

Contents

1	Introduction	1
1.1	This Book in a Nutshell	1
1.2	Contribution of this Research	3
2	Automatic Speech Recognition	5
2.1	Relevant Keywords from Probability Theory and Statistics	5
2.1.1	Discrete and Continuous Probability Distribution	6
2.1.2	A Hidden Markov Models	6
2.1.3	Estimating HMM Parameters	7
2.2	Phonemes	8
2.3	Prosody	9
2.3.1	Duration	9
2.3.2	Power	10
2.3.3	Pitch	10
2.3.4	Glottal Characteristics	10
2.4	Speech to Text	10
2.4.1	Pre-Processing	10
2.4.2	Feature Extraction	11
2.4.3	Decoding	12
2.4.4	Post-processing	14
2.5	Applying Speech Recognition	14
2.5.1	Evaluation Measures	14
2.5.2	Speech Dialog Systems	15
2.5.3	Focus of This Work	17
3	Properties of Non-native Speech	19
3.1	Accent	19
3.2	Phoneme Confusion	20
3.3	Non-phonemic Differences	21

4	Pronunciation Variation Modeling in the Literature	25
4.1	Types of Pronunciation Variations	26
4.2	Data-Driven and Knowledge-Based Methods	26
4.3	Representing Pronunciation Variations	27
4.3.1	Phoneme Confusion Rules	27
4.3.2	Pronunciation Networks	28
4.4	Other Related Topics	28
4.4.1	Computer Assisted Language Learning (CALL)	28
4.4.2	Multilingual Speech Recognition	29
4.5	Relevance for This Work	30
5	Non-native Speech Database	31
5.1	Existing Non-native Speech Databases	31
5.1.1	Speech Operated Travel Assistance	35
5.1.2	Military Communications	36
5.1.3	Presentation Transcription	37
5.1.4	Computer Assisted Language Learning	37
5.2	The ATR Non-native Speech Database	38
5.2.1	Speakers	38
5.2.2	Data Contents and Properties	39
5.2.3	Speaker Anxiety	41
5.2.4	Skill Annotation	42
5.2.5	Comparison to Existing Databases	44
6	Handling Non-native Speech	47
6.1	Rule-Based Lattice Processing	47
6.1.1	Rule Generation	48
6.1.2	Lattice Processing	49
6.1.3	Extension to Word Level	51
6.1.4	Applying Rules on the Pronunciation Dictionary	52
6.2	Multilingual Weighted Codebooks	52
6.3	Automatic Scoring of Non-native Pronunciation Skills	54
6.3.1	Sentence Level Pronunciation Scoring	54
6.3.2	Word Level Pronunciation Scoring	57
6.3.3	Scoring Experiments	59
6.4	Classification of Non-Verbal Utterances	62
6.4.1	Data	63
6.4.2	Features	64
6.4.3	Experimental Setup	65
6.4.4	Classes and Features	66
6.4.5	Clustering Methods	67
6.4.6	Results	69
6.5	Lessons Learned	69

7	Pronunciation HMMs	71
7.1	Why HMMs?	72
7.2	Generation and Initialization	72
7.3	Training	73
7.4	Application	74
7.5	Experimental Setup	75
7.5.1	Data and Software	75
7.5.2	Acoustic and Language Models	77
7.6	Phoneme Recognition	78
7.7	Rescoring with Pronunciation HMMs	80
7.7.1	Word HMM Initialization	80
7.7.2	Training of Word HMMs	80
7.7.3	Rescoring with Word HMMs	81
8	Outlook	85
	Appendix A: Hotel Reservation Dialog	89
	Appendix B: Confusion Matrices	91
	Appendix C: Speaker Information	95
	Appendix D: Human Evaluation	99
	Glossary	103
	References	105

Statistical Pronunciation Modeling for Non-Native
Speech Processing

Gruhn, R.E.; Minker, W.; Nakamura, S.

2011, X, 114 p., Hardcover

ISBN: 978-3-642-19585-3