List of Figures

1.1 Speech production process [1] 2
1.2 Schematic diagram of nasal tract and vocal tract [1] 3
1.3 Expanded view of middle and inner ear [2] 4
1.4 Physiological model of the human ear [3] 5
1.5 Block diagram of a summary of the approaches to our research work . . . 12

2.1 Hamming window for different values of β 15
2.2 Illustration of Framing and Windowing [4] 16
2.3 Plot of mel-scale converted from frequencies 17
2.4 The plot of Daubechies-6 wavelet function 19
2.5 Discrete wavelet transform: wavelet coefficients at different levels 20
2.6 Steps of determining feature vectors. 21
2.7 Plot of a sample signal x 22
2.8 Plot of autocorrelations for various values of n 23
2.9 Alignment of two sequences of unequal lengths 25
2.10 Steps of determining a Dynamic time warping distance between vectors X and Y of unequal lengths: (a) Arrangement of two vectors of unequal lengths as a matrix. (b) Absolute difference between all the elements of X and Y. (c) Determining the elements of first row. (d) Determining the elements of first column. (e) Determining the elements of second row. (f) Determining the elements of the third and fourth column. The element $D(x_5, y_4) = 5$ is the dynamic time warping distance between these two vectors. 27
2.11 Optimal alignment between X and Y to determine minimum distance . . 28
2.12 Multilayered perceptron network architecture 31
2.13 Radial basis function network architecture 35
2.14 Calculations of Forward Algorithm 41
2.15 Calculations of Backward Algorithm 44
2.16 Summary of values of forward and backward variables 46
2.17 Summary of calculated values of $\gamma_t(i)$ 47
2.18 Summary of calculated values of variables $\delta_t(i)$ and $\psi_t(i)$ 51

2.19 Determination of $\xi_t(i,j)$ for Baum-Welch algorithm 53
2.20 Confusion matrix of 4-class classification problem 59
2.21 Identifying True positive, True Negative, False Positive and False Negative in the confusion matrix of 4-class classification problem 60

3.1 A recording of digits one to ten spoken by one speaker in the Gujarati language 65
3.2 The plot of first word extracted from the recording shown in the Figure 3.1 66
3.3 The plot of first word extracted from the recording of another speaker . . 67
3.4 DTW distances between two speakers using MFCC 70
3.5 Architecture of the multilayered perceptron used in the model 72
3.6 Output of the recognised digit 2 73
3.7 Architecture of the radial basis function network used in the model . . . 74

4.1 Original speech signal consisting of one word 78
4.2 Speech signal after applying pre-emphasising with α=0.97 78
4.3 Thirteenth frame of the speech signal 79
4.4 Hamming window of the thirteenth frame of the speech signal 79
4.5 Fourier transform and Power spectrum of the thirteenth frame of the speech signal 80
4.6 Mel-frequency discrete wavelet coefficients of thirteenth frame 80
4.7 Fourier transform and Power spectrum of entire speech signal 81
4.8 Mel-frequency discrete wavelet coefficients of entire speech signal 81
4.9 The plot of Coiflet-1 wavelet function 84
4.10 A multilayered network architecture used in Model 2 86
4.11 The plot of Daubechies-6 wavelet 87
4.12 Decay in training loss with increase in iterations for Model 2 88
4.13 Confusion matrix of 10 test samples of feature vectors for Model 2 89
4.14 Confusion matrix with leave-one-out procedure for Model 2 90
4.15 A radial basis function network architecture used in our work for Model 3 91
4.16 Decay in training loss with increase in iterations for Model 3 92
4.17 Confusion matrix for the classification using HMM 95
4.18 Confusion matrix for 200 test patterns for the augmented dataset 96
4.19 Confusion matrix using HMM for augmented dataset 98
4.20 A plot of a speech signal having 17 words 100
4.21 Short-term auto-correlation (normalised) of the signal shown in the Figure 4.20. The spikes shows the voiced part corresponding to 17 sentence. . . 101
4.22 Flowchart of process of ASR for continuous sentences 101
4.23 5 sentences having 17 words used for the models of continuous speech recognition using multilayered perceptrons 102

4.24 Plot of recording of word extracted using short-term autocorrelation . . 102
4.25 Multilayered perceptron architecture for the speech recognition of continuous words 103
4.26 5 sentences having 32 words used for the models of continuous speech recognition using hidden Markov models 106
4.27 Confusion matrix for 16 words in the model of continuous speech recognition using HMM 107

5.1 Block diagram to understand Bagging ensemble learning method 111
5.2 5 sentences having 32 words used for the ensemble learning models 112
5.3 Sentences chosen for the validation of models 112
5.4 Validation of 5 ANN models on sentence-1 113
5.5 Validation of 5 ANN models on sentence-2 114
5.6 Confusion matrix of validation of sentence-1 on first hidden Markov model 114
5.7 Confusion matrix of validation of sentence-2 on first hidden Markov model 115
5.8 Validation of 9 HMM models on sentence-1 115
5.9 Validation of 9 HMM models on sentence-2 116

6.1 Speech recogniser interface 118
6.2 Updated speech recogniser interface 119