

Contents

Declaration	ii
Certificate by the supervisor(s)	iii
Acknowledgement	v
Abstract	vi-vii
List of Acronyms	viii-xi
List of Figures	xii-xiii
List of Tables	xiv-xv
Notations	xvi
Chapter 1: Prologue	1-14
1.1 Motivation	1
1.2 Automatic speech recognition (ASR) system	2
1.2.1 Feature extraction phase	2
1.2.2 Acoustic feature classification phase	3
1.2.3 Pronunciation modeling phase	4
1.2.4 Language modeling phase	4
1.2.5 Decoding phase	5
1.3 Issues in automatic speech recognition	5
1.4 Research issues	6
1.5 Performance metrics and ASR tools	8
1.5.1 Kaldi ASR toolkit	9
1.5.2 Sphinx ASR toolkit	9
1.5.3 HTK ASR toolkit	9
1.6 Corpus design and associated challenges	10
1.6.1 Key aspects of a speech corpus	10
1.6.2 Challenges in designing a speech corpus	11
1.7 Genesis & contribution of thesis	12

Chapter 2: Literature Review	15-34
2.1 Introduction	15
2.2 Feature extraction phase	17
2.2.1 Filterbank optimization	18
2.2.2 Low variance feature estimation	19
2.3 Acoustic modelling classifier	20
2.3.1 Hidden markov model	21
2.3.1.1 Optimization of HMM approach	21
2.3.2 Deep neural network approach and its optimization	24
2.4 Analysis of Indian ASR corpus	27
2.5 State of the art research in Punjabi speech recognition	28
2.6 Comparison of different Indian ASR engine	31
2.7 Research gap	32
2.8 Objectives of the research	33
2.9 Summary	33
Chapter 3: Refinement of HMM based classifier with optimized features	35-61
3.1 Theoretical background	35
3.1.1 Hidden markov model	35
3.1.1.1 Key components of HMM	36
3.1.1.2 Deficiency in baseline HMM technique	38
3.1.2 Genetic algorithm	38
3.1.3 Differential evolution	39
3.2 An overview of system architecture strategy	41
3.2.1 Front end module	43
3.3.2 Back end module	44
3.3.2.1 Acoustic modelling	44
3.3.2.2 GA based feature refinement	45

3.2.2.3 DE based feature refinement	46
3.2.2.4 HMM based feature classification	47
3.2.2.5 Lexicon and language modeling	48
3.2.2.6 Decoding module	49
3.3 Punjabi isolated dataset preparation	51
3.4 Results and discussion	53
3.4.1 Result analysis obtained with speakers variability	54
3.4.2 Result analysis obtained with various modeling classifier techniques	55
3.4.3 Result analysis obtained with two hybrid classifier and one baseline HMM approach at various SNR level	56
3.4.4 Result analysis obtained with different environment conditions	56
3.4.5 Result analysis obtained with varying speech utterances	58
3.5 Discussions and performance comparison of various hybrid classifiers	59
3.6 Summary	61
Chapter 4: Heterogeneous feature extraction and optimized HMM classifiers	62-79
4.1 Introduction	62
4.2 Heterogeneous feature vector approaches	64
4.2.1 Speech feature vector generation using MFCC	64
4.2.2 Speech feature vector generation using PLP	65
4.2.3 Speech feature vector generation using RASTA-PLP	66
4.3 Hybrid HMM classifier for feature refinement	67
4.4 Combining of heterogeneous feature extraction approaches with hybrid modeling classifiers	68
4.5 Dataset selection	73
4.6 Proposed system simulation	74

4.6.1 Accuracy rate for heterogeneous speech feature vector generation combination	74
4.6.2 Experiments with speaker variability and training complexity	77
4.7 Summary	78
Chapter 5: Improved filter bank on multitaper framework	80-97
5.1 Introduction	80
5.2 Robust speech analysis approaches	81
5.2.1 Gammatone frequency cepstral coefficients	81
5.2.2 Mel frequency cepstral coefficient	82
5.2.3 Multitaper spectrum estimation	84
5.2.4 Selection of the tapers	86
5.3 Principal component analysis based filter bank optimization	86
5.4 Proposed integrated Punjabi-ASR system with PCA optimized filter on multitaper framework	87
5.5 Corpus setup	91
5.6 Experimental analysis	92
5.7 Summary	97
Chapter 6: Continuous Speech Recognition using DNN-HMM Classifier	98-111
6.1 Introduction	98
6.2 An overview of acoustic modeling classifier	99
6.2.1 GMM-HMM model	99
6.2.2 Deep neural network model	100
6.2.3 DNN-HMM model	101
6.3 An overview of heterogeneous system architecture strategy	102
6.3.1 Baseline system implementation using GMM-HMM and DNN-HMM based acoustic modelling	103
6.4 Punjabi dataset preparation	104

6.5 Results and discussion	105
6.5.1 Performance evaluation on different modeling units	106
6.5.2 Performance evaluation on varying range of feature dimension with LDA approach	107
6.5.3 Performance evaluation with varying number of hidden layer	108
6.5.4 Performance evaluation with varying Gaussian mixture	109
6.6 Summary	110
 Chapter 7: Summary of Contribution and Future Directions	 112-115
7.1 Significant contribution	112
7.2 Future direction	114
 Bibliography	 116-132
List of Publications	133