

Contents

<i>List of contributors</i>	xi
<i>Preface</i>	1
1 Introduction	7
1.1 Constraints on speech synthesis	7
1.2 Synthesis techniques	9
1.3 Functional outline of MITalk	12
I Analysis	
2 Text preprocessing	16
2.1 Overview	16
2.2 Input	17
2.3 Output	18
2.4 Formatting operations	18
3 Morphological analysis	23
3.1 Overview	23
3.2 Input	27
3.3 Output	27
3.4 The algorithm	28
3.5 An example of a decomposition	35
3.6 The lexicon	36
4 The phrase-level parser	40
4.1 Overview	40
4.2 Input	41
4.3 Output	41
4.4 Parts of speech	41
4.5 The part-of-speech processor	43
4.6 The parser algorithm	45
4.7 Some examples	51
5 Morphophonemics and stress adjustment	52
5.1 Overview	52
5.2 Input	52
5.3 Output	52
5.4 Morphophonemic rules	52
5.5 Stress modification rules	54
5.6 An example	54
6 Letter-to-sound and lexical stress	57
6.1 Overview	57
6.2 Letter-to-sound	57
6.3 Lexical stress placement	61
6.4 An example	69

II Synthesis	
7 Survey of speech synthesis technology	71
7.1 Overview	71
7.2 Background	72
7.3 Synthesis techniques	73
7.4 Applications	79
8 The phonological component	81
8.1 Overview	81
8.2 Input representation for a sentence	81
8.3 Comparison between ideal synthesis input and system performance	85
8.4 Stress rules	86
8.5 Rules of segmental phonology	87
8.6 Pauses	88
8.7 Evaluation of the analysis modules	89
9 The prosodic component	93
9.1 Overview	93
9.2 Segmental durations	93
10 The fundamental frequency generator	100
10.1 Overview	100
10.2 Input	101
10.3 Output	102
10.4 The O'Shaughnessy fundamental frequency algorithm	103
10.5 Adjustments to the O'Shaughnessy algorithm	107
10.6 Potential improvements from additional syntactic information	107
11 The phonetic component	108
11.1 Overview	108
11.2 "Synthesis-by-analysis" of consonant-vowel syllables	109
11.3 General rules for the synthesis of phonetic sequences	116
11.4 Summary	122
12 The Klatt formant synthesizer	123
12.1 Overview	123
12.2 Vocal tract transfer functions	139
12.3 Radiation characteristic	150
13 Some measures of intelligibility and comprehension	151
13.1 Overview	151
13.2 Phoneme recognition	152
13.3 Word recognition in sentences	157
13.4 Comprehension	161
13.5 General discussion and conclusions	167
14 Implementation	172
14.1 Conceptual organization	172
14.2 Development system	173
14.3 Performance system	174
14.4 UNIX implementation	174
14.5 Using the system	175

Appendixes

A	Part-of-speech processor	177
B	Klatt symbols	179
C	Context-dependent rules for PHONET	181
D	Sample test trials from the Modified Rhyme Test	202
E	Sample test materials from the Harvard Psychoacoustic Sentences	203
F	Sample test materials from the Haskins Anomalous Sentences	204
G	Sample passage used to test listening comprehension	205
	<i>References</i>	207
	<i>Index</i>	215

List of figures

2-1	Example of FORMAT processing	18
3-1	State transition diagram for the morph sequence FSM	31
3-2	Decomposition of "scarcity"	37
4-1	Noun group ATN listing	47
4-2	Verb group ATN listing	48
4-3	ATN diagram for verb groups	49
4-4	ATN diagram for noun groups	50
4-5	Example of PARSER operation	51
5-1	Input to and output from SOUND1	55
6-1	Suffix detection in the word finishing	58
6-2	Application of letter-to-sound rules to caribou	60
6-3	Application of letter-to-sound rules to subversion	60
6-4	Example of letter-to-sound and stress rule operation	69
7-1	Synthesis blocks of the MITalk system	72
7-2	An example of the differences between words spoken in isolation and words spoken as a continuous utterance	74
8-1	Example of PHONO1 and PHONO2 processing	82
9-1	Example of the processing performed by PROSOD	94
10-1	Example of F0 contours	105
11-1	Spectrum analysis of a speech waveform	111
11-2	First and second formant motions in English vowels	112
11-3	Linear prediction of plosive bursts before vowels	113
11-4	Frequency of the lowest three formants measured at voicing onset for syllables involving BB, DD, and GG	114
11-5	Synthesis strategy for a CV syllable	115
11-6	Templates for smoothing adjacent phonetic segment targets	117
11-7	Constants used to specify the inherent formant and durational characteristics of a sonorant	120
12-1	Interface between synthesizer software and hardware	123
12-2	Components of the output spectrum of a speech sound	125
12-3	Parallel and cascade simulation of the vocal tract transfer function	126
12-4	Cascade/parallel configurations supported by MITalk	127
12-5	Block diagram and frequency response of a digital resonator	129
12-6	Block diagram of the cascade/parallel formant synthesizer	131
12-7	Four periods from voicing waveforms	135
12-8	Waveform segment and magnitude spectrum of frication noise	137
12-9	Magnitude of the vocal tract transfer function	141
12-10	Nasalization of the vowel IH in the syllable "dim"	143
12-11	Effect of parameter changes on the vocal tract transfer function	146
12-12	Preemphasized output spectra from cascade and parallel models	148
12-13	Spectra from two different parallel synthesis configurations	149
12-14	Transfer function of the radiation characteristic	150
13-1	Average percent errors across various manner classes	154
13-2	Distribution of errors and most frequent perceptual confusions	155

13-3	Percent correct comprehension scores for reading and listening groups	165
14-1	Sample MITalk session	176
C-1	Pre-aspiration parameter smoothing	189
C-2	Diphthong transition smoothing	194

List of tables

2-1	Abbreviation translations performed by FORMAT	19
3-1	Morph spelling change rules for vocalic suffixes	36
8-1	Klatt symbols used in the synthesis modules	84
9-1	Minimum and inherent durations in msec for each segment type	96
10-1	Relative peak levels of words according to their parts of speech	101
11-1	Parameter values for the synthesis of selected vowels	119
11-2	Parameter values for the synthesis of selected components of English consonants before front vowels	121
11-3	Variable control parameters specified in PHONET	122
12-1	List of control parameters for the software formant synthesizer	132
13-1	Characteristics of the passages used to measure comprehension	163
B-1	Klatt symbols for phonetic segments	179
B-2	Klatt symbols for nonsegmental units	180
C-1	Parameter targets for nonvocalic segments	186
C-2	Parameter targets for vocalic segments	187
C-3	Default values for duration of forward smoothing (Tcf)	188
C-4	Default values for Bper	188
C-5	Diphthong transition parameters	194
C-6	Duration of forward smoothing for obstruents (Tcobst)	196
C-7	Default plosive burst duration	197