

Contents

- 1 Introduction 1**
 - 1.1 State-of-the-Art and Previous Work 2
 - 1.2 Motivation and Contribution 4
 - 1.3 Overview 5
- 2 Fundamentals 7**
 - 2.1 Introduction 7
 - 2.2 Automatic Speech Recognition 7
 - 2.2.1 Speech Recognition Architecture 7
 - 2.2.2 Language Modeling 8
 - 2.2.3 Acoustic Modeling 9
 - 2.2.4 Training 10
 - 2.2.5 Evaluation 10
 - 2.3 The Arabic Language 11
 - 2.3.1 Modern Standard Arabic 15
 - 2.3.2 Classical Arabic 17
 - 2.3.3 Dialectal Arabic 18
 - 2.3.4 Morphological Complexity 19
 - 2.3.5 Diacritization 19
 - 2.3.6 Challenges for Dialectal Arabic Speech Recognition 22
 - 2.4 Summary 23
- 3 Speech Corpora 25**
 - 3.1 Introduction 25
 - 3.2 Modern Standard Arabic Corpus 25
 - 3.2.1 Corpus Specifications 26
 - 3.2.2 Post-Processing 27
 - 3.2.3 Corpus Key Facts 27
 - 3.3 Egyptian Colloquial Arabic Corpus 27
 - 3.3.1 Corpus Specifications 28
 - 3.3.2 Recording Setup 28
 - 3.3.3 Phonetic Transcription 29

3.3.4	Corpus Key Facts	30
3.4	Levantine Colloquial Arabic Corpus	30
3.4.1	Corpus Specifications	30
3.4.2	Post-Processing	31
3.5	Summary	32
4	Phonemic Acoustic Modeling	33
4.1	Introduction	33
4.2	Phonemic Baseline	33
4.2.1	Baseline System Description	33
4.2.2	Baseline Results	34
4.3	MSA as a Multilingual Acoustic Model	35
4.3.1	MSA and ECA Phonetic Comparison	35
4.3.2	Phoneme Sets Normalization	36
4.3.3	Recognition Results	39
4.4	Data Pooling Acoustic Modeling	39
4.5	Supervised Cross-Lingual Acoustic Modeling	41
4.5.1	Maximum Likelihood Linear Regression (MLLR)	42
4.5.2	Maximum A-Posteriori (MAP)	42
4.5.3	Adaptation Results	43
4.6	Effect of MSA Speech Amount on Supervised Adaptation	44
4.7	Unsupervised Adaptation	44
4.7.1	Adaptation Approach	46
4.7.2	Baseline and System Description	47
4.7.3	Adaptation Results	47
4.8	Effect of MSA Speech Data Size on Unsupervised Adaptation	47
4.9	Summary	49
5	Graphemic Acoustic Modeling	53
5.1	Introduction	53
5.2	Graphemic Modeling for MSA	54
5.2.1	Graphemic Lexicon	55
5.2.2	Assumptions and Procedure	55
5.2.3	System Description	56
5.2.4	Effect of Gaussians per State	57
5.2.5	Effect of Training Data Amount	58
5.3	Cross-Lingual Graphemic Modeling for ECA	60
5.3.1	ECA Baseline	60
5.3.2	MSA as a Multilingual Acoustic Model	60
5.3.3	Data Pooling Acoustic Modeling	61
5.3.4	Supervised Cross-Lingual Acoustic Modeling	61
5.3.5	Unsupervised Graphemic Acoustic Modeling	63
5.3.6	Spelling Variants	63
5.4	Cross-Lingual Graphemic Modeling for LCA	67
5.5	Discussion	68
5.6	Summary	69

6	Phonetic Transcription Using the Arabic Chat Alphabet	71
6.1	Introduction	71
6.2	The Arabic Chat Alphabet	72
6.3	Speech Corpus	73
6.3.1	ACA Transcription Layer	73
6.4	System Description and Baselines	73
6.4.1	Phoneme-Based Baseline	74
6.4.2	Grapheme-Based Baseline	75
6.5	ACA-Based Acoustic Modeling	76
6.5.1	Initial ACA-Based Model	76
6.5.2	Modeling ACA Common Transcription Errors	76
6.5.3	Modeling ACA Ambiguity	78
6.5.4	ACA-Based Final Results	78
6.6	ACAST: A New Phonetic Notation for Arabic	79
6.7	Summary	80
7	Conclusions and Future Directions	81
7.1	Summary	81
7.2	Contributions	83
7.2.1	Theoretical	83
7.2.2	Practical	83
7.2.3	Experimental	84
7.3	Future Work	84
Appendix A	Buckwalter Transliteration	87
Appendix B	Egyptian Colloquial Arabic Lexicon	89
Appendix C	Seinnheiser ME-3 Specifications	99
Appendix D	Buddy 6G USB Specifications	101
References		103
Index		109

Novel Techniques for Dialectal Arabic Speech
Recognition

Elmahdy, M.; Gruhn, R.; Minker, W.

2012, XXII, 110 p., Hardcover

ISBN: 978-1-4614-1905-1