

CONTENTS

	PAGE
ENGLISH ABSTRACT	ii
THAI ABSTRACT	iii
ACKNOWLEDGEMENTS	v
CONTENTS	vi
LIST OF TABLES	viii
LIST OF FIGURES	ix
LIST OF SYMBOLS	xiv
LIST OF TECHNICAL VOCABULARY AND ABBREVIATIONS	xv
 CHAPTER	
1. INTRODUCTION	1
1.1. Motivation	1
1.2. Contribution	3
 2. BACKGROUND AND RELATED WORK	 5
2.1. Importance of Automatic Tonal Speech Recognizer	5
2.2. Speech Recognition System for Hardware Implementation	6
2.3. Tone Recognition	13
2.4. High Thesis Statement	19
 3. THEORY AND SOLUTION OF LOW CALCULATION COMPLEXITY TONE RECOGNITION	 21
3.1 Overview of Reduced Complexity Tone Recognition	21
3.2 Feature Extraction	21
3.3 F_0 Estimation	30
3.4 Tone Decision	32
 4. ARCHITECTURE CALCULATION COMPLEXITY OF TONE RECOGNITION	 35
4.1 Low Power Tone Recognition	37
4.2 Parallel Architecture of a Low Power Tone Recognition	41
 5. RESULT AND EVALUATION	 43
5.1. Recognition Accuracy	43
5.2. Calculation Complexity	47
5.3. Processing Time	52
 6. DISCUSSION	 54
6.1. Recognition Accuracy	46
6.2. Vowel-MDF Segmentation	46
6.3. ATSR Without Tone Recognition	47
 7. CONCLUSION AND FUTURE WORK	 58
7.1 Limitations and Future Work	58
 REFERENCES	 59

	PAGE
APPENDIX	
A. Fundamental Frequency Trajectories of Five Thai Tones	67
B. Phonetic Alphabets and Test Words	71
C. Fundamental Frequency Trajectories	80
 CURRICULUM VITAE	 118

LIST OF TABLES

TABLE		PAGE
1.1	Tone influence by using the same consonant and vowel phonemes but different tones	1
4.1	The number of operations in a frame	37
4.2	Processing order of PE1	42
4.3	Number of cycles processed in a frame	44
5.1	Percentage improvement of tone recognition accuracy	48
5.2	Percentage improvement of speech recognition accuracy	49
5.3	Number of frames	49
B.1.1	Thai initial and final consonant phonetic alphabets	71
B.1.2	Vowel phonetic alphabet	72
B.1.3	Five Thai tone and four Mandarin Chinese tone diacritics	72
B.2.1	110 test words for tone recognition	73
B.3.1	45 test words for automatic tonal speech recognition	77

LIST OF FIGURES

FIGURE	PAGE
1.1 Standard F_0 contours of four Mandarin Chinese tones	2
1.2 Standard F_0 contours of five Thai tones	2
2.1 List of the 30 most populous countries in 2013	5
2.2 StoredBPP based VLSI architecture	7
2.3 Flowchart of general HMM computation	8
2.4 Complete recognition system	9
2.5 The proposed output probability calculation (OPC) flow chart	10
2.6 Two-stage language model search	10
2.7 Viterbi cache architecture	11
2.8 Proposed speech recognition	11
2.9 GMM computation flow	12
2.10 Viterbi computation flow	12
2.11 Actual and the synthesized F_0 contours (top) A plot of the amplitude and temporal locations of the tone commands (bottom)	14
2.12 (a) baseline tone features (b) baseline tone features + tonal assimilation features (c) first-half tone features (d) second-half tone features	14
2.13 The example of the F_0 contour model parameter extraction for an utterance of Standard Chinese read by Speaker B	15
2.14 F_0 contours and underlying tone command pattern for four lexical Chinese tones of the word “yi”	16
2.15 Tone nucleus extraction algorithm	17
2.16 Examples of extracted onset course, tone nuclei, and offset course	18
2.17 Tone nucleus modeling and proposed V_{MDF} comparison	19
3.1 An overview of reduced complexity tone recognition	21
3.2 Digital filter low-pass filters (a) Butterworth low-pass filter with $P = 26$, (b) Chebyshev type 1 low-pass filter with $P = 9$, (c) Chebyshev type 2 low-pass filter with $P = 9$, (d) Elliptic low-pass filter with $P = 5$	22
3.3 Specifications of digital band-pass filter	23
3.4 Frequency response of digital band-pass filter	23
3.5 The word “ <i>chûu</i> ” before (top) and after (bottom) elliptic low-pass filter	24
3.6 Example of the word “ <i>thōo / rā / sàp</i> ” with input speech (top), zero crossing (middle), speech energy (bottom)	26
3.7 Example of the word “ <i>thōo / rā / sàp</i> ” syllable segmentation	26
3.8 Vowel segmentation of the word “ <i>lûak</i> ” before applying vowel threshold (top) after applying vowel threshold (bottom)	27
3.9 F_0 feature waveforms of the word “ <i>lûak</i> ” where $j = 31$ and $P_v = 48$ (a) AC (b) $AMDF$ (c) V_{MDF}	29
3.10 Association between the current, past, and consecutive samples where a frame $j = 6$ and $l_l(6) = 44$	31
3.11 Example of the word “ <i>hnâa</i> ”; vowel signal (top), V_{MDF} -based pitch detector (middle), and F_0 characteristic of falling tone (bottom)	32
3.12 Decision tree of 5 Thai tones	33

LIST OF FIGURES

FIGURE	PAGE	
3.13	F_0 trajectories of the word “ <i>lûak</i> ”; AC (top), V_{MDF} (middle), and $AMDF$ (bottom)	34
4.1	Comparison of architectures for programmable hardware	35
4.2	Series computation of tone recognition	39
4.3	Parallel computation of tone recognition	40
4.4	Parallel architecture of proposed tone recognition	41
4.5	Structure of process element 1 (PE1) and data allocation	41
4.6	Process element 2 (PE2)	43
4.7	Process element 3 (PE3)	43
5.1	Comparison of tone recognition accuracy of the word 1-25	46
5.2	Comparison of tone recognition accuracy of the word 26-50	46
5.3	Comparison of tone recognition accuracy of the word 51-75	47
5.4	Comparison of tone recognition accuracy of the word 76-100	47
5.5	Comparison of tone recognition accuracy of the word 101-110	47
5.6	Comparison of ATSR recognition accuracy of the word 1-23	48
5.7	Comparison of ATSR recognition accuracy of the word 24-45	48
5.8	Data allocation performance	55
5.9	Total processing time	55
6.1	F_0 affected by the preceding syllable	56
6.2	A very low pitch point value detected by feature extraction process	57
6.3	V_{MDF} and AC after mesh 5	57
6.4	Vowel region error of syllable 1	58
6.5	Vowel segmentation error of syllable 3	58
6.6	Vowel segmentation errors	58
6.7	Zero-crossing rate errors	59
6.8	Overview of ATSR	59
A.1	Mid tone of the word <i>khāa</i>	68
A.2	Low tone of the word <i>kaʼi</i>	68
A.3	Falling tone of the word <i>khâa</i>	68
A.4	Falling tone of the word <i>khâa</i>	69
A.5	Rising tone of the word “ <i>sǝǝŋ</i> ”	69
C.1	F_0 trajectory of “ <i>māa</i> ”	81
C.2	F_0 trajectory of “ <i>m`aa</i> ”	81
C.3	F_0 trajectory of “ <i>mâa</i> ”	81
C.4	F_0 trajectory of “ <i>ma'a</i> ”	82
C.5	F_0 trajectory of “ <i>mãa</i> ”	82
C.6	F_0 trajectory of “ <i>mîi</i> ”	82
C.7	F_0 trajectory of “ <i>mii</i> ”	83
C.8	F_0 trajectory of “ <i>mîi</i> ”	83
C.9	F_0 trajectory of “ <i>mi'i</i> ”	83
C.10	F_0 trajectory of “ <i>mîi</i> ”	84
C.11	F_0 trajectory of “ <i>māy</i> ”	84
C.12	F_0 trajectory of “ <i>m`ay</i> ”	84

LIST OF FIGURES

FIGURE	PAGE
C.13 <i>F</i> ₀ trajectory of “ <i>mây</i> ”	85
C.14 <i>F</i> ₀ trajectory of “ <i>ma'y</i> ”	85
C.15 <i>F</i> ₀ trajectory of “ <i>măy</i> ”	85
C.16 <i>F</i> ₀ trajectory of “ <i>khāy</i> ”	86
C.17 <i>F</i> ₀ trajectory of “ <i>kh`ay</i> ”	86
C.18 <i>F</i> ₀ trajectory of “ <i>khây</i> ”	86
C.19 <i>F</i> ₀ trajectory of “ <i>kha'y</i> ”	87
C.20 <i>F</i> ₀ trajectory of “ <i>khăy</i> ”	87
C.21 <i>F</i> ₀ trajectory of “ <i>kāy</i> ”	87
C.22 <i>F</i> ₀ trajectory of “ <i>k`ay</i> ”	88
C.23 <i>F</i> ₀ trajectory of “ <i>kây</i> ”	88
C.24 <i>F</i> ₀ trajectory of “ <i>ka'y</i> ”	88
C.25 <i>F</i> ₀ trajectory of “ <i>kăy</i> ”	89
C.26 <i>F</i> ₀ trajectory of “ <i>khāa</i> ”	89
C.27 <i>F</i> ₀ trajectory of “ <i>kh`aa</i> ”	89
C.28 <i>F</i> ₀ trajectory of “ <i>khâa</i> ”	90
C.29 <i>F</i> ₀ trajectory of “ <i>kha'a</i> ”	90
C.30 <i>F</i> ₀ trajectory of “ <i>khăa</i> ”	90
C.31 <i>F</i> ₀ trajectory of “ <i>nāa</i> ”	91
C.32 <i>F</i> ₀ trajectory of “ <i>năa</i> ”	91
C.33 <i>F</i> ₀ trajectory of “ <i>nâa</i> ”	91
C.34 <i>F</i> ₀ trajectory of “ <i>n`aa</i> ”	92
C.35 <i>F</i> ₀ trajectory of “ <i>na'a</i> ”	92
C.36 <i>F</i> ₀ trajectory of “ <i>chûu</i> ”	92
C.37 <i>F</i> ₀ trajectory of “ <i>m`uu</i> ”	93
C.38 <i>F</i> ₀ trajectory of “ <i>sṵy</i> ”	93
C.39 <i>F</i> ₀ trajectory of “ <i>kh`eet</i> ”	93
C.40 <i>F</i> ₀ trajectory of “ <i>lām / phōŋ</i> ”	94
C.41 <i>F</i> ₀ trajectory of “ <i>hēen / frii</i> ”	94
C.42 <i>F</i> ₀ trajectory of “ <i>sīaŋ / riak</i> ”	94
C.43 <i>F</i> ₀ trajectory of “ <i>pəət / sīaŋ</i> ”	95
C.44 <i>F</i> ₀ trajectory of “ <i>pīt / sīaŋ</i> ”	95
C.45 <i>F</i> ₀ trajectory of “ <i>tām / bōn</i> ”	95
C.46 <i>F</i> ₀ trajectory of “ <i>ṵām / phṵ</i> ”	96
C.47 <i>F</i> ₀ trajectory of “ <i>lōŋ</i> ”	96
C.48 <i>F</i> ₀ trajectory of “ <i>khwăa</i> ”	96
C.49 <i>F</i> ₀ trajectory of “ <i>th`at / pāy</i> ”	97
C.50 <i>F</i> ₀ trajectory of “ <i>hmṵan</i> ”	97
C.51 <i>F</i> ₀ trajectory of “ <i>thēew</i> ”	97

LIST OF FIGURES

FIGURE	PAGE
C.52 <i>F</i> ₀ trajectory of “ <i>l̥h̥ak</i> ”	98
C.53 <i>F</i> ₀ trajectory of “ <i>kh̥am</i> ”	98
C.54 <i>F</i> ₀ trajectory of “ <i>t̥at</i> ”	98
C.55 <i>F</i> ₀ trajectory of “ <i>p̥ê</i> ”	99
C.56 <i>F</i> ₀ trajectory of “ <i>ɔ̄aw / ɔ̄ɔk</i> ”	99
C.57 <i>F</i> ₀ trajectory of “ <i>t̥ok / t̄oŋ</i> ”	99
C.58 <i>F</i> ₀ trajectory of “ <i>s̄uu ŋ</i> ”	100
C.59 <i>F</i> ₀ trajectory of “ <i>t̥am</i> ”	100
C.60 <i>F</i> ₀ trajectory of “ <i>l̄ot</i> ”	100
C.61 <i>F</i> ₀ trajectory of “ <i>ch̄oŋ</i> ”	101
C.62 <i>F</i> ₀ trajectory of “ <i>t̄ɛɛn / t̄ii</i> ”	101
C.63 <i>F</i> ₀ trajectory of “ <i>m̄ak / s̄ut</i> ”	101
C.64 <i>F</i> ₀ trajectory of “ <i>n̄ɔy / s̄ut</i> ”	102
C.65 <i>F</i> ₀ trajectory of “ <i>r̄ā / d̄ap / s̄i aŋ</i> ”	102
C.66 <i>F</i> ₀ trajectory of “ <i>r̄uup / b̄ɛɛp</i> ”	102
C.67 <i>F</i> ₀ trajectory of “ <i>r̄āay / k̄aan</i> ”	103
C.68 <i>F</i> ₀ trajectory of “ <i>s̄āt / s̄uan</i> ”	103
C.69 <i>F</i> ₀ trajectory of “ <i>hȳut / ph̄ap</i> ”	103
C.70 <i>F</i> ₀ trajectory of “ <i>t̄ām / sa'm</i> ”	104
C.71 <i>F</i> ₀ trajectory of “ <i>ph̄ɛɛn / t̄ii</i> ”	104
C.72 <i>F</i> ₀ trajectory of “ <i>c̄ii / phi / ɔ̄ɛt</i> ”	104
C.73 <i>F</i> ₀ trajectory of “ <i>pl̄āay / th̄a aŋ</i> ”	105
C.74 <i>F</i> ₀ trajectory of “ <i>n̄ām / th̄ā a ŋ</i> ”	105
C.75 <i>F</i> ₀ trajectory of “ <i>p̄āy / t̄ii</i> ”	105
C.76 <i>F</i> ₀ trajectory of “ <i>s̄êen / th̄ā a ŋ</i> ”	106
C.77 <i>F</i> ₀ trajectory of “ <i>r̄oŋ / r̄ɛɛm</i> ”	106
C.78 <i>F</i> ₀ trajectory of “ <i>c̄ūt / ph̄āk / ph̄ɔn</i> ”	106
C.79 <i>F</i> ₀ trajectory of “ <i>ph̄āt / t̄ā a / kh̄ā a ŋ</i> ”	107
C.80 <i>F</i> ₀ trajectory of “ <i>kh̄r̄aŋ / pr̄ap / ā a / k̄aat</i> ”	107
C.81 <i>F</i> ₀ trajectory of “ <i>c̄ut / t̄at</i> ”	107
C.82 <i>F</i> ₀ trajectory of “ <i>c̄ut / ch̄ōm / w̄iw</i> ”	108
C.83 <i>F</i> ₀ trajectory of “ <i>r̄ót / t̄it</i> ”	108
C.84 <i>F</i> ₀ trajectory of “ <i>th̄ōo / r̄ā / s̄ap</i> ”	108
C.85 <i>F</i> ₀ trajectory of “ <i>th̄ōo / d̄uan</i> ”	109
C.86 <i>F</i> ₀ trajectory of “ <i>r̄āay / k̄aan / th̄ōo</i> ”	109
C.87 <i>F</i> ₀ trajectory of “ <i>w̄āan / s̄āau</i> ”	109
C.88 <i>F</i> ₀ trajectory of “ <i>th̄ōo</i> ”	110
C.89 <i>F</i> ₀ trajectory of “ <i>sr̄ā a ŋ</i> ”	110
C.90 <i>F</i> ₀ trajectory of “ <i>p̄ii</i> ”	110
C.91 <i>F</i> ₀ trajectory of “ <i>pr̄ā / ch̄ūm / t̄ā a ŋ / th̄ōo / r̄ā / s̄ap</i> ”	111
C.92 <i>F</i> ₀ trajectory of “ <i>r̄āp / s̄āay</i> ”	111
C.93 <i>F</i> ₀ trajectory of “ <i>m̄āy / r̄āp / s̄āay</i> ”	111

LIST OF FIGURES

FIGURE	PAGE
C.94 F_0 trajectory of “ <i>wēe/lāa/thōo</i> ”	112
C.95 F_0 trajectory of “ <i>pāan / klāa η</i> ”	112
C.96 F_0 trajectory of “ <i>cùt / rōy / tō</i> ”	112
C.97 F_0 trajectory of “ <i>nāy / m̄an</i> ”	113
C.98 F_0 trajectory of “ <i>khō / mūun / kāan / cā / rāa / cōn</i> ”	113
C.99 F_0 trajectory of “ <i>dūan</i> ”	113
C.100 F_0 trajectory of “ <i>tū / rā / kī t</i> ”	114
C.101 F_0 trajectory of “ <i>th̄w / sāay / rō</i> ”	114
C.102 F_0 trajectory of “ <i>sūan / tūaw</i> ”	114
C.103 F_0 trajectory of “ <i>d̄an</i> ”	115
C.104 F_0 trajectory of “ <i>pā / t̄ / th̄n</i> ”	115
C.105 F_0 trajectory of “ <i>t̄an / khāam / cām</i> ”	115
C.106 F_0 trajectory of “ <i>ḡāan</i> ”	116
C.107 F_0 trajectory of “ <i>wān / k̄t</i> ”	116
C.108 F_0 trajectory of “ <i>sāp / dāa</i> ”	116
C.109 F_0 trajectory of “ <i>gēm</i> ”	117
C.110 F_0 trajectory of “ <i>sā / m̄t / thōo / s̄p</i> ”	117

LIST OF SYMBOLS

SYMBOL		UNIT
$E()$	Energy	-
F_0	Fundamental frequency	Hz
j	Frame number	-
j_v	Frame that includes the possible pitch frequency range	-
k	Frame ending	-
l_i	Local minimum positions	-
l_l	Integer-valued pitch period	-
l_0	Optimal pitch period coefficient	-
L	Number of silence frame	-
M	Frame shift	-
M_w	Maximum F_0 of each individual word	-
M_{h1}	Maximum F_0 of the first half syllable	-
M_{h2}	Maximum F_0 of the second half syllable	-
N	Frame length	-
$N_0()$	Normalized sum of squared error	-
O	Arithmetic operations	-
P_v	Pitch period	-
S	Slope	-
$ S $	Absolute slope	-
T_e	Energy threshold	-
$T_p()$	Pitch period threshold	-
$T_{sil}()$	Silence threshold	-
$T_z()$	Zero crossing threshold	-
$V_{MDF}()$	Vowel magnitude difference function	-
$V1$	Value from the conventional method	-
$V2$	Value from the proposed method	-
$Z()$	Zero crossing	-

LIST OF TECHNICAL VOCABULARY AND ABBREVIATIONS

AC	=	Autocorrelation
ASR	=	Automatic speech recognizer
ASICs	=	Application-specific integrated circuits
ATSR	=	Automatic tonal speech recognizer
CHMM	=	Continuous hidden Markov model
DA	=	Data allocation
DSP	=	Digital signal processor
FPGAs	=	Field-programmable gate arrays
HMM	=	Hidden Markov's model
PE	=	Process element
SOC	=	System on chip
StoredBPP	=	Stored-based block parallel processing
StreamBPP	=	Stream-based block parallel processing
V_{MDF}	=	Vowel magnitude difference function