESEARCH WORK	18

CHAPTER NO.	TITLE	PAGE NO.
2	TEXT-TO-SPEECH SYNTHESIS	19
2.1	INTRODUCTION	19
2.2	NATURAL LANGUAGE GENERATION	20
2.3	EXPERTS' FINDINGS IN PREVIOUS WORKS	21
2.4	APPROACHES TO PROSODY MODELING	22
2.4.1	Prosodic Phrasing	23
2.4.2	Pitch Modeling	24
2.4.3	Duration Modeling	26
2.4.4	Intensity Modeling	27
2.5	CONCATENATIVE SPEECH SYNTHESIS	29
2.5.1	Pitch and Duration Modification	30
2.5.2	Unit selection Synthesis	30
2.6	SUMMARY	32
3	TEXT TO SPEECH SYNTHESIS FOR MOBILE DEVICES – ARCHITECTURE AND DESIGN	34
3.1	INTRODUCTION	34
3.2	DESIGN OF THE DISTRIBUTED SPEECH SYNTHESIS SYSTEM USING THE CONVENTIONAL FESTIVAL SOFTWARE	40
3.3	DESIGN OF THE DISTRIBUTED SPEECH SYNTHESIS SYSTEM USING FLITE SOFTWARE	44
3.4	SUMMARY	46

CHAPTER NO.	TITLE	PAGE NO.
4	TEXT-TO-SPEECH SYNTHESIS WITH AUTOMATIC GENERATION OF SPEECH UNITS	48
4.1	INTRODUCTION	48
4.2	DIPHONE SYNTHESIS	48
4.3	UNIT SELECTION SYNTHESIS	51
4.4	MANUAL EXTRACTION OF SYLLABLES FROM TAMIL WORDS	54
4.5	SEGMENTATION USING MINIMUM PHASE GROUP DELAY FUNCTIONS	54
4.6	SUMMARY	56
5	ARCHITECTURE OF THE PROPOSED SYSTEM	57
5.1	INTRODUCTION	57
5.2	CORPUS ENTRY FOR SYLLABLES AND DIPHONES	58
5.3	MEETING THE FUNCTIONAL COMPONENTS OF PROSODY	60
5.3.1	Phrase Modeling	61
5.3.2	Duration Modeling	64
5.3.3	Intonation Modeling	66
5.4	SYSTEM ARCHITECTURE	67
5.5	SYLLABLE AND DIPHONE CLUSTERING	70
5.6	CONTEXT SWITCHING BETWEEN DATABASES	73
5.7	SUMMARY	75

CHAPTER NO.	TITLE	PAGE NO.
6	DESIGN OF THE SPEECH DATABASES FOR THE TEXT TO SPEECH SYNTHESIS IN TAMIL	76
6.1	INTRODUCTION	76
6.2	GENERATING POLYSYLLABLE UNITS USING GROUP DELAY BASED SEGMENTATION ALGORITHM	76
6.3	CREATING SAMPLE SYLLABLE AND DIPHONE DATABASES	79
6.4	EVALUATION OF THE SYLLABLE AND DIPHONE UNITS	82
6.4.1	General Evaluation of the Speech Units	82
6.4.2	Prosodic Evaluation of Syllables and Diphones	84
6.5	UNRESTRICTED SPEECH SYNTHESIS	87
6.6	SUMMARY	89
7	MODELING AND EVALUATION OF PROSODY FEATURES	90
7.1	INTRODUCTION	90
7.1.1	Modeling Pause Duration Between Syllables	90
7.1.2	Pause Duration Between Velar Sounds and Other Sound Units	93
7.1.3	Pause Duration Between Palatal Sounds and Other Classes	94

CHAPTER NO.	TITLE	PAGE NO.
7.1.4	Pause Duration Between Retroflex Sounds and Other Classes	94
7.1.5	Pause Duration Between Dental Sounds and Other Classes	95
7.1.6	Pause Duration Between Labial Sounds and Other Classes	96
7.1.7	Pause Duration Between Approximant, Fricative Sounds and Other Classes	97
7.1.8	Pause Duration Between Vowel Sounds and Other Classes	98
7.1.9	Pause Duration Between Syllables and Diphones	99
7.2	CART MODEL FOR PAUSE DURATIONS	100
7.3	SUMMARY	101
8.	RESULTS AND DISCUSSION	102
8.1	SUMMARY AND CRITICISM	102
8.2	FUTURE DIRECTIONS	104
	APPENDIX	105
	REFERENCES	107
	LIST OF PUBLICATIONS	118

LIST OF TABLES

TABLE NO.	TITLE	PAGE NO.
2.1	ToBI Transcription for pitch accent patterns	26
4.1	Details of the manual syllabification of the word ' <i>arasangam</i> '	54
5.1	POS Tags and their descriptions	62
5.2	Predicting occurrences of morphemes in Tamil	64
5.3	Standard pitch accent patterns	67
6.1	Various speech units studied and evaluated using Group Delay based Segmentation Algorithm	77
6.2	MOS for sentences synthesized using Diphones, Syllables, and dual database	83
6.3	Comparative number of units synthesized using Diphones, Syllables, and Dual database	84
6.4	Predicting realizations of morphemes in Tamil	88
7.1	Assigning Class Ids for various manners of articulations in Tamil (Syllables)	92
7.2	Assigning Class Ids for various manners of articulations in Tamil (Diphones)	92
A1.1	Comparative statement of various speech units	105

LIST OF FIGURES

FIGURE NO.	TITLE	PAGE NO.
1.1	Structure of a syllable	3
1.2	Two main components of TTS	5
1.3	Block diagram of a TTS system	6
1.4	A Typical Text to Speech Synthesis System	7
1.5	Evolution of Text to Speech Synthesis system	7
1.6	Text-to-speech synthesis cycle	9
2.1	Typical Natural Language Generation System	20
2.2	An example CART tree	28
3.1	Survey of the number of mobile phone users in India	34
3.2	Middleware client-server architecture	35
3.3	Role of middleware in client-server architecture	36
3.4	TTS cycle in Distributed Environment	37
3.5 (i)	User Interface proposed for the DSS system	39
3.5 (ii)	User Interface on receiving the transliterated Tamil text input	39
3.6	Block Diagram of the DSS system	41
4.1	Transition between two phones in a diphone unit	49
4.2	Block diagram of TTS system using diphone synthesis in the Festival speech synthesis framework	50
4.3	Architecture of TTS system using unit selection based synthesis in Festival	52
4.4	Decision tree that clusters various entries of the speech unit <i>'/bi/</i>	53

FIGURE NO.	TITLE	PAGE NO.
5.1	Modeling a dual database for Tamil Speech Synthesis	59
5.2	An example CART tree	66
5.3	Proposed System Architecture	69
5.4	Text to Speech Synthesis System	71
5.5	Various phases of a Text to Speech Synthesis System	72
5.6	Context switching between databases	73
5.7	Effect of Context Switching between databases in time taken for overall synthesis	74
6.1	The speech waveform for the phrase <i>‘vAzhga vaLamudan.’</i>	87
7.1	Average Pause Duration between velar sounds and other sound units	93
7.2	Average Pause Duration between palatal sounds and other sound units	94
7.3	Average Pause Duration between Retroflex sounds and other sound units	95
7.4	Average Pause Duration between Dental sounds and other sound units	96
7.5	Average Pause Duration between labial sounds and the other sound units	97
7.6	Average Pause Duration between approximants, fricative sounds, and the other sound units	98
7.7	Average Pause Duration between vowels and the other sound units	99
7.8	Pause Duration taken by syllables and Diphones	100