Contents

1 Introduction **5**
- 1.1 Summary . . . 6
- 1.2 Outline . . . 6

2 Background Study **8**
- 2.1 Assamese Language . . . 8
- 2.2 Definition . . . 9
  - 2.2.1 Mel Frequency Cepstrum . . . 9
  - 2.2.2 Sound . . . 10
  - 2.2.3 Transcription . . . 10
  - 2.2.4 Pitch . . . 10
  - 2.2.5 Prosody . . . 11
  - 2.2.6 Syllable . . . 11
  - 2.2.7 Intensity . . . 11
  - 2.2.8 Waveform . . . 11
  - 2.2.9 Phonemes . . . 11
  - 2.2.10 Phones . . . 12
  - 2.2.11 Articulation . . . 12
  - 2.2.12 ASCII . . . 12
- 2.3 IPA . . . 12
  - 2.3.1 Consonant Chart . . . 12
  - 2.3.2 Hidden Markov Model Tool Kit . . . 16

3 Data Collection and Tools Used **17**
- 3.1 Data Collection . . . 17
  - 3.1.1 Mode of Recording . . . 17
  - 3.1.2 Environment of Recording . . . 18
  - 3.1.3 Device Used . . . 18
  - 3.1.4 Encoding . . . 18
  - 3.1.5 Male Female Ratio . . . 18
  - 3.1.6 Age Group . . . 18
- 3.2 Tools Used . . . 18
  - 3.2.1 Wavesurfer . . . 18

3.2.2 IPA Typing Tool 19

4 Syllabification Rules **20**

5 IPA Transcription **22**
5.1 IPA for Assamese 22
5.1.1 Consonants and Vowels 22
5.2 Frequency 24
5.2.1 IPA Symbols 24
5.2.2 Assamese Words 26
5.3 Issues and Interesting Observation 26
5.3.1 Consonants 27
5.3.2 Vowels 28
5.3.3 Clustering 29

6 Automatic Transcription **30**
6.1 Manual ASCII Transcription 30
6.2 Configuration File 31
6.3 Dictionary 32
6.4 Master Label File 32
6.5 MFC 33
6.6 Training 34
6.7 Testing 34

7 Results **35**
7.1 Overall Results 36
7.2 Comparison 36

8 Conclusion and Future Work **40**

List of Figures

2.1 Assamese Language Tree 9
2.2 IPA 2005 13

6.1 Automatic Transcription Process 30
6.2 HCopy Operation 33

7.1 Fixing Training Files 37
7.2 Fixing Training Files 38

List of Tables

5.1 IPA Vs Assamese 23
5.2 Number of IPA Phonemes Used 24
5.3 IPA SYMBOLS FOR ASSAMESE 25
5.4 Most Frequent Assamese Word 27
5.5 Less Frequent Assamese Word 28

6.1 Assamese Vs ASCII 31

7.1 Accuracy List 36
7.2 IPA TRANS 39