Contents

List of Figures xvi

List of Tables xix

1 Introduction 1

1.1 Background ................................................. 1
1.2 Motivation ................................................. 3
1.3 Outline of the Thesis Organisation .......................... 4

2 Review of Previous Works 11

2.1 Introduction .................................................. 11  
2.2 Review on Speech Processing based on Allophonic Variations and Phone Duration Modeling .............................................. 13  
2.3 Review on Grapheme-to-Phoneme transcription methods ........ 15  
2.4 Review on Keyword Spotting Techniques ...................... 16  
2.4.1 Acoustic Keyword Spotting Methods ..................... 19  
2.4.2 LVCSR based Keyword Spotting Methods ................ 21  
2.5 Review on audio classification using Multiple Instance Learning .... 23  
2.6 Review on Speaker Spotting methods with emphasis on nonlinear properties of the speech signal .......................... 25  
2.7 A special review on Indian language speech recognition ........ 26  
2.8 Conclusion .................................................... 41
3 A Study on Malayalam Phonology based on Durational and Spectral Characteristics of Allophonic Variations in Vowel Phonemes 43

3.1 Introduction .................................................. 43

3.2 Malayalam Phoneme Set and Allophones ......................... 45

3.2.1 Malayalam Vowel Phones ................................ 45

3.2.2 Malayalam Consonant Phones .............................. 47

3.3 Founding an Exhaustive Rule Set for Malayalam Allophone Formation 48

3.3.1 Rule set for the formation of Malayalam Vowel Allophones based on the Position and Neighbourhood Information ............. 49

3.3.2 Rule Set for the formation of Malayalam Consonant Allophones .... 57

3.3.3 TEMU Malayalam Phonetic Dataset ........................ 61

3.4 Durational Properties of Malayalam Vowel Allophones ............ 62

3.5 Formant Frequency Analysis of Malayalam Vowel Allophones ....... 68

3.6 Clustering of Vowel Allophones using K-means clustering .......... 77

3.7 Conclusion ..................................................... 80

4 A Comprehensive Grapheme-to-Phoneme Transcription Algorithm for Malayalam with Application to Speech Processing 82

4.1 Introduction ..................................................... 82

4.2 Malayalam Orthography and Categorisation of Grapheme Units .... 84

4.2.1 Malayalam Vowels ........................................... 84

4.2.2 Malayalam Diphthongs ....................................... 86

4.2.3 Anusvaram and Chandrakkala ............................... 86

4.2.4 Malayalam Consonant Classes ............................... 87

4.2.5 Formation of Malayalam Compound Letters .................. 88

4.2.6 Chillukkal ................................................. 90

4.3 A Complete Rule based Automatic G2P Transcriptor for Malayalam .. 91

4.3.1 Pre-Processing Stages ...................................... 91

4.3.2 Implementation of the Proposed Malayalam G2P Transcription Algorithm ................................................. 92
4.3.3 Malayalam Phoneme to IPA Mapping .................................. 99
4.4 Statistical Analysis of Malayalam Phonemes ............................... 100
  4.4.1 Malayalam Word and News Sentence Text Corpora .................. 101
  4.4.2 Phoneme Statistical Analysis Results ................................. 102
4.5 Conclusion .............................................................................. 111

5 Implementation of Keyword Spotting in Malayalam Speech using
Continuous Hidden Markov Modelling ........................................ 113
  5.1 Introduction ........................................................................... 113
  5.2 Data Preparation ................................................................. 116
    5.2.1 Data Preparation for KWS System Training ....................... 116
    5.2.2 Data Preparation for KWS System Evaluation .................... 118
  5.3 Knowledge Base Generation for the Implementation of HMM Decoder . 120
    5.3.1 Knowledge Base Preparation Tool (KBPT-M) ....................... 121
  5.4 Architecture of the HMM based Automatic Speech Recognition (ASR)
    System ................................................................................. 126
    5.4.1 MFCC Feature Extraction Process ................................. 127
    5.4.2 Development of Acoustic Models for HMM ..................... 129
    5.4.3 Generation of n-gram Language Models ......................... 136
    5.4.4 HMM Decoding and Word Lattice Generation ................ 137
  5.5 Proposed Keyword Spotting System Architecture for Malayalam .... 140
    5.5.1 ASR based Keyword Spotting Technique ....................... 140
    5.5.2 Filler Model based Acoustic Approach for Keyword Spotting(FMA-KWS) ......................................................... 142
  5.6 Experimental Results ............................................................ 143
  5.7 Conclusion .............................................................................. 147
## 6 Automatic Content based Classification of Speech Audio using Multiple Instance Learning Approach

6.1 Introduction ......................................................... 148

6.2 Feature Extraction from News Audios for Classification ............... 150
  6.2.1 Mel Frequency Cepstral Coefficient (MFCC) Feature Extraction 150
  6.2.2 Perceptual Linear Prediction (PLP) Feature Extraction ........... 151

6.3 Content based Audio Classification using MIL ........................ 152
  6.3.1 MIL for News Audio Classification .......................... 152
  6.3.2 mi-Graph based Classification Method ....................... 154
  6.3.3 mi-SVM based Classification Method .......................... 155

6.4 Simulation Experiments and Results ............................... 156

6.5 Conclusion ......................................................... 159

## 7 Effective Speaker Spotting based on Nonlinear Properties of Vocal Tract

7.1 Introduction ......................................................... 161

7.2 Segmentation of Vowel Units from Continues Speech .................. 163

7.3 Nonlinear Dynamics of Vocal Tract ................................ 164
  7.3.1 Nonlinear Features used in Speaker Modelling ................. 166
  7.3.2 Eigen Value of Reconstructed Phase Space .................... 170
  7.3.3 Speaker Modelling based on Chaotic Properties of the Power Spectrum ......................................................... 172

7.4 Speaker Spotting Experiments based on Nonlinear Features and ANN . 179

7.5 Conclusion ......................................................... 181

## 8 Conclusions and Future Research Directions

8.1 Conclusion ......................................................... 182

8.2 Contributions ..................................................... 185

8.3 Future Direction .................................................. 187

References ............................................................ 189
<table>
<thead>
<tr>
<th>Contents</th>
<th>xv</th>
</tr>
</thead>
<tbody>
<tr>
<td>Appendix A</td>
<td>212</td>
</tr>
<tr>
<td>Appendix B</td>
<td>214</td>
</tr>
<tr>
<td>List of Publications of the Author</td>
<td>218</td>
</tr>
</tbody>
</table>