

University of M'Hamed Bouguerra
Boumerdes
Algeria

Engineering Faculty
Department of Electrical and Electronic Engineering
(DGEE)

Continuous Speech Recognition
Using
Hidden Markov Models

Application to colloquial Algerian Arabic

Reggab Mourad

In the partial fulfilment of the degree of "Magister"
In "Ingénierie des Systèmes Electroniques"
October 2004

Assessors:

Pr. L. Reffoufi	Professor	U.M.B.B	President
Pr. D. Berkani	Professor	E.N.P	Examiner
Dr. L. Hamami	Maître de Conférence	E.N.P	Examiner
Dr. M. Noureddine	Maître de Conférence	U.M.B.B	Supervisor
Mr. H. Dahimene	Chargé de Coures	U.M.B.B	Co-Supervisor

Contents

Abstract	i
Acknowledgments	ii
Dedication	iii
List of Figures	iv
List of Tables	v
1-Introduction:	1
1.1 History of speech recognition.....	2
1.2 Users and areas of use.....	2
1.3 Recognition models.....	3
1.4 Variability of speech signals.....	3
1.4.1 Variability of Speech by Speakers.....	4
1.4.2 Variability from Environments.....	5
1.5 The Arabic Language.....	5
1.5.1 Linguistic Varieties.....	5
1.5.2 Problems for Automatic Speech Recognition in Algerian Arabic.....	6
1.5.3 Previous Work in Arabic Speech Recognition.....	6
1.6 Research Purposes.....	7
1.7 Organization of the Thesis.....	7
2-The Speech Signal:	8
2.1 Speech Production.....	9
2.1.1 Articulators.....	9
2.1.2 The Voicing Mechanism.....	10
2.1.3 Spectrograms and Formants.....	12
2.2 Speech Perception.....	13
2.2.1. Physiology of the Ear.....	13
2.3 Speech Representations.....	14
2.4 Phonemics and Phonetics.....	15
3- From Speech To Feature Vectors	16
3.0 Overview: what is a feature.....	16
3.1 Preprocessing.....	17
3.1.1 Noise Canceling.....	17
3.1.2 Preemphasis.....	17
3.1.3 Voice Activation Detection (VAD).....	18
3.2 Frame Blocking and Windowing.....	20
3.3 Feature Extraction.....	21
3.3.1 Linear Prediction.....	22
3.3.2 The Mel-Cepstrum.....	24
3.3.2.1 The Cepstrum.....	24
3.3.2.2 The Mel scale.....	26
3.3.2.3 The Mel-Cepstrum.....	28
3.3.3 Energy measures.....	31
3.3.4 Delta and Acceleration Coefficients.....	32
3.3.5 Summary.....	34
3.4 Post-processing.....	35
4-Hidden Markov Models	37
4.1 Discrete-Time Markov Chains.....	38
4.1.1 The Weather Forecast Markov Model.....	39
4.2 The Discrete-Time Hidden Markov Model.....	41

4.2.1 The Urn and Ball Model.....	41
4.2.2 Discrete Observation Densities.....	42
4.2.3 Continuous Observation Densities.....	42
4.2.4 Hidden Markov Model Types.....	44
4.2.5 Summary.....	45
4.2.5.1 Elements of a HMM.....	45
4.2.5.2 HMM Constraints for Speech Recognition Systems.....	46
4.3 The three Basic Problems for Hidden Markov Models.....	47
4.4 Solution to Problem 1 - Probability Evaluation.....	47
4.4.1 The Forward Algorithm.....	48
4.4.2 The Backward Algorithm.....	50
4.4.3 Scaling the Forward and Backward Variables.....	51
4.5 Solution to Problem 2 - “Optimal” State Sequence.....	54
4.5.1 The Viterbi Algorithm.....	55
4.5.2 The Alternative Viterbi Algorithm.....	56
4.6 Solution to Problem 3 - Parameter Estimation.....	57
4.6.1 Multiple Observation Sequences.....	62
4.6.2 Initial Estimates of HMM Parameters.....	64
4.6.3 Numerical Issues.....	64
5- The HTK Software Architecture	66
5.1 HTK Software Architecture.....	67
5.2 Generic Properties of a HTK Tool.....	68
5.3 Script Files.....	69
5.4 Configuration Files.....	69
5.5 Standard Options.....	70
5.6 Speech Input/Output.....	71
5.7 Speech Signal Processing.....	73
5.7.1 Linear Prediction Analysis.....	74
5.7.2 Cepstral Features.....	75
5.7.3 Perceptual Linear Prediction.....	75
5.7.4 Copying and coding using Hcopy.....	76
5.8 Transcriptions and Label Files.....	77
5.8.1 Label File Structure.....	77
5.8.2 HTK Label Files.....	77
5.8.3 Master Label Files.....	78
5.9 HMM Definition Files	79
5.10 Tied-Mixture Systems.....	81
5.11 HMM Parameter Estimation	81
5.11.1 Training Strategies.....	81
5.12 Networks and Dictionaries.....	83
5.13 Constructing a Dictionary.....	85
5.14 Decoding.....	86
6-System implementation and Results.....	88
6.1. Description.....	88
6.2. Compiling HTK under Windows.....	88
6.3. Data Preparation.....	89
6.3.1 Step 1 – The Task Grammar.....	89
6.3.2 Step 2 – The Dictionary.....	93
6.3.3 Step 3 – Recording Data.....	93
6.3.4 Step 4 – Creating the Transcription Files.....	94

6.3.5 Step 5 – Coding the Data.....	96
6.4. Creating Monophone HMMs.....	97
6.4.1 Step 6 – Creating Flat Start Monophones.....	97
6.4.2 Step 7 – Fixing the Silence Model.....	100
6.4.3 Step 8 - Realigning the Training Data.....	101
6.5. Creating Tied-State Triphones.....	101
6.5.1 Step 9 – Making Triphones from Monophones.....	102
6.5.2 Step 10 - Making Tied-State Triphones.....	104
6.6. Recognizer Evaluation.....	107
6.6.1 Step 11 – Recognising the Test Data.....	107
6.6.2 Running the Recogniser Live.....	108
6.7 Summary.....	110
Conclusions and Further work suggestions.....	113
Appendix A: Arabic phonemes	
Appendix B: The K-Means Algorithm	