

## **LIST OF FIGURE**

1. Figure 1.1: Multifactor Authentication
2. Figure 1.2 [A]: User identification
3. Figure 1.2 [B]: User verification/authentication
4. Figure 1.3: Speaker authentication approaches
5. Figure 1.4: An example of verbal information authentication process
6. Figure 1.5: The system by combining VIV with speaker authentication
7. Figure 1.6: Representation of the training phase of a speaker authentication system
8. Figure 1.7: Modular representation of the testing phase of a speaker authentication applied in the present work
9. Figure 2.1: Speech recognition classification
10. Figure 2.2: Transmission Theory of Information for recognition system with Statistical modeling framework of speech
11. Figure 2.3: Block diagram of acoustic phonetic speech recognition system
12. Figure 2.4: Block diagram of pattern-recognition speech recognition system
13. Figure 2.5: Hidden Markov model state diagram
14. Figure 2.6: Vector quantization example
15. Figure 2.7: Identification, verification or authentication- 2 basic goals of SRS
16. Figure 2.8: A generic speaker verification or authentication system
17. Figure 2.9: Authentication approach used in this research work
18. Figure 2.10 [A]: FRR and FAR as a function of a threshold  $\theta$ . The intersection point determines the EER
19. Figure 2.10 [B]: Illustration of Equal Error Rate for a given set of true and impostor scores
20. Figure 2.11: Examples of a ROC curve (a) and the corresponding DET curve (b)
21. Figure 3.1: Different Stages of Speech Process
22. Figure 3.2: Human vocal system
23. Figure 3.3: Mechanism of human speech production system and characteristics identification
24. Figure 3.4: Snap shot of BODO vowel /ʒ/ /o/ spectrogram

25. Figure 4.1: Block diagram of recording setup
26. Figure 4.2: Database tree
27. Figure 4.3: Labeling of speech signal
28. Figure 4.4: Bodo word बं-बं थां bong[AJ] bong[AJ] thAng[VT] - female
29. Figure 4.5: Bodo word बं-बं थां uttered by male
30. Figure 4.6: Database structure
31. Figure 5.1: Hierarchy of Bodos
32. Fig 5.2: Bodoland Territory
33. Figure 6.1: Physical interpretation view points of features
34. Figure 6.2: Speech signal pre-processing diagram
35. Figure 6.3: Speech signal digitization procedures
36. Figure 6.4: Speech signal /बं-बं थां/ with silence and same signal after silence removal
37. Figure 6.5[A]: Spectrogram of Bodo word /फे-फे फु-फु गाब /
38. Figure 6.5[B]: pitch, intensity, formants and, pulses of Bodo vowel औ /a<sup>w</sup>/
39. Figure 6.6 [A]: Bodo vowel इ /i/ before and after normalization
40. Figure 6.6 [B]: Bodo sentence /फे-फे फु-फु गाब / before and after normalization
41. Figure 6.7: Speech signal for /बं-बं थां/ before and after pre-emphasis and filtration
42. Figure 6.8[A]: Speech signal for आ /a/ before and after pre-emphasis
43. Figure 6.8 [B]: Bodo consonant sound /ra/: Original wave, after pre-emphasis and then after Filtration in time domain
44. Figure 6.8 [C]: Bodo sentence /फे-फे फु-फु गाब / before and after pre-emphasis
45. Figure 6.9: Framing
46. Figure 6.10: short time analysis of speech signal
47. Figure 6.11: Pitch and formants of Bodo words /जा-जा जा/ and / फे-फे फु-फु गाब

48. *Figure 6.12[A]: Screen shots of Fundamental frequency estimation of six Bodo vowels uttered by male speakers in time domain*
49. *Figure 6.12[B]: Screen shots of Fundamental frequency estimation of six Bodo vowels uttered by female speakers in time domain*
50. Figure 6.13[A]: Feature vectors of Bodo vowel pitch data for male and female speakers
51. Figure 6.13[B]: Comparison chart of Bodo vowel pitch data for male and female speakers
52. Figure 6.14[A]: Bodo cv (consonant/vowel) pattern word pitch estimation for male informant
53. Figure 6.14[B]: Bodo cv pattern word pitch feature vector for male informant
54. *Figure 6.15 [A]: Fundamental frequency of Bodo cv pattern word from female speakers in time domain*
55. Figure 6.15 [B]: Bodo cv pattern word pitch feature vector for female informant
56. Figure 6.16 [A]: Formant frequency feature vector for Bodo vowel औ /a<sup>w</sup>/ (female)
57. Figure 6.16 [B]: Formant frequency feature vector for Bodo vowel औ /a<sup>w</sup>/ (male)
58. Figure 6.17: figure [A] through [E] along with the estimated values of first three formants for the Bodo Vowel Consonant pattern words of female informants
59. Figure 6.18[a] -[c]: Comparing the formants f1, f2, f3 for male and female through vc pattern words
60. Figure 6.19: Block diagram of MFCC
61. Figure 6.20: Flow diagram of MFCC computation
62. Figure 6.21: Graphs detailing outputs at each stage of the processing of the MFCC feature extraction
63. Figure 6.22: Hz-Mel Scale
64. Figure 6.23: MFCC calculation taking weight in y axis and frequency in x axis
65. Figure 6.24: Snapshot of MFCC feature vector for frame 1 through frame 8
66. Figure 6.25[A]: MFCC graph generated for female speaker औ /a<sup>w</sup>/

67. Figure 6.25[B]: MFCC graph generated for male speaker औ /a<sup>w</sup>/
68. Figure 6.26: feature vector for male speaker उ /u/ frame 8 through 12
69. Figure 6.27: LPC estimation block diagram as envisioned by the author
70. Figure 6.28: The feature vector, 20 coefficients of frames 16<sup>th</sup> and 17<sup>th</sup> of Bodo vowel अ /o/ -female
71. Figure 6.29: LPCC generated from the 8<sup>th</sup>/12<sup>th</sup>/16<sup>th</sup> and 24<sup>th</sup> frame of BODO vowel “अ” /o/ -uttered by female informant
72. Figure 6.30: LPCC generated from the 8<sup>th</sup>/12<sup>th</sup>/16<sup>th</sup> and 24<sup>th</sup> frame of BODO vowel “औ” /a<sup>w</sup>/for a female speaker and feature vector for frame 8<sup>th</sup> and 9<sup>th</sup>
73. Figure 6.31: Cepstral Coefficients extracted from the 8<sup>th</sup>, 12<sup>th</sup>, 16<sup>th</sup> and 24<sup>th</sup> frame of BODO vowel /इ/ /i/ for a female speaker and output feature vector of frame 24<sup>th</sup> and 25<sup>th</sup>
74. Figure 6.32: 16<sup>th</sup> frame LPC of Bodo vowels आ /a/, ए /e/, and उ /u/ for female speaker
75. Figure 6.33: 16<sup>th</sup> frame LPC of the Bodo cvc pattern word /सान/ / san/ for [a]male speaker and [b] female speaker
76. Figure 6.34: 16<sup>th</sup> frame LPC of the Bodo vc pattern word / अन/ / san/ for [a] female speaker and [b] male speaker
77. Figure 6.35: 16<sup>th</sup> frame LPC of the Bodo vc pattern word / खाओ/ / khao/ for [a] female speaker and [b] male speaker
78. Figure 6.36: Zero crossings of a speech signal
79. Figure 6.37[a]: ZCR feature vector estimation screenshot for Bodo अ /o/, आ /a/, इ /i/, ए /e/, उ /u/ and औ /a<sup>w</sup>/ vowels uttered by female speaker
80. Figure 6.37[b]: ZCR feature vector estimation screenshot for Bodo अ /o/, आ /a/, इ /i/, ए /e/, उ /u/ and औ /a<sup>w</sup>/ vowels uttered by male speaker

81. Figure 6.38: Comparison graph of ZCR for male and female with reference to Bodo vowels
82. Figure 6.39: ZCR feature vector screenshot-Bodo consonant /ja/ of female and female speaker
83. Figure 6.40: ZCR feature vector screenshot for Bodo consonant /ra/ uttered by female and male speaker
84. Figure 6.41: tabulated values and the graphical representation of ZCR for male and female informants with respect to selected Bodo consonants
85. Figure 6.42: ZCR for male and male informants for Bodo sentence /बेरेया अं-अं गाबो/ of female and male informants
86. Figure 6.43[a]: ZCR comparison of three different utterances of the same speaker (first 60 frames) uttered by female informants
87. Figure 6.43[b]: ZCR comparison of three utterances of three different speakers (first 60 frames) for the same utterance
88. Figure 6.44: ZCR of a BODO speech wave /बं-बं थां/
89. Figure 6.45: Comparison of voice and unvoiced segments of the Bodo word /बं-बं थां/(female)
90. Figure 6.46: STE feature vector
91. Figure 6.47[a]: STE of Boro vowel आ /a/ (female) and the generated feature vector
92. Figure 6.47[b]: STE of Boro vowel औ /a<sup>w</sup>/ (female) and the generated feature vector
93. Figure 6.47[c]: STE graph and feature vector of BODO vowel phonemes /इ//i/ of female speaker
94. Figure 6.47[d]: STE graph and feature vector of BODO vowel phonemes अ /o/-male
95. Figure 6.48: STE of BODO sentence /बेरेया अं-अं गाबो/ of female informant

96. Figure 6.49[a]: STE comparison of BODO consonant /ख//k<sup>h</sup>/ uttered by male (green stars) and female (black stars) informants
97. Figure 6.49[b]: STE comparison of BODO sentence / बेवहाय दाब-दाब जायगानिफ्राय फैनाय मानसिफ्रा दाब-दाब जायगायाव थायो/ uttered by male (red stars) and female (green stars) informants
98. Figure 6.50: End-point detection in a BODO speech Sentence
99. Figure 6.51: Voice and unvoiced segments energy comparison of the Bodo word /बं-बंथां/ uttered by female speaker
100. Figure 6.52[a]: K means clustering algorithm
101. Figure 6.52[b]: Snapshot of 2-Clusters using K-means
102. Figure 6.52[c]: Clustering N-phoneme word. (KMC: K-means clustering algorithm)
103. Figure 6.53: Block diagram of the Authentication system
104. Figure 7.1: Histogram of the comparative results for 5 utterances
105. Figure 7.2[a][b][c]: Comparison of male and female vc pattern word first three formants.
106. Figure 7.3: Extracted MFCC values (N \* 12 order matrix) of 12 MFCCs for every frame
107. Figure 7.4[A]: MFCC (12<sup>th</sup>order) coefficients of Bodo speech /ja/ of 3 different informants
108. Figure 7.4 [B]: MFCC (12<sup>th</sup>order) coefficients of Bodo speech /ja/ of 3 different informants
109. Figure 7.5: MFCC graph of five speech signals used in the experiment. औ /a<sup>w</sup>/for a female speaker and feature vector for frame 8<sup>th</sup> and 9<sup>th</sup>
110. Figure 7.6: Cepstral Coefficients extracted from the 8<sup>th</sup>, 12<sup>th</sup>, 16<sup>th</sup> and 24<sup>th</sup> frame of BODO vowel
111. Figure 7.7: STE feature vector
112. Figure 7B.1: Taxonomy of 8 neural net classifiers. Most similar pattern classification algorithms to neural network algorithms are shown at the bottom layer
113. Figure 7B.2: Supervised learning model

114. Figure 7B.3: Unsupervised learning model
115. Figure 7B.4[A]: A feed forward neural network
116. Figure 7B.4 [B]: Single layer feed-forward network
117. Figure 7B.5: Feed-forward MLP
118. Figure 7B.6: Recurrent ANN model
119. Figure 7B.7: KMC screenshot
120. Figure 7B.8[a]: Feed-Forward Topology of the speech recognition/ authentication ANN
121. Figure 7B.8[b]: Feed-Forward Topology of the speaker recognition/ authentication ANN
122. Figure 7 B.9: An input feature vector
123. Figure 7 B.10: Comparison of conventional classifiers and Neural Net classifiers. a) In traditional classifier input/output are passed serially and processing is serial. b) In NN input/output are passed in parallel and processing is also parallel
124. Figure 7 B.11: A single-layer perceptron forming two decision regions for classes A and B applying 80 trials
125. Figure 7 B.12 [a]: Creating NEURAL NETWORK with input layer, one hidden layer and the output layer feeding feature vectors pitch, formants, MFCC, LPC, ZCR and STE
126. Figure 7B.12 [b]: ANN training performance plot: Speaker dependent Speech authentication
127. Figure 7B.12 [c]: ANN training state: Speaker dependent Speech authentication
128. Figure 7B.12 [d]: ANN regression plot: Speaker dependent Speech authentication
129. Figure 7B.12 [e]: Confusion matrix generated for 25 sample utterances keeping the hidden neuron as 80 (5 speakers X 5 repetitions each)
130. Figure 7B.13: Test results of classification for 10 samples (1 utterance each of 10 speakers from the same database used for training the ANN)
131. Figure 7B.14 [a]: Graphical comparison of speech recognition performance
132. Figure 7B.14 [b]: Graphical comparison of speech recognition performance
133. Figure 7B.15 [a]: Neural Network speech dependent Speaker recognition/ Authentication

134. Figure 7B.15 [b]: Neural Network training state plot (tested with the same word)
135. Figure 7B.15 [c]: Neural Network training performance plot (tested with the same word)
136. Figure 7B.15 [d]: Neural Network regression plot (tested with the same word)
137. Figure 7B.16: Test results of classification for 10 samples (10 utterances by only 1 speaker, here speaker 4 from the same database used for training the ANN)
138. Figure 7B.17 [a]: ANN design for speech independent speaker recognition/Authentication
139. Figure 7B.17 [b]: Speech/text Independent Speaker Authentication: Neural Network training performance plot for speech independent speaker recognition
140. Figure 7B.18 [a]: Five classes perfectly visible for five speakers
141. Figure 7B.18 [b]: ANN training performance for five speakers
142. Figure 7B.18 [c]: ANN test performance for five different speakers which are clearly rejected by producing far away results. Only in one occasion our ANN produces close output which matches the output classes during training
143. Figure 7B.18 [d]: ANN test performance for five different speakers using registered words (pre existing templates) which are clearly classified by producing matching results during training our ANN whereas it narrowly missed the fourth word
144. Figure 7B.19 (A): Text dependent and text independent Speaker authentication comparison
145. Figure 7B.19 (B): Text dependent and text independent Speaker authentication comparison
146. Figure 7B.20: Performance diagram of SPERIA B