

# Contents

<b>List of Figures</b>	<b>xii</b>
<b>List of Tables</b>	<b>xv</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Speech Recognition . . . . .	1
1.2 Speech Production Process . . . . .	2
1.3 Human and Machine Speech Perception . . . . .	3
1.3.1 Human Speech Perception . . . . .	3
1.3.2 Machine Speech Perception . . . . .	4
1.4 States of Speech Signal . . . . .	4
1.5 Types of Speech Recognition . . . . .	5
1.6 Applications of Speech Recognition . . . . .	6
1.7 Speech Recognition of Indian Languages . . . . .	7
1.8 Literature Survey . . . . .	8
1.8.1 Indian Languages: . . . . .	8
1.8.2 Gujarati Language: . . . . .	9
1.9 Objectives . . . . .	10
1.10 Conclusion . . . . .	11
<b>2 Preliminaries and Methodologies</b>	<b>13</b>
2.1 Feature Extraction using MFCC . . . . .	13
2.2 Feature Extraction using MFDWC . . . . .	18
2.2.1 Wavelets . . . . .	18
2.3 Short-Term Autocorrelation . . . . .	22
2.4 Normalisation . . . . .	23
2.5 Dynamic Time Warping . . . . .	25
2.6 Multilayered Perceptrons . . . . .	28
2.7 Radial Basis Function Networks . . . . .	34
2.8 Hidden Markov Model . . . . .	37
2.8.1 Evaluation Problem and Forward/Backward Algorithm . . . . .	39

2.8.2	Determination of Optimal Hidden State using Viterbi Algorithm . . . . .	45
2.8.3	Determination of Optimal Model Parameters using Baum-Welch Algorithm . . . . .	52
2.9	Performance Measures . . . . .	58
2.9.1	Confusion Matrix . . . . .	58
2.9.2	Recall . . . . .	60
2.9.3	False Positive Rate . . . . .	60
2.9.4	Specificity . . . . .	61
2.9.5	Precision . . . . .	61
2.9.6	F-Measure . . . . .	61
2.9.7	Accuracy . . . . .	61
2.10	Conclusion . . . . .	62
<b>3</b>	<b>Speech Recognition using MFCC Features</b>	<b>63</b>
3.1	Recording . . . . .	63
3.2	Noise Removal . . . . .	65
3.3	Extracting Words . . . . .	66
3.4	Feature Extraction . . . . .	67
3.5	Making the Lengths of Feature vector equal . . . . .	69
3.6	Model 1: Dynamic Time Warping . . . . .	70
3.7	Model 2: Multilayered Perceptrons . . . . .	71
3.8	Model 3: Radial Basis Function Network . . . . .	73
3.9	Accuracy Comparison for Isolated Words based on MFCC . . . . .	75
3.10	Conclusion . . . . .	75
<b>4</b>	<b>Speech Recognition using MFDWC Features</b>	<b>76</b>
4.1	Feature Extraction using MFDWC . . . . .	77
4.2	Making the Lengths of Feature Vector Equal . . . . .	82
4.3	Models for Recognition of Isolated Words . . . . .	82
4.3.1	Pre-processing, Word Extraction and Feature Extraction . . . . .	83
4.3.2	Model 1: Dynamic Time Warping . . . . .	83
4.3.3	Model 2: Multilayered Perceptron . . . . .	85
4.3.4	Model 3: Radial Basis Function Network . . . . .	89
4.3.5	Model 4: Hidden Markov Models . . . . .	92
4.3.6	Models for Augmented Dataset . . . . .	94
4.3.7	Accuracy Comparison for Isolated Words based on MFDWC . . . . .	97
4.4	Models for Recognition of Continuous Sentences . . . . .	99
4.4.1	Word Extraction using Short-Term Autocorrelation . . . . .	99
4.4.2	Model 1: Multilayered Perceptron . . . . .	100

4.4.3	Model 2: Hidden Markov Models . . . . .	105
4.4.4	Accuracy Comparison for Continuous Sentences based on MFDWC	108
4.5	Conclusion . . . . .	109
<b>5</b>	<b>Ensemble Learning</b>	<b>110</b>
5.1	Ensemble Learning . . . . .	110
5.2	Approach 1: ELMP . . . . .	111
5.3	Approach 2: ELHMM . . . . .	113
5.4	Conclusion . . . . .	115
<b>6</b>	<b>Graphical User Interface</b>	<b>117</b>
<b>7</b>	<b>Conclusions</b>	<b>120</b>
7.1	Future scope: . . . . .	121
	<b>References</b>	<b>122</b>