

Contents

Part I Background

1	Introduction	3
1.1	Background	3
1.2	Phoneme Boundary Segmentation: Present Technique	6
1.3	ANN as a Speech Processing and Recognition Tool	11
1.3.1	Speech Recognition Using RNN	12
1.3.2	Speech Recognition Using SOM	14
1.4	Motivation	15
1.5	Contribution	16
1.6	Organization	16
	References	17
2	Speech Processing Technology: Basic Consideration	21
2.1	Fundamentals of Speech Recognition	21
2.2	Speech Communication Chain	24
2.3	Mechanism of Speech Perception	28
2.3.1	Physical Mechanism of Perception	28
2.3.2	Perception of Sound	30
2.3.3	Basic Unit of Speech Perception	32
2.4	Theories and Models of Spoken Word Recognition	33
2.4.1	Motor Theory of Speech Perception	35
2.4.2	Analysis-by-Synthesis Model	37
2.4.3	Direct Realist Theory of Speech Perception	37
2.4.4	Cohort Theory	38
2.4.5	Trace Model	40
2.4.6	Shortlist Model	42
2.4.7	Neighborhood Activation Model	43
2.5	Conclusion	44
	References	44
3	Fundamental Considerations of ANN	47
3.1	Introduction	47
3.2	Learning Strategy	50

3.3	Prediction and Classification Using ANN	52
3.4	Multi Layer Perceptron	53
3.5	Recurrent Neural Network	55
3.6	Probabilistic Neural Network	63
3.6.1	Architecture of a PNN Network	64
3.7	Self-Organizing Map	66
3.7.1	Competitive Learning and Self-Organizing Map (SOM)	69
3.8	Learning Vector Quantization	71
3.8.1	LVQ Approach	72
3.8.2	LVQ Algorithm	72
3.9	Conclusion	74
	References	74
4	Sounds of Assamese Language	77
4.1	Introduction	77
4.2	Formation of Assamese Language	78
4.3	Phonemes of Assamese Language	79
4.3.1	Vowels	80
4.3.2	Diphthongs	84
4.3.3	Stop Consonant	84
4.3.4	Nasals	86
4.3.5	Fricatives	88
4.3.6	Affricates	89
4.3.7	Semi Vowels	90
4.4	Some Specific Phonemical Features of Assamese Language	91
4.5	Dialects of Assamese Language	92
4.6	Conclusion	92
	References	92
5	State of Research of Speech Recognition	95
5.1	Introduction	95
5.2	A Brief Overview of Speech Recognition Technology	97
5.3	Review of Speech Recognition During the Last Two Decades	98
5.4	Research of Speech Recognition in Indian Languages	101
5.4.1	Statistical Approach	102
5.4.2	ANN-Based Approach	106
5.5	Conclusion	109
	References	109

Part II Design Aspects

6	Phoneme Segmentation Technique Using Self-Organizing Map (SOM)	117
6.1	Introduction	117
6.2	Linear Prediction Coefficient (LPC) as Speech Feature	118
6.3	Application of SOM and PNN for Phoneme Boundary Segmentation	119
6.4	DWT-Based Speech Segmentation	121
6.5	Proposed SOM- and PNN-Based Segmentation Algorithm	123
6.5.1	PNN Learning Algorithm	124
6.5.2	SOM Weight Vector Extraction Algorithm	125
6.5.3	PNN-Based Decision Algorithm	125
6.6	Experimental Details and Result	126
6.6.1	Experimental Speech Signals	126
6.6.2	Preprocessing	127
6.6.3	Role of PNN Smoothing Parameter	129
6.6.4	Comparison of SOM- and DWT-Based Segmentation	132
6.7	Conclusion	134
	References	135
7	Application of Phoneme Segmentation Technique in Spoken Word Recognition	137
7.1	Introduction	137
7.2	Linear Prediction Model for Estimation of Formant Frequency	139
7.2.1	Human Vocal Tract and Linear Prediction Model of Speech	140
7.2.2	Pole or Formant Location Determination	142
7.3	LVQ and Its Application to Codebook Design	142
7.4	Phoneme Segmentation for Spoken Word Recognition	143
7.4.1	RNN-Based Local Classification	144
7.4.2	SOM-Based Segmentation Algorithm	145
7.4.3	PNN- and F1-Based Vowel Phoneme and Initial Phoneme Recognition	145
7.4.4	LVQ Codebook Assisted Last Phoneme Recognition	146
7.5	Experimental Details and Results	146
7.5.1	Experimental Speech Signals	148
7.5.2	RNN Training Consideration	148
7.5.3	Phoneme Segmentation and Classification Results	150
7.6	Conclusion	151
	References	152

8	Application of Clustering Techniques to Generate a Priori Knowledge for Spoken Word Recognition	153
8.1	Introduction	153
8.2	K-Means Clustering (KMC).	154
8.3	KMC Applied to Speech Data	155
8.4	Experimental Work.	157
8.4.1	Experimental Speech Samples	157
8.4.2	Role of RNN in Decision Making of the Proposed Technique	158
8.4.3	Result and Limitation	159
8.5	Conclusion.	161
	References	162
9	Application of Proposed Phoneme Segmentation Technique for Speaker Identification	163
9.1	Introduction	163
9.2	Certain Previous Work Done	165
9.3	Linear Prediction Residual Feature	169
9.4	EMD Residual-Based Source Extraction	169
9.5	LVQ Codebook and Speaker Identification	170
9.6	Speaker Database	170
9.7	Experimental Details and Results	171
9.8	System Description.	171
9.8.1	Vowel Segmentation Results	172
9.8.2	Speaker Identification Results Using LP Residual	174
9.8.3	Speaker Identification Results Using EMD Residual.	176
9.9	Conclusion.	180
	References	180
10	Conclusion	183
10.1	Conclusion.	183
10.2	Limitation	184
10.3	Future Scope	184
	Index	185

Phoneme-Based Speech Segmentation using Hybrid
Soft Computing Framework

Sarma, M.; Sarma, K.K.

2014, XXI, 187 p. 62 illus. in color., Hardcover

ISBN: 978-81-322-1861-6