

# A Phonetic Description of Assam Sora



by  
Luke Horo

A Dissertation  
Submitted in Partial Fulfillment of the Requirements  
For the Degree of Doctor of Philosophy in Linguistics

Department of Humanities and Social Sciences  
Indian Institute of Technology Guwahati  
Guwahati, Assam-781039  
November-2017



# Declaration

I declare, the dissertation entitled, “A Phonetic Description of Assam Sora” submitted by me to the Indian Institute of Technology Guwahati, for the award of the degree of Doctor of Philosophy in Linguistics, is an original work carried out by me under the supervision of Dr. Priyankoo Sarmah. I have not submitted the dissertation in any form to another University or Institute for the award of diploma or degree.

All external sources used for the completion of this dissertation have been acknowledged and cited according to the rules and regulations given by Indian Institute of Technology, Guwahati.

Luke Horo  
Department of Humanities and Social Sciences  
Indian Institute of Technology Guwahati  
Guwahati-781039, Assam, India

November 28, 2017

# Certificate

This is to certify that the dissertation entitled, “A Phonetic Description of Assam Sora” submitted by Mr. Luke Horo (Registration Number: 11614120) is an authentic work carried out by him under my supervision. The dissertation fulfills all the requirements for the award of Doctor of Philosophy in Linguistics as prescribed by the Indian Institute of Technology, Guwahati. The results in this thesis have not been submitted to any other University or Institute.

Dr. Priyankoo Sarmah  
Department of Humanities and Social Sciences  
Indian Institute of Technology Guwahati  
Guwahati-781039, Assam, India

November 28, 2017

*This thesis is dedicated to my family  
and  
the Sora community of Assam*



## Acknowledgements

Completing my Phd has been a long and challenging journey that I could not have done on my own. I would like to extend my gratitude and thanks to the many people who supported me and encouraged me throughout the past 6 years. Also, I acknowledge that my entrance into the PhD program and completion of it, would not have been possible without God's grace and presence in my life. I thank God for granting me the strength, courage, and hope needed to complete my PhD study.

Secondly, I would like to thank the person most influential in my Phd journey, my advisor Dr. Priyankoo Sarmah. I thank Dr. Sarmah for his continuous support, for his patience, motivation, and vast knowledge of phonetics. His guidance and knowledge helped me immensely all throughout the research and writing of this thesis.

Besides my advisor, I would like to thank the rest of my Doctoral Committee: Dr. Shakuntala Mahanta for her knowledge on phonology and for teaching me the course work necessary at the start of my PhD program; Prof. Arupjyoti Saikia for his knowledge on the history of labour migration in Assam; Prof. Rohini Mokashi Punekar for her encouragement during seminar presentations; and Dr. Bidisha Som for her unfailing support.

My sincere thanks also goes out to SIL international (Liza Guts, Ken Pretol, and Eric Andwik) for funding my early research from 2011 to 2013, and to PAJHRA for allowing me to do part time research from 2011-2013.

Without their precious support it would not have been possible to conduct this research.

I thank my fellow labmates in the Phonetics and Phonology Laboratory and friends in the Department of HSS for their companionship and camaraderie during meals, tea breaks, and organising conferences. In particular I am thankful to Kalyan Das, my batch mate and friend, for all his advice and encouragement through classes and studying together. I also thank the Phonetics and Phonology Laboratory for providing lab facilities 24x7, accessibility to computers, recording equipment, and high speed internet. Additionally, I would like to thank Lakshminath Bezbaroa central library, IIT Guwahati for giving access to essential books that made my research possible.

I am also thankful to Mr. Gitartha Chaliha for providing me the important Sora contacts in Koilamari Tea Estate. Last but not the least I would like to thank my Sora informants, who provided food and shelter wherever I went for my fieldwork. I thank Budbari Sawra of Singrijhan Tea Estate for sitting patiently for several days for data elicitation. I also thank Chandan Sawra of Singrijhan Tea Estate who provided contacts in different regions for me to carry out my field work. I also thank those who accompanied me on field visits: Binud Sawra and Shyamanta Sawra for coming with me to help record informants in Sonitpur district; Amar Nath Sawraw who came to the field locations in Assam and Orissa; Magon Sawra for coming to Udalguri district; and Bikram Sawra for helping me with the data collection in Lakhimpur district. Without their support, collection of field data and completion of this thesis would not have been possible.

## Abstract

This work is a phonetic description of Sora, a South Munda language of the Austroasiatic language family. Sora is lesser studied language spoken by the tribes of the same name, residing in parts of Orissa and Andhra Pradesh. In the 19<sup>th</sup> century groups of Sora people migrated from Orissa to Assam as indentured tea labourers. Significantly, the emigrant Sora people did not return to their homelands after the completion of their indenture period. As a result, a considerable number of Soras are also found in Assam. This work particularly describes the Sora language as it is spoken in Assam. Therefore, in order to distinguish it from Sora variety of Orissa, this work uses the names Assam Sora and Orissa Sora. This work also treats the migration of Assam Sora as an important event as the migration caused a complete geographic separation between Assam Sora and Orissa Sora. Additionally, the migration resulted in uprooting Assam Sora from its familiar linguistic environment and transplanting it in a new linguistic environment. Therefore, apart from describing the phonetic properties of Assam Sora, this work also compares the phonetic properties of Assam Sora and Orissa Sora.

Phonetic description of Assam Sora is based on field studies conducted in four Sora villages of Assam, where five Sora men and five Sora women were recorded from every village. Subsequently, a field study was also conducted in one Sora village of Orissa in order to make a synchronic comparison between Assam Sora and Orissa Sora. The phonetic proper-

ties of Assam Sora described in this work includes description of vowels, consonants and syllables. The description of vowels, stop consonants, nasals and syllable structure are also substantiated with acoustic analysis. Thus, this work provides evidence that (C)V(C) is the basic syllable structure in Assam Sora, and the phoneme inventory of Assam Sora includes six vowels and eighteen consonants. Also, the acoustic analysis in this work confirms that Assam Sora has an iambic stress pattern where, the second syllable is more prominent than the first syllable in disyllables. Thus, prominence of second second syllable is explained with the help of three phonetic cues, namely, vowel duration, fundamental frequency and vowel intensity. Additionally, this work proposes that Assam Sora vowel properties such as vowel height and vowel frontness/backness change as a function of syllable position. Thus, the change in vowel property across syllables is made clear with the help of vowel formants, perceptual distances between adjacent vowels and vowel area spaces. Among consonant phonemes, phonetic description with the help of acoustic analysis is provided for stop consonants and nasal consonants in Assam Sora. This mainly includes classification of voicing contrast in word initial stop consonants using VOT measures, classification of word final nasal sounds using nasal formants and anti-formants and classification of word medial glottal stops using spectral measures.

Subsequently, by comparing Assam Sora phonetic properties with existing descriptions of Orissa Sora it is observed that while the consonant inventories are fairly similar, there are differences in the vowel inventory. However, existing descriptions of Orissa Sora are inconsistent. For that reason comparison between Assam Sora and existing descriptions of Orissa Sora is inadequate. Therefore, this work compares the synchronic speech

data of Assam Sora and Orissa and the comparison is substantiated with acoustic analysis. This work reveals that the phonetic properties of vowels, stop consonants and word stress are essentially similar in Assam Sora and synchronic Orissa Sora. Thus, this work proposes that even after staying separate for more than a hundred years in two distant geographical locations, the Sora community in Assam has effectively preserved their linguistic characteristics. In this regard, migration history of the Sora community provide insights into the factors that helped the Sora community to preserve their linguistic peculiarities. The Soras suffered years of subjugation and social seclusion. As a result of that, the Sora community never got enough opportunities to interact with other linguistic communities of Assam. Also, other communities never showed any inclination to establish social contacts with the Sora community of Assam. Hence, this work concludes that, in spite of being geographically distant from each other, synchronically Assam Sora and Orissa Sora have similar phonetic properties.

# Contents

Declaration . . . . .	i
Certificate . . . . .	ii
Dedication . . . . .	iii
Acknowledgements . . . . .	iv
Abstract . . . . .	vi
List of Figures . . . . .	xiii
List of Tables . . . . .	xv
<b>1 Sora language and its speakers</b>	<b>1</b>
1.1 Introduction . . . . .	1
1.2 Ethnolinguistic and linguistic background of Sora . . . . .	2
1.2.1 Phylogenetic classification of Sora . . . . .	4
1.2.2 Sora nomenclature . . . . .	7
1.2.3 Varieties of Sora . . . . .	8
1.2.4 Sora in the literature . . . . .	9
1.2.5 Sora orthography . . . . .	11
1.3 Historical background of tea labourers' migration to Assam . . . . .	14
1.3.1 Scarcity of tea labour in Assam . . . . .	15
1.3.2 Indentured tea labourers in their homeland . . . . .	17
1.3.3 Gross recruitment of indentured tea labourers . . . . .	18
1.3.4 Migration of Soras to Assam . . . . .	20
1.3.5 Subjugation and isolation of indentured tea labourers in Assam . . . . .	26
1.4 Population movement and languages . . . . .	30
1.4.1 Linguistic characteristics of transplanted languages . . . . .	32
1.4.2 Sociology of transplanted languages . . . . .	35
1.5 Research objectives . . . . .	38
1.6 Methodology . . . . .	38
1.6.1 Areas of field study . . . . .	39
1.6.2 Composition of the data . . . . .	41

1.6.3	Data recording and annotation . . . . .	41
1.6.4	Participants . . . . .	42
1.6.5	Acoustic and temporal measurements . . . . .	42
1.6.6	Data representation and statistical measurements . . . . .	43
1.7	Organisation of the dissertation . . . . .	44
<b>2</b>	<b>Segmental properties of Orissa Sora and Assam Sora</b>	<b>46</b>
2.1	Introduction . . . . .	46
2.2	Literature review . . . . .	47
2.2.1	Segmental properties of Austroasiatic languages . . . . .	50
2.2.2	Segmental properties of Munda languages . . . . .	59
2.2.3	Segmental properties of Orissa Sora . . . . .	70
2.2.4	Orissa Sora vowel phonemes . . . . .	73
2.2.5	Orissa Sora consonant phonemes . . . . .	77
2.2.5.1	Stops . . . . .	78
2.2.5.2	Fricatives . . . . .	79
2.2.5.3	Nasals . . . . .	79
2.2.5.4	Liquids and Approximant . . . . .	80
2.3	Overview of Assam Sora vowel phonemes . . . . .	81
2.4	Overview of Assam Sora consonant phonemes . . . . .	83
2.4.1	Stops . . . . .	84
2.4.2	Nasals . . . . .	86
2.4.3	Fricative and Affricate . . . . .	88
2.4.4	Liquids and Approximants . . . . .	89
2.5	Conclusions . . . . .	91
<b>3</b>	<b>Syllable structure in Assam Sora</b>	<b>95</b>
3.1	Introduction . . . . .	95
3.2	Syllable structure in Assam Sora . . . . .	99
3.3	Word stress in Austroasiatic and Munda languages . . . . .	103
3.4	Acoustic analysis of word stress in Assam Sora . . . . .	105
3.4.1	Vowel duration in first and second syllable of Assam Sora di- syllables . . . . .	106
3.4.2	Fundamental frequency in first and second syllable of Assam Sora disyllables . . . . .	110
3.4.3	Vowel intensity in first and second syllable of Assam Sora di- syllables . . . . .	116

3.5	Conclusions . . . . .	119
<b>4</b>	<b>Acoustic phonetic properties of vowels in Assam Sora</b>	<b>122</b>
4.1	Introduction . . . . .	122
4.2	Vowel systems and theories . . . . .	124
4.2.1	Quantal Theory . . . . .	124
4.2.2	Adaptive Dispersion theory . . . . .	125
4.2.3	Dispersion-Focalization theory . . . . .	127
4.3	Interpretation of Assam Sora vowel formants . . . . .	129
4.3.1	Formant frequency of front vowels /i/ and /e/ in Assam Sora	132
4.3.2	Formant frequency of central vowels /ə/ and /a/ in Assam Sora	134
4.3.3	Formant frequency of back vowels /u/ and /o/ in Assam Sora	136
4.4	Assam Sora vowel plots . . . . .	139
4.5	Assam Sora vowel formants in individual syllables of disyllables . . .	142
4.5.1	F1 and F2 of similar vowels in individual syllables of disyllables	143
4.5.2	F1 and F2 of adjacent vowels in individual syllables of disyllables	147
4.5.3	Assam Sora vowel space in individual syllables of disyllables .	150
4.6	Distinctiveness of Assam Sora vowels . . . . .	154
4.6.1	Euclidean distance and statistical distinctiveness . . . . .	154
4.7	Conclusions . . . . .	156
<b>5</b>	<b>Acoustic phonetic properties of stops and nasals in Assam Sora</b>	<b>158</b>
5.1	Introduction . . . . .	158
5.2	Nasal consonants in Assam Sora . . . . .	162
5.2.1	Acoustic characteristics of nasal consonants in Assam Sora . .	162
5.3	Voicing contrast in Assam Sora stop consonants . . . . .	166
5.3.1	Acoustic characteristics of voicing contrast in stop consonants	167
5.3.2	VOT of Assam Sora stop consonants . . . . .	169
5.4	Glottal stop /ʔ/ in Assam Sora . . . . .	175
5.4.1	Phonetic realization of /ʔ/ in Assam Sora . . . . .	177
5.5	Conclusions . . . . .	180
<b>6</b>	<b>Synchronic comparison between Assam Sora and Orissa Sora</b>	<b>184</b>
6.1	Introduction . . . . .	184
6.2	Vowels in Assam Sora and Orissa Sora . . . . .	186
6.2.1	Vowel formants in Assam Sora and Orissa Sora . . . . .	188

6.3	Assam Sora and Orissa Sora vowel formants in individual syllables of disyllables . . . . .	195
6.3.1	F1 and F2 of similar vowels in individual syllables of Assam Sora and Orissa Sora disyllables . . . . .	196
6.3.2	F1 and F2 of adjacent vowels in individual syllables of Assam Sora and Orissa Sora disyllables . . . . .	199
6.3.3	Assam Sora and Orissa Sora vowel space in individual syllables of disyllables . . . . .	202
6.4	Word stress in Assam Sora and Orissa Sora . . . . .	206
6.4.1	Vowel duration across syllables in Assam Sora and Orissa Sora	208
6.4.2	Fundamental frequency across syllables in Assam Sora and Orissa Sora . . . . .	210
6.4.3	Vowel intensity across syllables in Assam Sora and Orissa Sora	213
6.5	Stop consonants in Assam Sora and Orissa Sora . . . . .	216
6.5.1	VOT of stop consonants in Assam Sora and Orissa Sora . . . . .	218
6.6	Conclusion . . . . .	222
<b>7</b>	<b>Conclusions and future research prospects</b>	<b>225</b>
7.1	General conclusions . . . . .	225
7.2	Segmental properties of Assam Sora . . . . .	226
7.2.1	Assam Sora syllable structure . . . . .	227
7.2.2	Assam Sora vowel phonemes . . . . .	228
7.2.3	Assam Sora consonant phonemes . . . . .	229
7.3	Comparing Assam Sora and Orissa Sora . . . . .	231
7.4	Closing remarks and future research prospects . . . . .	235
<b>A</b>	<b>Background information of participants</b>	<b>238</b>
<b>B</b>	<b>Data sets collected for the dissertation</b>	<b>240</b>
	<b>Glossary of terms</b>	<b>254</b>
	<b>References</b>	<b>255</b>

# List of Figures

1.1	Map of Sora speaking areas in Assam and Orissa . . . . .	3
1.2	Phylogenetic classification of Sora by N. H. Zide (1969) . . . . .	6
1.3	Phylogenetic classification of Sora by Anderson (1999) . . . . .	6
1.4	<i>Sora Sompeng</i> : Sora traditional script . . . . .	12
1.5	Illustration of Sora migration (Duarah, 2005) . . . . .	21
1.6	A sample of the Certificate of Emigration issued for tea labourers emigrating to Assam . . . . .	22
1.7	<i>Phulmai Sobor</i> A Sora migrant in Assam . . . . .	23
1.8	Sora villages in Assam . . . . .	40
1.9	Sora village in Orissa . . . . .	40
3.1	Speaker normalized vowel duration in first and second syllable . . . . .	108
3.2	Distribution of speaker normalized vowel duration in first and second syllable . . . . .	109
3.3	Speaker normalized average $f_0$ in first and second syllable . . . . .	112
3.4	Speaker normalized maximum $f_0$ in first and second syllable . . . . .	113
3.5	Distribution of speaker normalized average $f_0$ in first and second syllable	114
3.6	Distribution of speaker normalized maximum $f_0$ in first and second syllable . . . . .	115
3.7	Speaker normalized average vowel intensity in first and second syllable	117
3.8	Distribution of speaker normalized average vowel intensity in first and second syllable . . . . .	118
4.1	Region wise Assam Sora vowel plots . . . . .	141
4.2	Assam Sora vowels in 1st and 2nd syllable . . . . .	144
4.3	Assam Sora vowel space in first and second syllable . . . . .	152
5.1	Average nasal spectra of Assam Sora with formants indicated with up arrows and anti-formants indicated with down arrows . . . . .	164
5.2	Nasal spectra of Assam Sora as produced by a single speaker . . . . .	165

5.3	Wave forms of word initial /p/ and /b/ . . . . .	171
5.4	Wave forms of word initial /t/ and /d/ . . . . .	171
5.5	Wave forms of word initial /k/ and /g/ . . . . .	172
5.6	VOT of voiced and voiceless stops in Assam Sora . . . . .	173
5.7	Wave form showing complete glottal stop in the word <i>iʔi</i> . . . . .	177
5.8	Wave form showing voiced glottal stop in the word <i>iʔi</i> . . . . .	178
5.9	Wave form showing creaky voiced glottal stop in the word <i>iʔi</i> . . . . .	179
5.10	Two types of pitch patterns in creaky voiced glottal stop . . . . .	180
6.1	Assam Sora and Orissa Sora vowel plots . . . . .	193
6.2	Assam Sora and Orissa Sora vowel plots in individual syllables of disyllables . . . . .	197
6.3	Assam Sora and Orissa Sora vowel space in disyllables . . . . .	203
6.4	Speakers normalized vowel duration in first and second syllable . . . . .	209
6.5	Speakers normalized average <i>f</i> 0 in first and second syllable . . . . .	212
6.6	Speakers normalized maximum <i>f</i> 0 in first and second syllable . . . . .	212
6.7	Speakers normalized average vowel intensity in first and second syllable . . . . .	215
6.8	VOT of oral stops in Assam Sora and Orissa Sora . . . . .	221
7.1	Vowel correlation between Assam Sora and Orissa Sora . . . . .	233
7.2	Stop consonants correlation between Assam Sora and Orissa Sora . . . . .	234

# List of Tables

1.1	Language wise population of tea labourers in Assam (Registrar General of India, 2001) . . . . .	27
2.1	Austroasiatic consonants (Jenny et al., 2014)Part-A . . . . .	52
2.2	Austroasiatic consonants (Jenny et al., 2014)Part-B . . . . .	53
2.3	Proto-AA and Proto-Munda initial consonants (Sidwell & Rau, 2014)	55
2.4	Austroasiatic vowels (Jenny et al., 2014) . . . . .	57
2.5	Proto-Austroasiatic and Proto-Munda vowel inventory (Sidwell & Rau, 2014). Undetermined Proto-Munda vowels are underlined . . . . .	57
2.6	Traditional Munda consonants . . . . .	60
2.7	Modern Munda consonants (Anderson, 2014) . . . . .	65
2.8	Modern Munda vowels (Anderson, 2014) . . . . .	69
2.9	Orissa Sora consonants inventory . . . . .	71
2.10	Orissa Sora vowel inventory . . . . .	72
2.11	Assam Sora vowel phonemes (Horo & Sarmah, 2015) . . . . .	82
2.12	Phonemic distribution of Assam Sora vowels (Horo & Sarmah, 2015)	82
2.13	Assam Sora consonant phonemes . . . . .	84
2.14	Voicing contrast in Assam Sora stop consonants . . . . .	85
2.15	Glottal stop in Assam Sora . . . . .	85
2.16	Nasal consonants in Assam Sora . . . . .	87
2.17	Fricative and affricate consonants /s/ and /ɖʒ/ in Assam Sora . . . . .	88
2.18	Liquid consonants in Assam Sora . . . . .	90
3.1	Assam Sora onset and coda consonants . . . . .	102
3.2	Average vowel duration in first and second syllable . . . . .	107
3.3	Average of average $f_0$ in first and second syllable . . . . .	111
3.4	Average of maximum $f_0$ in first and second syllable . . . . .	111
3.5	Average of average vowel intensity in first and second syllable . . . . .	116
4.1	Contrasting pairs for vowels /i/ and /e/ . . . . .	132

4.2	Average formant frequency of Assam Sora vowels /i/ and /e/ . . . . .	133
4.3	Contrasting pairs for vowels /ə/ and /a/ . . . . .	134
4.4	Average formant frequency of Assam Sora vowels /ə/ and /a/ . . . . .	135
4.5	Contrasting pairs for vowels /u/ and /o/ . . . . .	136
4.6	Average formant frequency of Assam Sora vowels /u/ and /o/ . . . . .	137
4.7	Euclidean distance between similar vowels across syllables . . . . .	146
4.8	Euclidean distance between front vowels in first and second syllable .	148
4.9	Euclidean distance between central vowels in first and second syllable	148
4.10	Euclidean distance between back and central vowels in first and second syllable . . . . .	149
4.11	Vowel area estimation in first and second syllable . . . . .	153
4.12	Vowel euclidean distance between Assam Sora vowels . . . . .	155
5.1	Word initial voicing contrast in Assam Sora stop consonants . . . . .	170
5.2	Average VOT of voiced stops . . . . .	171
5.3	Average VOT of voiceless stops . . . . .	172
5.4	VOT of dental [d̪] and aspirated [kʰ] . . . . .	174
6.1	Orissa Sora /i/ and /u/ (Anderson & Harrison, 2011) compared with Assam Sora . . . . .	187
6.2	Orissa Sora vowel minimal set (Donegan & Stampe, 2002) compared with Assam Sora . . . . .	187
6.3	Average vowel formants of Assam Sora and Orissa Sora . . . . .	190
6.4	Euclidean distance between similar vowels across syllables . . . . .	198
6.5	Euclidean distance between front vowels in first and second syllable .	200
6.6	Euclidean distance between central vowels in first and second syllable	200
6.7	Euclidean distance between back and central vowels in first and second syllable . . . . .	200
6.8	Vowel area estimation in first and second syllable of Assam Sora and Orissa Sora . . . . .	204
6.9	Average vowel duration in first and second syllable . . . . .	208
6.10	Average <i>f</i> 0 in first and second syllable . . . . .	211
6.11	Maximum <i>f</i> 0 in first and second syllable . . . . .	211
6.12	Average vowel intensity in first and second syllable . . . . .	214
6.13	Average VOT of voiced stops in Assam Sora and Orissa Sora . . . . .	219
6.14	Average VOT of voiceless stops in Assam Sora and Orissa Sora . . . .	220

7.1	Assam Sora vowel phonemes . . . . .	228
7.2	Assam Sora consonant phonemes . . . . .	229
A.1	Information of female participants . . . . .	238
A.2	Information of male participants . . . . .	239
B.1	Assam Sora consonant minimal sets . . . . .	240
B.2	Assam Sora vowel minimal sets (Horo & Sarmah, 2015) . . . . .	241
B.3	Assam Sora disyllables (1) . . . . .	242
B.4	Assam Sora disyllables (2) . . . . .	243
B.5	Assam Sora disyllables (3) . . . . .	244
B.6	Assam Sora disyllables (4) . . . . .	245
B.7	Assam Sora disyllables (5) . . . . .	246
B.8	Assam Sora disyllables (6) . . . . .	247
B.9	Assam Sora disyllables (7) . . . . .	248
B.10	Assam Sora disyllables (8) . . . . .	249
B.11	Assam Sora disyllables (9) . . . . .	250
B.12	Assam Sora disyllables (10) . . . . .	251
B.13	Assam Sora disyllables (11) . . . . .	252
B.14	Assam Sora disyllables (12) . . . . .	253



# Chapter 1

## Sora language and its speakers

### 1.1 Introduction

This dissertation is a phonetic description of the speech sounds in Sora as it is spoken by the Sora community living in Assam and Orissa. This work examines Sora in the context of language migration and explains the synchronic relation between Sora spoken in Assam and Orissa on the basis of their phonetic properties. Sora is a South Munda language of the Austroasiatic family that is mainly spoken in Orissa and Andhra Pradesh. A variety of the Sora language emerged in Assam due to the migration of Sora speakers from Orissa to Assam in the 19<sup>th</sup> century. While the Sora variety of Orissa has been studied to some extent in the past, the Sora variety of Assam has been barely studied. Therefore, the primary objective of this research is to provide a phonetic description of the Sora variety spoken in Assam. Moreover, literature on migrated languages suggest that migrated languages are distinct from the original language because they lose their idiosyncratic features after coming in contact with other languages in the new location. Therefore, the secondary objective of this research is to make a synchronic comparison of the phonetic features in the two Sora varieties to examine whether the Sora variety of Assam has been effected by language migration or not. The following section (§1.2) discusses the linguistic and ethnolinguistic background of the Sora community. Thereafter §1.3 describes

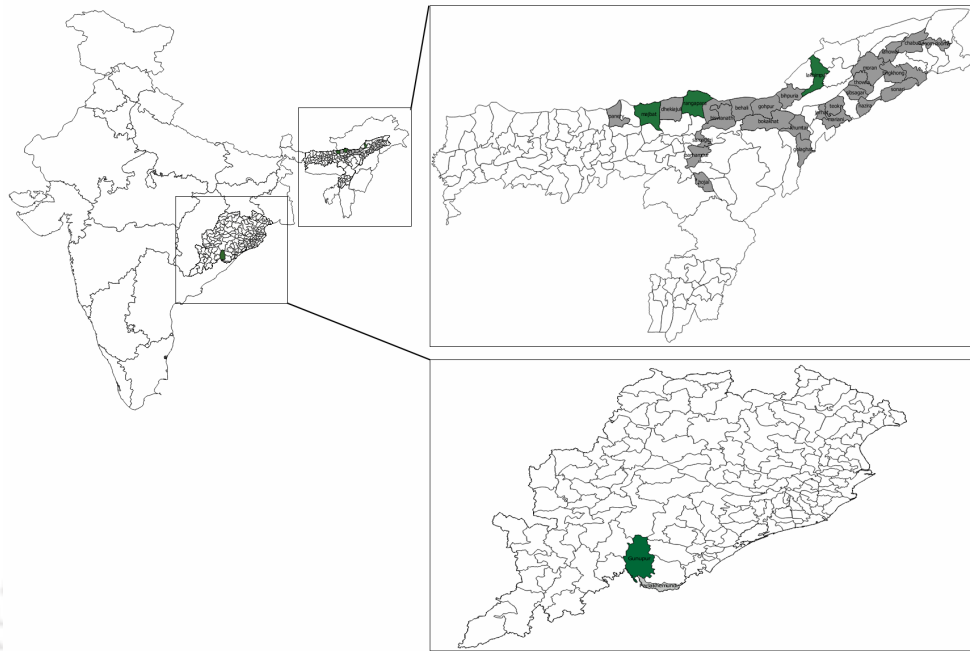
the historical background of Sora migration from Orissa to Assam and discusses the circumstances that led to the emergence of Sora in Assam. Subsequently, section §1.4 reviews the impact of human migration on languages and discusses the linguistic and sociolinguistic consequences of language migration observed cross linguistically. Then, the research objectives are presented in §1.5 and the methodology adopted for achieving the stated objectives are discussed elaborately in §1.6. Finally, §1.7 briefly describes the organization of this dissertation.

## 1.2 Ethnolinguistic and linguistic background of Sora

Officially, Sora is a non-scheduled language of Orissa, however some Sora groups migrated to Assam in the 19<sup>th</sup> century. Therefore, this work uses the term Assam Sora and Orissa Sora in order to distinguish between Sora varieties spoken in two different geographical locations (See Figure 1.1). According to the Registrar General of India (2001) there are 252,519 Soras in India of which 1,72,288 are recorded in Orissa and 406 are recorded in Assam. However, in Assam the community estimates that currently there are at least 40,000 Soras, and this research has discovered more than 2000 Soras in just four villages of Assam. Therefore, the total population of Soras in Assam appears to be much greater than the official records estimate.

Regarding the origin of Soras in India, etymological evidences suggest Sora is one of the oldest tribes of eastern and central India. Studies in the past have drawn historical and archaeological evidences to relate Sora etymology with the *Epics* and *Puranas* of Hindu mythology (B. Sharma, 1992) and (Lakshmi Narayana, 2013). Likewise, Soras are related with the Sabaras of Vedic and Sanskrit literature (Grierson, 1906). The origin of *Jagannath Cult* in Orissa is associated with the Sora tribe (Dash, 2014) of Orissa. However, currently most Soras have animistic beliefs, only some Soras have

Figure 1.1: Map of Sora speaking areas in Assam and Orissa



adopted Hinduism, and a few others have adopted Christianity.

Soras in Orissa are found in both plains and hills and they are often categorized on the basis of their habitation, occupation, food, social status, customs and habits (A. Mohanty, 1996). Generally Soras are an agrarian society but, some of them are also hunters and gatherers. Moreover, some Soras are skilled in making tools and utensils and such groups are generally considered to be a sub-class of the Sora community in Orissa. On the other hand, in Assam the Soras started living as indentured tea labourers. Currently, although the indentured period is completed, and some Soras have different means of livelihood, a large number of Sora migrants and their subsequent generations are still labouring in the tea gardens of Assam. As a result, their original social stratification is abandoned, however their beliefs, cultural practices and heritage language are still preserved. Development and documentation of the Sora language in Assam is different from the Sora language of Orissa. Therefore, in order to provide a background about Sora in general some linguistic information about the language are introduced in the following subsections.

### 1.2.1 Phylogenetic classification of Sora

Sora is a South Munda language of the Austroasiatic language family, but classification of Sora, as well as of Munda and Austroasiatic languages have changed and developed in many ways. According to Diffloth & Zide (1992) the Austroasiatic language family is the most 'archaic' language family found in Southeast Asia and East India. Likewise, Sidwell (2015) argues that, "the Austroasiatic phylum is the oldest recognisable language family in the mainland South and Southeast Asia". However, there has always been a dispute among scholars of Austroasiatic comparative linguistics and historical linguistics regarding the classification of the Austroasiatic language family. Disagreement arises mainly due to methodological differences and due to limited availability of adequate data on individual languages of the Austroasiatic language family. Moreover, it is well known that a vast majority of Austroasiatic languages such as Mon, Khmer, Vietnamese, Khmu etc. are spoken in Southeast Asia and other languages such as Munda, Nicobarese and Khasian are spoken in South Asia (eastern and central India and the Nicobar islands). As a result, Austroasiatic languages are assumed to have adopted areal features that lead to a general classification of the languages on the basis of their geographical distributions. Also, individual Austroasiatic languages exhibit diverse features that are often difficult to identify in the cognate languages. As a result, there are often confusions in the classification of Austroasiatic languages. Therefore, this section discusses major trends observed in the phylogenetic classification of Sora, Munda, and Austroasiatic languages in general.

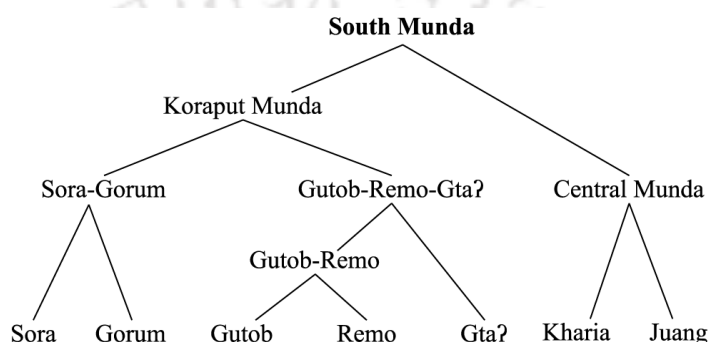
The Austroasiatic language family was first established by Schmidt (1906), he proposes the existence of an '*Austriac*' language family including two major sub-families namely Austronesian and Austroasiatic. In the Austroasiatic family Schmidt (1906) proposes the existence of seven sub-families, namely Mon-Khmer, Senoi, Palaung-Wa-Riang, Khasi, Nicobarese, Munda and Tscham. However, in the subsequent period different scholars propose different classifications of the Austroasiatic phylum. Con-

sequently, newer classifications of the Austroasiatic phylum continue to emerge. One such classification of Austroasiatic language family is proposed by Pinnow & Kuiper (1963). They propose that Austroasiatic languages can be divided into two major branches, the Western group and the Eastern group. While the Western group is further sub-divided into Nahali and Munda, the Eastern group is divided into Nicobarese and Palaung-Khmer. Subsequently, another classification of Austroasiatic languages is proposed by Diffloth & Zide (1992). This classification is based on the typological division of languages and it proposes Mon-Khmer and Munda as the two major sub-branches of the Austroasiatic language family. Moreover, it further divides the Mon-Khmer branch into Northern, Eastern and Southern Mon-Khmer. However, following this, Diffloth (2005) reconsidered Nicobarese to be the third branch of the Austroasiatic language family which was earlier grouped under Southern Mon-Khmer sub-branch of the Austroasiatic language family. Currently, the Mon-Khmer branch of Austroasiatic language family has been again restructured. Consequently, from Sidwell (2013) it is known that there are at least 13 primary branches of the Austroasiatic language family. The 13 primary branches include Aslian, Bahnaric, Katuic, Khasian, Khmer, Khmuic, Mangic/Pakanic, Monic, Munda, Nicobaric, Palaungic, Pearic and Vetic. However, Sidwell (2013) mentions, discriminating features of the individual branches are still not known and branching nodes of the overall language tree are also not resolved. Thus, there are still possibilities of further research into the phylogenetic classification of Austroasiatic languages.

The Munda branch was once discarded as a member of the Austroasiatic language family. However, Pinnow & Kuiper (1963) has convincingly established that the Munda branch is a distinctive member of the Austroasiatic language family. Thus, although the Munda branch appears to be typologically diverged from the other branches of Austroasiatic family, there are no doubts that Munda is a core member of the Austroasiatic language family. Sora belongs to the Munda branch of the

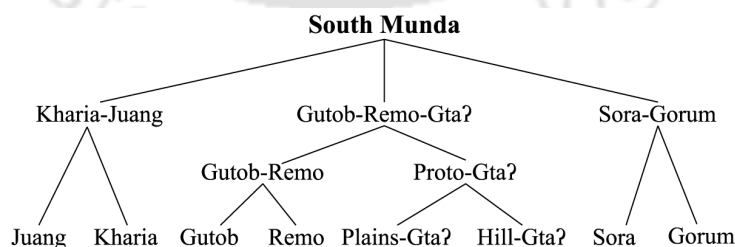
Austroasiatic language family and it is further classified under the South Munda sub-branch (Grierson, 1906), (Stampe, 1965), (Starosta, 1976, 1992), (A. R. Zide, 1976), and (Diffloth & Zide, 1992). Traditional classification of South Munda by N. H. Zide (1969), propose a further sub-branching of South Munda into Koraput Munda and Central Munda, Sora is included in the Koraput Munda sub-branch (see Figure 1.2).

Figure 1.2: Phylogenetic classification of Sora by N. H. Zide (1969)



Koraput Munda and Central Munda sub-classification of South Munda is also supported by other Austroasiatic scholars such as Stampe (1965) and Diffloth & Zide (1992). However, Anderson (1999) and Anderson (2001) reject the Koraput Munda sub-branching of South Munda. They propose a new sub-classification of South Munda into Kharia-Juang, Gutob-Remo-Gata and Sora-Gorum on the basis of verb morphology of South Munda languages (see Figure 1.3).

Figure 1.3: Phylogenetic classification of Sora by Anderson (1999)



Anderson (1999) and Anderson (2001) discarded the Koraput Munda sub-branching mainly because they found that Koraput Munda is only an areal name of languages spoken in Koraput district of Orissa. They argue that a language named Koraput

Munda was never spoken by any group in the proto stage. Hence, from these studies the sub-classification of South Munda has changed and Sora is now regarded as a South Munda language instead of a Koraput Munda language. This dissertation follows the convention proposed by Anderson (2001) and considers Assam Sora to be a member of the South Munda subfamily and regards Gorum as the closest cognate language of the Sora language.

### **1.2.2 Sora nomenclature**

In the Sora literature it is observed that the name of the language as well as name of the ethnic group appears differently in different literature. Census of India uses the name Savara, and in some cases Grierson (1906) also uses the name Savara. Additionally, Grierson (1906) uses the name Sawara to refer to the same Sora language spoken in Orissa, Chota Nagpur, West Bengal, Madras and Central Province and he uses the name Sobor to refer to a speculative “purest” race of the Sora tribe living in these places. Another nomenclature of Sora is found in A. Mohanty (1996), he uses the name Sora or Soralang to refer to the language but uses the name Saora to describe the ethnic group living in parts of Orissa, Bihar, Madhya Pradesh, West Bengal and Andhra Pradesh. He also considers Savara, Sabar and Saur as alternative names for the same group of people. Remarkably, Rao (1996) prefers to use the name Savara to describe the ethnic group living in parts of Andhra Pradesh only. However, despite the existence of different nomenclature of Sora, Ramamurti (1931) and Ramamurti (1938) strongly favour using the name Sora to refer the language as well as the ethnic group. Subsequently, the vast majority of Sora literature has consistently used the name Sora and this work also follows the same naming convention.

### 1.2.3 Varieties of Sora

The Sora language is spoken in different regions of Southern Orissa and there are claims that Sora has regional varieties based on the geographical distribution of its speakers. Research has revealed a few dialects of Sora in Orissa, but existence of Sora dialects in Assam are not reported. Reporting on the existence of Sora dialects in Orissa, an unpublished work by Stampe (1965) made a comparison between two Sora dialects namely Sogeda and Sundung. He suggests that difference in vowel height is a distinguishing feature between the two Sora dialects. Likewise, Ramamurti (1938), proposes six possible regional dialects of Sora namely Gumma, Ramagiri, Udayagiri, Kapu, Serung, and Jirang on the basis of lexical differences. However, he does not include the dialects identified in Stampe (1965). Also, Ramamurti (1938) proposes that Gumma is the standard dialect of Sora, but in his dictionary he chose to record Sora speakers of a village called Maliha because he found them to be monolingual Sora speakers. However, he does not mention whether this variety should be counted as a separate dialect or not. Additionally, he comments that there could be phonemic differences even within Sora speakers of the same dialect. Recently, Anderson & Harrison (2008b) studied the Gumma dialect and speculated that there are more distinguishable Sora dialects in Orissa. Although they do not document the other significant dialects of Sora, they report that Sora speakers are aware of phonetic, prosodic and morpho-syntactic differences across different Sora dialects. Moreover, they claim that although Juray is often regarded as a separate South Munda language, it may actually be a divergent Sora dialect. Thus, it is evident dialects of Sora are not clearly documented. A number of dialects are postulated, but there are no conclusive descriptions so far.

Assam Sora speakers are aware that there are lexical differences among Sora speakers living in different regions of Assam. However, of all the Sora dialects mentioned above, they recognize only the Juray dialect. Assam Sora speakers are aware that

Assam Sora and Juray have lexical differences, but they do not consider Juray to be a separate language. Thus, it is likely that Juray is a dialect of Assam Sora. However, Juray speakers of Assam are not recorded in this dissertation and therefore its relation with Assam Sora is not known at this point of time. Additionally, Assam Sora speakers of Sessa and Singrijhan region report that there are lexical differences in the speech varieties of the two regions. However, it is found that lexical difference in the speech varieties of Sessa and Singrijhan are mainly due to difference in semantic domains. For example, the word for 'what' in Sora is *itin* but, Sora speakers of Singrijhan use the word *zin*. In two other areas, namely Koilamari and Lamabari both words are used interchangeably. It is found that the two words vary mainly because they are used in two different semantic domains. While *itin* is used in polite interrogative sentences, *zin* is used in less polite interrogative sentences. Thus, it is likely that there may be some regional varieties of Assam Sora, therefore Assam Sora speech data recorded from four regions of Assam are subjected to phonetic comparison in this work. Hence, this will help in verifying if Assam Sora has any regional variation or not, on the basis of its phonetic properties.

#### **1.2.4 Sora in the literature**

There has been a considerable amount of research on the Sora language as well as on the Sora community. As a result, academic and non-academic literature on Sora is available. Existing Sora literature is mainly based on Orissa Sora whereas, literature based on Assam Sora is limited. Printed Sora literature based on Orissa Sora includes journal articles, books, book chapters, dictionaries, academic dissertations and language manual. One of the earliest existing books on Sora is Grierson (1906). This work documents Sora among other Indian languages and includes Sora in the volume on Munda and Dravidian languages of India. In the subsequent period scholars have continued researching the Sora language and its speakers. As a result, different forms

of Sora literature have been produced at various point of time. Some of the significant Sora literature are listed and reviewed in this section.

Journal articles relating to linguistic research on Orissa Sora include works such as Stampe (1965), A. R. Zide (1976), Starosta (1976, 1992), Donegan (1993) and N. Zide (1999). These works describe some linguistic features of Orissa Sora in the domains of morphology, syntax and phonology. There are also journal articles describing anthropological researches conducted on the Sora tribe. Such works include Vitebsky (1980, 1993, 1998). These works document some of the important rituals and beliefs of the Sora tribe in Orissa. Also, some linguistic and ethnolinguistic research on Sora is available in the form of book chapters in A. Mohanty (1996) and Rao (1996). Similarly, a sketch grammar of Orissa Sora by Anderson & Harrison (2008b) is also available in the form of book chapter in an edited volume by Anderson (2008b). Additionally, there are academic dissertations by Stampe (1963), Starosta (1967) and A. R. Zide (1982) on Sora and related languages that describe aspects of phonology and syntax in those languages. However, these works are not published and are often difficult to access. Apart from the above mentioned research works on Sora there are also three bilingual dictionaries of Sora compiled by the Ramamurti. They include Ramamurti (1914) (Telugu Savara dictionary) and Ramamurti (1933, 1938) (English-Sora and Sora-English dictionary). Another important contribution by the same author is Ramamurti (1931), which is a language manual for learning the Sora language. These works by Ramamurti are important contributions to the language and he has shown great passion in documenting the Orissa Sora language in these works. On the other hand, research based on Assam Sora includes only one book on the anthropological description of Assam Sora by Kar (1981) and a linguistic journal article on the phonetic features of Assam Sora vowels by Horo & Sarmah (2015). Hence, it is evident that Orissa Sora has been studied and documented to some extent in the past, but Assam Sora is largely undescribed. However, a review

of literature on Orissa Sora speech sound characteristics reveals that the existing descriptions are inconsistent and not described adequately.

A review of the available literature shows that scholars of anthropology and linguistics have studied various aspects of Sora in general. The anthropological researchers have mostly described the peculiarities of the Sora tribe in Orissa and Assam. On the other hand, the linguistic research describing various aspects Sora morphology, syntax, phonetics and phonology, are mostly based on Orissa Sora. Reviewing further the anthropological findings and examining the morphological and syntactic peculiarities of Sora are not within the scope of this research. Rather Sora phonetics and phonology are focused in this dissertation because there is no consensus among the existing literature on Orissa Sora Segmental properties. Hence, Chapter 2 of this dissertation further discusses the irregularities present in the existing descriptions of Orissa Sora segmental properties.

### **1.2.5 Sora orthography**

Sora literature is mostly written in English, but some literature written in Sora language is also available. However, Sora literature written in Sora language is mainly based on Sora speakers living in Orissa and Andhra Pradesh whereas, writings in Sora language based on Assam Sora is very limited. Moreover, Sora writings of Orissa and Andhra Pradesh are not standardized yet. In both Orissa and Andhra Pradesh Sora is written using four different scripts including Oriya, Telugu, Roman and Sompeng (a traditional Sora script). It is evident that Oriya, Telugu and partially modified Roman script are commonly used to write Sora in Orissa and Andhra Pradesh. Writings in Sora language using Oriya and Telugu script include some pedagogical and cultural texts. Besides, writings in Sora language using Roman script include the New Testament published by Bible Society of India in 1965 in Bangalore and a pictorial handbook of Sora written by Piers Vitebsky (a British anthropologist) and Monoshi

Raika (a Sora Scholar of Gunupur) published in 2001 at Visakhapatnam. Additionally, a traditional Sora script (Figure 1.4) is used to write Sora mostly in Orissa and Andhra Pradesh.

Figure 1.4: *Sora Sompeng*: Sora traditional script

𑌒	𑌓	𑌔	𑌕	𑌖	𑌗
SA	TA	BA	CA	DA	GA
𑌘	𑌙	𑌚	𑌛	𑌜	𑌝
MA	NGA	LA	NA	VA	PA
𑌞	𑌟	𑌠	𑌡	𑌢	𑌣
YA	RA	HA	KA	JA	NYA
𑌤	𑌥	𑌦	𑌧	𑌨	𑌩
A	E	I	U	O	AI MRAI
𑌪	𑌫	𑌬	𑌭	𑌮	𑌯
1	2	3	4	5	6 7 8 9 10

The traditional Sora script is called Sora Sompeng and it is mostly used to write guide books on how to use the script. It is believed that a Sora scholar named Mangei Gomango devised the Sora Sompeng in 1936 and inscribed it on a piece of rock. Although Oriya and Telugu scripts were preferred in the past over Sora Sompeng, currently Sora Sompeng is extensively promoted among Sora speakers of Orissa and Andhra Pradesh. Moreover, a religious belief is associated with Sora Sompeng in Orissa as well as in Andhra Pradesh. The Soras believe that Sora Sompeng is provided to them supernaturally through a spirit. Therefore the traditional script has gained significant importance among the Soras of Orissa and Andhra Pradesh. Linguistic (N. Zide, 1999) and typographic (Priyadarshini, 2016) features of Sora Sompeng have been studied to some extent and currently Sora Sompeng is also available on Unicode (Everson, 2009). Sora Sompeng has 25 characters including 18 consonants, 6 vowels

and 1 auxiliary letter. It also includes 10 numerals. It is believed that the first 24 letters represent 24 gods in Sora. However, the rationality behind the arrangement of letters is not known and the sound symbol correlation is also not known. Moreover, while examining Sora Sompeng, N. Zide (1999) suggests the script lacks linguistic efficiency and does not represent the full phoneme inventory of Sora.

In contrast, Sora is not commonly written in Assam and there are only a few printed Sora books written by the Sora speakers of Assam. Two such Sora books written in Assam are (Duarah, 2005) and (Obed, 2015). Both these works are language handbooks written in Assamese script. Some literate Soras have attempted to write Sora in Roman script, but they have not been able to produce any printed literature so far. Thus writing in Sora language in Assam is not standardized yet. Production of Sora literature in Sora language is mostly promoted by either individual writers or by an apex body named “All Assam Saora Samaj”. The apex body represents a literature society and there have been attempts in the past to write Sora language in Assam on various occasions. However, there are no significant Sora literature in Sora language in Assam other than the ones mentioned above.

In summary, there are evidences that Soras are an indigenous tribe of Orissa and there are claims that they have existed since pre-historic India. There are also evidences that Soras migrated from Orissa to Assam in the 19<sup>th</sup>. Studies in the past have documented the Sora language to some extent however, existing literature on Sora language is mostly based on the Sora variety spoken in Orissa. Whereas, the Sora variety spoken in Assam is hardly described. Therefore, the focus of this work is the Sora variety of Assam. Moreover, since the origin of Assam Sora is the Sora variety spoken in Orissa, finding the co-relation between Assam Sora and Orissa Sora is of particular interest of this dissertation. Also, since Assam Sora emerged due to the migration of people, it is the interest of this dissertation to examine the circumstances of migration as well as the consequences of migration on the language

variety. Hence, the following section provides a historical background about the migration of indentured tea labourers' migration to Assam of which the Soras were an integral part.

### **1.3 Historical background of tea labourers' migration to Assam**

Studies on migration suggest, population movement or the migration of people can be motivated by a variety of factors. Some communities voluntarily migrate from poorer areas to urban areas in search of better economic opportunities, while some communities are forcefully migrated by a dominant group. Migration of people from European countries to the United States is an example of voluntary migration, and migration of African slaves to the Caribbean and United States is an example of forced migration. Migration of people can also be temporary as well as permanent. Some migrants return to their place of origin, either after achieving their economic goals or on completion of their contract period in the migrated land. On the other hand, some migrants permanently settle in the new territory. Migration of the Turkish 'guest workers' to Germany who work in Germany during winter and return to Turkey in summer is an example of temporary migration. Labour migration from former European colonies of Africa, India and Pakistan to various parts of Europe is an example of migrants who permanently settled in the new territory.

In the 19<sup>th</sup> century several tribal groups of central and eastern India including the Soras started migrating to Assam as indentured as tea labourers. Although the tea labourers migrated for a contracted period, a large number of them permanently settled in Assam. Currently several generations of the indentured tea labourers including the Soras have lived in Assam for more than hundred years. Consequently, most of them have completely abandoned their ancestral lineages. In the post mi-

gration period, the indentured labourers including the Soras have faced numerous challenges while relating to the host communities in Assam. Thus, understanding the migration pattern of indentured tea labourers to Assam is an important aspect of this dissertation, since migration patterns of indentured tea labourers including the Soras has a significant effect on their linguistic peculiarities. Therefore, this section specifically describes the circumstances that led to the migration of indentured tea labourers to Assam. Also, the nature of their settlement in the new location in the subsequent period is described in this section.

Migration of indentured tea labourers is also an important event in the history of Assam and therefore it has been extensively discussed in several historical narratives. Some of the key narratives include works such as Kar (1979, 1981), Chatterjee & Gupta (1981), Behal (1985), Goswami (1999), and J. Sharma (2006, 2009). These studies have shown that migration of indentured tea labourers in Assam started at the beginning of 19<sup>th</sup> century, and the labourers came from various parts of India such as Bihar, Orissa, Madhya Pradesh, Andhra Pradesh, and Madras Presidency. These studies suggest, there were two factors that led to the migration of indentured tea labourers to Assam. The first factor was the growing need of plantation labour in Assam and the second factor was the difficult living conditions of the migrants in their homeland. The following subsections discuss the two factors in detail. Thereafter, migration of Soras is discussed more elaborately with the help of personal narratives.

### **1.3.1 Scarcity of tea labour in Assam**

An acute scarcity of plantation labour in the tea industries of Assam in the nineteenth century was the main reason for recruiting indentured tea labourers from different parts of India. According to Goswami (1999), nineteenth century Assam was a period of significant economic change, primarily centered around the setting up of an ambitious tea industry in Assam. Significantly, Chatterjee & Gupta (1981), Gupta

(1959) and Guha (1977) as cited in Goswami (1999) view the discovery of wild tea plants in Assam in 1823 as the primary motivation behind the colonization of Assam. Hence, discovery of tea and setting up the tea industries in Assam were important undertakings of the colonial rule in Assam in the nineteenth century. However, there are evidences that the tea industry did not grow rapidly in the beginning, due to an acute scarcity of plantation labourers in the state. The new industrial setup had failed to attract local labourers of Assam in the initial days (Chatterjee & Gupta, 1981), (Gogoi & Goswami, 2004). The company adopted policies such as increasing land taxes, prohibiting opium cultivation, and releasing government controlled opium at higher price in order to force local people into plantation work (Goswami, 1999). As a result, some Bodo Kachari and some Naga farmers worked in the tea industries for a short period mainly to pay the increased land taxes, or to pay their debts (Goswami, 1999). However, the harsh working conditions, particularly in tea firing, became the primary reason for in-stable labour in the new industrial setup (Goswami, 1999). Eventually, when the companies failed to force local people to work in the tea gardens, they stereotyped the local people as lazy and non industrious. Chatterjee & Gupta (1981), Gogoi & Goswami (2004), Goswami (1999) and J. Sharma (2009) oppose such views and suggest that the simple livelihood of Assamese society and a favorable demography of Assam did not motivate the farmers to do wage labouring in the nineteenth century. The fertile land in the Brahmaputra valley provided for the local farmers abundantly, so there was no need for additional earning sources. Moreover, cultivation was preferred over working in plantations which required working under the harsh conditions (Goswami, 1999). As a result, the industry was not able to grow as rapidly as it was originally projected. Subsequently, these circumstances provoked the planters to look for more labourers from other parts of the country (J. Sharma, 2009). The planters started looking for labourers that could readily adapt to the working conditions in the tea industries of Assam. They became partic-

ularly interested in recruiting immigrant labourers from distant and poorer areas of India. This would prevent the labourers from leaving the industry even if they were not satisfied with the working conditions (Goswami, 1999). The quest for suitable labour force ended with the tribes of eastern and central India and a large number of them migrated to Assam in different phases. In this regard, scholars have suggested that conditions of the tribes in these areas were very deplorable and therefore it was easy for requiters to convince them to migrate to Assam. Hence, the following section discusses the condition of indentured tea labourers in their homeland which motivated them to migrate to Assam.

### **1.3.2 Indentured tea labourers in their homeland**

The homeland of indentured tea labourers in Assam identified by Kar (1981), Chatterjee & Gupta (1981), Goswami (1999) and Gogoi & Goswami (2004) include places such as Bengal, Bihar, Orissa, Chotanagpur, Uttar Pradesh, Madras and Maharashtra. These studies show, in the nineteenth century these states were deeply affected by poverty, famine, droughts, landlessness and forced commercialization. Therefore, J. Sharma (2009) referring to the German Lutheran Mission reports of 1906, suggests the tribal population living in these states were compelled to migrate to Assam to find better opportunities. Additionally, it is found that, expansion of the colonial frontier in the Eastern and Central India in the nineteenth century also required the colonial rule to relocate the tribal population living in these areas to other places. Hence, while the people were themselves compelled to migrate, it was also suitable for the colonial rule to move them to Assam as indentured tea labourers.

However, Chatterjee & Gupta (1981) presents a different perspective. They suggest, although living conditions in the above mentioned states were not favorable for the tribal population, unlawful labour recruitment in those areas was the actual reason for the migration of indentured tea labourers to Assam. Evidences shows, in the

nineteenth century when tea labourers migrated to Assam, there were other places, such as the coal industries in Ranchi, that had better opportunities and working conditions for the labourers than in Assam. However, the labour recruitment system was improper and the tribes were driven to Assam through false promises. Thus, they indicate that migration was not voluntary, but the people were deceptively recruited and then forced to migrate to Assam. Moreover, Kar (1981) also proposes that after the first batch of labourers were brought to Assam, the company visualized them as the most suitable workers and tried hard to bring the same group of people in larger numbers. As a result, unfair means were adopted to recruit indentured tea labourers. Therefore, in order to understand how a large number of tribes from eastern and central India were deceitfully migrated to Assam, the following section discusses some of the faulty practices of indentured tea labourer recruitment in Assam in the nineteenth century.

### **1.3.3 Gross recruitment of indentured tea labourers**

Recruitment of indentured tea labourers to Assam started in 1841 utilizing three systems of labour recruitment: (1) through contractors (2) through the tea garden 'Sardari' system and (3) through labour recruiting organizations (Kar, 1981). Recruitment through contractors is referred as 'Arakati' in Goswami (1999), Chatterjee & Gupta (1981), and Behal (1985). The process described by Kar (1981) and Goswami (1999) suggests that Arakati were recruiting agents or contractors appointed by the planters. The planters paid them heavily to travel and procure immigrant labourers from the labour recruiting places. Motivated by large sums of money, they tempted tribal men and women to migrate to Assam by promising easy and stable life (Goswami, 1999). Under 'arakati' system, 400 labourers migrated to Assam between 1858 and 1859 and the number increased to 84,915 labourers in 1863 and to 18,076 labourers in 1873.

Significantly, Chatterjee & Gupta (1981) reveal, after recruiting the immigrant labourers, their transport to Assam was accomplished using inhuman methods. Evidences reveals, large number of people were shipped together in small boats without sufficient food or medical facilities. Consequently, out of 84,915 labourers shipped to Assam between 1863 and 1866, the death toll was about 30,000 people. Even after such high mortality rates, migration of immigrant tea labourers to Assam continued. Their procurement in large numbers through any means became the prime interest of the tea companies in Assam in the subsequent period.

However, since the death toll of tea labourers continued to increase the government tried to intervene through the Act III of 1863. As a result, a licensed system of recruiting indentured tea labourers was introduced, resulting in labour recruitment to be checked at certain points in Kolkata (Chatterjee & Gupta, 1981). The new system was termed Sardari system, but the only thing that had changed was the recruiting agent (Kar, 1981). Unlike the contractor system, where the recruiting agent was an outsider, in Sardari system the recruiting agent was an employee of the plantation industry, who had already migrated to Assam. These men and women, often referred as 'Sardar', were given extra privilege by the planters, such as seasonal bonus and gifts, to persuade them to travel back to their home states and bring their friends and relatives to work in the tea industry of Assam (Goswami, 1999). Motivated by special treatment of planters, the Sardars travelled to their homeland with good gifts and attractive descriptions about the benefits of working in the tea gardens of Assam. They also tried to entice their relatives by describing the beautiful demography of Assam. As a result, procurement of indentured tea labourers through the Sardars proved easier and effective, since now the tribes were tempted by the people they already knew. Hence, under the Sardari system it was easier for the Sardars to persuade the tribal people than it was at the time of the contractors. Consequently, the Sardars were also successful in bringing additional indentured tea

labourers to Assam in the subsequent period (Goswami, 1999).

However in 1951, recruitment of immigrant labourers was banned through the contractors as well as through the Sardars. The imposition was made since recruitment and migration of indentured tea labourers to Assam was not properly documented under both the labour recruitment systems. The planters did not maintain records regarding the number and identity of migrant labourers before their settlement in Assam. Hence, a third system of labour recruitment was introduced, whereby some labour recruiting organizations were formed according to the guidelines given by the government. These guidelines gave proper instructions for recruiting indentured tea labourers to Assam and unlike the earlier two systems, where the labourers were brought for an indefinite period of time, under the new system an agreement was made between the employer and the employee to work for a duration of 5 to 6 years. Under this system, there was a reduction in the number of immigrant labourers between 1960 and 1961 (Kar, 1981). In summary, the facts described so far are relevant for all the indentured tea labourers in Assam. However, since the focus of this dissertation is particularly the migration of Soras, the following section describes some of the personal narratives Sora speakers regarding their experience of migrating from Orissa to Assam along with other indentured tea labourers.

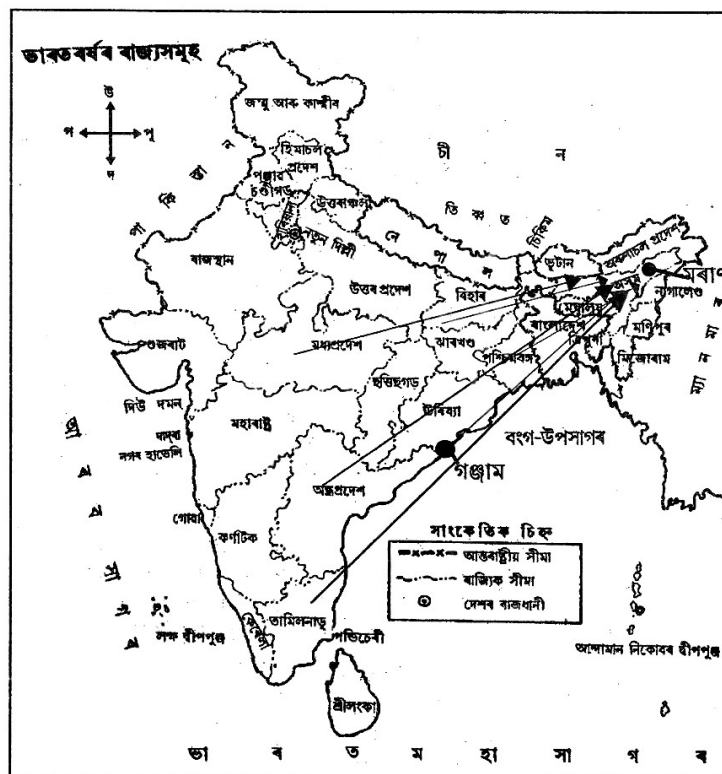
#### **1.3.4 Migration of Soras to Assam**

Sora speakers are an integral part of the indentured tea labourer's community in Assam. According to Kar (1981), Soras are mainly found in the Dibrugarh district of Assam. However, this work has found that currently Soras are living in at least 10 different districts of Assam (Figure 1.1). In these districts the Soras are mostly concentrated within a tea garden periphery or are inhabiting in areas adjacent to a tea garden. Regarding the migration of Soras, Kar (1981) has shown that Soras in Assam originated from the Ganjam district in Orissa and the same origin of their

migration is shown by an Assamese Scholar Duarah (2005) (see Figure 1.5).

Figure 1.5: Illustration of Sora migration (Duarah, 2005)

দেৱ দূৰবাৰ - চাঁওৰা ভাষাৰ আভাস




Migration of Soras from Orissa to Assam is confirmed through the emigration certificate of a Sora migrant living in Koilamari tea estate of Lakhimpur district in Assam. It is evident from the emigration certificate that, the person had migrated from Ganjam district in Orissa to Koilamari in Assam in 1941 (see Figure 1.6)

. The emigration certificate states that the person migrated to Assam with his family particularly for labouring in the Koilamari tea estate. An Emigration certificate of most other Soras in Assam is a rare thing to find. Many of them have either lost their emigration certificate or the first generation migrants had never passed it on to the following generations. However, from field studies it is found that Soras in Assam acknowledge the fact that their forefathers once lived in Orissa and that they migrated to Assam mainly for labouring in the tea gardens of Assam.

Figure 1.6: A sample of the Certificate of Emigration issued for tea labourers emigrating to Assam

FORM B. 4 Rs.

**ASSAM TEA DISTRICTS**



**EMIGRANT LABOUR CESS**



**चार रुपया**  
**CERTIFICATE OF EMIGRATION**

**Part I.** Kollamari T.B. Serial number **161604** Valid for the year ending 30th September 1941

**Part II.** Issued to THE TEA DISTRICTS LABOUR ASSOCIATION.

*Personal particulars of emigrant.*

Name Tola. Father's Name Sumbra.  
 Age 36 Sex M Caste SOWRA.  
 Village Barangsingi. Post Office Serango.  
 Police station Serango. District Ganjam.  
 Name of recruiter (if any) Ramana G.73.

Particulars of height and distinguishing marks.	Right and left thumb impressions.
Height:- <u>4'-10½"</u> One mole on the back & inner side of left calf muscle	 

Members of family \* accompanying emigrant :-

Name.	Age.	Sex.	Relationship.
Gosadi.	32	F	Wife.
Lakia.	11	M	Son.
Sansae.	5	F	Daughter.
Guduni.	1	F	"

Serial number of way bill or list of certificates 123.

Place of making entries Berhampore (Ganjam) Local Forwarding Agency.

Date 7-4-41. (Sd.) [Signature]  
 Designation Local Forwarding Agent, T.D.L. Association.

Factors leading to the migration of Soras to Assam are typically similar to the factors that led other indentured tea labours to migrate to Assam. As it has been

seen in the case of other indentured tea labourers, many Soras migrated to Assam primarily due to economic insufficiency in their homeland. At the same time, many others migrated as a result of intense alluring by return migrants who were sent by the planters with the sole purpose of bringing the same tribe in larger numbers. Kar (1981) presents case studies of five Sora tea garden workers in Mancotta tea estate of Assam in Dibrugarh District, describing the reasons for their migration. Three respondents reported that they migrated to Assam mainly because they were not able to support their family; and their relatives in Assam had told them that there were better opportunities for livelihood in Assam. The other two respondents reported that they had migrated at an early age with a view that survival was easier in Assam than it was in their homelands.

Case studies of three migrant Soras in Assam are recorded in this dissertation. Although it is very difficult to find a first generation Sora migrant in Assam at this point of time, it was possible to record one first generation migrant's personal migration history. The other case studies are obtained from second generation migrants.

Figure 1.7: *Phulmai Sobor* A Sora migrant in Assam



**Case study 1:** Phulmai Sobor (60 years female, Figure 1.7) is a first generation Sora migrant living in Koilamari tea estate of Lakhimpur district in Assam. She

started living in Koilamari tea estate at the age of 20 years and her place of origin was a village in Orissa named 'Rupe' in Ganjam district. Her father had already migrated to Koilamari through 'Chalan' when she was very young and so she grew up with her mother and her grandparents. When her mother died suddenly her grandparents did not have enough money to support her so they married her to somebody in the village. Soon after her marriage, her husband left her in Orissa and migrated to Arunachal Pradesh. Later, when her husband did not return for a long time, her father-in-law started abusing her and therefore she ran away from her in-laws house at night through some unknown forest terrains. Following this incident, she tried to find out how she could go live with her father in Assam. At that time she came to know about an agent, she refers to as Sardar, who told her that he could take her to her father in Assam. Eventually, the Sardar helped her to migrate from Orissa to Assam, even paying for her travel and food. She traveled to Assam by train but does not remember the traveling route. However unlike she was promised by the Sardar, she was first taken to Arunachal Pradesh to work as a road construction labourer. Later, she was encouraged by some fellow workers to go to Koilamari and after some days she eventually met her father. She mentioned that after a few years she also met her first husband, but by then she had remarried in Koilamari and had a family of her own. Phulmai Sobor remained a monolingual Sora speaker for several years and even when this elicitation was done she was not able to speak fluently any language other than Sora.

**Case studies 2 and 3:** Monoj Sora (55 years male) and Bijay Sora (60 years male) migrated with their parents at a very early age, they do not remember where they came from or where they lived when they first migrated to Assam. When they were teenagers, a contractor brought them to Singrijhan to clear forest area in order to establish a new tea garden there. After the forest was cleared, a British officer came and trained them to plant tea plants, and also carry out the processing of tea leaves.

Eventually, some thatched houses were built and they started living in the newly established tea garden, in Singrihan. Based on what they learned from their parents, Monoj Sora and Bijay Sora provided a narrative about the history of their migration to Assam. They told that the British officers often sent contractors from Assam to Orissa, referred to as 'Telenga', to bring the Soras to Assam. The contractors allured their parents and kinsmen by telling them about abundant work and money in Assam. They told them they could travel to Assam without paying anything. Some contractors even took with them seeds of tea plant and enticed their parents and kinsmen to come to Assam by saying things such as, "This is one such seed that will yield a tree that can sustain your family for many generations". Also, Monoj Sora and Bijay Sora mention that before their family migrated to Assam they were facing several challenges in their homeland. The temperature was very high and people in their village lived in wet lands to keep themselves cold. Additionally, the land was not fertile and the only agricultural product that would grow was *bassia latifolia* also known as madhuca or mahua. Hence, being distressed by the challenges at home and being persuaded by the contractors repeatedly, Monoj Sora's and Bijay Sora's family along with many other Soras of their village started migrating to Assam. Moreover, since they were not allowed to migrate as singles many of their kinsmen even faked their wedding for migrating to Assam. After migrating to Assam, Monoj Sora's and Bijay Sora's family were never allowed to return to their homelands. Moreover, they could not return since they neither had enough money, nor they knew how they could return to their homeland. Also, since they were not able to speak any local language, the situation was even more challenging. Thus, these facts narrated by Monoj Sora and Bijay Sora provide evidence about how the families of these two Sora men migrated from Orissa and permanently settled in Assam.

Hence, it is ascertained that groups of Sora people migrated from Orissa to Assam mainly for labouring in the tea gardens. The case studies also reveal that very few

Soras returned to Orissa, but settled permanently in Assam and are engaged in the same occupation for several years for which they had originally migrated. Under these circumstances, Assam Sora speakers became permanently separated from their land of origin and started living in a completely different sociolinguistic environment. As a result, it can be hypothesized that their language was also affected due to their migration from Orissa to Assam. However, it is found, in the post migration period a dominating attitude of the tea planters largely contributed towards the preservation of language of the indentured tea labourers in Assam including the Soras. After their migration to Assam, the indentured tea labourers were completely under the mercy of the tea planters. Evidences suggest, the tea planters dealt authoritatively with the indentured tea labourers and working conditions in the tea garden were extremely harsh. Thus, although the people migrated with an aim to improve their livelihood, they eventually ended up being in another difficult situation. Also, there was a significant change in the demography of the migrant labourers and their social setup was severely affected. However, it is significant that the people were able to retain their ethnicity and linguistic identities. In this regard, historical reports indicate that subjugation and isolation of the indentured tea labourers in the post migration period eventually helped them to retain their linguistic identities. Hence, in order to understand how the indentured tea labourers in Assam managed to retain their language even after being detached from their place of origin and settling in a new sociolinguistic environment, the following section discusses the situation of indentured tea labourers in the post migration period.

### **1.3.5 Subjugation and isolation of indentured tea labourers in Assam**

Indentured tea labourers in Assam include different linguistic communities. By examining the data from Registrar General of India (2001) it can be postulated that

speakers of 19 different languages belonging to three different language families (see 1.1) constitute the immigrant tea labourers in Assam. Additionally, when the migrant labourers arrived in Assam speakers of different Indo Aryan and Tibeto Burman languages were already residing in the state. Thus, migration of indentured tea labourers resulted in the creation of a multilingual society in the tea gardens of Assam. Consequently, the context was very suitable for language contact and language change.

Table 1.1: Language wise population of tea labourers in Assam (Registrar General of India, 2001)

	Family	Language	India	Assam
1.	Austroasiatic	Santali	6,469,600	242,886
2.	Austroasiatic	Munda/Mundari	469,357/1,061,352	93,088/32,718
3.	Austroasiatic	Kharia	239,608	6,108
4.	Austroasiatic	Bhumij	47,443	2113
5.	Austroasiatic	Savara	252,519	406
6.	Austroasiatic	Koda/Kora	43,030	308
7.	Austroasiatic	Korwa	34,586	90
8.	Austroasiatic	Gadaba	26,262	7
9.	Austroasiatic	Korku	574,481	6
10.	Dravidian	Kurux/Kurukh	1,751,489	72,311
11.	Dravidian	Gondi	2,713,790	8,468
12.	Dravidian	Kisan	141,088	1,475
13.	Dravidian	Parji	51,216	601
14.	Dravidian	Khond/Kondh	118,597	425
15.	Dravidian	Malto	224,926	328
16.	Dravidian	Konda	56,262	270
17.	Dravidian	Koya	362,070	146
18.	Dravidian	Kui	916,222	29
19.	Indo Aryan	Bhili	9,582,957	212

However, it is significant that these migrant languages are still preserved in Assam to a considerable extent. In this regard, review of historical reports helps in understanding the factors that contributed towards language preservation. It is found that, subjugation and isolation of indentured tea labourers largely contributed in preserving their linguistic identities. There were no provisions for language contact and solidarity of migrant labourers eventually helped them in preserving their languages for several generations. After their arrival in Assam the migrant labourers were essentially over-

powered by the tea planters through authoritative treatment. Also, the indenture legislation provided the planters legitimacy over the migrant labourers making the labourers become dependent on the planters for all their basic needs. Studies such as Behal (1985) and Behal (2014) have extensively reported the dominance of tea planters over the indentured labourers.

Evidences suggest, in order to retain the migrant labourers, the planters deliberately isolated them and restricted their movement to within the tea garden periphery only. The labourers were compulsorily settled in small huts called 'Labour line' and watchmen were inducted to prevent their movement beyond the tea garden territory. As a result, daily life of the migrant labourers was confined within the tea garden area. Also, the tea gardens were located in isolated places far from urban settlements and guarded with metal wires and watchmen to prevent anyone from travelling through the tea garden. Thus, communities outside the tea garden were restricted from making contact with the migrant labourers.

Subsequently, the planters started extracting maximum work from the migrant labourers since they had invested in their recruitment and transportation. Working hours of the migrant labourers were fixed and they were paid only a minimum wage. The labourers were obligated to do the tea garden work and if the workers did not comply the planters would discipline them through violent means. Failing to maintain the work routine made the labourers liable to be punished and penalized severely. Hence, life of migrant labourers in the tea gardens of Assam was very difficult.

Additionally, the tea labourers' individual identities were also effected by the authoritative behaviour of the tea planters. According to Behal (2014), the migrant labourers were often treated as bonded slaves. The local non-labouring tea planters community, started considering the Labour Lines as socially and culturally taboo places. Consequently, the labourers lost their self-esteem and started viewing themselves as inferior human beings in comparison to the tea garden owners. Thus, the

migrant labours suffered severely in the tea gardens of Assam and it became impossible for them to escape due to strict surveillance. Moreover, the labourers could not abscond from the tea gardens because they did not have enough money to travel and they did not speak any language other than their mother tongues. Also, they were kept illiterate for many years and many of them remained as illiterate monolingual speakers for several generations.

Thus several factors motivated the indentured tea labourers, including the Soras, to migrate to Assam. Many people were unlawfully driven to Assam for several years in the nineteenth century. It is also revealed that although the migration of indentured to Assam was contractual, very few people returned to their homelands and a large majority of the people, including the Soras, settled permanently in Assam. Also, it is evident that after the migrants started working in the tea gardens of Assam, the tea planters exploited the workers and subjected them to inhumane working conditions. Moreover, reviewing the historical accounts reveal that migration of the indentured tea labourer to Assam placed them in a new demography, in a confined territory. Thus, the people were not only detached from their familiar environments they were also unwillingly exposed to new industrial and social setups. Over the years, the indentured tea laborers have lived within the same confined tea garden territory and have remained isolated from other communities in the vicinity for a long time. These factors have largely contributed towards the preservation of the languages of the indentured labourers including the Soras. Moreover, initially although the Soras were mixed with other groups, when new indentured labourers stopped migrating, they were reshuffled and grouped on the basis of their first language. Therefore, Soras did not get enough opportunity to interact with speakers of other migrated groups for a long time and some Sora generations remained monolinguals for several years. Thus, although the Soras were completely separated from their place of origin and settled in an entirely new location, their subjugation and isolation contributed towards the

preservation of their language. Therefore, it is likely that there will be less differences between the the Sora variety of Assam and the Sora variety of Orissa. However, prior to examining the similarities and dissimilarities between Assam Sora and orissa Sora, it is important to review a few other migrated languages of the world. This will help in understanding how people's migration generally effects their language. Therefore, the following section reviews the studies which deal with population movement and languages.

## 1.4 Population movement and languages

Migration studies mainly focus on socio-political, economic and anthropological perspectives of population movement or people's migration. It is evident from such studies that, migration of people affects the migrating community as well the recipient community. The migrating community is affected since it is uprooted from its familiar social setup and the recipient community is affected since it has to relate to the new community within its existing social setups. Additionally, it is known that, when a group of people migrate, their language also migrates from one place to another. As a result, the language is separated from its place of origin and is exposed to a new sociolinguistic environment. Consequently, in some cases the migrant languages emerge as a new language variety in the location where the people have newly migrated. Such languages are commonly known as 'Transplanted language' and transplanted languages are usually formed when speakers of a language variety migrate from one place to another due to various reasons. Thus, overseas Hindi and overseas Bhojpuri varieties spoken in countries such as Fiji, Mauritius and South Africa are examples of some transplanted languages that emerged due the migration of Indian immigrant labourers to these countries in the 19<sup>th</sup> century (Bhatia, 1981). Thus, it is evident that in addition to effecting the social lives of the people migration

also effects their languages. Therefore, it is important to examine languages that migrate from one place to another due to the migration of its speakers.

In this regard, studies such as Bhatia (1981), Domingue (1981), Avram (2014) and Siegel (1998) have largely focused of language migration and they propose that migration of people severely effects the language of the migrating community. Evidences suggest that transplanted languages are usually distinct from the original language and research has shown that transplanted languages undergo language change after migrating from its place of origin. Also, language contact and dialect contact are identified as two important factors of language change in transplanted languages (Kerswill, 2006). Typically, language contact occurs when the migrating community comes into contact with languages of the recipient community and dialect contact occurs in cases where the migrating communities are speakers of different dialects of the same language (Trudgill, 1986). Linguistic and sociolinguistic processes of language change related to language contact including sound change, borrowing, pidginisation, creolisation, multilingualism, second language acquisition and language drift are mostly observed in transplanted languages. Also, linguistic and sociolinguistic processes of language change related to dialect contact including second dialect acquisition, accommodation, mixing, simplification, leveling, hyperdialectalism and reallocation are also found in transplanted languages. For these reasons, studies have proposed that transplanted languages are usually distinct from the original language.

However, language maintenance or language change in transplanted language depend on the nature of the migrating community (Kerswill, 2006). It is found that although language change due to language contact is inevitable in transplanted languages, language contact can be prevented if the migrating community moves to an isolated location (Kerswill, 2006). Similarly, a migrating community can prevent language contact by maintaining an isolating attitude in the new location (Trudgill, 2004). Likewise, if the recipient community systematically segregates the migrating

community, language contact will not take place between the migrating community and the recipient community. Therefore, these conditions indicate that linguistic and sociolinguistic changes, due to language and dialect contact in transplanted languages, largely depends on the contact relationship between the host community and migrating community.

Thus, studies on language migration confirms that people's migration also causes migration of languages and a migrant languages are generally known as transplanted languages. Also, it is found that transplanted languages tend to change and become distinct from the original language. However, studies have also shown that some sociological factors can help transplanted languages to avoid language change. Hence, in order to observe the nature of language change in transplanted languages the following section discusses the linguistic characteristics of transplanted languages. Also, in order to understand the forces that can prevent language change in transplanted languages, the subsequent section discusses the sociology of language migration.

#### **1.4.1 Linguistic characteristics of transplanted languages**

Studies such as Avram (2014), Mufwene (2007), Lee (2014) and Pattanayak (1985) have shown that language migration causes language contact sociolinguistic processes such as language death, pidginization, creolization, and second language acquisition in transplanted languages. Additionally, studies such as Siegel (1998) and Mesthrie (1995) propose that linguistic changes such as lexical change, morphological change and phonological change are commonly found in transplanted languages. Thus, evidences suggest that migration of language communities causes language contact and language contact makes transplanted languages typically distinct from the original language. Therefore, Kerswill (2006) proposes that migration of language communities is an important extra-linguistic factor of language change. Thereby, transplanted languages such as overseas Hindi and overseas Bhojpuri spoken in countries such as

Fiji, Mauritius and South Africa provide evidences for language change caused due to language migration. Evidences suggest that vocabulary relating to Indian caste system and religious practices are not found in transplanted Fijian Hindi and South African Bhojpuri since the migrating community have abandoned the practices. Similarly, loss of gender distinction in the verb morphology, loss of honorific suffix in the pronominal forms and simplification of number suffixes in the pronominal forms are some morphological changes reported in transplanted Mauritian Bhojpuri, Fijian Hindi and South African Bhojpuri. Likewise, Kerswill (2002) provides evidence of phonological change in transplanted Norwegian dialects of Odda and Tyssedal including replacement of the vowels [ɐ] and [ɛ:] to [ɔ] and [ɔ:] as in the word for 'floor' where /gɛlv/ is changed to /gɔlv/ and the word for 'been' where /vɛ:rə/ is changed to /vɔ:rə/; loss of morphophonemic velar palatal alternation as in the word for 'root' where /ta:k/ did not alter to /ta:çə/ even after addition of the definite suffix /ə/; and vowel shortening as in the word /çot/ which changed to /çot/. Thus, studies on these transplanted languages suggest that linguistic changes are found in transplanted languages causing transplanted languages to be different from the original language.

Moreover, beside the evidences mentioned above, review of literature on transplanted languages reveal that there is an additional linguistic phenomena generally related to transplanted languages. The process is called 'Koineisation' and a language emerging out of Koineisation is called Koine (Kerswill, 2002). Typically Koineisation occurs when people from different parts of a single language area migrate and start living together in a new settlement. Therefore, Ramyeard (1988) as cited in Barz & Siegel (1988) argues that Koineisation is a general phenomenon applicable to all transplanted languages. Koineisation has been defined by different scholars in slightly different ways, but the main idea about Koineisation is same in all the descriptions. For instance, according to Gambhir (1981) as cited in Siegel (1988) the process of Koineisation includes elimination of idiosyncratic features of the dialects in contact

and preservation of the common shared features. Similarly, Trudgill (1986) defines Koineisation as a process of mixing different dialects followed by leveling of common shared features and elimination of universally marked features. Likewise, Siegel (1985), Kerswill (2002) and Kerswill (2006) also define Koineisation as a process of simplification and stabilization of transplanted languages. Moreover, in addition to these definitions Mesthrie (1995) further divides the process Koineisation into three stages as described below:

1. Stage I **Variation**: Presence of idiosyncratic forms in the speech of elderly speakers
2. Stage II **Simplification**: Reduction of complex forms and increase in regularity in certain paradigms
3. Stage III **Deletion**: Deletion of marked forms in the contacting languages

Koineisation is a linguistic process that explains contact induced language change particularly in transplanted languages. Koineisation basically proposes that transplanted languages are different from the original language since they are less complex and have linguistic characteristics that are universally unmarked. Also, Koineisation can occur at every structural level of a language including lexical, phonemic, morphemic as well as syntactic. Consequently, it is proposed that transplanted languages become distinct from the original language at every structural level. Therefore, scholars have extensively discussed Koineisation in the description of transplanted languages.

Hence, the linguistic characteristics of transplanted languages provide evidence that migration of people also effects their languages. Also, it is evident that migrating languages emerge as transplanted languages in the location where the people have newly migrated. Additionally, in some cases it is observed that transplanted languages have a tendency to change and become distinct from the original language.

However, it cannot be denied that language change largely depends on different social factors surrounding the migrating community. Therefore, in order to understand how different social factors effect the languages of migrating community the following section discusses the sociology of transplanted languages.

### **1.4.2 Sociology of transplanted languages**

Studies such as Gambhir (1981), Barz & Siegel (1988), Siegel (1985, 1998), Mesthrie (1995), and Kerswill (2002, 2006) have extensively described the typical characteristics of transplanted languages and have affirmed that migrant languages develop as transplanted languages. Evidences suggest, depending on the nature of the community involved, there can be two types of transplanted languages:

1. Transplanted colonial languages: American English, Canadian French, Afrikaans and Icelandic (Reinecke, 1969) as cited in (Barz & Siegel, 1988)
2. Transplanted immigrant languages: Overseas Hindi and Bhojpuri varieties in Fiji, South Africa and Mauritius (Barz & Siegel, 1988)

It is found that transplanted colonial languages are easy to maintain in new settlements since a large number of people speak those languages. Therefore, speakers of transplanted colonial languages can easily retain their languages even after being detached from their place of origin. On the other hand, transplanted immigrant languages are usually spoken by socially marginalized people and are therefore difficult to maintain. Such languages often do not survive in the new settlement as can be seen in case of African migrants in North America. However, there are some other transplanted immigrant languages that have survived in new environments. A significant example of such transplanted immigrant language are the overseas Hindi and Bhojpuri varieties spoken in countries like Fiji, South Africa, and Mauritius (Barz & Siegel, 1988). It is evident that transplanted overseas Hindi varieties are still spoken

in these countries despite the migrating community being a marginalized and completely isolated from its place of origin. In this regard, studies suggest that linguistic composition of the migrating community contributes toward the survival of transplanted overseas Hindi varieties in different countries. For instance, Barz & Siegel (1988); Siegel (1998) suggest that Fijian Hindi survived in the new settlement mainly because all the Indian immigrants in Fiji spoke at least a variety of Hindi. Therefore, they did not have to acquire a new language and abandon their mother tongues. Similarly, Ramyead (1988) suggests that Mauritian Bhojpuri survived in the new environment mainly because all the Indian immigrants in Mauritius belonged to the same language community. In contrast, (Mesthrie, 1995) argues South African Bhojpuri survived with some difficulty since the Indian immigrants in South Africa spoke both Indo-Aryan and Dravidian languages. Thus, South African Bhojpuri survived after speakers of Dravidian languages lost their languages. Additionally, Fijian Hindi received government support much earlier than Mauritian and South African Bhojpuri which also helped in standardizing the language earlier than Mauritian Bhojpuri and South African Bhojpuri. Thus, the two basic categories of transplanted languages suggest that survival of transplanted languages primarily depends on the social status of the migrating community as well as on their linguistic composition.

Additionally, after transplanted languages survive in a new location language change through language contact is often observed in those languages. However, research has shown that language change in transplanted languages largely depends on the contact relationship between the migrating community and the host community. In this regard, Coleman (1997) as cited in Kerswill (2006) suggests, migrants usually have two types of relationship with the host community namely participation or segregation. According to Coleman (1997) while a participating group 'attempts to maximize social mobility', a segregating group 'tries to minimize contact with the outside world'. Hence, Kerswill (2006) argues that it is more likely for a participat-

ing group to assimilate with the host community over time. Whereas, a segregating attitude will lead the group to maintain their own language. Consequently, a participating group will emerge as a multilingual community and there will be an active sharing of linguistic features between the communities in contact. On the other hand, segregation of the migrant community will prevent assimilation with the host community and there will be no sharing of linguistic features between the communities in contact. Thus, linguistic processes related to language contact are more likely to be found in transplanted languages of a participating community and less likely in a segregating community. Hence, there are evidences that migration of people also causes migration of languages and migrant languages emerge as transplanted languages. However, depending on the nature of the migrating community and their contact relationship, a transplanted language can be distinct from the original language or it can successfully preserve the features of the original language.

Thus, in the review of literature on transplanted languages it has been observed that when speakers of a particular language migrate, their languages are also transplanted. Thereafter, the survival and development of such languages are associated with various social conditions. It was also observed that in the post migration period transplanted languages have a general tendency to stabilize through a process of simplification known as Koineisation, whereby idiosyncratic features of transplanted languages are eliminated. Hence, considering the fact that Assam Sora speakers are a migrated community this research counts Assam Sora as a transplanted variety of Sora and it is hypothesized that the general tendency to stabilize a transplanted language is also applicable for Assam Sora. At the same time, considering the migration history of the Soras in Assam, it is assumed that their subjugation and isolation within the tea gardens may have helped them preserve their linguistic peculiarities. Therefore, this dissertation is particularly interested in examining Assam Sora in the light of language migration.

## 1.5 Research objectives

The first objective of this dissertation is to provide a detailed description of the distinctive speech sounds in Assam Sora. This includes identification of the distinctive phonemes in Assam Sora and a detailed description of their distribution in the language. Acoustic phonetic analysis is provided based on speech data collected from native Assam Sora speakers living in four regions in Assam and one region in Orissa.

Secondly, Assam Sora spoken in four different geographical locations in Assam is compared to discover any linguistic variation arising due to geographical difference or due to the difference in their place of origin.

Thirdly, from the review of transplanted languages it is observed that when languages are transplanted they undergo language change in the new environment by being in contact with other languages. Descriptions of transplanted languages have a clear evidence for such changes. Thus, this work also investigates how Assam Sora fits the characteristics described for transplanted languages.

Fourthly, since previous phonetic descriptions of Orissa Sora are also highly inconsistent, the objective of the present work is to investigate the phonetic properties of Orissa Sora. Subsequently, the relation between Assam Sora and Orissa Sora, with the help of synchronic speech data, is explored.

## 1.6 Methodology

The description of Assam Sora speech sounds is based on speech data collected from native Assam Sora speakers through field studies. Also, field studies were conducted in Orissa, in order to collect synchronic Orissa Sora speech data. The field studies were initiated by identifying the major concentrations of Assam Sora speakers in Assam and Orissa Sora speakers in Orissa. This is followed by identifying native Assam Sora and Orissa Sora speakers in the specific locations and recording their

speech while eliciting the data sets. Data sets used for eliciting and recording Assam Sora and Orissa Sora speakers are compiled by reviewing some relevant literature on the language. Subsequently, the digital speech data are subjected to acoustic analysis of different speech sound categories, and interpretation of the data is represented with the help of different visualizations. Also, in order to confirm the evidences, the data is subjected to statistical analysis. These methods used in this research are elaborated in the subsequent sections.

### **1.6.1 Areas of field study**

Data used in this dissertation consists of speech data collected from native Sora speakers living in Assam and Orissa. Field studies were conducted in four villages of Assam and in one village of Orissa. The four villages in Assam include Singrijhan and Sessa tea estate under the Rangapara state assembly constituency in Sonitpur district, Lamabari tea estate under the Majbat state assembly constituency in Udalguri district and Koilamari tea estate under the Lakhimpur state assembly constituency in Lakhimpur district. Figure 1.8 shows the map of Assam, highlighting Sora villages in light grey and highlighting areas of data collection in dark green.

The map of Assam is created by marking the 126 assembly constituencies of Assam. There are approximately 1000 Soras in Singrijhan, 550 Soras in Sessa, 800 Soras in Lamabari and 300 Soras in Koilamari. These are the major Sora speaking areas in the three districts of Assam. It was found that Sora speakers in all the four villages originated from Ganjam district in Orissa and lived in the these villages for more than 100 years. In Orissa, the field study was conducted in Rayagada district of southern Orissa that is adjacent to parts of Vizianagaram and Srikakulam district of Andhra Pradesh. Initially few interviews were done in a village named Marichaguda under Padmapur block but, speech data was recorded in a village named Raiguda under Jagannathpur Gaon Panchayat of Gunupur Block in Rayagada district.

Figure 1.8: Sora villages in Assam

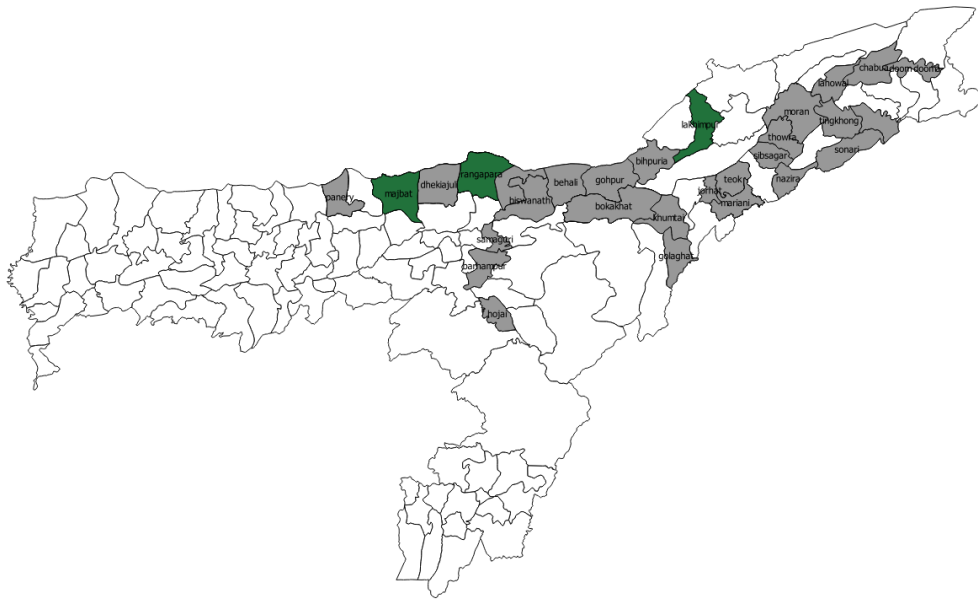


Figure 1.9: Sora village in Orissa



Figure 1.9 shows the map of Orissa highlighting a Sora village in light grey and highlighting the area of data collection in dark green. The map of Orissa is also created by marking the 147 assembly constituencies of Orissa. Raiguda is a Sora dominated village and there are approximately 85 houses in the village. The area was selected in consultation with the Assam Sora informants and one Assam Sora

informant also participated in the study as a Sora interpreter in Orissa.

### **1.6.2 Composition of the data**

Speech data recorded during the field studies include data sets that are generated from various data sources. However, while Orissa Sora has a good number of data sources, the only Assam Sora data source available is Horo & Sarmah (2015) that was developed from the basic Swadesh wordlist. Hence, primary data sources used in this dissertation are drawn from existing data sources on Orissa Sora particularly from Ramamurti (1938) and Anderson & Harrison (2011). Thus, the data set compiled from all three sources mainly consists of a vowel minimal set compiled from Horo & Sarmah (2015); differently transcribed vowels of Orissa Sora compiled from Anderson & Harrison (2011); data for consonantal minimal sets compiled from Ramamurti (1938); word initial oral stop consonants compiled from Ramamurti (1938); and derived and non derived disyllabic words compiled from Ramamurti (1938). Taking all these data together, a total of 1010 lexical items are recorded during field studies in every location and the compiled data set is presented in Appendix B of this dissertation.

### **1.6.3 Data recording and annotation**

In the field studies that were conducted in Assam, interviews were conducted in ‘Sadri’ (an Indo Aryan lingua franca) and the speakers were asked to produce targeted Sora words from the data set mentioned above. All the participants were recorded saying every word once in isolation and once in the sentence frame ‘*jen ---- gamlai*’ translated as ‘I ---- said’. On the other hand in Orissa, the Assam Sora informant helped in interpreting the conversation between the interviewer and interviewee but the data elicitation procedure was kept same. The speakers were recorded with a Shure unidirectional head-worn microphone connected to a Tascam linear PCM recorder via xlr

jack. The sampling frequency was 44.1 kHz, 24 bit in WAV format. Subsequently, digital speech data has been manually annotated in Praat (Boersma & Weenink, 1992) for (a) word boundary (b) syllable boundary and (c) phoneme boundaries. While vowel phonemes are marked between steady state formants, word initial oral stop consonants are marked between release of the stop consonant and onset of the glottal pulse of the following vowel.

#### **1.6.4 Participants**

Ten native Sora speakers were interviewed in every village mentioned above and the ten speakers in each village included 5 male and 5 female speakers between 20 to 60 years of age. All speakers recorded for this dissertation are either bilinguals or trilinguals. While the Assam Sora speakers could speak Indo Aryan languages such as Assamese and/or Sadri, Orissa Sora speakers could speak only an Indo Aryan language namely Oriya or could also speak a Dravidian language, namely, Telugu. Table A.1 and A.2 in Appendix lists all the Sora speakers who participated in the field study undertaken for this dissertation.

#### **1.6.5 Acoustic and temporal measurements**

Acoustic analysis of vowels is based on formant frequencies of the first two formants (F1 and F2). Formant frequencies for F1 and F2 are extracted at vowel midpoints for steady state formants using a script for Praat and Hertz values for F1 and F2 are converted to Mel using an inbuilt function in Praat. Analysis of word initial oral stop consonants is based on the VOT (Voice Onset Time) values of the stop consonants that is calculated by measuring the duration between the release of the oral stop consonant and the onset of the glottal pulse of the following vowel. Acoustic analysis of nasal consonants is based on nasal formants and anti-formants. For this purpose, spectrum is extracted at the middle of the total nasal duration by transforming the nasal speech

data into spectrogram and calculating the power of the spectrum at the middle of the nasal duration. Finally, word stress in disyllabic words is analyzed by measuring fundamental frequency, vowel duration and vowel intensity in first and second. While fundamental frequency is calculated from the average and maximum fundamental frequency of the vowels between beginning and end of glottal pulses, vowel duration is calculated by measuring the duration between the beginning and end of steady state vowel formants and vowel intensity is calculated for acoustic energy of the vowels between same steady states vowel formants. Methods of data representation and their interpretation in this dissertation are explained in the following section.

### **1.6.6 Data representation and statistical measurements**

Numerical data generated from the recorded speech data has been graphically represented in this dissertation for different purposes. Firstly, formant frequencies of Sora vowels are plotted across F1 and F2 axis using scatter plots, Lobanov normalized plots and vowel space plots. Each of these plots have helped in interpreting Sora vowel phonemes categorically and have also helped in visualizing the acoustic space of vowel phonemes in Assam Sora and Orissa Sora. Secondly, bar diagrams are plotted for graphically representing the temporal measures of VOT, vowel duration, fundamental frequency and vowel intensity to visualize the consistency of these measures in Assam Sora as well as in Orissa Sora. For nasal sounds, spectral plots are used to represent nasal formants and anti-formants observed in nasal consonants. Additionally, considering the fact that data are liable to have faulty representations, kernel density plots are also presented mainly for showing the difference in syllable prominence. Finally, in order to confirm the evidences, statistical measurements such as one way ANOVA test, euclidean distance measurement and bonferroni test are also run for the numerical data used in this dissertation.

## 1.7 Organisation of the dissertation

This dissertation is divided into seven chapters in the manner described below.

- **Chapter 1** provides a linguistic and ethnolinguistic background of Sora. Followed by this, it describes the migration of Sora speakers from Orissa to Assam from a historical perspective. It also explains the linguistic and sociolinguistic consequences of language migration from a typological perspective. The research objectives and the methods adopted in this dissertation are also explained in this chapter.
- **Chapter 2** elaborates the segmental properties of Sora that includes identifying the contrastive phonemes in Assam Sora. Orissa Sora segmental properties are also reviewed in this chapter. While discussing the segmental features of Orissa Sora and Assam Sora, this chapter examines some of the typical segmental features of Sora that are considered to be core features of Munda languages of the Austroasiatic language family.
- **Chapter 3** discusses the syllable structure of Assam Sora. This includes identifying the basic syllable structure of Assam Sora and describing the phonotactic constraints. Also, the phonetic realization of word stress in Assam Sora is described in this chapter with the help of acoustic analysis.
- **Chapter 4** presents the acoustic properties of vowels in Assam Sora. This includes examining and interpreting Assam Sora vowel formants recorded from native Assam Sora speakers living in four major Sora villages in Assam. Moreover, acoustic analysis of Assam Sora vowels also includes examination of vowel dynamic features such as varying vowel formants and vowel space across syllables.
- **Chapter 5** presents the acoustic properties of two consonant categories in

Assam Sora namely stop consonants and nasal consonants. Stop consonants are divided into two types including six oral stops and one glottal stop. Analysis of oral stops is based on VOT of the six stop consonants and analysis of glottal stop describes its varying phonetic realization in Assam Sora. For the nasal consonants, this chapter mainly classifies the physical properties of Assam Sora nasals in terms of their acoustic phonetic characteristics.

- **Chapter 6** makes a synchronic comparison of Assam Sora and Orissa Sora phonetic features. The comparison is based on acoustic analysis of vowels, word stress and stop consonants in the two Sora varieties.
- **Chapter 7** discusses the final conclusion of this research work. The discussion here highlights the major findings of the research and provides the justification. The discussion contextualizes the findings of this research with related languages of Assam Sora as well as with languages that have similar ethnolinguistic histories.

## Chapter 2

# Segmental properties of Orissa Sora and Assam Sora

### 2.1 Introduction

The segmental properties of Orissa Sora and Assam Sora are discussed in this chapter. While Orissa Sora segmental properties are discussed from the review of relevant Orissa Sora literature, description of Assam Sora segmental properties are based on the speech data collected from native Assam Sora speakers living in four villages of Assam, namely, Singrijhan, Sessa, Koilamari and Lamabari. Also, segmental properties of Austroasiatic and Munda languages are reviewed in this chapter in order to discuss the typological relation of Orissa Sora and Assam Sora with other Austroasiatic languages in general and with Munda languages in particular. In this regard, existing studies propose that Orissa Sora has preserved the typical Munda features. For instance, Stampe (1965) argues that while most Munda languages have acquired Indo-Aryan features such as retroflexion and small vowel system, these features are not found in Orissa Sora. However, although segmental properties of Orissa Sora have been studied in the past, different scholars have proposed different opinions. As a result, the basic inventory of the language is proposed differently in different studies. Hence, in order to highlight the commonalities and irregularities present in the existing descriptions, a review of Orissa Sora segmental properties is presented in this

chapter. On the other hand, Assam Sora segmental properties are not described so far. Therefore, this chapter provides an overview of the distinctive speech sounds of Assam Sora using primary data collected during this work. This, includes identifying the distinctive speech sounds and describing their phonemic distributions.

Additionally, studies on segmental properties of Austroasiatic and Munda languages argue that Munda languages have significantly drifted away from Austroasiatic languages. For instance, (Donegan, 1993) proposes that there are less similarities between Munda and other branches of Austroasiatic languages. However, the study by Anderson & Zide (2002) propose that, the drift is not as great as it has been argued and there are significant similarities between Munda and other Austroasiatic languages. Therefore, while reviewing the segmental properties of Austroasiatic and Munda languages, some of these disputing issues are also discussed in this chapter. Thus, prior to the description of Assam Sora segmental properties, this chapter reviews the literature on Austroasiatic, Munda and Orissa Sora segmental features in §2.2 and then an overview of Assam Sora segmental inventory is described in §2.3 and §2.4. Finally, the observations and findings of the chapter are concluded in §2.5.

## 2.2 Literature review

Segmental properties of Austroasiatic languages, Munda languages and Orissa Sora are discussed in this section. Descriptions here show the typological relation of Orissa Sora with Munda languages and the relation of Munda languages with Austroasiatic languages, on the basis of their segmental properties. Regarding Orissa Sora, Stampe (1965) proposes that the following segmental features are typical to Orissa Sora and these features account for Orissa Sora to be included in the typological core of Munda phonology.

1. (C)V(C) syllable structure.

2. Glotalization of stops.
3. Asymmetry between voiceless dental stop /t/ and voiced alveolar stop /d/.
4. Pronunciation of /r/ as a trill consonant.
5. Diphthongization of vowels before palatals.

Similarly, Diffloth & Zide (1992) identifies the following segmental features as common to Munda and Mon-Khmer branches of the Austroasiatic language family. Therefore, these segmental features account for Munda to be included in the typology of Austroasiatic language.

1. Non release of final stop.
2. Inclusion of a complete set of stops and nasals in root-final position in four places of articulation, namely, labial, alveolar, palatal and velar.
3. Deficiency of spirantization.
4. Loss of final /ʔ/ allowing open syllables.
5. Frequently occurring word initial consonant sequences

Thus, the studies by Stampe (1965) and Diffloth & Zide (1992) make it clear that Orissa Sora shares the segmental typology of Munda languages and Munda languages share the segmental typology of Austroasiatic languages. However, the segmental inventory of Orissa Sora, particularly the vowel inventory, is often considered to be unique among other Munda languages. Evidences suggest that, while most Munda languages have a smaller vowel system, Orissa Sora has a relatively larger vowel system. In this regard, the study by Donegan (1993) suggests that Orissa Sora is a preservative Munda language that has preserved the vowel system of Proto-Munda.

Therefore, she hypothesizes that the vowel system of Orissa Sora adequately resembles to the vowel system of other Austroasiatic languages such as Mon-Khmer, that normally have larger vowel systems. Thus, considering this hypothetical relation between Orissa Sora and Austroasiatic languages, it appears that Orissa Sora is a unique Munda language.

Similarly, evidences suggest that, beside the common features, there are some unique features in Munda and non Munda languages, such as Mon-Khmer, of the Austroasiatic language family. For instance, (Stampe, 1965) suggests that many Munda languages have segmental features such as consonant aspiration, five vowel system and distinction between retroflex and non-retroflex stop consonants. These segmental features are not common to the Mon-Khmer branch of the Austroasiatic language family. Likewise, the typical sesquisyllabic word prosody present in Mon-Khmer languages is lacking in Munda languages. In this regard, past studies have mostly related the unique features of Munda and Mon-Khmer languages with areal typologies. For this reason, Munda languages are related to the South Asian typology and Mon-Khmer languages are related to Southeast Asian typology (Jenny et al., 2014). Therefore, some segmental features of Munda and Mon-Khmer languages are treated as secondary developments. Thus, It is generally accepted that Austroasiatic languages have acquired areal features of the areas where they are spoken. In this regard, Donegan & Stampe (2002, 2004) suggest that, the features of Munda and Mon-Khmer languages are not only distinct from one another, but are structurally opposite in nature. Consequently, due to the hypothetical influence of areal typology, Munda and Mon-Khmer languages are treated as very diverged from one another.

Thus, although there are clear evidence regarding the typological relation between Orissa Sora and Munda languages and between Munda languages and Austroasiatic languages, it appears that some unique features in Orissa Sora and Munda languages. Also, such unique features result in hypothesizing Orissa Sora as distinct from other

Munda languages and considering the Munda languages as significantly deviated from other branches of Austroasiatic languages. Therefore, in order to elaborate on these typological relations the following subsections review some of the relevant literature that describe the segmental features of Austroasiatic languages, Munda languages and Orissa Sora.

### **2.2.1 Segmental properties of Austroasiatic languages**

Segmental properties of Austroasiatic languages discussed in this section are mainly reported from Jenny et al. (2014). A typological review of Austroasiatic segmental properties reveals that, while most languages have a similar inventory of speech sounds, there are variations in size and structure of inventories of individual languages. Variation in speech sounds is observable in the consonant inventory as well as in the vowel inventory. In the consonant inventory, evidences suggest that there are both small and large consonant inventories in Austroasiatic languages. For example, languages with small consonant inventories include Car Nicobarese and Khmer (17 consonants) and languages with large consonant inventories include languages such as Kharia (34 consonants), Bugan (36 consonants) and Sedand (39 consonants). With regard to vowels, languages with small vowel inventories include Kharia (5 vowels) and Khasi (5 vowels with length distinctions and 1 short vowel) and languages with large vowel inventories include languages such as Bugan (10 vowels), Car (10 vowels) and Chong (9 vowel with length distinctions). Additionally, vowel length is found contrasting in many Austroasiatic languages and the inventory of diphthongs is also reported to be large in some Austroasiatic languages.

Taking the consonant inventory first, it is found that size of consonantal inventory in Austroasiatic languages depends on the syllable position of the consonants. Also, syllable weight determines the size of consonantal inventory in Austroasiatic languages. It is reported that consonant inventory of Austroasiatic languages is largest

in the first onset position of the main (stressed) syllable (Jenny et al., 2014). For instance in Khasi, while only four oral stops [p,t,t̪,k] can occur in the coda position, all other oral stops can occur in the onset position. Similarly, in Khmer the consonant inventory is larger in the onset position of the major syllable than at the end of a word. Thus, phonemic distinction of consonants in these languages is better represented at the onset position and not at the end of the word. A comparative segmental inventory of Austroasiatic languages accumulated from Jenny et al. (2014) is presented in Table 2.1 and 2.2.



Table 2.1: Austroasiatic consonants (Jenny et al., 2014)Part-A

Khmer	p	b	t	d	c	k	m	n	ɲ	r	l	j	h	?	(f)	v	s	
Chong	p	b	t	d	c	k	m	n	ɲ	r	l	w	j	h	?	(f)	s	
	p <sup>h</sup>		t <sup>h</sup>		c <sup>h</sup>	k <sup>h</sup>								?				
Kammu	p	β	t	d	c	k	m	n	ɲ	r	l	w	j	h	?		s	
	p <sup>h</sup>		t <sup>h</sup>		c <sup>h</sup>	k <sup>h</sup>					<sup>?</sup> w	<sup>?</sup> j						
Car	p		t		c	k	m	n	ɲ	r	l	j	h	?	f	v	s	
										t								
Kharia	p	b	t	d	c	k	m	n	ɲ	r	l	w	j	h	(ʔ)	ʒ	g	s
	b <sup>β</sup>		t	d	c <sup>h</sup>	k <sup>h</sup>				(r)					ʒ <sup>β</sup>	g <sup>β</sup>		
			t <sup>h</sup>	d <sup>β</sup>						(r <sup>β</sup> )								
			t <sup>h</sup>	d <sup>β</sup>					(n)									
Khasi	p	b	t	d		k	m	n	ɲ	r	l	w	j	h	?		s	dʒ
	p <sup>h</sup>	b <sup>β</sup>	t <sup>h</sup>	d <sup>β</sup>		k <sup>h</sup>											ʒ	dʒ <sup>β</sup>
Pnar	p	b	t	d		k	m	n	ɲ	r	l	w	j	h	?		s	tʃ
	p <sup>h</sup>		ṭ	ḍ														dʒ
			t <sup>h</sup>			k <sup>h</sup>												

Table 2.2: Austroasiatic consonants (Jenny et al., 2014)Part-B

Katu	p	b	t	d	c	k	m	n	ɲ	ʋ	r	l	w	j	h	ʔ	ʒ	g	s						
	p <sup>h</sup>	β	t <sup>h</sup>	d <sup>h</sup>	k <sup>h</sup>	k <sup>h</sup>												f							
Danau	p	b	t	d	c	k	m	n	ɲ	ʋ	r	l	w	j	h	ʔ	ʒ	g	f	v	s	ç	x		
	p <sup>h</sup>	β	t <sup>h</sup>	d <sup>h</sup>	c <sup>h</sup>	k <sup>h</sup>													s <sup>h</sup>						
Mon	p	β	t	d	c	k	m	n	ɲ	ʋ	r	l	w	j	h	ʔ			s	ç					
	p <sup>h</sup>		t <sup>h</sup>		c <sup>h</sup>	k <sup>h</sup>	m̃	ñ	ɲ̃		l̃	l̃	w̃												
Bunong	p	β	t	d	c	k									h	ʔ		(f)	ç						
	p <sup>h</sup>		t <sup>h</sup>		c <sup>h</sup>	k <sup>h</sup>				r	l	w	j												
	m <sup>p</sup>		n <sup>t</sup>		n <sup>c</sup>	ɲ <sup>k</sup>				ʔ <sup>w</sup>															
Bugan	p	b	t	d	c	k	m	n	ɲ	ʋ	r	l	w	j	h	ʔ	g		f	s	ts	ç	x	ɣ	q
	p <sup>h</sup>		t <sup>h</sup>		k <sup>h</sup>	k <sup>h</sup>	z												θ		ts <sup>h</sup>	z			
Sedang	p	b	t	d	c	k	m	n	ɲ	ʋ	r	l	w	j	h	ʔ	ʒ	g	s						
	β		d <sup>h</sup>		ʔ <sup>m</sup>	ʔ <sup>n</sup>	ʔ <sup>n</sup>	ʔ <sup>n</sup>	ʔ <sup>n</sup>	ʔ <sup>ɲ</sup>	ʔ <sup>r</sup>	ʔ <sup>l</sup>	ʔ <sup>w</sup>	ʔ <sup>j</sup>					(f)						
					m̃	ñ	ɲ̃	ʋ̃	l̃	l̃	w̃	j̃							s						

According to Jenny et al. (2014), Austroasiatic languages usually have 20 to 30 consonants in the consonant inventory. However, from Table 2.1 and 2.2 it is observed that the number of commonly occurring consonants is fairly small. Commonly occurring consonants in most Austroasiatic languages include stop consonants such as /p, b, t, d, c, k, ʔ/; nasal consonants such as /m, n, ɲ, ŋ/; liquid consonants such as /r, l, j, w/ and fricative consonants such as /f, s, h/. Additionally, many Austroasiatic languages also include voiceless aspirated stops such as /p<sup>h</sup>, t<sup>h</sup>, c<sup>h</sup>, k<sup>h</sup>/ whereas, the voiced counterparts are found only in a few languages. Thus, regarding the phonemic distribution of the Austroasiatic consonant inventory, it can be postulated that commonly occurring stops in most Austroasiatic languages are produced at five places of articulation namely bilabial, dental/alveolar, palatal, velar and glottal. The common nasals are produced at four places of articulation namely bilabial, dental/alveolar, palatal and velar and the common fricatives are produced at three places of articulation namely, labiodental, alveolar/post alveolar and glottal whereas, further fricatives are rare in Austroasiatic languages. Among liquid consonants in Austroasiatic languages common manners of articulation include approximants and trills and common places of articulation include labial, alveolar, lateral and palatal.

Considering variations in Austroasiatic consonant inventory, a remarkable divergence can be observed by comparing the consonants of South and Southeast Asian Austroasiatic languages, particularly the Munda branch of South Asian Austroasiatic languages. It is noticed that there are some unique consonants in Austroasiatic languages of each region that are not found in the other region. Firstly, voiced implosive sounds such as /ɓ, ɗ, ɠ, ʄ/ occur in a number of Southeast Asian Austroasiatic languages but, are not found in Austroasiatic languages of South Asia. Similarly, retroflex sounds such as /ʈ, ɖ, ʂ/ occur in many Munda languages of South Asian Austroasiatic languages but, are not found in Austroasiatic languages of Southeast Asia. Usually, it is believed that implosive sounds are typical to the typology of Southeast Asian

languages and retroflexion is typical to South Asian languages. Moreover, while the implosive sounds are regarded as native to the Proto-Austroasiatic language, retroflexion is regarded as a later development in Munda branch of South Asian Austroasiatic languages. This is also reflected in the reconstruction of consonant inventory of the proto languages. Table 2.3 presents the Proto-Austroasiatic and Proto-Munda initial consonant inventory as proposed by Sidwell & Rau (2014).

Table 2.3: Proto-AA and Proto-Munda initial consonants (Sidwell & Rau, 2014)

pAA	p	b	ɓ	t	d	ɗ	c	ɟ	f	k	g	ʔ	m	n	ɲ	ŋ	w	r	l	j	s	h
pMunda	p	b		t	d		c	ɟ		k	g		m	n	ɲ	ŋ	w	r	l	j	s	h

Table 2.3 suggests that initial consonant inventory of Proto-Munda and Proto-Austroasiatic are fairly similar. The only difference is the inclusion of implosive consonants /ɓ, ɗ, f/ and glottal consonant /ʔ/ in Proto-Austroasiatic that are lacking in Proto-Munda. This indicates that after branching away from Proto-Austroasiatic, Proto-Munda languages lost the implosive consonants as well as the glottal stop. In modern Munda, while implosive consonants are still lacking, a glottal stop is found in many modern Munda languages. In this regard, it is argued that Proto-Munda lost the glottal stop at an early stage but, it reemerged in modern Munda languages in different forms such as preglotalized stops, allophone of oral stops in word final position and also as a phoneme in some languages. However, except Gorum (Rau, 2011) the phonemic distribution and phonetic realization of /ʔ/ is not adequately described in many Munda languages. Therefore, it is difficult to generalize the phonemic nature of /ʔ/ in Munda languages. Similarly, retroflex consonants found in Munda branch of South Asian Austroasiatic languages are unattached in Proto-Austroasiatic as well as in Proto-Munda. It is believed that retroflex consonants developed in Munda languages at a later stage when they came in contact with Dravidian and Indo-Aryan languages. Also, Arsenault (2017) proposes that retroflexion is an areal feature in South Asia that occurs across genetic boundaries. However, a detailed description of

the phonemic distribution and phonetic realization of retroflex consonants in individual South Asian Austroasiatic languages is lacking. Thus, from the typology of the Austroasiatic consonantal inventory it is observed that while a number of individual languages have large consonantal inventories, size of the shared inventory is relatively small. Also, it is observed that a major distinction between Austroasiatic languages of South and Southeast Asia emerged mainly from the presence of two consonantal categories namely implosives and retroflexes. Significantly, the Munda languages lost the native Austroasiatic implosive consonants and acquired the non-native retroflex consonants after separating from the Austroasiatic parent language.

Considering the vowel inventory next, evidences suggest that full vowel inventory of Austroasiatic languages is also available in the main (stressed) syllable only. Unstressed syllables usually have the short vowel /ə/ that can also occur in the main (stressed) syllable. However, other vowels usually do not occur in the unstressed syllable. Thus, syllable position and syllable weight effectively determines the phonemic status of vowels in many Austroasiatic languages. Moreover, size and structure of vowel system in Austroasiatic languages can also vary considerably across languages. Hence, based on the size and structure of the vowel system there can be two broad groups of Austroasiatic languages. The first category includes languages with nine or more vowels and the second category includes languages with five or six vowels. Usually Austroasiatic languages of Southeast Asia such as Chong, Car, Bagan and Bunong have large vowel systems whereas, languages of South Asia such as Kharia, Mundari and Khasi have small vowel systems. A comparative vowel inventory of Austroasiatic languages compiled from Jenny et al. (2014) is presented in Table 2.4.

From Table 2.4 it is observed that, an inventory of nine or more vowels usually include the vowels /i, e, ε, ɨ, ə, a, u, o, ɔ/ whereas, an inventory of five or six vowels include the vowels /i, e, a, o, u/ and /ə/ or /ɨ/. Thus, it is noticed that, variation in size and structure of vowel systems occurs mainly due to the inclusion

Table 2.4: Austroasiatic vowels (Jenny et al., 2014)

Chong	i/i:	e/e:	a/a:	o/o:	u/u:	ɛ/ɛ:	ɔ/ɔ:	ə/ə:	ɨ/ɨ:		
Kharia	i	e	a	o	u						
Khasi	i/i:	e/e:	a/a:	o/o:	u/u:				ɨ		
Car	i	e	a	o	u	ɛ	ɔ	ə	ɨ	ɣ (œ)	
Koho	i	e	a	o	u	ɛ	ɔ	ə	ɨ		ɑ
Bugan	i	e	a	o	u	ɛ	ɔ	ə			y ʊ
Bunong	i	e	a/a:	o/o:	u/u:	ɛ	ɔ/ɔ:	ə/ə:	ɨ/ɨ:		
Kammu	i/i:	e/e:	a/a:	o/o:	u/u:	ɛ	ɔ	ə/ə:	ɨ/ɨ:		ʌ

of mid and central vowels in the vowel inventory of Southeast Asian Austroasiatic languages. Southeast Asian Austroasiatic languages normally have two mid vowels /e, ɛ/ at the front, two mid vowels /o, ɔ/ at the back and three central vowels /ɨ, ə, a/. Whereas, South Asian Austroasiatic languages have just one front mid vowel /e/, one back mid vowel /o/ and one central vowel /a/ or two central vowels /a, ɨ/ or /a, ə/ (except languages such as Santali, Ho and Gta?). Thus, Southeast Asian Austroasiatic languages are able to maintain three distinctive vowel heights in front, back as well as in the mid region. Whereas, South Asian Austroasiatic languages can distinguish only two vowel heights in front and back and one or two vowel heights in the central region. Significantly, historical-comparative studies suggest that, the vowel system of Southeast Asian Austroasiatic languages closely represent the Proto-Austroasiatic vowel system. Whereas, vowel system of South Asian Austroasiatic languages have a reduced vowel system. Table 2.5 presents the Proto-Austroasiatic and Proto-Munda vowel inventory as proposed by Sidwell & Rau (2014).

Table 2.5: Proto-Austroasiatic and Proto-Munda vowel inventory (Sidwell & Rau, 2014). Undetermined Proto-Munda vowels are underlined

pAA Vowels	i	i:	e	e:	ɛ	ɛ:	ə	ə:	a	a:	u	u:	o	o:	ɔ	ɔ:	ie	uo	iə	uə
pMunda Vowels	i					<u>e</u>	ə		a		u				<u>ɔ</u>					
	I					<u>e</u>	<u>ə</u>		<u>a</u>		<u>u</u>				<u>ɔ</u>					

Table 2.5 suggests that the vowel inventory of Proto-Munda was reduced at some point after being separated from Proto-Austroasiatic. However, Sidwell & Rau (2014)

suggest that some vowels of Proto-Munda are underdetermined (underlined) and therefore, the status of those vowels is unclear. Moreover, while the vowel system of Southeast Asian Austroasiatic languages are fairly well described and it is believed that many Southeast Asian Austroasiatic languages have restructured their vowel inventory by splitting or merging due to register genesis, the motivation for vowel restructuring in Munda languages is not very clear. It is only believed that the vowel system reduced mainly because of a rhythmic drift in Munda languages that is caused by language contact (Donegan, 1993). However, it is not conclusive whether small vowel system is actually an evidence of vowel reduction in the Munda branch of South Asian Austroasiatic languages or not. There are no conclusive results yet mainly because a lot of individual languages are not described in detail so far. Hence, in order to achieve a better comparative study in Austroasiatic languages, it is essential that individual Munda languages are described adequately.

A review of segmental properties of Austroasiatic languages reveals that while most Austroasiatic languages have similar segmental features, there is a typological differences between the Munda branch of South Asia and the non Munda branch of Southeast Asia. While the consonant inventory is fairly similar, the vowel inventory shows a significant diversion between Austroasiatic languages of the two regions. Studies so far have also indicated that while Southeast Asian Austroasiatic languages closely relate to the Proto-Austroasiatic segmental inventory, the South Asian Austroasiatic languages have greatly diverged from the proto language. This phenomena is particularly reported in the Munda branch of South Asian Austroasiatic languages. However, in order to determine whether all Munda languages are equally diverged and if some Munda languages are actually preservative, it is important to review the typology of Munda languages within the sub-branch. Therefore, the following section reviews the typology of Munda segmental features from available relevant literature.

## 2.2.2 Segmental properties of Munda languages

Sub-Branch level comparative descriptions of Munda segmental properties can be found in studies such as Konow (1908), Pinnow & Kuiper (1963), N. H. Zide (1969), Bhattacharya (1975), Sidwell & Rau (2014) and Anderson (2014). Additionally, sketch grammar of some major Munda languages is also available in an edited volume by Anderson (2008b). However, it should be noted that complete segmental description of some individual languages are still lacking. Therefore, the available literature only sparsely provides a typology of segmental features in Munda language. From a typological point of view it is observed that while all Munda languages have similar segmental properties only some languages have slightly different properties. Hence, in order to identify typical and unique Munda segments this section discusses the segmental properties in Munda languages from a typological point of view.

Considering the Munda consonants first, it is observed that there are some conflicting views in the historical-comparative literature on Munda languages. For instance, Konow (1908) and N. H. Zide (1969) propose that Munda consonants include four manners of articulation namely stop, fricative, nasal and liquid and five places of articulation namely labial, dental, retroflex, palatal and velar. On the other hand Bhattacharya (1975) proposes that Munda consonants include five manners of articulation namely stop, affricate, nasal, flap and liquid and five places of articulation namely labial, dental, retroflex, palatal and velar. A comparative consonant inventory of Munda languages proposed in these works is presented in Table 2.6.

Table 2.6: Traditional Munda consonants

Konow (1908)	p	b	t	d	t	d	ʈ	ɖ	k	g	s	m	n	ɲ	j	r	l	v	w	h	
	p <sup>h</sup>	b <sup>h</sup>	t <sup>h</sup>	d <sup>h</sup>	t <sup>h</sup>	d <sup>h</sup>	ʈ <sup>h</sup>	ɖ <sup>h</sup>	k <sup>h</sup>	g <sup>h</sup>			ɲ								
	ʔp		ʔt				ʔc		ʔk												
N. H. Zide (1969)	p	b	t	d		d	ʈ	ɖ	k	g	s	m	n	ɲ		r	l				
	ʔb					ʔd	ʔj			ʔg											
Bhattacharya (1975)	p	b	t	d	t	d	ʈ	ɖ	k	g	m	n	ɲ		r	l			h	ʔ	
	ʔp	ʔb	ʔt	ʔd	ʔt	ʔd	ʔc	ʔj	ʔk	ʔg											

Munda consonants presented in Table 2.6 reveal that while some consonants are uniformly proposed by all scholars, other consonants are proposed by individual scholars only. Therefore, variations are observed in the consonant inventory of Munda consonants. In this case, variation in the Munda consonants is mainly observed in retroflex consonants, post-alveolar affricates, pre-glottalized consonants, aspirated consonants, fricatives and approximants. However, these variations are only found in historical-comparative Munda literature whereas, the status to these sounds are quite clear in the synchronic literature of modern Munda languages. Therefore, the diachronic development of these sounds in synchronic Munda varieties are briefly discussed below.

It is mentioned earlier that retroflex consonants are commonly found in Munda languages, but they are not native to the Munda consonant system. Also, presence of retroflex consonants in Munda is reported as one of the main reasons for divergence between Munda and Southeast Asian Austroasiatic languages. Thus, development of retroflex consonants is significant to the typology Munda and Austroasiatic languages. According to Pinnow & Kuiper (1963) Proto-Austroasiatic did not have retroflex consonants and it is likely that retroflex consonants developed under the influence of Indo-Aryan and Dravidian languages in the Proto-Munda stage. Consequently, retroflex consonants are recorded in Munda languages and studies such as Konow (1908) propose that Munda has four retroflex consonants including /t̪, d̪, ɽ, ɳ/. However, later descriptions such as N. H. Zide (1969) and Bhattacharya (1975) include only one retroflex consonant /d̪/. In this regard, N. H. Zide (1969) indicates that /t/ and /d̪/ have voicing asymmetry in Munda languages due to loan phonology whereby, /t/ and /d̪/ correspond to the dental /t/ and retroflex /d̪/ of Indo-Aryan languages. Likewise, Bhattacharya (1975) also reports that only /d̪/ is a frequently occurring retroflex consonant in Munda whereas, /t̪/ is mostly found in Indo-Aryan loan words and /ɽ/ is only found as an allophone of the liquid consonant /r/. The

nasal retroflex consonant /ŋ/ was not recorded in any of the subsequent studies. Thus, it is found that although retroflex consonants were identified in Munda languages, they were classified in different ways by different scholars. Therefore, there were different accounts regarding the existence of retroflex consonants in Munda languages for a long time. However, the synchronic descriptions have shown that all four retroflex sounds identified by Konow (1908) are phonemic consonants in a number of modern Munda languages.

Post-alveolar affricates in Munda languages were previously transcribed as /c/ and /j/. However, the traditional transcription often causes difficulty in classifying the two sounds in modern Munda languages. For instance, works such as Konow (1908) also use the letters /c/ and /j/ to transcribe palatal stop consonants /c/ and /j/ in Munda languages. Consequently, it is often difficult to find the exact IPA equivalent for /c/ and /j/ in modern Munda languages. In this regard, from the English examples given by (Konow, 1908) and from the phonetic description given by Bhattacharya (1975) it is clear that /c/ and /j/ are post-alveolar affricates in Munda and /tʃ/ and /dʒ/ are their IPA equivalents. Moreover, there are no clear evidences regarding the presence of palatal stops in modern Munda languages. Therefore, through careful examination of available data, this work considers /c/ and /j/ as post-alveolar affricates and not palatal stops in Munda languages. Additionally, while Bhattacharya (1975) suggests that only /dʒ/ is a commonly occurring affricate consonant in Munda languages and /tʃ/ is not uniform, synchronic descriptions reveal that both these sounds are phonemic consonants in a number of modern Munda languages.

Glottal constriction is a significant phenomenon in Munda consonant systems. Glottal constriction in Munda consonants includes some pre-glottalized consonants /<sup>h</sup>p, <sup>h</sup>b, <sup>h</sup>t, <sup>h</sup>d, <sup>h</sup>k, <sup>h</sup>g, <sup>h</sup>c, <sup>h</sup>j/ and a glottal stop /ʔ/. Pre-glottalized consonants are often referred to as checked consonants in Munda languages. Pre-glottalized consonants generally include stop consonants that have a glottal constriction in addition to the

oral occlusion. In Munda languages it is agreed that pre-glottalized stops occur only word finally. However, a variation is observed in the inclusion of pre-glottalized stops in Munda consonants. While Konow (1908) includes only voiceless pre-glottalized stops and N. H. Zide (1969) includes only voiced pre-glottalized stops, Bhattacharya (1975) argues that both voiced and voiceless pre-glottalized stops are frequently found in Munda languages. However, evidences suggest that pre-glottalized stops are better represented as allphones of the Munda stop consonants that tend to be glottalized in the word final position only. Also, as mentioned earlier, since there is no clear evidence for the presence palatal stops /c/ and /ɟ/, existence of pre-glottalized /<sup>ʔ</sup>c/ and /<sup>ʔ</sup>ɟ/ is more unlikely. On the other hand Bhattacharya (1975) has shown that unlike pre-glottalized stops a full glottal stop occurs both word medially and word finally in Munda languages. Thus, it is more likely that a full glottal stop is the actual glottal phoneme in modern Munda languages and this is also evident in the synschronic description of many modern Munda languages.

Regarding aspirated consonants, Konow (1908) proposes the presence of aspirated stops for all possible stop consonants in Munda languages. However, N. H. Zide (1969) does not approve of aspirated consonants in Munda. Likewise, although Bhattacharya (1975) includes three aspirated consonants, he suggests that there is no uniformity in the production of aspirated consonants in Munda languages. Similarly, Pinnow & Kuiper (1963) suggests that aspiration is a secondary development in Munda languages that is influenced by Indo-Aryan languages. However, aspirated consonants are recorded in non-Munda Austroasiatic languages such as Mon and Khasi and therefore the Indo-Aryan influence on Munda aspirated consonants requires more research. Thus, although there are indications that aspirated consonants are not native to Munda languages, there are evidences that some Munda languages have adopted aspirated stop consonants probably due to language contact. However, presence of phonemic aspirated consonants in Munda languages requires more evidences.

The next set of consonants that cause variation in Munda consonant systems are the fricative sounds /v/, /s/, /z/ and /h/. In the historical-comparative literature it is found that, while Konow (1908) includes three fricative consonants /s, h, v/, N. H. Zide (1969) includes only /s/ and Bhattacharya (1975) includes only /h/ in the inventory Munda consonant system. Moreover, Bhattacharya (1975) claims that although a dental sibilant /s/ is often found in some Munda languages, it is usually produced in place of the affricate consonants /tʃ/ and /dʒ/. However, it is found from the literature that fricative consonants including /s/ and /h/ commonly occur in many modern Munda languages, and languages such as Remo (Anderson & Harrison, 2008a) also include a labio-dental fricative /v/ and a voiced alveolar fricative /z/. Thus, it is evident that although fricative consonants were traditionally underrepresented, the synchronic data reveals that fricative sounds are important phonemic constituents in Munda consonant system. Similarly, the approximants /j/ and /w/ appear to be insignificant in traditional Munda consonant system but, both sounds are recorded as distinct phonemes in modern Munda languages.

Thus, from the discussion so far it is evident that some Munda consonants were not described uniformly in the historical-comparative descriptions of Munda consonant system. However, the lack of uniformity in Munda consonants has been cleared in synchronic descriptions to a considerable extent. Therefore, a more uniform consonant system of Munda languages compiled by Anderson (2014) can be seen in Table 2.7.

Table 2.7: Modern Munda consonants (Anderson, 2014)

Santali	p	b	t	d	tʃ	dʒ	t	ɖ	k	g	m	n	ɲ	ɳ	ʀ	ʁ	ʎ	ʑ	h	s	ʔ
Mundari	p	b	t	d	tʃ	dʒ	t	ɖ	k	g	m	n	ɲ	ɳ	ʀ	ʁ	ʎ	w	h	s	ʔ
Keraʔ	p	b	t	d	tʃ	dʒ	t	ɖ	k	g	m	n	ɲ	ʔ	ʀ	ʁ	ʎ	wʔ	h	s	ʔ
Ho	p	b	t	d	tʃ	dʒ	t	ɖ	k	g	m	n	[ɲ]	ʀ	ʁ	ʎ	w		h	s	ʔ
Korku	p	b	t	d	tʃ	dʒ	t	ɖ	k	g	m	n	ɲ	ɳ	ʀ	ʁ	ʎ	v,w	h	s	ʔ
Juang	p	b	t	d	tʃ	dʒ	t	ɖ	k	g	m	n	ɲ	ɳ	ʀ	ʁ	ʎ		s		ʔ
Kharia	p	b	t	d	tʃ	dʒ	t	ɖ	k	g	m	n	ɲ	ʀ	ʁ	ʎ	w		h	ʔ	ʔ
Sora	p	b	t	d	tʃ	dʒ		d	k	g	m	n	ɲ	ʀ	ʁ	ʎ			s		ʔ
Gorum	p	b	t		s	z	t	ɖ	k	g	m	n		ʀ	ʁ	ʎ			s		ʔ
Remo	p	b	t	d	tʃ,ts,s	z,dz	t	ɖ	k	g	m	n	ɲ	ʀ	ʁ	ʎ	w,v		s		ʔ
Gutob	p	b	t		tʃ,ts	dʒ,dʒ	tʔ	ɖ	k	g	m	n	ɲ	ʀ	ʔ	ʁ	dʒʔ		ʔ	c	ʔ
Gtaʔ	p	b	t	ʔ	tʃ	dʒ	t	ɖ	k	g	m	n		ʀ	ʁ	ʎ	w		h	s	ʔ
																					C <sup>h</sup>

The comparative list of Munda consonants in Table 2.7 confirms that, most of the consonant phonemes are typologically similar in all Munda languages. Also, it is evident that inclusion of retroflex consonants in Munda languages is a major distinguishing factor between the consonant system of Munda and non-Munda Austroasiatic languages. Whereas, the other Munda consonants have clear correspondence with the consonant system of non-Munda Austroasiatic languages. Therefore, although a complete segmental description of some individual Munda languages is still lacking, it can be accepted that the Munda consonant system is fairly well understood. However, this assumption does not work for the Munda vowel system and therefore some traditional as well as synchronic variations in Munda vowel system are discussed below.

Scholars in the past often had difficulties in classifying Munda vowels and N. H. Zide (1969) admits that the Munda vowel system is comparatively more complicated than the Munda consonant system. Difficulty in classifying the Munda vowel system can be seen in some of the comparative-historical studies as well as in some of the synchronic studies dealing with Munda vowel system. Comparative studies dealing with Munda vowel system includes a main branch level study by Bhattacharya (1975) and some sub-branch level studies by Munda (1968, 1969) on Kherwarian, by N. H. Zide (1966) on Korcu-Kherwarian, by Osada (1996) on Proto-Kherwarian, by N. H. Zide (1965) on Gutob-Remo-Gta? and on Sora-Juray-Gorum by A. R. Zide (1982). Additionally, a few other studies have also examined Munda vowel system from a comparative-historical point of view and such works include Konow (1908), Donegan (1993), Anderson (2014) and Sidwell & Rau (2014). Similarly, brief sketches about the synchronic vowel system in some of the major Munda languages is also available in a edited volume by Anderson (2008b). The historical-comparative studies mentioned above suggest that Munda languages typically have the following five vowels /i, e, a, o, u, a/. Similarly, the synchronic descriptions also suggest that all major Munda

languages except Santali and Orissa Sora have the same five vowels in their vowel system. Santali (Ghosh, 2008) has 8 vowels including /i, e, ε, ə, a, u, o, ɔ/ and Orissa Sora (Donegan, 1993) has nine vowels including /i, e, ε, ə, i, a, u, o, ɔ/.

Significantly, Donegan (1993) relates the exceptionally large vowel system of Orissa Sora with the vowel inventory of Mon-Khmer languages and proposes that Orissa Sora has actually preserved the Proto-Munda vowel system. This indicates that in addition to the five Munda vowels including /i, e, o, u, a/, Proto-Munda also had the four Orissa Sora vowels including /ε, ɔ, ə, i/. However, scholars in the past had different opinions regarding the existence of these mid and central vowels in Munda languages. For instance, Konow (1908) mentions the lower mid vowels [ε] and [ɔ] are not easily distinguishable in Munda languages. Likewise, Bhattacharya (1975) mentions that although [ε] and [ɔ] occur in some Munda languages they are not Proto-Munda vowels. Rather, he claims that /e/ and /o/ are usually lowered to [ε] and [ɔ] in loan words or due to some phonological processes in some Munda language. Similarly, regarding the central vowels while Bhattacharya (1975) completely excludes the high central vowel /i/, he mentions that the mid central vowel /ə/ is not native to Munda and is borrowed from Indo-Aryan languages. Therefore, Bhattacharya (1975) classifies Munda vowels into three categories namely primary, secondary and peripheral vowels. According to Bhattacharya (1975) the primary vowels include the five uniform vowels /i, e, a, o and u/ and the secondary vowels include the three non uniform vowels /ε, ə and ɔ/ whereas, the peripheral vowels are not well described. Hence, he argues that, only the five primary vowels occur in all Munda languages whereas, the secondary vowels occur only in some languages either in loanwords or when a phonological process changes the primary vowel into a secondary vowel. On the other hand (Donegan, 1993) suggests that /ə/ and /i/ originally occurred in Proto-Munda but, are lost in many modern Munda languages except Sora, mainly due to the intrinsic shortness of the two vowel. She also proposes that the two vowels have either lowered or fronted independently

in every Munda language except Sora. For this reason, Pinnow & Kuiper (1963) also suggests that although central vowels are not common to Munda languages, the two central vowels /i/ and /ə/ should be reconstructed in Proto-Munda. Thus, it is evident that conflicting views regarding the Munda vowel system in the Proto-Munda stage arises mainly due to the inclusion of the two central vowel /i/ and /ə/ and two mid vowels /ɛ/ and /ɔ/. Hence, although there are claims that these vowels existed in the Proto-Munda stages there are conflicting views about the hypothesis. Also, a detailed discussion about the vowel inventory of Orissa Sora in the following sections will reveal that the resemblance between Orissa Sora vowel system and Proto-Munda vowel system requires more research.

Additionally, the Munda vowel system with only five vowels also appears significantly diverged from the large vowel systems of Proto-Austroasiatic and modern Austroasiatic languages of Southeast Asia. In this regard, it is hypothesized that originally Proto-Munda had a large vowel system but, later it was reduced to five vowel system in the modern Munda languages. Therefore, it is assumed that vowel reduction in Munda languages is geographically effected under the influence of Indo-Aryan and Dravidian languages. However, it is observed that there are conflicting views regarding Proto-Munda vowel system as well. Firstly, the sub-branch level studies mentioned above indicate that Proto-Munda had seven vowel phonemes including /i, e, ɪ, ə, u, o/. On the other hand, the main branch level study by Bhattacharya (1975) argues that Proto-Munda also had the same five vowels that are found in modern Munda languages. Yet another Proto-Munda vowel system proposed by Sidwell & Rau (2014) include some well defined and some underdetermined vowels such as /i-I, u-U, ə-Ə, a-A, E, O/ where the capital letters represent the underdetermined vowels in Proto-Munda vowel system. Hence, following the assumption that Munda vowel system has reduced after separating from Proto-Austroasiatic, different scholars have tried to reconstruct Proto-Munda vowel system in different ways. However, exist-

ing results are conflicting with each other and therefore reconstructing the actual Proto-Munda vowel system still requires more research. Moreover, the vowel system of many synchronic Munda languages lacks clear correspondences with each other as can be seen from the comparative list of Munda vowels compiled by Anderson (2014) presented in Table 2.8.

Table 2.8: Modern Munda vowels (Anderson, 2014)

Santali	a	i	u	o	e	ɔ	ɛ	ã	ə		ɨ	<sup>n</sup> D				
Mundari	a	i	u	o	e			ã								
Kera?	a	i	u	o	e			ã								
Ho	a	i	u	o	e	ɔ	ɛ	ã	a:	a <sup>?</sup>						
Korku	a	i	u	o	e			?	a:		à					
Juang	a	i	u	o	e	ɔ		ã	a:		?					
Kharia	a	i	u	o	e			?			?					
Sora	a	i	u	o	e			?	a:	a <sup>?</sup>	ə	i	?	?	ɨ	ɪ
Gorum	a	i	u		e	ɔ		ã	?	?		ə				
Remo	a	i	u		e	ɔ		ã	?	?			ɨ	<sup>n</sup> D		
Gutob	a	i	u	o	e			?		a <sup>?</sup>						
Gta?	a	i	u	o	e	ɔ	ɛ/œ	ã		a <sup>?</sup>	i	?	ɨ	<sup>n</sup> D		

Hence, it is evident that unlike the consonant system of Munda languages, the vowel system has more variations in the historical-comparative descriptions as well as in the synchronic descriptions. As a result, while the Proto-Munda consonant system is fairly well understood, the Proto-Munda vowel still requires more researching. Also, in the Segmental inventory of Munda languages, it is observed that while the consonant inventory of Orissa Sora is typologically uniform with Munda consonant system, the vowel inventory of Orissa Sora is unique to the Munda vowel. Also, it is proposed that Orissa Sora is a preservative Munda language. However, a closer look at the descriptions of Orissa Sora segmental inventory reveals that there are significant inconsistencies in the available literature on Orissa Sora segmental properties. Therefore, in order to identify the inconsistencies, the following section discusses in detail the typical characteristics of Orissa Sora segmental features.

### 2.2.3 Segmental properties of Orissa Sora

This section provides a detailed review of Orissa Sora segmental properties as it is represented in the available literature. Studies that have partially or elaborately described Orissa Sora segmental properties include works such as Ramamurti (1931, 1938), Stampe (1963, 1965), A. R. Zide (1982), Donegan (1993), P. Mohanty (1997), N. Zide (1999) and Anderson & Harrison (2008b). In addition to these studies, there are a few unpublished conference papers such as Donegan (1980, 1981, 2009) that are probably a result of the authors' field study in Sora speaking areas of Orissa, undertaken between 1980 and 2004. A review of the available literature reveals that Orissa Sora has been considerably studied in the past, but its segmental properties are proposed differently by different scholars. Consequently, the descriptions are not consistent and there are conflicting views within the existing descriptions. Thus, this section primarily highlights the inconsistency in Orissa Sora segmental properties.

The Sora Manual by Ramamurti (1931) is one of the earliest detailed descriptions of the Orissa Sora. The manual consists of different sections including phonetic and grammatical description of the Sora language, basic Sora vocabulary and a guide to sociocultural practices of the Sora community. The manual is mainly dedicated for guiding Sora language learners with basic Sora vocabulary, common Sora phrases and pronunciation of Sora speech sounds. However, the manual also includes some substantial description of the Sora language such lexical, grammatical and phonetic characteristics of the language. Phonetic description provided in the Sora Manual was the first description of the speech sounds in Orissa Sora. It proposes that Orissa Sora has 27 phonemes including 17 consonants and 10 vowels (see Table 2.9 and 2.10). The inventory is based on phonetic letters as (Ramamurti, 1931) uses IPA characters to write Orissa Sora and the same inventory is also used in the Sora-English Dictionary by Ramamurti (1938). In both these works the author strongly advocates the use of phonetic letters for writing Orissa Sora over other writing systems, and he states:

“The symbols employed in this Manual to represent the speech-sounds of So:ra: belong to the alphabet of the International Phonetic Association, which is now widely known. The So:ra:s have no traditional alphabet of their own. No Indian script is perfect” (Ramamurti, 1931).

Thus, the phonetic alphabets used to write Sora in the Sora Manual and in the Sora-English Dictionary is accepted as the phoneme inventory of Orissa Sora proposed by Ramamurti (1931). In the subsequent period, a second phoneme inventory of Orissa Sora is presented by Stampe (1965). It proposes that Orissa Sora has 27 speech sounds that includes 18 consonants and 9 vowels (see Table 2.9 and 2.10). Also, similar to Stampe’s phoneme inventory of Orissa Sora, A. R. Zide (1982) presents a third phoneme inventory for Orissa Sora. It identifies 26 phonemes including 17 consonants and 9 vowels (see Table 2.9 and 2.10). Additionally, A. R. Zide (1982) presents a surface phonetic inventory of Orissa Sora phonemes including some allophonic variations that are not included in the phoneme inventory of Orissa Sora. In addition to these phoneme inventories of Sora, more recently Anderson & Harrison (2008b) suggest that Orissa Sora has 26 speech sounds including 18 consonants and 8 vowels (see Table 2.9 and 2.10) however, they do not propose that these sounds are the distinctive phonemes of Orissa Sora.

Table 2.9: Orissa Sora consonants inventory

1	p	b	t	d		k	g	s	ʃ	m	n	ɲ	ɳ	r	l	j	ɽ	ʔ			
2	p	b	t		ɖ		ʃ	k	g	s		m	n	ɲ	ɳ	r	l	j	w	ɽ	ʔ
3	p	b	t	d			ʃ	k	g	s		m	n	ɲ	ɳ	r	l	j		ɽ	ʔ
4	p	b	t	d		c	ʃ	k	g	s		m	n	ɲ	ɳ	r	l	j		ɽ	ʔ

1 (Ramamurti, 1931) 3 (A. R. Zide, 1982)

2 (Stampe, 1965) 4 (Anderson & Harrison, 2008b)

Thus, it is evident that there are inconsistencies in the proposed phoneme inventories of Orissa Sora. However, it is noteworthy that, of all the phoneme inventories that are proposed for Orissa Sora, the phoneme inventory proposed by Stampe (1965)

Table 2.10: Orissa Sora vowel inventory

1	a	i	ɪ	u	ʊ	e	o	ü	ö	ə		
2	a	i		u		e	ɛ	o	ɔ	ə	ɨ	
3	a	i		u		e	ɛ	o	ɔ		ɨ	ʌ
4	a	i		u		e		ɔ				
5	a	i		u	ʊ	e	o			ə	ɨ	

1 (Ramamurti, 1931) 2 (Stampe, 1965) 3 (A. R. Zide, 1982)  
4 (P. Mohanty, 1997) 5 (Anderson & Harrison, 2008b)

is mostly cited by other scholars who have studied Orissa Sora in the subsequent period. Moreover, the vowel description of Ramamurti (1931) is often criticized by later scholars. For instance, while discussing the vowel length in Orissa Sora Stampe (1963) as cited in A. R. Zide (1982) states:

“If Ramamurti’s data were completely reliable, this (vowel length distinction in Orissa Sora) might have been one plausible interpretation.”  
(Stampe, 1963) cited in (A. R. Zide, 1982).

Thus, scholars in the past have raised concerns regarding the reliability of Ramamurti’s data evidences. Therefore, later studies by Donegan and Stampe have largely followed the inventory proposed by Stampe (1963, 1965). However, although Stampe (1963) raised concerns regarding the reliability of Ramamurti’s description, it is observed that much of Stampe’s work is also available in manuscript form only, and his data evidences are not available for reexamination. Also, much of Donegan’s relevant work such as Donegan (1980, 1981, 2009) are not published. Therefore, the data evidences of these works too cannot be re-examined. Moreover, A. R. Zide (1982) suggests that there are unresolved issues in Donegan (1981) particularly in the use of tense and lax vowel classifications. These classifications may not be phonetically accurate for the vowels in Orissa Sora. However, commenting on her own work A. R. Zide (1982) states that it is only a brief sketch of the Sora phonemes and does not provide a full phonological description. As a result, the inconsistency in the phoneme inventory of Orissa Sora continues to exist. Following this, although

Anderson & Harrison (2008b) identified the speech sounds of Orissa Sora, they too abstain from considering the sounds as a distinct phonemes of Orissa Sora. Thus, it is evident that different scholars described Orissa Sora segmental properties in different ways. There is a general disagreement regarding the size of the phoneme inventory as well as regarding the position of individual phonemes. Therefore, in order observe the inconsistencies more closely, the following sections review the significance of individual speech sounds of Orissa Sora. The discussion below mainly reviews the two works by Ramamurti (1931) and A. R. Zide (1982), since only these two works have elaborate descriptions with data evidences. Whereas, other works of Orissa Sora are either not accessible or do not have elaborate descriptions.

#### **2.2.4 Orissa Sora vowel phonemes**

As mentioned earlier, there are three different vowel inventories of Orissa Sora. Firstly, Ramamurti (1931) in the Sora Manual suggests that Orissa Sora has 10 vowel phonemes. However, the description given by Ramamurti (1931) indicates that only six vowels /a, i, u, e, o, ə/ are phonemic vowels and the remaining four vowels [ɪ, ʊ, ö, ü] are allophonic variations. For instance, Ramamurti (1931) describes the vowel [ɪ] as a mid-front-short vowel but reports that the sound is also perceived as high front vowel /i/ or mid front /e/. Similarly, he describes [ʊ] as a mid-back-unrounded vowel but reports that, the sound is often perceived as high back vowel /u/ or mid back /o/. Additionally, he describes [ö] and [ü] (equivalent to [ø] or [ɣ] and [y] or [ɥ] in IPA) as mixed vowels and reports that while [ö] may be perceived as /e/ or /o/, [ü] is often perceived as /i/ or /o/. Thus, the four vowel sounds identified by Ramamurti (1931) are not very clear and from the author's description it appears that these vowel sounds are more likely to be allophonic variations of the remaining six vowels /a, i, u, e, o, ə/. Moreover, in Ramamurti (1938) although Orissa Sora vowel inventory is extended to different lengths and variations, the primarily vowel grouping includes

only six vowel categories /a, i, u, e, o, ə/. Moreover, Ramamurti (1938) indicates that, the four vowel sounds might represent dialectal variation of Orissa Sora, but it is difficult to confirm since there are no appropriate data to re-examine the claim. Therefore, from the descriptions of Ramamurti (1931, 1938) it is apparent that Orissa Sora actually has six vowel sounds including /a, i, u, e, o, ə/. Whereas, the other sounds are considered to be either allophonic variations of the six main vowel or result of dialectal variation that needs further verification.

On the other hand, A. R. Zide (1982) reports that all modern scholars agree that Orissa Sora has 9 vowel phonemes including /a, i, u, e, o, ə, ɨ, ɛ, ɔ/. However, she admits that there are some disagreements in Orissa Sora vowel transcriptions that are not fully resolved. According to A. R. Zide (1982), most notable disagreement arises in the transcription of /o/ vs. /u/ and /e/ vs. /ɨ/. Explaining the discrepancies in transcription, she suggests that in case of /o/ vs. /u/ there may be a dialectal variation, whereas in case of /e/ vs. /ɨ/ she suggests that there may be a free variation within the same dialect. Moreover, A. R. Zide (1982) suggests that Orissa Sora vowels classification in Donegan (1981) is adequate for phonemic representation but is not phonetically accurate. A. R. Zide (1982) argues that although the phonemic features are attributed correctly, there is some obscurity in their phonetic realizations. A. R. Zide (1982) specifically mentions that the use of tense and lax phonetic feature by Donegan (1981) is problematic. According to A. R. Zide (1982), tense and lax vowels are phonological terms that are largely used to describe English vowel oppositions, whereas phoneticians have not found a genuine acoustic or articulatory correlate for such distinctions. Additionally, A. R. Zide (1982) suggests that it is the roundedness feature in Orissa Sora that is more appropriate in distinguishing back vowels than the tense and lax feature. Regarding the roundedness feature, A. R. Zide (1982) mentions that, Orissa Sora back vowels often appear to be produced with a compression of the lips that lacks spreadedness or flattening. Donegan (1981) as

cited in A. R. Zide (1982) calls the phenomenon ‘in-rounding’. However, according to A. R. Zide (1982) it is not clear whether the feature ‘in-rounding’ is the actual cue involved in distinguishing /u/ and /o/ in Orissa Sora, but she agrees that additional features are needed to account for the difference in articulation of the two back vowels beside their height features. She also argues that there are various degrees of rounding as well as some features of labiality in Sora vowels that are not handled properly in standard IPA symbol, but she admits that such characteristics are yet to be described adequately.

Moreover, an unpublished work by Starosta (1964) has indicated that Orissa Sora vowel pairs [i]/[e], [u]/[o], [ə]/[a], [ɨ]/[ɐ] and [e]/[ɨ] do not have clear phonetic distinctions and they may actually appear in free variation. Also, two other recent descriptions of Orissa Sora vowel sound reveal some additional conflicting views. For instance, P. Mohanty (1997) suggests that Orissa Sora has only five vowel phonemes including /i, e, a, ə, u/. It also proposes that Orissa Sora vowel inventory is influenced by an areal typology whereby three language namely Kui (a Dravidian language), Orissa Sora (a Munda language) and Oriya (a Indo Aryan language) have developed similar vowel inventories. Similarly, although Anderson & Harrison (2008b) do not propose a distinct vowel inventory of Orissa Sora, the eight vowel sounds /i, e, o, u, a, ə, ɨ, ʊ/ that are identified by them also stand unique. Thus, the discussion above reveals that there are some conflicting views in the nine vowel system of Orissa Sora particularly regarding the mid vowels and the central vowels. Hence, although many modern scholars have agreed that Orissa Sora has nine vowel phonemes, it will be useful to re-examine the nine vowel system of Orissa Sora.

Additionally, it is noteworthy that a nine vowel system of Orissa Sora, proposed by Stampe (1965), A. R. Zide (1982), Donegan (1993) and Donegan & Stampe (2002) makes Orissa Sora very unique in the vowel typology of Munda languages since other Munda languages mostly have five or six vowels. Regarding the uniqueness of Orissa

Sora vowel inventory, Donegan (1993) argues that Orissa Sora is a preservative language and it has successfully maintained the nine vowel system of Proto-Munda. For instance, Donegan (1993) reports that while most Munda languages have eliminated the high central vowel [i] and mid central vowel [ə] due to their intrinsic shortness, Orissa Sora has preserved those vowels in the closed syllables. Therefore, she strongly favours that Orissa Sora has nine vowels and assumes that the Orissa Sora vowel system may have similarities with languages of the Mon-Khmer branch of the Southeast Asian Austroasiatic languages. However, it is known that although Austroasiatic languages have large vowel inventories (Jenny et al., 2014), Munda languages have a relatively small vowel inventory with 5 or 6 vowels. The closest languages to Orissa Sora namely Gorum (Anderson & Rau, 2008), Gutob (Griffiths, 2008), and Remo (Anderson & Harrison, 2008a) have 5 vowels each, whereas Gta? (Anderson & Harrison, 2008a) has 6 vowels in the vowel inventory. Therefore, it is less likely that Orissa Sora will have a vowel system of nine vowels and the discrepancies discussed above provide evidences for this.

Hence, it is evident from the literature review that while Ramamurti (1931, 1938) extensively discusses the Orissa Sora vowels, later scholars such as Stampe (1963, 1965), Donegan (1981) and A. R. Zide (1982) have largely rejected his proposal on account of poor evidences. However, it is evident that, later scholars too do not provide elaborate descriptions of Orissa Sora vowel phonemes, and much of their works are either unpublished or are available in manuscript form only. Additionally, there are conflicting views even regarding the later descriptions of Orissa Sora vowel phonemes. Thus, it is clear that although there have been attempts to describe the vowel characteristics of Orissa Sora in the past, the descriptions are largely inconsistent and inconclusive.

### 2.2.5 Orissa Sora consonant phonemes

Similar to vowel phoneme inventory, there are also different consonant phoneme inventories proposed for Orissa Sora. However, unlike the vowel inventory, most of the consonants in Orissa Sora appear to be consistent and uniform and only a few consonants tend to vary across different studies. Firstly, while both Ramamurti (1931, 1938) and A. R. Zide (1982) propose that Orissa Sora has 17 consonant phonemes, the two inventories differ mainly in their selection of two consonant phonemes in Orissa Sora. For instance, while Ramamurti (1931, 1938) proposes that Orissa Sora has a voiced palatal fricative /j/, the sound is not reported in A. R. Zide (1982). On the contrary, A. R. Zide (1982) proposes that Orissa Sora has a voiced palatal stop /ɟ/ that changes to a palatal affricate due to some phonological process. However, the process is not clearly defined and therefore, the status of the palatal sounds is not clear in Orissa Sora in these two studies. In addition to these two inventories, the inventory proposed by Stampe (1965) identifies two additional consonants including a voiced retroflex stop /ɖ/ and a labial approximant /w/. However, since the data evidences of Stampe (1965) are not accessible, it is not possible to re-examine the consonant phonemes proposed in his work. Subsequently, although Anderson & Harrison (2008b) do not propose the actual consonant inventory of Orissa Sora, consonant sounds identified by them also stand unique as it was in case of the vowel sounds identified by them. Moreover, it is found that Orissa Sora consonant phonemes are discussed in some detail only in two of the above mentioned works namely Ramamurti (1931, 1938) and A. R. Zide (1982). Hence, similar to the review of vowel phonemes, this section also reviews Orissa Sora consonants mostly from these two works and discusses the consonant phonemes according to their place and manner of articulation.

### 2.2.5.1 Stops

Ramamurti (1938) suggests that Orissa Sora has seven stop consonants. This includes voiced and voiceless bilabial stops /b/ and /p/, a voiceless dental stop /t/, a voiced alveolar stop /d/, voiced and voiceless velar stops /g/ and /k/, and a glottal stop /ʔ/. Moreover, Ramamurti (1938) argues that in Orissa Sora there is neither a voiced dental stop [d̪] nor a voiceless alveolar stop [t̪] and therefore /t/ and /d/ have a voicing asymmetry in Orissa Sora. Also, Ramamurti (1938) claims that there are no retroflex stops in Orissa Sora. In addition to the the seven stop consonants, A. R. Zide (1982) suggests that Orissa Sora also has a voiced palatal stop /ɟ/. She mentions that the palatal stop /ɟ/ often changes to a palatal affricate [tʃ] or an alveolar fricative [z] due to some phonological process, but she does not elaborate the processes in her work. Also, contrary to Ramamurti's description, A. R. Zide (1982) mentions that in Orissa Sora the consonant /d/ varies between dental, alveolar and retroflex articulation. A. R. Zide (1982) proposes that word initially and word medially /d/ is produced as a retroflex sound and word finally the pronunciation varies between dental and alveolar articulation. Thus evidences suggest that the voicing asymmetry between /t/ and /d/ in Orissa Sora is not consistent and there may be a tendency to bring a symmetry between the two consonants. Additionally, A. R. Zide (1982) proposes that all voiced stops tend to be glottalized or devoiced in the word final position. This is also reported in Ramamurti (1938) where he uses the term checked consonants to refer to the glottalized stops in word final position. Thus, it is evident the while most of the stop consonants are uniformly described in the works by A. R. Zide (1982) and Ramamurti (1938), there are some inconsistencies particularly regarding alveolar, palatal and retroflex stops.

### 2.2.5.2 Fricatives

Both Ramamurti (1938) and A. R. Zide (1982) identify only one fricative consonant in Orissa Sora namely a dental sibilant sound /s/. Ramamurti (1938) mentions that the sound occurs only word initially and word medially, but never word finally. Moreover, a voiced fricative consonant [z] is also reported in Orissa Sora, but while Ramamurti (1938) considers it to be an allophone of the voiced palatal fricative /j/, A. R. Zide (1982) considers it to be a borrowed sound or an allophone of the voiced palatal stop /ɟ/. Hence, it is found that while a voiceless fricative sound /s/ is considered to be a distinct phoneme in Orissa Sora, a voiced fricative is often treated as an allophonic variation or a borrowed phoneme.

The second fricative consonant in Orissa Sora proposed by Ramamurti (1931, 1938) is the voiced palatal fricative /j/. According to Ramamurti (1938) while /j/ occurs word initially, medially and finally in word final position it is often checked or glottalized. Moreover, Ramamurti (1938) suggests that, /j/ may also realize as a voiced alveolar fricative [z] in some dialects. Commenting upon the development of /j/ in Orissa Sora, Ramamurti (1938) speculates that it has developed from the velar plosive sounds /k/ and /g/. To illustrate this, Ramamurti (1938) relates the plural suffix *-ji* in Orissa Sora with *-gi* in Pareng and as *-ki* in Kharia and Juang.

### 2.2.5.3 Nasals

The descriptions of both Ramamurti (1938) and A. R. Zide (1982) suggest that Orissa Sora has four nasal consonants including a bilabial nasal /m/, an alveolar nasal /n/, a palatal nasal /ɲ/ and a velar nasal /ŋ/. Also, regarding the alveolar /n/ both studies suggest that /n/ changes to a dental sound in the environment of another dental sound such as /t/. Moreover, Ramamurti (1938) mentions that while /m/, /n/ and /ɲ/ can occur word initially, word medially as well as word finally, /ŋ/ can occur only word medially and word finally but, never word initially. Thus, it is evident that the

four nasal consonants are the most consistent consonant sounds in Orissa Sora that have the least differences across different descriptions of the language.

#### 2.2.5.4 Liquids and Approximant

Similar to nasal consonants, liquids and approximants are consistent in Orissa Sora. Thus, both Ramamurti (1938) and A. R. Zide (1982) propose three liquid consonants and one approximant in Orissa Sora. The liquid consonants in Orissa Sora include a trill consonant /r/, a lateral consonant /l/ and a retroflex tap consonant /ɽ/ and the approximant consonant includes a palatal approximant /j/. Additionally, although the three liquid consonants are uniformly present in the consonant inventory of Orissa Sora, both Ramamurti (1938) and A. R. Zide (1982) mention some phonotactic and ideolectal variations for these consonants. Regarding the trill consonant /r/ and lateral consonant /l/, Ramamurti (1938) suggests that their place of articulation varies between dental and alveolar when the two sounds occur in the environment of a dental sound such as /t/ or an alveolar such as /d/. Similarly, regarding the retroflex tap /ɽ/ Ramamurti (1938) mentions that the sound is distinctly pronounced only in remote areas whereas, most Sora speakers tend to replace it with /r/ or /d/. Regarding the trill consonant /r/ A. R. Zide (1982) mentions that a cluster consonant [dr] is frequently produced as a variation of trill consonant /r/ in word initial position. However, it is evident that although the three liquid consonants have some phonotactic and ideolectal variation, the phonemes are consistently present in both the descriptions of Orissa Sora discussed above. Also, variation regarding the approximant /j/ is not reported in any of the works mentioned here.

Hence, it is found that similar to the vowel phonemes, Orissa Sora consonant phonemes also lack a consistent description. As a result, more than one consonant inventory is proposed for the same language. However, there are similarities in a majority of the consonant phonemes proposed by different scholars in their individual

consonant inventories. Thus, consonant phonemes commonly identified by all scholars include the five oral stop consonants /p, b, t, k, g/; the four nasal consonants /m, n, ɲ, ŋ/; a voiceless fricative consonant /s/; a trill consonant /r/; a lateral consonant /l/; a retroflex tap consonant /ɽ/, a glottal stop consonant /ʔ/ and a palatal approximant consonant /j/. Additionally, the common consonants identified in Orissa Sora are typologically similar to the Munda consonant system. Also, typical Munda consonant feature such as the voicing asymmetry between /t/ and /d/ is observed in Orissa Sora. On the other hand, some variation in Orissa Sora consonant system is observed in distinguishing consonant pairs such as a voiced alveolar stop [d] and a voiced retroflex stop [ɖ]; a voiced palatal stop [j], a voiced palatal fricative [ɟ] and voiced alveolar fricative [z] and a labial approximant consonant [w]. Significantly, it is observed that variation in these consonants are observed mainly due to the lack of proper description as well as due to the lack of adequate evidences. Hence, it is concluded that, although Orissa Sora segmental properties are studied in the past it is not sufficient to understand the segmental properties of Assam Sora, particularly considering the fact that Assam Sora speakers are separated from their place of origin. Therefore, in order to have a general understanding about the segmental properties of Assam Sora the following sections provide an overview of the vowels and consonant phonemes of Assam Sora.

### 2.3 Overview of Assam Sora vowel phonemes

Based on Assam Sora speech data and evidences of Horo & Sarmah (2015) it is known that Assam Sora has six vowel phonemes. Table 2.11 presents the six vowel phonemes of Assam Sora.

Thus, the vowel inventory of Assam Sora includes two front vowels whereby one is a high front vowel /i/ and the other is a mid front vowel /e/. Similarly, in the two

Table 2.11: Assam Sora vowel phonemes (Horo & Sarmah, 2015)

	Front	Central	Back
High	i		u
Mid	e	ə	o
Low		a	

back vowels one is a high back vowel /u/ and the other is a mid back vowel /o/. The two central vowels include a mid central vowel /ə/ and a low central vowel /a/. Data evidences for these six vowel sounds in Assam Sora are presented in Table 2.12 which also includes the phonemic distribution of the six vowels in Assam Sora.

Table 2.12: Phonemic distribution of Assam Sora vowels (Horo & Sarmah, 2015)

Vowel	Sora	English	Initial	Meaning	Medial	Meaning	Final	Meaning
/a/	ɕaʔa	snake	aŋgi	battle axe	araŋ	sour	baŋsa	good
/e/	ɕeʔe	red	eŋsi	finger ring	areŋ	stone	sese	choose
/i/	ɕiʔi	tooth	iŋsa	plough shaft	gariŋ	alliance	tusi	push
/o/	ɕoʔo	fruit	oŋro	kind of honey	iroŋ-or	myna	kisso	dog
/ə/	aŋəm	name	əʔ-na	to fan	urəŋ	bamboo	bəsə	salt
/u/	aŋum	urine	uŋru	heating	uruŋ-a	bring	asu	sick

The phonemic distribution of Assam Sora vowels suggest that, while the five peripheral vowels /i, e, o, u, a/ can occur in all phonemic environments, the mid central vowel /ə/ mostly occurs word medially and only rarely word initially and word finally. The same vowel distribution is observed in Assam Sora speakers of all four regions that have been recorded in this dissertation. Thus, this work did not find any regional variation of Assam Sora based on the vowel phonemes.

Significantly, Assam Sora vowel inventory appears distinct from the vowel inventory of Orissa Sora proposed by different scholars. Assam Sora vowel inventory may appear similar to Orissa Sora vowel inventory proposed by Ramamurti (1931) only if the four allophonic vowels [ɪ, ʊ, ö, ü] are excluded from his proposed inventory. Also, unlike Ramamurti speculates, phonemic vowel length is not found in Assam Sora, rather a phonetic vowel length is seen in Assam Sora that is typically used for

postulating word stress in the second syllable of disyllabic words (Horo & Sarmah, 2015). The vowel inventory of Orissa Sora proposed in P. Mohanty (1997) may also appear similar to Assam Sora vowel inventory, except that P. Mohanty (1997) does not include the mid central vowel /ə/ in Orissa Sora vowel inventory. Hence, smaller vowel inventories of Orissa Sora have some resemblance with the vowel inventory of Assam Sora. Whereas, the exceptionally large inventory with nine vowels proposed by Stampe (1963, 1965), which has gained consensus over Ramamurti's inventory, appears to be very distinct from the vowel inventory of Assam Sora. If it is accepted that Orissa Sora actually has nine vowels, it may be postulated that, Assam Sora vowel inventory has reduced and some phonemic distinction of Orissa Sora vowel phonemes are lost and have merged with the closest vowel phonemes in Assam Sora. This will also imply that the vowel inventory of Assam Sora has significantly changed after the speakers migrated from Orissa to Assam in the 19<sup>th</sup> century. Moreover, it will indicate that Assam Sora vowel inventory is moving towards the typical small system of Munda languages. However, this can be examined only by comparing the vowel data of Orissa Sora and Assam Sora and therefore Chapter 6 presents the comparison of vowel inventory in the two Sora varieties. Additionally, the chapter also compares synchronic speech data of Orissa Sora and Assam Sora to explain the the present day vowel phenomenon in Orissa Sora and Assam Sora with measurable evidences.

## **2.4 Overview of Assam Sora consonant phonemes**

Assam Sora consonant phonemes have not been described in the past, therefore the consonant inventory proposed here is primarily based on Assam Sora speech data recorded in this dissertation. The speech data recorded from Assam Sora speakers of all four regions reveal that Assam Sora has 18 consonant phonemes. Assam Sora consonant phonemes arranged according to their manner and place of articulation are

presented in Table 2.13.

Table 2.13: Assam Sora consonant phonemes

	Labial	Dent	Alveo	Post-Alveo	Retro	Palat	Velar	Glott
Plosive	p	b	t	d			k	ʔ
Fricative			s					
Nasal		m		n		ɲ	ŋ	
Lateral				l				
Trill				r				
Tap					ɽ			
Approx	w					j		
Affricate					ɟ			

From Table 2.13 it is found that Assam Sora consonants represent eight places of articulation including labial, dental, alveolar, post-alveolar, retroflex, palatal, velar and glottal and eight manners of articulation including stop, fricative, nasal, lateral, trill, tap, approximant and affricate. Data evidences for these classifications and a brief description of Assam Sora consonant phonemes are given bellow according to their manners of articulation.

### 2.4.1 Stops

Assam Sora has seven stop consonants representing five places of articulation namely labial, dental, alveolar, velar and glottal. Assam Sora stop consonants include two labial stops /p/ and /b/; a dental stop /t/; an alveolar stop /d/; two velar stops /k/ and /g/ and a glottal stop /ʔ/. Phonemic distribution of the stop consonants reveal that all stop consonants, except the bilabial stops /p/ and /b/, can occur only word initially and word medially. Whereas, the bilabial stops /p/ and /b/ can occur word initially, word medially as well as word finally. Also, all stop consonants, except the glottal stop /ʔ/, have voicing contrasts in Assam Sora. Thus, the consonants /p, t, k/ are voiceless stops and the consonants /b, d, g/ are voiced stops in Assam Sora. Voicing contrast in the six oral stop consonants of Assam Sora are exemplified in

Table 2.14.

Table 2.14: Voicing contrast in Assam Sora stop consonants

Stops	Sora	English
/b/	bənum	ant hill
/d/	dənum	hatching (create)
/g/	gənum	dreaming
/p/	pənum	warm
/t/	tənum	collecting, heap, deposit
/k/	kənum	handfull

Significantly, the voicing contrast of /t/ and /d/ in Assam Sora is similar to the asymmetric voicing contrast of /t/ and /d/ in Orissa Sora. Thus, the phoneme /t/ in Assam Sora is a voiceless dental stop and the phoneme /d/ is a voiced alveolar stop. However, it is observed that the asymmetric voicing between /t/ and /d/ is sometimes irregular in Assam Sora. As a result, some Assam Sora speakers tend to produce a voiced dental stop [d̪] in place of the voiced alveolar stop /d/. Evidences suggest that it is only a free variation and there are no environmental influences for this variation. On the other hand, production of the voiceless dental stop /t/ is more uniform in Assam Sora.

In addition to the six oral stops, Assam Sora also has a phonemic glottal stop /ʔ/ that occurs only word medially and word finally but, never word initially. Examples illustrating the phonemic distinction of the glottal stop in Assam Sora are presented in Table 2.15.

Table 2.15: Glottal stop in Assam Sora

Stops	Sora	English
/ʔ/	iʔi	louse
/t/	iti	what
/d/	idi	cut
/ʔ/	aniʔ	written
/b/	anib	root
/ʔ/	uʔba	brother
#	uba	pull out hand

The glottal stop also has phonetic variations in Assam Sora. In word medial

position a glottal stop is realized either as a complete voiceless glottal stop, a creaky voiced glottal stop, or a voiced glottal stop (Kalita et al., 2016). Similarly, in word final position a glottal stop may realize either as a complete glottal stop or as a creaky voiced glottal stop Kalita et al. (2017). These phonetic features of the glottal stop in Assam Sora are further explained in Chapter 5 of this work. Additionally, it is observed that some Assam Sora speakers also tend to aspirate the voiceless velar stop consonant /k/ to [k<sup>h</sup>] in the word initial position only. It is apparent that some Assam Sora speakers who frequently interact in Sadri with non Assam Sora speakers are starting to acquire the aspiration feature that is commonly found in Sadri. However, Chapter 5 of this work illustrates that, the pattern is not consistent among the participants of the this work and therefore it is proposed that the aspirated [k<sup>h</sup>] is found mostly as a free variation of the voiceless stop consonant /k/ in Assam Sora.

Thus, Assam Sora speech data used in this chapter has illustrated that Assam Sora has six oral stop consonants /p, b, t, d, k, g/ and one glottal stop consonant /ʔ/. Hence, it is evident that Assam Sora stop consonants have maximum similarity with the stop consonants of Orissa Sora proposed by Ramamurti (1938). Also, there is similarity in the /t/ and /d/ voicing asymmetry feature in the two Sora varieties. This also confirms a significant typological similarity with the Munda consonant system. However, unlike A. R. Zide (1982) proposes, a voiced palatal stop /ɟ/ is not found in Assam Sora, but from a review of Munda segmental features, it is known that a voiced palatal stop is also not common in Munda languages.

## 2.4.2 Nasals

Assam Sora has four nasal consonants representing four distinct places of articulation namely labial, alveolar, palatal and velar. Thus, the nasal consonants in Assam Sora include a bilabial nasal /m/, an alveolar nasal /n/, a palatal nasal /ɲ/ and a velar

nasal /ŋ/. Examples illustrating distinction between the four nasal consonants in Assam Sora is presented in Table 2.16.

Table 2.16: Nasal consonants in Assam Sora

Nasals	Sora	English
/m/	mama	get away!
/n/	nama	this
/ɲ/	ɲama	touch
/m/	ajim	to remember
/n/	ajin	he
/ŋ/	ajiŋ	sister
/m/	ɲama	touch
/ŋ/	ɲaŋa	bow
/n/	ɲanaŋ	learn

Phonemic distribution of the four nasal consonants suggest that, while the bilabial nasal /m/ and the alveolar nasal /n/ occur in all phonemic environments, the palatal nasal /ɲ/ occurs only word initially and word medially, but never word finally. The velar nasal /ŋ/ occurs only word medially and word finally, but never word initially. Also, in word final position it is observed that, some nasal consonants do not distinguish their place of articulation. For example, the word for ‘boil’ in Assam Sora is produced as *asim* as well as *asin*. Similarly, the the word for ‘what’ is produced as *itin* as well as *itiŋ*. In this regard, studies such Ramamurti (1931) and Starosta (1976) suggest that a word final nominal nasal /-n/ is often dropped in Orissa Sora due to various morphophonological processes. However, the motivation for alternation between nasal place of articulation of the nasal consonant in the word final position is not yet known. Therefore, data in this chapter has mainly identified the four nasal consonants in Assam Sora and Chapter 5 has described some acoustic characteristics of the nasal sounds in Assam Sora. Hence, from the characteristics of the four nasal phonemes of Assam Sora it is evident that the nasal phonemes of Orissa Sora proposed by both Ramamurti (1938) and A. R. Zide (1982) are similar to the nasal phonemes of Assam Sora. Also, presence of the four nasal phonemes in Assam Sora

confirms their uniformity in the typology of Munda consonant system.

### 2.4.3 Fricative and Affricate

Assam Sora has one fricative consonant /s/ and one affricate consonant /ɕ/. The fricative consonant /s/ is a alveolar sound in Assam Sora and the affricate consonant is a post-alveolar sound in Assam Sora. The two consonants also differ in their voicing feature such that, /s/ is a voiceless fricative consonant and /ɕ/ is a voiced affricate consonant in Assam Sora. Table 2.17 exemplifies the voicing contrast of the two consonants /s/ and /ɕ/ in Assam Sora.

Table 2.17: Fricative and affricate consonants /s/ and /ɕ/ in Assam Sora

	Sora	English
sə	səkur	rough
ɕə	ɕəkur	form into lump like cooked rice
sa	sabda	to pound ragee millet
ɕa	ɕabda	to suck or munch with noise
se	sedsed	to throw in disorder
ɕe	ɕedɕed	to importune, pressingly (as beggars)
si	siʔi	hand
ɕi	ɕiʔi	tooth
so	soʔo	foul smell
ɕo	ɕoʔo	fruit
su	subsub	to tell lies
ɕu	ɕubɕub	sound of foot steps

Phonemic distribution of the two consonants suggest that both /s/ and /ɕ/ can occur word initially and word medially but, they never occur word finally in Assam Sora. Moreover, it is observed that the voiceless alveolar fricative /s/ is sometimes produced as a voiceless post-alveolar affricate [tʃ] by some Assam Sora speakers. For instance, the word for ‘dog’ is sometimes produced as *kittʃo* instead of *kisso*. Thus, it is likely that the fricative consonant /s/ was originally and an affricate consonant [tʃ] in Assam Sora which is also common in the typology of Munda language. However, the participants in this work have mostly produced the voiceless alveolar fricative,

therefore this chapter considers /s/ as the primary consonant in Assam Sora. Moreover, it is observed that sometimes the voiced post-alveolar affricate is also produced as a voiced alveolar fricative [z]. For example, the word for ‘fruit’ is sometimes produced as zoʔo instead of ɕoʔo. However, since a voiced post-alveolar affricate is more consistently produced, this chapter considers /ɕ/ as the primary consonant in Assam Sora.

Hence, Assam Sora fricative and affricate consonants are only partly similar to the fricative and affricate consonants of Orissa Sora. While the voiceless alveolar fricative /s/ is reported by most scholars of Orissa Sora, the voiced post-alveolar affricate /ɕ/ is not reported in Orissa Sora. Rather, Ramamurti (1931, 1938) proposes the presence of a voiced palatal fricative /j/ in Orissa Sora, which is not found in the Assam Sora data used in this chapter. However, a voiced alveolar fricative [z], that is often treated as an allophonic sound in Orissa Sora, also occurs in place of the voiced post-alveolar affricate consonant /ɕ/ in Assam Sora. Additionally, the fricative and the affricate consonant are also unique consonants in the Munda consonant system. It is known that, except Gorum (Anderson & Rau, 2008) and Remo (Anderson & Harrison, 2008a), a voiced alveolar fricative is not common to Munda consonant system and usually a voiced post-alveolar affricate /ɕ/ is present in most of the modern Munda languages. Hence, the presence of these two consonants in Assam Sora indicate an important resemblance between Assam Sora and Munda languages.

#### 2.4.4 Liquids and Approximants

Speech data reported in this chapter provides evidence that Assam Sora has three liquid consonants and two approximants. The liquid consonants include an alveolar trill /r/, an alveolar lateral /l/ and a retroflex tap /ɽ/ and the approximants include a labial approximant /w/ and a palatal approximant /j/. Examples of these liquids and approximants in Assam are presented in Table 2.18.

Table 2.18: Liquid consonants in Assam Sora

Liquids	Sora	English
/r/	kuru	hair
/ɽ/	kuɽu	rice
/r/	karu	monkey
/ɽ/	kaɽu	blue
/j/	jaŋ	mother
/r/	raŋa	deny
/l/	laŋga	good
/w/	waŋ	father

In the liquid consonants it is observed that, while the alveolar trill /r/ and the alveolar lateral /l/ can occur in all three phonemic environments, the retroflex tap /ɽ/ occurs only word medially but never word initially or word finally in Assam Sora. Similarly, among the approximant consonants it is observed that palatal approximant mainly occurs word initially and word medially and the labial approximant /w/ occurs only word initially. Moreover, in word final position the palatal approximant /j/ often forms an off-glide with the preceding vowel. Thus, it is evident that Assam Sora and Orissa Sora have similar liquid and approximant consonants. However, the labial approximant /w/ found in Assam Sora does not occur consistently in Orissa Sora. It is seen on in the consonant inventory proposed by Stampe (1965), but since the data evidences of his work are not accessible the distinctive nature of /w/ in Orissa Sora is not known.

Hence, the description above provides evidence that Assam Sora has 18 consonant phonemes representing eight places of articulation namely Bilabial, Dental, Alveolar, Post-Alveolar, Retroflex, Palatal, Velar and Glottal and eight manners of articulation namely Stops, Fricative, Nasals, Lateral, Trill, Tap Approximant and Affricat. The 18 consonants presented in Table 2.13 are uniformly produced by Assam Sora speakers of all four regions. Therefore, this work does not distinguish any regional variation of Assam Sora on the basis of its consonant phonemes. Moreover, from the description of the consonant inventory in Assam Sora it is observed that most

of the consonant phonemes are similar to the consonant phonemes of Orissa Sora, particularly the consonants that are commonly found in all the consonant inventories proposed by different scholars. Also, similar to the vowel phonemes, Assam Sora consonant phonemes have maximum similarity with the typology of Munda consonant system. Hence, there are no significant indications that there is a diachronic change in the consonant inventory of Assam Sora from the consonant inventory of Orissa Sora. However, considering the fact that Assam Sora speakers are completely separated from their place of origin it is important to examine the synchronic relation between the two Sora varieties. Therefore, phonetic features of some common consonants in Assam Sora are instrumentally examined in Chapter 5 and later those features are compared with the synchronic speech data of Orissa Sora in Chapter 6.

## 2.5 Conclusions

This chapter has mainly reviewed the segmental properties of Austroasiatic languages, Munda languages and Orissa Sora as it is represented in some of the relevant literature on these languages. Additionally, this chapter has presented an overview of the segmental properties of Assam Sora based on speech data recorded from native Assam Sora speakers living in four villages of Assam. From the literature review this chapter has concluded that, besides their historical classifications, Austroasiatic languages are also grouped into two geographical categories namely Southeast Asian Austroasiatic languages and South Asian Austroasiatic languages. In this regard, past studies have concluded that Austroasiatic languages exhibit areal features of the geographical locations where they are spoken (Diffloth & Zide, 1992). As a result, Austroasiatic languages of Southeast Asia and South Asia are considered to be typologically diverged from each other. Thus, evidences based on segmental properties suggest that Southeast Asian Austroasiatic languages such as Mon-Khmer have

large segmental inventories. Whereas, South Asian Austroasiatic languages such as Munda have small segmental inventories. Hence, it is found that Southeast Asian Austroasiatic languages usually have 9 or 10 vowels in their vowel system whereas, South Asian Austroasiatic languages such as Munda and Kharia have only 5 or 6 vowels in their vowel systems. Also, studies have proposed that while the Southeast Asian Austroasiatic languages are typically similar to the Proto-Austroasiatic language, the South Asian languages have diverged from the parent language. Thus, the segmental properties of Mon-Khmer languages are considered as native to Proto-Austroasiatic speech segments whereas, the segmental properties of Munda languages are considered to be very diverged from the Proto-Austroasiatic segmental properties (Donegan, 1993). For instance, implosive consonants present in Southeast Asian Austroasiatic languages are considered as native Austroasiatic sounds and are argued to be lost in South Asian Austroasiatic languages. On the other hand, the retroflex consonants present only in South Asian Austroasiatic languages are argued to be non-native Austroasiatic sounds. Thus, based on this evidences most of the studies conclude that the present day Southeast Asian Austroasiatic languages best represent the Proto-Austroasiatic segmental properties (Sidwell & Rau, 2014).

From reviewing the segmental properties of Munda languages, this chapter has concluded that while the consonant inventory is fairly well understood, there are some complexities in the vowel inventory. Consequently, it is found that the consonant system of all Munda languages are typologically similar, except the retroflex consonants, most Munda consonants have clear correspondences with consonant system of other Austroasiatic languages. On the other hand, vowel system of Munda languages appear significantly diverged from the vowel system of other Austroasiatic languages of Southeast Asia. It is found that most Munda languages lack vowels such as mid front and back vowels [ɛ] and [ɔ] and central high and mid vowels [i] and [ə]. These vowels are typical found in Southeast Asian Austroasiatic languages. There-

fore, their absence in Munda languages make the Munda vowel system typologically diverged from the vowel system of non-Munda Austroasiatic languages. However, some studies claim that Orissa Sora is a preservative Munda languages and these vowel sounds are well preserved in Orissa Sora. Therefore, it is argued that unlike the typical five vowel system of Munda languages, Orissa Sora has a vowel system of nine vowel phonemes. Also, it is hypothesized that the nine vowel system corresponds to the Proto-Munda vowel systems as well as the vowel system of modern Austroasiatic languages of Southeast Asia (Donegan, 1993). Thus, past studies have concluded that although Munda segmental properties appear diverged from non-Munda Austroasiatic languages, there are evidences that Munda languages are still significant members of the Austroasiatic language family (Anderson & Zide, 2002).

However, from reviewing the segmental properties of Orissa Sora it is concluded that, Orissa Sora segmental properties are inadequately described, therefore there are inconsistencies in the existing descriptions. As a result, the literature review in this chapter revealed that more than one segmental inventory of Orissa Sora is proposed by different scholars. Although the segmental inventory of Orissa Sora is proposed by more than one scholar, a description about their characteristics is very briefly discussed in Ramamurti (1931, 1938) and only roughly presented in A. R. Zide (1982). Therefore, Orissa Sora segmental features are reviewed mostly from the contributions of these two works. These studies reveal that while the consonant inventory of Orissa Sora is fairly uniform, the vowel inventory is significantly different. Therefore, this chapter concludes that existing studies do not provide a conclusive description of Orissa Sora segmental properties.

Finally, from the overview of Assam Sora segmental properties, this chapter has concluded that Assam Sora has 6 vowels and 18 consonants. Also, it is concluded from this chapter that Assam Sora speakers living in the four villages of Assam have the same segmental inventory. Therefore it is confirmed that there are no regional

variations of Assam Sora on the basis of its segmental inventory. Moreover, this chapter has concluded that while the consonant phonemes of Assam Sora and Orissa Sora are fairly similar, the vowel inventory of Assam Sora appears to be distinct from Orissa Sora. Specifically, considering the claim that Orissa Sora has nine vowels, the Assam Sora vowel system appears to be reduced in size. However, it also indicates that Assam Sora vowel system has become typologically similar to the small vowel system of Munda languages. Thus, this chapter concludes that segmental properties of Orissa Sora and Munda languages are mostly preserved in Assam Sora and some variation is observed mainly due to the inconsistent descriptions of Orissa Sora. As a result, the diachronic relation of Orissa Sora and Assam Sora is slightly confusing and therefore, this chapter motivates the comparison of the synchronic speech data of Assam Sora and Orissa Sora in the subsequent chapter.

## Chapter 3

# Syllable structure in Assam Sora

### 3.1 Introduction

This chapter describes the syllable structure and phonotactic constraints of Assam Sora and discusses in detail the acoustic characteristics of syllable nucleus in disyllabic words. Describing syllable structure is essential to the description of segmental properties because a syllable is the basic structural unit of a meaningful word and therefore syllable structure has an important role in the phonology of a language. Syllables consists of sequences of sound segments identified as consonants (C) and vowels (V) and a typical syllable usually consists of a vowel segment which may or may not be preceded and/or followed by one or more consonant segments. In a syllable structure the vowel segment is known as syllable *nucleus*, consonant segments preceding the vowel nucleus is called syllable *onset* and consonant segments following the vowel nucleus are called *coda* consonant. Moreover, in a syllable structure only the nucleus is an obligatory constituent, whereas the onset and coda consonants are both optional. Therefore, the typical structure of a syllable is (C)V(C) where ‘V’ represents the vowel nucleus and ‘(C)’ represents the optional onset and coda consonants. Typologically a ‘CV’ syllable structure is more common in languages of the world whereas syllable structures such as ‘VC’ or ‘CCV’ are less common (Davenport et al., 2010). This indicates that the smallest meaningful word in a language can

have a CV syllable structure. The universal tendency can be understood by considering the phonetic and phonological principles that determine the syllable structure in different languages.

The phonetic basis of a syllable structure is often defined by the notion of sonority peak, since sonority peaks help in identifying the number of syllables in a given utterance (Mannell et al., 2009). Sonority is a speech sound phenomena and speech sounds generally have different degrees of sonority. Factors determining sonority in speech sounds include degree of vocal constriction of a speech sound, sustainability of a speech sound, and the loudness of a speech sound. Thus, sonorous sounds have least vocal constriction, can be produced for longer duration, and are louder than non-sonorous sounds. Therefore, based on these properties it is known that vowels are the most sonorous and obstruents are the least sonorous speech sounds. Moreover, nasal and liquid consonants occupy an intermediate position between the most sonorous and the least sonorous speech sounds. Hence, since vowels occur in the nucleus position of a syllable, it is possible to identify a syllable from its sonority peak. This indicates that sonority in a syllable gradually increases from the onset position and reaches the highest peak at the nucleus and then gradually declines in coda position. Thus, sonority peaks can provide information about the number of syllables in a speech utterance.

However, sonority peaks in a syllable does not give any information about syllable boundaries. Therefore, it is often difficult to identify syllable boundaries particularly in case of polysyllabic words. A polysyllabic word has more than one syllable and it is difficult to mark the beginning and end of individual syllables in such words. Generally native speakers of a language are able to identify the number of syllables in a word without understanding the notion of sonority peaks and they are also able to identify syllable boundaries intuitively. However, this is not applicable to non-native languages. Therefore, for identifying syllable boundaries of a non-native language a

phonological principle known as *onset maximization* is generally applied (Davenport et al., 2010). The principle of onset maximization suggests that in case of polysyllabic words a medial consonant should be syllabified as a onset consonant rather than as a coda consonant. Thus, the principle of onset maximization reinforces the fact that languages generally prefer to have open syllables, and therefore a CV syllable structure is a frequently occurring syllable structure in languages of the world. However, in languages that allow both onset and coda consonants to occur, such languages also have a CVC syllable structure. Thus, beside a single onset, many languages also allow a single coda in their syllable structure.

Syllable structure with single onset consonant and single coda consonant is typologically more common. However, some languages also allow more than one consonant in the onset position as well as in the coda position. A maximum of two consonants is usually permissible in a number of languages, but some languages allow consonant sequences up to six consonants. However, there are constraints regarding the positioning and number of phonemes in a syllable structure and such constraints are known as phonotactic constraints (Mannell et al., 2009). Phonotactic constraints basically limit the number of permissible phonemes in a syllable structure and also define the sequence of permissible consonants in onset and coda positions of a syllable structure. For instance, in English, while there can be three consonants including /spr/ in the onset position of a word such as 'sprint', there can be no words that has onset consonants /dm/ or /bm/. Thus, phonotactic constraints not only control the number of possible consonants but also define the specific consonants that can form a sequence in a syllable structure. However, phonotactic constraints are generally language specific. This indicates that consonant sequences allowed in the onset position of a particular language may not be permissible in another language. Additionally, consonant sequences in a syllable structure also abide by the sonority principle. This ensures that, a least sonorous consonant should always precede the

more sonorous consonant in the onset position and vice versa in the coda position. Thus, both phonotactic constraints and sonority principles can explain the internal composition of a syllable structure and can also determine whether a particular sound sequence is a valid syllable or not.

Thus, it is clear that describing syllable structure primarily includes identifying permissible syllable templates and determining the shape and length of minimal words in languages. Also, the phonological basis of forming longer words can be understood from the constraints that regulate the sequencing of two or more syllables within a single word. Significantly, a relative relation between adjacent syllables of longer words is also an important feature of syllable structure across languages. In some languages the longer words exhibit a tendency to make one syllable stronger or more prominent than the adjacent syllable. This phenomena is known as stress (Kager, 2007), it indicates that the stronger syllable bears stress and the weaker syllable is unstressed. However, stress is not an inherent feature of the syllable, it is a relative relation of strength or prominence between stressed and unstressed syllables. Thus, stress is a binary feature that is composed of a stressed and a unstressed syllable and the binary unit of stress is generally known as a foot (Davenport et al., 2010). Moreover, depending on the position of stressed and unstressed syllable within a foot there can be two types of foot structures. If the stressed syllable precedes the unstressed syllable it is known as *trochee* foot, and if the stressed syllable follows the unstressed syllable it is known as *iambic* foot. Thus, stress in many languages are primarily categorized on the basis of these two foot structures. While some languages such as Czech, French and Turkish have a consistent foot structure, in languages such as English the foot structure may vary depending on the morphophonological function of stress. However, the focus of this chapter is to examine the physical or phonetic properties of word stress and explaining further the morphophonological functions of stress are not within the scope of this work.

Examining the phonetic properties of word stress is primarily useful in distinguishing stressed syllables from unstressed syllables. Also, it provides significant information about the quality of vowel nuclei in stressed syllables. Therefore, it is important to examine the phonetic properties of word stress particularly the features that attribute prominence to stressed syllables. Prominence is generally attributed to stressed syllables through additional effort in its production that is perceivably distinct from unstressed syllables. This indicates that, stressed syllables are produced with greater articulatory movement that results in making the stressed syllables sound longer, louder and pitched higher than the unstressed syllables. These features are extensively studied to describe the phonetic properties of stressed syllables across languages and evidences suggest stressed syllables are effectively distinguished on the basis of these properties. Therefore, it is widely accepted that the phonetic basis of stressed syllables include features such as duration, intensity and fundamental frequency. Hence, in addition to describing the syllable structure and phonotectic constraints of Assam Sora in §3.2, this chapter also explores the typology of word stress in Austroasiatic and Munda languages in §3.3 and describes the acoustic phonetic analysis of word stress in Assam Sora in §3.4. Finally, the findings and observations of the chapter are concluded in §3.5.

## **3.2 Syllable structure in Assam Sora**

According to Jenny et al. (2014) Austroasiatic languages usually have a CV(C) syllable and both monosyllabic and disyllabic words are commonly found in Austroasiatic languages. However, monosyllabic words are mostly found in Austroasiatic languages of Southeast Asia and rarely in Austroasiatic languages of South Asia. For instance, Khasi, a Mon-Khmer language of the Austroasiatic family, has frequently occurring monosyllabic words that have (C)CV(C) syllable structure. On the other hand,

Munda languages usually have a (C)V(C) syllable structure and monosyllabic words are rarely found in Munda languages. Similarly, regarding the syllable structure of Orissa Sora Stampe (1965) reports that the shape of Sora syllable is (C)V(C) and that a Sora morpheme usually consists of one or two syllables but rarely three syllables. Likewise, data reported in this chapter also reveals that the shape of Assam Sora syllable is (C)V(C). However, unlike Orissa Sora, it is observed that monosyllabic words are rare rather disyllabic words appear to be the minimal word template in Assam Sora. Also, shape of both the syllables in disyllables are observed to be (C)V(C). Examples 1 to 12 illustrate the frequently occurring syllable combinations in Assam Sora disyllables.

1. CV.CV    *ɕelu*    'meat'
2. CV.CVC    *datul*    'trap'
3. CVC.CV    *gulɕi*    'seven'
4. CVC.CVC    *kampuŋ*    'stomach'
5. CVC.V    *daʔa*    'water'
6. CVC.VC    *taŋor*    'road'
7. V.CV    *əli*    'rice beer'
8. V.CVC    *usal*    'skin'
9. VC.V    *iʔi*    'louse'
10. VC.VC    *eŋer*    'gooseberry'
11. VC.CV    *arsi*    'long tailed monkey'
12. VC.CVC    *onder*    'finger'

Additionally, onset clusters are also observed in second syllables where the first segment is always a stop consonant, generally voiced, and the second segment is either a trill consonant /r/ or a retroflex consonant /ɽ/. Examples 1 to 6 illustrate some of the onset clusters found in Assam Sora disyllables.

1. CVC.CCV *kumbɽi* ‘buttock’
2. CVC.CCV *sindri* ‘cloth’
3. VC.CCV *engre* ‘a garden herb’
4. VC.CCV *antre* ‘look’
5. CVC.CCV *ɕampra* ‘disheveled’
6. CV.CCV *bokɽi* ‘a while ago’

However, it is observed that cluster consonants are rare in onset position of first syllable and consonant clusters never occur in the coda position of either syllables. Regarding the phonotactic constraints in Orissa Sora Stampe (1965) proposes that all consonants, except velar nasal /ŋ/, can occur in syllable onsets, whereas for coda consonant he proposes that only /b, d, ʃ, m, n, ɲ, l, r and j/ can occur postvocally. Moreover, Stampe (1965) mentions that in coda position stop consonants are usually devoiced or glottalized and the palatal consonant /ʃ/ is produced as a glottalized palatal approximant [ʔj]. Also, the trill consonant /r/ is argued to be produced as a cluster consonant [dr] in the onset position. Similarly, in case of Assam Sora, data examined in this chapter reveals that while all six vowel segments /a, i, e, o, u, ə/ can occur in first and second syllable, some consonant segments are restricted in the syllable boundaries. Assam Sora consonants that can occur in onset and coda position of first and second syllables are presented in Table 3.1.

From Table 3.1 it is observed that all consonants, except /ŋ and ʔ/, can occur in the onset position, but only the consonants /p, b, l, m, n, ɲ, r, ʔ/ can occur in the

Table 3.1: Assam Sora onset and coda consonants

	b	d	g	p	t	k	s	ɕ	m	n	ɲ	j	r	l	w	ɽ	ʔ
Onset <sup>1</sup>	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	!	✓	✓	✓	!	!
Onset <sup>2</sup>	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	!	✓	✓	!	✓	!
Coda <sup>1</sup>	✓	!	!	✓	!	!	!	!	✓	✓	!	✓	!	✓	!	!	✓
Coda <sup>2</sup>	✓	!	!	✓	!	!	!	!	✓	✓	!	✓	!	✓	!	!	✓

<sup>1</sup> Syllable 1; <sup>2</sup> Syllable 2

coda position. However, two consonants including the labial approximant /w/ and the retroflex tap /ɽ/ exhibit an additional phonotactic constraint in Assam Sora. It is found that while the consonant /w/ as in *wap* meaning ‘leaf’ can occur only in the onset position of the first syllable, the consonant /ɽ/ as in *kaɽip* meaning ‘long sword’ can occur only in the onset position of the second syllable. On the other hand, the coda consonants occur uniformly in both the syllables.

Thus, the description of Assam Sora syllable structure reveals that the typical syllable structure of Assam Sora is (C)V(C). Moreover, evidences suggest that monosyllabic words are rarely found, and the minimal word template in Assam Sora is disyllabic. Therefore, all the analysis in this work is done by using only the disyllabic words in Assam Sora consisting of all disyllabic combinations that are exemplified in this chapter. Also, from the examination of phonotactic constraints in Assam Sora syllable structure it is evident that the inventory of vowel segments in first and second syllable of disyllabic words are always similar. Thus, all Assam Sora vowels can occur in both the syllables of a disyllabic word. Likewise, the inventory of onset and coda consonants in Assam Sora are also uniform in both the syllables of a disyllabic word. However, two Assam Sora consonants including the labial approximant /w/ and the retroflex tap /ɽ/ are found to be restricted in the onset position of either the first syllable or the second syllable. Moreover, it is noticed that consonant clusters are also found in Assam Sora, but they are mostly restricted in the onset position of the second syllable only. Also, consonant clusters in Assam Sora are usually formed by combining a stop consonants such as, /b, d, g, p, t, k/ and liquid consonants such as,

/r/ and /ɽ/.

Additionally, besides these characteristics of the syllable structure in Assam Sora, the study by Horo & Sarmah (2015) proposes that Assam Sora syllable structure is particularly characterized by the presence of an iambic stress pattern in disyllabic words. This implies that, in disyllabic words the second syllable is more prominent than the first syllable and the claim is substantiated with acoustic analysis. Therefore, it is the interest of this work to extend this finding by incorporating additional data, since the study by Horo & Sarmah (2015) is based on Assam Sora speaker of one particular region, namely, Singrijha. Thus, the subsequent sections reexamine the phonetic realization of word stress in Assam Sora with acoustic analysis and by incorporating speech data from three additional regions namely Sessa, Koilamari and Lamabari. Additionally, prior to analysis, the following section reviews some of the relevant literature on the description of word stress in Austroasiatic and Munda languages.

### **3.3 Word stress in Austroasiatic and Munda languages**

Word stress in Austroasiatic languages is often related to the presence of a unique syllable type commonly known as ‘sesquisyllable’. Sesquisyllabic words consist of two syllables where an initial unstressed (minor) syllable is always followed by a stressed (major) syllable. Moreover, while the major syllable is characterized as a normal syllable, the minor is characterized as phonologically and phonetically reduced. This implies that only short centralized vowels can occur in the minor syllable, whereas the major syllable can include the full vowel inventory (Matisoff, 1973). Therefore, the weak or the ‘minor’ syllable is often referred as sesquisyllable or a half syllable. Presence of sesquisyllabic word prosody is reported in many Austroasiatic languages

of Southeast Asia and it is well known that sesquisyllabic word prosody is atypical to Austroasiatic languages as well as to many other languages spoken in Southeast Asia. On the other hand, it is argued by Donegan (1993) and Donegan & Stampe (1983, 2004) that sesquisyllabic word prosody is lacking in Munda languages. Evidences suggest that sesquisyllabic words once existed in proto Munda, but the modern Munda languages lost it due to rhythmic and syntactic reorganization in various languages.

However, Anderson & Zide (2002) and Anderson (2015) propose that phonological words in Munda languages exhibit iambic stress and they argue that the iambic stress resembles to the sesquisyllabic word prosody of Southeast Asian Austroasiatic languages. Evidences suggest that weak-strong stress patterns in Munda languages are often maintained by a bimoraic constraint on the phonological word structure. According to Anderson & Zide (2002), proto Munda root-element was commonly monosyllabic and monomorphemic in form. Whereas, in modern Munda languages the free form is often a longer word. Therefore, Anderson & Zide (2002) proposes that, free forms in Munda languages are obligated by a bimoraic constraint. The bimoraic constraint necessitates the root-element to combine with some derivational process such as affixation, reduplication or compounding. As a result, monosyllabic words become rare and disyllabic words become the minimal word template in Munda languages. This is observed in Gta? as well as in Assam Sora. Subsequently, the bimoraic constraint facilitates the realization of iambic stress in Munda languages. Thus, Munda languages such as Santali (Ghosh, 2008) and Plains Remo (Anderson & Harrison, 2008a) are reported to have iambic stress and evidences suggest that in these languages the stress is always on the second syllable. Similarly, the study on Assam Sora Horo & Sarmah (2015) has also provided instrumental evidences regarding the presence of an iambic stress pattern in Assam Sora.

Additionally, Anderson (2015) argues that the phonological word in Munda languages has different word prosody and phrasal prosody and the iambic word prosody

of Munda languages resembles the sesquisyllabic word prosody of Austroasiatic languages. Also, Anderson (2015) clarifies that, the only difference between a weak syllable in Munda languages and a sesquisyllable in Austroasiatic languages of Southeast Asia lies in the syllable nuclei of the weak syllable. Unlike the weak sesquisyllable nuclei, that is always reduced to a centralized vowel, the weak syllable nuclei in Munda languages includes all contrasting vowel segments. Thus, these studies imply that presence of Iambic stress in Munda languages is an important segmental features. Evidences suggest the stress pattern helps in unifying the Austroasiatic languages of South and Southeast Asia, that are otherwise considered to be very diverge from each other. Therefore, the following section provide the relevant acoustic evidences for iambic stress pattern that has been observed in Assam Sora by incorporating speech data Assam Sora speakers living in different geographical locations.

### **3.4 Acoustic analysis of word stress in Assam Sora**

A previous study on word stress in Assam by Horo & Sarmah (2015) has shown that, in disyllables the syllable nuclei in second syllable is always longer, louder and pitched higher than the syllable nuclei in first syllable. Also, these properties of Assam Sora disyllables is validated by examining the acoustic features of syllable nuclei namely vowel duration, vowel intensity and fundamental frequency. Thus, the three acoustic features of syllable nuclei provide evidence that word stress in Assam Sora is iambic whereby, an unstressed syllable is always followed by a stressed syllable. However, the acoustic features of word stress in Assam Sora has been mainly described for Assam Sora speakers living in Singrijhan only and it is not known whether Assam Sora speakers living in different geographical locations of Assam have the same stress pattern or not. Therefore, in order to verify whether the iambic stress reported in Assam Sora is consistently maintained across different regions or

not, the following sections extend the investigation of word stress in Assam Sora disyllables by incorporating speech data recorded from Assam Sroa speakers of three other regions including Sessa, Koilamari and Lamabari. Hence, speech data recorded from these regions are examined separately in this section and the analysis mainly includes examination of the three acoustic characteristics of word stress in Assam Sora disyllables namely vowel duration, fundamental frequency and vowel intensity. Thus, the three acoustic characteristics in first and second syllable of Assam Sora disyllables are examined in the subsequent sections.

### **3.4.1 Vowel duration in first and second syllable of Assam Sora disyllables**

Presence of long vowels in the syllable nuclei is an important property of stressed syllables in languages of the world. Therefore, the acoustic correlate of vowel length, namely vowel duration, plays an important role in identifying stressed syllables crosslinguistically (Fry, 1955). Significantly, one of the primary distinctions between the syllable nucleus of first and second syllable in Assam Sora disyllables is also vowel duration. The study by (Horo & Sarmah, 2015) reports that Assam speakers generally tend to produce a longer vowel in the second syllable in comparison to the first syllable. Therefore, evidences suggest that vowel duration is an important cue of word stress in Assam Sora. However, this property of word stress has been mainly described for Assam Sora speakers living in Singrijhan region only and therefore it is know known whether Assam Sora speakers living in other geographical regions have the same distinction or not. Therefore, this chapter extends the examination of vowel duration in Assam Sora disyllables by incorporating speech data recorded from Assam Sora speakers living in three additional regions of Assam including Sessa, Koilamari and Lamabari. Thus, vowel duration in first and second syllable of Assam Sora disyllables is calculated separately for all these regions as well as for Sigrijhan.

This includes calculating the time interval between vowel start and vowel end of every vowel in first and second syllable of Assam Sora disyllables. Thus, Table 3.2 presents the region-wise average vowel duration distinction between first and second syllable with standard deviation in parenthesis.

Table 3.2: Average vowel duration in first and second syllable

	Singrijhan	Sessa	Koilamari	Lamabari
Syllable 1	83.72 (38.14)	77.13 (31.06)	84.6 (41.95)	74.27 (30.83)
Syllable 2	153.62 (58.47)	105.74 (44.99)	159.42 (70.37)	135.92 (49.34)

By incorporating speech data from Assam Sora speakers living in Sessa, Koilamari and Lamabari it is found that all Assam Sora speakers naturally lengthen the vowel nucleus in the second syllable of a disyllabic word. As a result, the vowel nucleus in first syllable becomes relatively shorter. This difference can be seen in the average vowel duration calculated separately for first and second syllable of Assam Sora disyllables. Thus, from Table 3.2 it is clearly evident that vowel duration is always longer in second syllable and relatively shorter in first syllable. Also, this distinction is found to be uniformly present across speakers of all four regions recorded in this chapter. Additionally, in order to reduce the speaker effect on vowel duration a speaker normalization method is adopted here by using the *z-score* equation as in (3.1).

$$z = \frac{x - \mu}{\sigma} \quad (3.1)$$

In equation (3.1)  $z$  represents the speaker normalized vowel duration value. Vowel duration produced by individual speakers is represented by  $x$ ,  $\mu$  represents the average vowel duration of all speakers and  $\sigma$  represents the standard deviation in the vowel duration of all speakers. Vowel durations are normalized because some Assam Sora speakers have a tendency to produce an extra long vowel in the second syllable. For instance, in Assam Sora speech data recorded in Koilamari, while average vowel

duration in the second syllable is 159.42 milliseconds, one female speaker produced vowels in the second syllable as long as 500 milliseconds. Similar speaker variations are observed in Assam Sora speech data recorded in the other three regions as well. Therefore, the speaker variation is normalized in order to visualize the vowel duration distinction based on syllable position only. A graphical representation of region-wise speaker normalized vowel duration distinction in first and second syllable is presented in Figure 3.1.

Figure 3.1: Speaker normalized vowel duration in first and second syllable

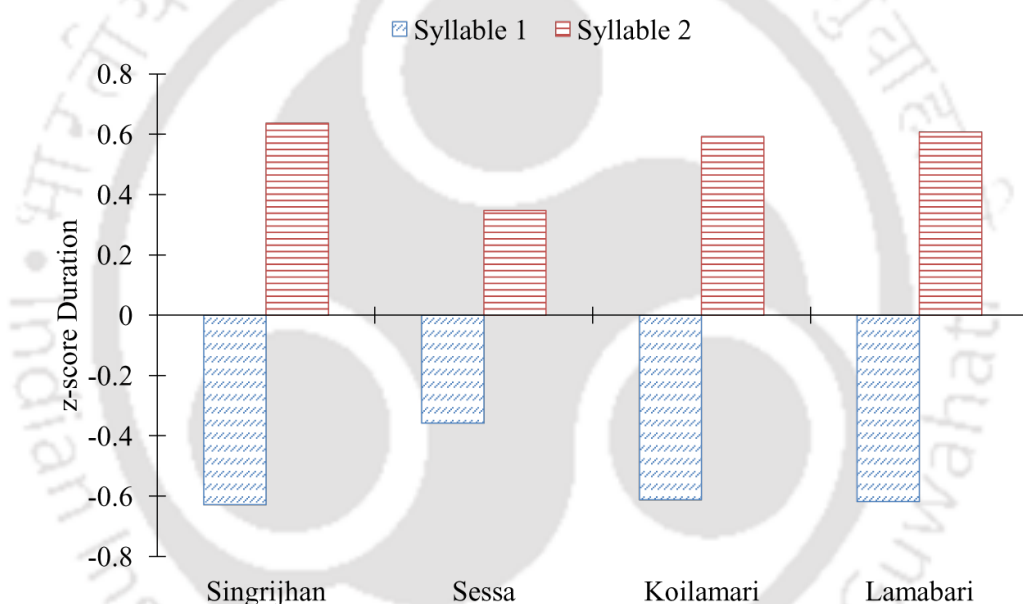
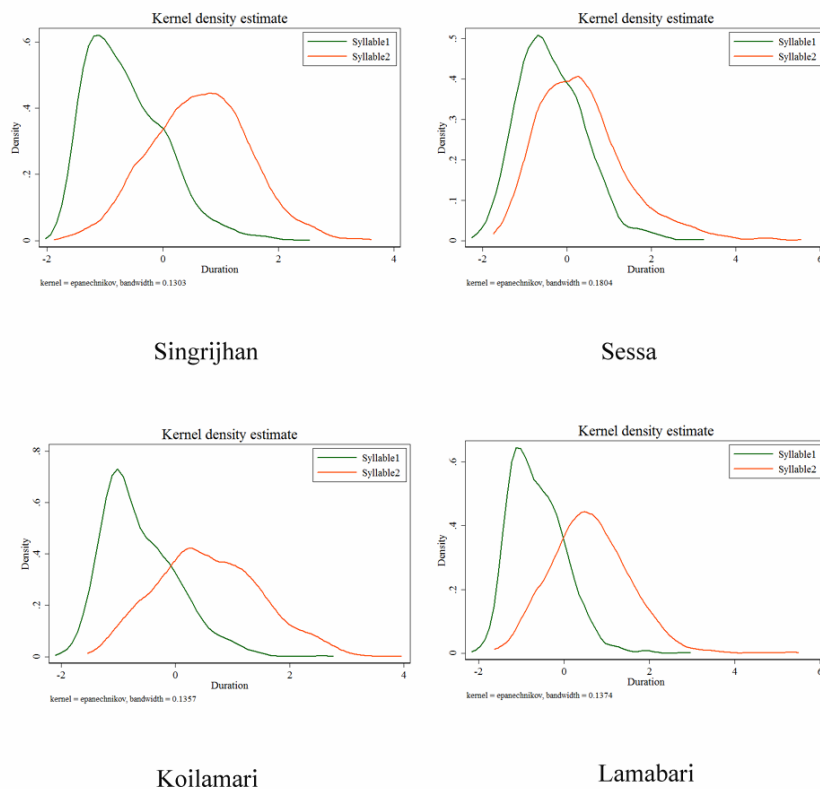


Figure 3.1 clearly reveals that vowel duration is always longer in the second syllable across Assam Sora speakers of all four regions recorded in this chapter. This indicates that vowels are generally longer in second syllable of Assam Sora disyllables. Moreover, in order further examine the vowel duration data, a kernel density estimation has been conducted. Kernel density estimation helps to review the distribution of vowel duration data separately in the two syllables. Kernel density distribution of speaker normalized vowel duration in first and second syllable is presented in Figure 3.2.

Figure 3.2: Distribution of speaker normalized vowel duration in first and second syllable



From Figure 3.2, it is evident that vowel duration distinction in first and second syllable of Assam Sora disyllables are clearly separated. Also, a similar amount of distinction is observed across Assam Sora speakers of all four regions recorded in this chapter. Additionally, in order to further confirm the results, a one-way ANOVA is conducted by considering the speaker normalized vowel duration as dependent variable and the syllable position as factor. From the statistical test it is found that, for all the four regions vowel duration in second syllable is significantly longer than in first syllable (Singrijhan,  $F(1,4710) = 3147.084$ ,  $p < 0.001$ ; Sessa,  $F(1,1907) = 271.653$ ,  $p < 0.001$ ; Koilamari,  $F(1,3080) = 1756.073$ ,  $p < 0.001$ ; Lamabari,  $F(1,2534) = 1531.623$ ,  $p < 0.001$ ). Hence, from the analysis of vowel duration in Assam Sora disyllables it is evident that second syllable is longer than first syllable and the statis-

tical test substantiates the fact that vowel duration is significantly longer in second syllable. Therefore, it is confirmed that vowel duration effectively contributes to the prominence of the second syllable in Assam Sora disyllables.

### **3.4.2 Fundamental frequency in first and second syllable of Assam Sora disyllables**

Pitch variation is another important cue for distinguishing stressed syllables from unstressed syllables. Usually, stressed syllables have higher pitch variations and unstressed syllables have lower pitch variations. This variation of pitch in stressed and unstressed syllable can be measured with the acoustic correlate of pitch namely fundamental frequency ( $f_0$ ). Therefore, fundamental frequency measurements are commonly adopted to investigate the difference between stressed and unstressed syllables in different languages of the world. Also, it is known that a stressed syllable is generally associated with higher fundamental frequency and an unstressed syllable is associated with lower fundamental frequency. Significantly, the study by (Horo & Sarmah, 2015) has also shown that fundamental frequency is always higher in second syllable than in first first syllable. However, the existing study is based on speech data recorded from Assam Sora speakers living in Singrijhan region only and it is known that, pitch variation is often sensitive to regional varieties of the same language. Therefore, in order to examine whether Assam Sora speakers of different geographical regions maintain the same pitch variation in first and second syllable or not, speech data recorded from three additional regions including Sessa, Koilamari and Lamabari are separately examined in this chapter. Thus, fundamental frequency in the speech data of Assam Sora speakers of these regions along with speech data from Singrijhan are examined and presented here. Additionally, for examining fundamental frequency distinction between first and second syllable of Assam Sora disyllables, first average fundamental frequency of the vowel nuclei in each syllable is examined.

This includes calculating the average pitch variation between vowel start and vowel end for every vowel in each syllable of the disyllables. However, considering that average fundamental frequency may ignore a unique pitch change within a given range, maximum fundamental frequency or maximum pitch variation within vowel start and vowel end in first and second syllable is also calculated here. Thus, considering the average fundamental frequency first, Table 3.3 presents the region-wise averages of average fundamental frequency distinction between first and second syllable in Assam Sora disyllables with standard deviation in parenthesis.

Table 3.3: Average of average  $f_0$  in first and second syllable

	Singrijhan	Sessa	Koilamari	Lamabari
Syllable 1	172.4 (51.09)	142.17 (34.15)	167.79 (59.15)	162.42 (51.63)
Syllable 2	181.61 (50.42)	153.27 (32.01)	181.27 (53.47)	161.94 (39.21)

From Table 3.3 it is observed that average fundamental frequency is always higher in the second syllable than in first syllable, except in case of Assam Sora speech data recorded in Lamabari region. This indicates that there is higher pitch variation in the second syllable than in the first syllable of Assam Sora disyllables. However, while this is true for Assam Sora data recorded from Singrijhan, Sessa, and Koilamari regions, the speech data recorded from Lamabari region does not show similar results. Significantly, by considering the maximum fundamental frequency next, a similar pattern is again found from the comparison of maximum fundamental frequency in first and second syllable of Assam Sora disyllables. Table 3.4 presents the region-wise average of maximum fundamental frequency distinction in first and second syllable of Assam Sora disyllables.

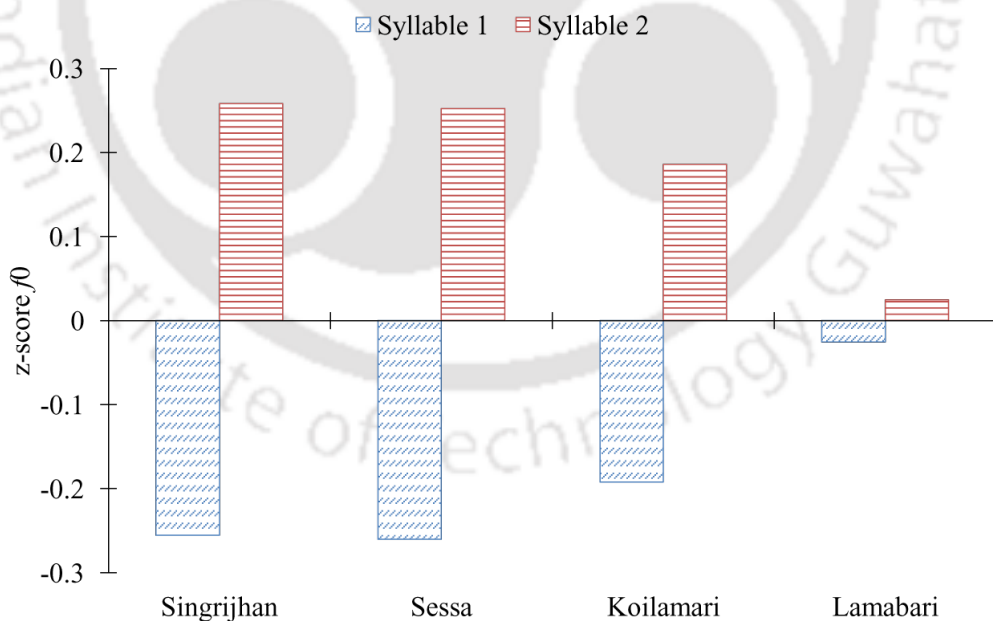
Table 3.4: Average of maximum  $f_0$  in first and second syllable

	Singrijhan	Sessa	Koilamari	Lamabari
Syllable 1	182.85 (61.18)	152.28 (43.78)	180.48 (67.24)	174.24 (61.41)
Syllable 2	198.43 (62.81)	164.68 (36.81)	196.4 (63.50)	173.58 (48.47)

From Table 3.4 it is observed that maximum fundamental frequency is also higher

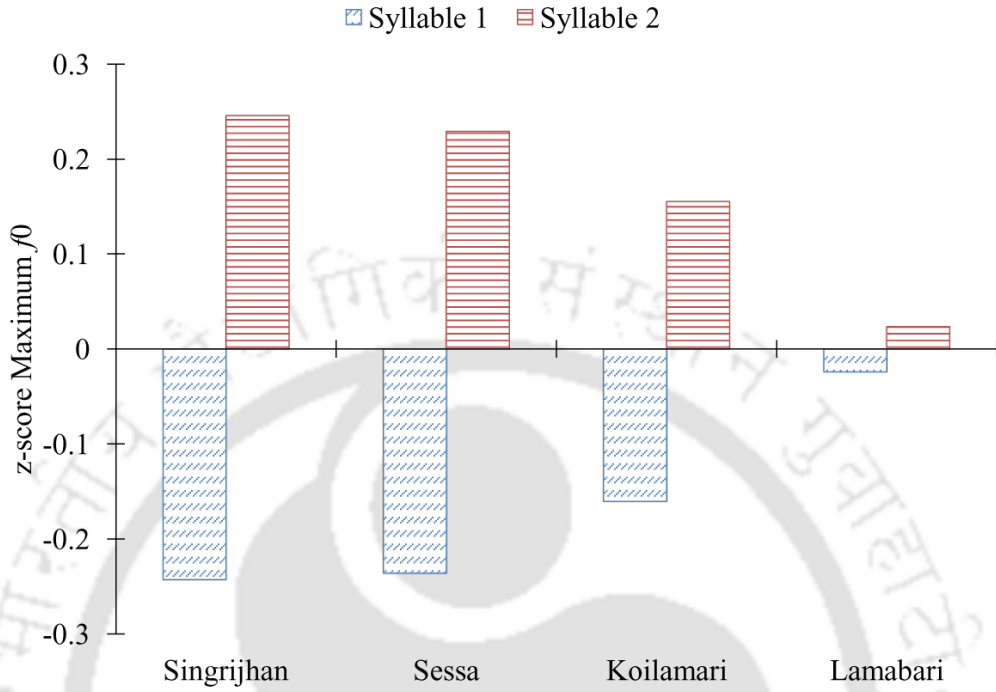
in second syllable of Assam Sora disyllables, except in case of speech data recorded from Lamabari region. Thus, it is evident that both average and maximum fundamental frequencies are higher in second syllable than in the first syllable, except in case of Assam Sora speech data recorded from Lamabari region. In this regard, it was assumed that there are speaker variations in the speech data of Lamabari region since fundamental frequency is sensitive to the physiological differences in vocal tract length of different speakers. Therefore, in order to reduce the speaker effect on fundamental frequency measurements, both average and maximum fundamental frequencies are normalized for all speakers across the four regions using the *z-score* equation. Hence, Figure 3.3 and 3.4 presents the graphical representations of speaker normalized average and maximum fundamental frequency distinction in first and second syllable of Assam Sora disyllables.

Figure 3.3: Speaker normalized average  $f_0$  in first and second syllable



Significantly, Figure 3.3 and 3.4 reveals that both average and maximum fundamental frequency are higher in the second syllable than in the first syllable of Assam Sora disyllables. Moreover, the normalized speech data shows that a similar pattern

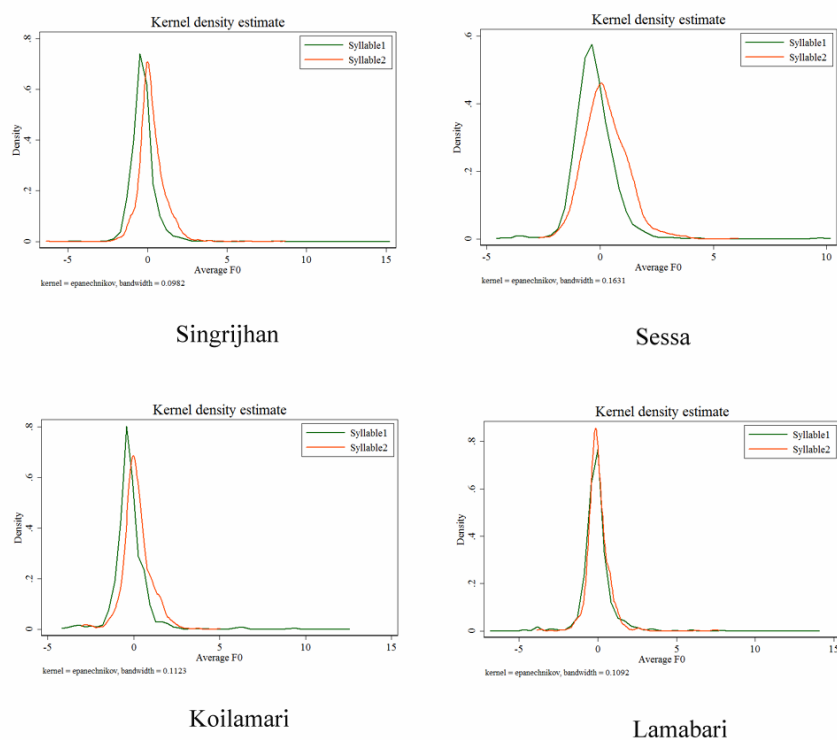
Figure 3.4: Speaker normalized maximum  $f_0$  in first and second syllable



is also observed in Assam Sora data recorded from Lamabari region that was previously not found while examining the non-normalized speech data. However, the distinction in Lamabari region is not as clearly visible as it is visible for the other three regions. On the contrary, a closer examination of the Lamabari speech data reveals that, the 10 Assam Sora speakers recorded in Lamabari region actually have a uniform fundamental frequency across first and second syllable of disyllables. This pattern can also be observed in the distribution of Lamabari data for average and maximum fundamental frequency measurements. Data distribution of average and maximum fundamental frequency measurements in first and second syllable of Assam Sora disyllables is estimated separately through kernel density estimation. Thus, Figure 3.5 and 3.6 presents the kernel density distribution of speaker normalized average and maximum fundamental frequency in first and second syllable of Assam Sora disyllables.

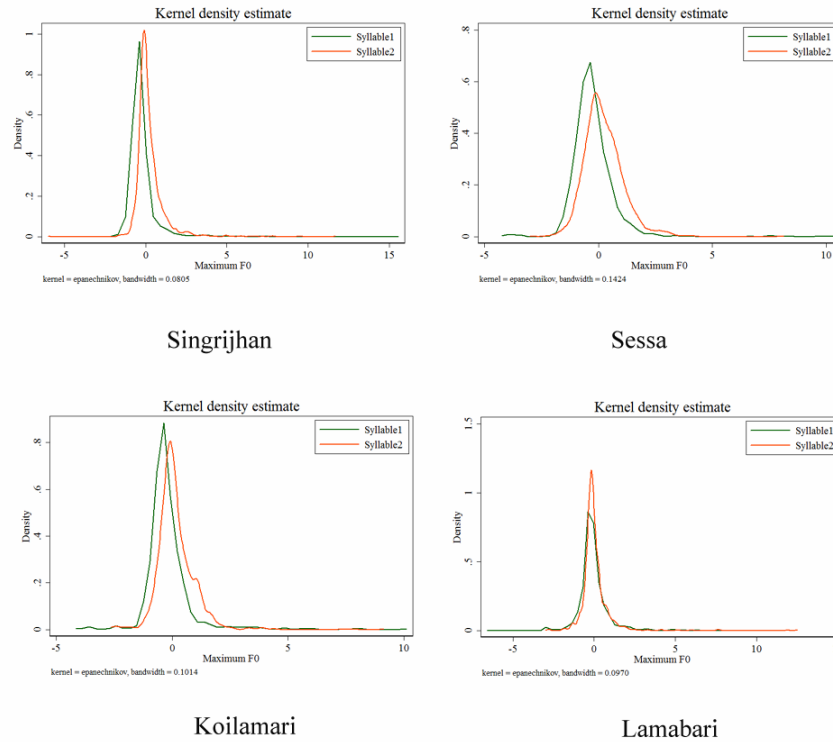
From Figure 3.5 and 3.6 it is evident that both speaker normalized average and

Figure 3.5: Distribution of speaker normalized average  $f_0$  in first and second syllable



maximum fundamental frequency are clearly separated in first and second syllable of Assam Sora disyllables, except in case of Assam Sora data recorded from Lamabari region. Thus, the analysis of fundamental frequency reinforces the fact that in Assam Sora disyllables there is higher pitch variation in second the syllable than in first syllable, except in case of Assam Sora speech variety spoken in Lamabari region. In case of Lamabari region, a closer examination of the speech data reveals that the pitch in the vowel nuclei is completely uniform in first and second syllable of Assam Sora disyllables. Additionally, a one-way ANOVA test conducted by considering normalized average  $f_0$  as dependent variable and syllable positions (first or second syllable) as factors also reveals the same variation. For all regions except Lamabari, average  $f_0$  is found significantly higher in the second syllable than in the first syllable (Singrijhan,  $F(1, 4710) = 333.114$ ,  $p < 0.001$ ; Sessa,  $F(1, 1907) = 134.417$ ,  $p < 0.001$ ; Koilamari,  $F(1, 3080) = 114.524$ ,  $p < 0.001$ ; Lamabari  $F(1, 2534) = 1.582$ ,  $p = 0.209$ ).

Figure 3.6: Distribution of speaker normalized maximum  $f_0$  in first and second syllable



Similarly, a one-way ANOVA test conducted by considering normalized maximum  $f_0$  as dependent variable and syllable position as factor also reveals the same pattern (Singrijhan,  $F(1,4710) = 298.997$ ,  $p < 0.001$ ; Sessa,  $F(1,1907) = 109.652$ ,  $p < 0.001$ ; Koilamari,  $F(1,3080) = 78.969$ ,  $p < 0.001$ ; Lamabari,  $F(1,2534) = 1.446$ ,  $p = 0.229$ ). Hence, from the analysis of fundamental frequency it is confirmed that, except for Assam Sora speech data recorded from Lamabari region, higher pitch variation in second syllable contributes to the prominence of second syllable in Assam Sora disyllables. Unfortunately, the reason for finding an exception in Assam Sora speech data recorded from Lamabari region is not known yet.

### 3.4.3 Vowel intensity in first and second syllable of Assam Sora disyllables

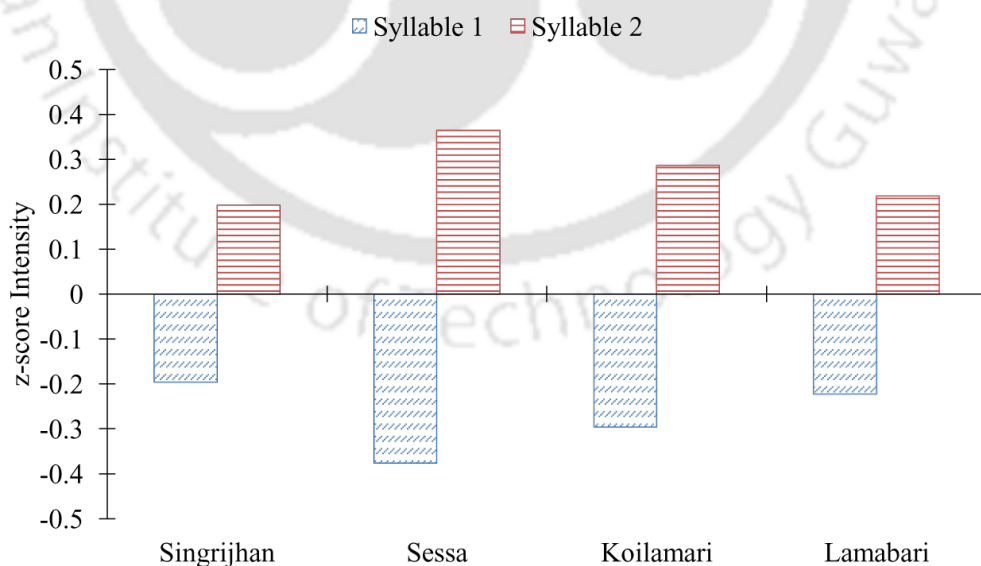
The third important cue for identifying stressed syllables in natural languages is the amplitude or loudness of a syllable nucleus and its acoustic correlate in vowel intensity. Vowel intensity measurements are crosslinguistically used to distinguish stressed syllables from unstressed syllables. In this regard, it is found that stressed syllables are generally characterized with greater vowel intensity than unstressed syllables. This implies that stressed syllables have greater acoustic energy and are able to be perceived louder than unstressed syllables. Previous study of word stress in Assam Sora disyllables by (Horo & Sarmah, 2015) also includes vowel intensity measurement and evidences suggest that the second syllable is always louder than the first syllable. However, similar to vowel duration and fundamental frequency measurements, vowel intensity measurements is also based on speech data recorded from Assam Sora speakers living in Singrijhan region only. Therefore, in order to examine whether vowel intensity distinction in first and second syllable is consistently maintained across regions, speech data from three additional regions including Sessa, Koilamari and Lamabari are subjected to the analysis of vowel intensity in Assam Sora disyllables in this section. For this purpose vowel intensity in first and second syllable is calculated separately for all four regions. This includes calculating the average amplitude of every vowel between vowel start and vowel end in first and second syllable of Assam Sora disyllables. Thus, Table 3.5 presents the region-wise average of average vowel intensity in first and second syllable of Assam Sora disyllables. Also, standard deviation for the same is given in parenthesis.

Table 3.5: Average of average vowel intensity in first and second syllable

	Singrijhan	Sessa	Koilamari	Lamabari
Syllable 1	72.86 (4.16)	74.1 (5.54)	61.28 (4.46)	58.28 (4.39)
Syllable 2	74.24 (3.17)	77.23 (5.56)	63.41 (4.63)	59.91 (4.83)

From Table 3.5 it is observed that average vowel intensity of the second syllable is generally greater than the average vowel intensity of the first syllable. Also, Assam Sora speech data of all four regions reveal that average vowel intensity of the second syllable is always greater than the first syllable. This, indicates that vowel nuclei of second syllable can be perceived as louder than the vowel nuclei of the first syllable in Assam Sora disyllables. However, the data also reveals that some Assam Sora speakers have a tendency to produce an extra loud vowel in the second syllable. For instance, in Singrijhan region, although average vowel intensity in second syllable is 74.24 decibels, one male speakers is recorded producing some vowels with an intensity of 80 decibels in in the second syllable. Therefore, in order to reduce the speaker effect, average vowel intensity in first and second syllable is normalized using the *z-score* equation. Thus, Figure 3.7 presents the region-wise graphical representation of speaker normalized average vowel intensity distinction between first and second syllable in Assam Sora disyllables.

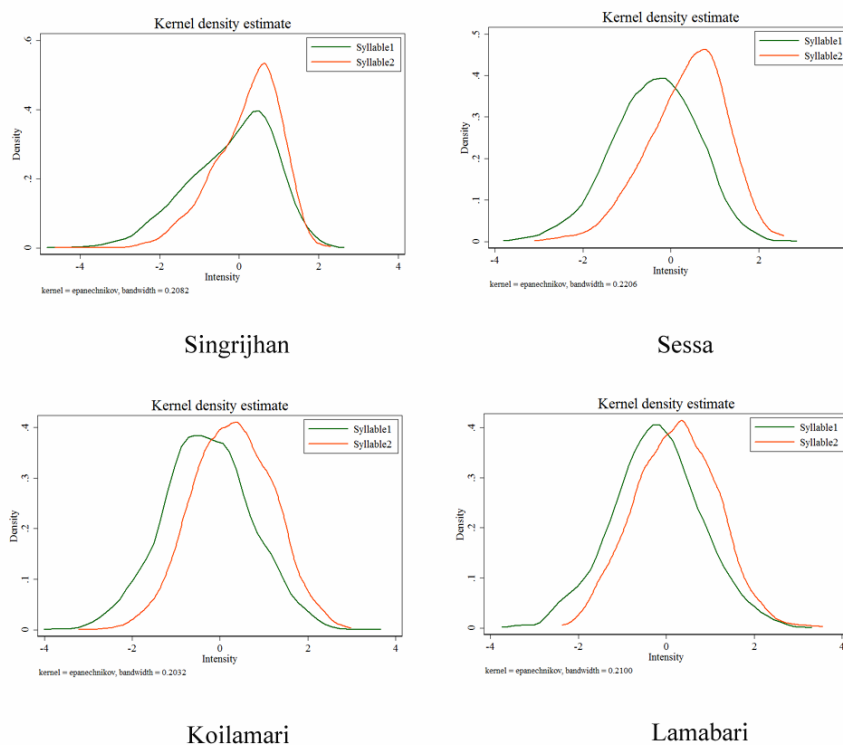
Figure 3.7: Speaker normalized average vowel intensity in first and second syllable



From Figure 3.7 it is observed that, speaker normalized average vowel intensity

data also exhibits similar relation between first and second syllable as was found in the non-normalized data. Thus, it is evident that average vowel intensity varies in first and second syllable of Assam Sora disyllables, such that average vowel intensity is always greater in the second syllable than in the first syllable. This, indicates that Assam Sora vowels have greater acoustic energy in the second syllable and therefore vowel nuclei of second syllable can be perceived louder than first syllable. Also, this variation in vowel intensity is uniformly observed in all Assam Sora speakers across the four regions. In order to confirm that average vowel intensity variation is consistently present in Assam Sora speech data, data distribution in first and second syllable is examined with the help of kernel density estimation. Thus, kernel density estimation of speaker normalized average vowel intensity in first and second syllable of Assam Sora disyllables is presented in Figure 3.8.

Figure 3.8: Distribution of speaker normalized average vowel intensity in first and second syllable



From Figure 3.8 it is observed that average vowel intensity is clearly separated in first and second syllables in Assam Sora of all regions except in the variety spoken in Singrijhan region. It is observed that average vowel intensity distinction in Assam Sora variety spoken in Singrijhan region has some overlaps. However, the distinction is not completely merged and a considerable amount of Singrijhan data shows that average vowel intensity is greater in second syllable than in first syllable of disyllabic words. The distinction is also observed in a one-way ANOVA test conducted by considering average vowel intensity as the dependent variable and syllable position as factor. For all four regions it is found that average vowel intensity is significantly greater in second syllable than in first syllable of Assam Sora disyllables (Singrijhan,  $F(1,4710) = 189.803$ ,  $p < 0.001$ ;  $F(1,1907) = 304.402$ ,  $p < 0.001$ ; Koilamari,  $F(1,3080) = 286.088$ ,  $p < 0.001$ ; Lamabari,  $F(1,2534) = 129.897$ ,  $p < 0.001$ ). Hence, the analysis of average vowel intensity confirms that vowel intensity has a significant role in contributing prominence to the second syllable of Assam Sora disyllables.

### 3.5 Conclusions

This chapter has primarily described the syllable structure of Assam Sora. This includes identifying the typical syllable template and the minimal word template in Assam Sora. Thus, Assam Sora data examined in this chapter reveals that (C)V(C) is a typical syllable template in Assam Sora. Regarding minimal word template, this chapter also reveals that monosyllabic words are rare in Assam Sora and disyllabic word is the minimal word template in Assam Sora. Also, it has been shown that in disyllabic words the syllable structure of both the syllables is (C)V(C). This has shown that the syllable structure of Assam Sora and Orissa Sora, as reported in Stampe (1965), are similar in shape. However, in case of minimal word length, Stampe (1965) mentioned that Orissa Sora has monosyllabic words, whereas in Assam Sora

it is found that monosyllabic words are rare and therefore disyllabic word appears to be the minimal word template. Significantly, the (C)V(C) syllable structure is typical to Munda languages and a disyllabic minimal word is also reported in Gta?, a Munda language spoken in parts of Southern Orissa. Thus, it is visible that despite being separated from its place of origin, Assam Sora still exhibits features of the parent language particularly in syllable structure, and also has close resemblance with related languages still spoken in its place of origin.

This chapter has also explored the phonotectic constraints of Assam Sora and has shown that in disyllabic words while all six Assam Sora vowels can occur on both syllables, there are restrictions on consonant sounds at the syllable boundaries. It has been shown that all consonants except the velar nasal /ŋ/ and the glottal stop /ʔ/ can occur in onset position whereas only the consonants /b, p, m, n, ŋ, r, l and ʔ/ can occur in the coda position. Moreover, an additional phonotetic constraint is found at the onset position whereby it is shown that the consonant /w/ can occur only in the onset position of the first syllable and /ɽ/ can occur only in the onset position of the second syllable. Additionally, it has been shown that consonant cluster in Assam Sora is mainly restricted in the onset position of the second syllable and cluster consonants are usually formed by combining stop consonants /p, b, t, d, k, g/ and liquid consonants /r/ and /ɽ/. Thus, the description of phonotectic constraints provides information about the composition of a syllable structure in Assam Sora and helps in identifying a valid syllable in the language.

Finally, this chapter has provided acoustic evidence that in Assam Sora disyllables the second syllable is more prominent than the first syllable. This indicates that words stress in Assam Sora is typically iambic. Acoustic evidences showing the prominence of second syllable are drawn from the examination of Assam Sora speech data recorded in four Assam Sora speaking regions in Assam, including Singrijhan, Sessa, Koilamari and Lamabari. Examination of prominence of the second syllable

is based on three phonetic cues of word stress, namely, vowel length, pitch variation and vowel amplitude. The Assam Sora speech data has been subjected to analysis of the acoustic correlates of these phonetics cues, namely, vowel duration, fundamental frequency and vowel intensity. Consequently, it has been shown that all three features contribute to the prominence of second syllable over first syllable in Assam Sora disyllables. Moreover, it is seen that while vowel duration is a robust cue for prominence of second syllable in Assam Sora, fundamental frequency and vowel intensity cues have certain variations. It has been shown that vowel duration is always longer in second syllable of Assam Sora disyllables across Assam Sora speakers of all four region. On the other hand, although the fundamental frequency and the vowel intensity cue shows that vowels in second syllable have higher pitch variation and sound louder, the pattern is inconsistent in Assam Sora variety of two region. Firstly it is found that the Assam Sora variety of Lamabari region does not make a distinction between the pitch variation in first and second syllable. Secondly only a small amount of Assam Sora data of Singrijhan region shows distinction in vowel intensity of first and second syllable. However, statistical analysis reveals that vowel duration and vowel intensity are significantly greater in second syllable in Assam Sora variety of all four region. Whereas, fundamental frequency is significantly higher in second syllable in Assam Sora variety of all the regions, except Lamabari. A closer look at the pitch variation of Lamabari data reveals that fundamental frequency is completely flat and uniform in first and second syllable in all disyllabic words. Ten Assam Sora speakers of Lamabari region have not shown any pitch variation and the factor responsible for this is not yet known. Hence, by excluding the Lamabari exception, it is concluded from this chapter that the disyllabic minimal word in Assam Sora has an iambic word stress that is realized by three phonetic cues namely vowel length, pitch variation and vowel amplitude.

## Chapter 4

# Acoustic phonetic properties of vowels in Assam Sora

### 4.1 Introduction

The Assam Sora vowel system has six vowels including /a, e, i, o, u ə/ as claimed in Horo & Sarmah (2015). This chapter provides a detailed description of the six vowels based on their acoustic phonetic properties. Assam Sora speech data is examined and the acoustic phonetic features of vowel phonemes, namely, formant frequency is analyzed in this chapter. The previous study on Assam Sora vowel phonemes by Horo & Sarmah (2015) revealed that the six vowels are categorically distinct from each other based on their acoustic phonetic properties. The study by Horo & Sarmah (2015) is based on speech data recorded from Assam Sora speakers of the Singrijhan region and it is unknown whether Assam Sora speakers living in other geographical areas have the same vowel features or not. Therefore, this chapter extends the existing findings by incorporating speech data of Assam Sora speakers living in three additional regions of Assam, namely, Sessa, Koilamari and Lamabari. Moreover, the previous study provided evidence that vowel properties including vowel height and vowel frontness/backness of Assam Sora vowel phonemes, change as a function of syllable position. Thereby, it has been proposed that the canonical vowels are differently realized in first and second syllable of disyllabic words due to a difference of

prominence between the two syllables. In this regard, Section 3.4 of Chapter 3 has provided evidence that the second syllable is significantly prominent in Assam Sora speech varieties of all four regions mentioned above. Therefore, this chapter extends the examination of variation in Assam Sora vowel properties including vowel height and vowel frontness/backness across syllables by incorporating speech data from the three additional regions of Assam, namely, Sessa, Koilamari and Lamabari. Subsequently, comparison of vowel features of the four regions will reveal whether Assam Sora vowel inventory exhibits any regional variation or not. Hence, besides providing a detailed acoustic phonetic description of Assam Sora vowels produced by speakers of four different regions of Assam, this chapter also examines the likelihood of having regional variations of Assam Sora on the basis of its vowel inventory.

Additionally, vowel system studies in the past have proposed acoustic phonetic theories on vowel system that generalize the organization of vowel systems crosslinguistically. Vowel system theories claim to predict the selection of vowel segments in an inventory based on the size of the vowel systems. Also, the theories explain why certain vowel sounds are more favoured over certain other vowel sounds in a vowel system. As a result, it is evident that only some languages in the world such as Turkish, Thai, English etc. have large vowel systems with 9 or more vowels. Whereas, languages with only 5 or 6 vowels are more common crosslinguistically (Liljencrants & Lindblom, 1972). Thus, it is fair to assume that Assam Sora has a vowel system that is more commonly occurring crosslinguistically. However, in order to justify this assumption, it is important to examine how Assam Sora vowel system relates to some of the theories on vowel system. Therefore, prior to the examination of Assam Sora vowel features and qualities this chapter reviews some of the theories on vowel system in the subsequent section (§4.2) in order to describe the crosslinguistic typology of vowel systems. This is followed by the acoustic phonetic description of Assam Sora vowel phonemes in §4.3 and the Assam Sora vowel plots are discussed in §4.4.

Subsequently, the dynamic nature of Assam Sora vowel properties in disyllables is described in §4.5 and the distinctiveness of Assam Sora vowels is discussed in §4.6. Finally, findings and conclusions of the chapter are presented in §4.7.

## 4.2 Vowel systems and theories

Two major questions addressed by theories on vowel systems are to find how vowel systems function as a whole and how individual vowels interact within a given vowel system. Schwartz et al. (1997a) argues '*if speaker-listener interaction can shape phonological inventories then phonological inventories should also provide information about the mechanism of speaker-listener interaction*'. Therefore, by defining the articulatory-auditory relation of vowel segments, scholars also attempt to predict universal tendencies in vowel systems across languages. In doing so, scholars often claim to predict vowel system sizes depending on the selection of individual vowel segments in the overall vowel system. In this regard Schwartz et al. (1997b) states '*various predictions of vowel system are primarily aimed at exploiting the auditory or articulatory arguments of vowel classifications to derive their most likely individual positions in the articulatory-acoustic space*'. Hence, based on articulatory-auditory relation of vowel segments some of the theories that claim to predict vowel system shape and size include the *Quantal Theory* (Stevens, 1972, 1989), the *Adaptive Dispersion Theory* (Liljencrants & Lindblom, 1972) and (Lindblom, 1986) and the *Dispersion-Focalization Theory* (Schwartz et al., 1997a). These three theories on vowel system are discussed below.

### 4.2.1 Quantal Theory

Quantal theory of vowel system is based on the relation between articulatory configuration of a vowel segment and its acoustic output. Stevens (1989) suggests that when

articulatory configurations are modified, the acoustic output also changes. However, Stevens (1989) also argues that the relation between the articulatory configuration and its acoustic output is not monotonic. Therefore, he proposes that for certain articulatory configurations there are more stable acoustic outputs. Whereas, for certain other articulatory configurations there are abrupt changes in the acoustic outputs. The stable regions are considered as insensitive to articulatory modifications and are therefore considered to be quantal regions. Thus, he argues that precise articulatory configurations give rise to well defined patterns of auditory responses, that are not sensitive to small perturbations or inaccuracies in articulation. However, if an articulatory configuration makes a major modification, there will be changes in the auditory response as well. Hence, according to Stevens (1989), quantal regions define the contrasting vowel segments in a vowel system and quantal relation between articulatory-auditory parameters is the principal factor for shaping the inventory of phonetic features. Stevens (1989) also argues that in phonology, those quantal regions actually correlate with the distinctive features, such that an acoustic output changes only in response to an addition of a particular distinctive feature in the articulatory configuration. Additionally, Stevens (1989) argues that the vowels /i, a, u/ are best examples of vowel segments produced in the quantal region. Hence, quantal theory predicts that these three point vowels are the most common vowels found crosslinguistically.

#### **4.2.2 Adaptive Dispersion theory**

Adaptive Dispersion theory (Liljencrants & Lindblom, 1972), (Lindblom, 1986) emphasizes that vowel inventories are structured to create contrast between vowels by maximally dispersing vowel sounds in the acoustic-perceptual vowel space. The theory proposes a model for maximizing the perceptual contrast in vowel inventories based on the structural properties of the inventory and the idea that vowel acoustic

space and vowel inventory size are correlated. The model proposes that vowel categories shift acoustically to escape crowded regions and fill in less crowded regions in order to maximize perceptual contrast between the vowels. For instance, vowels often shift when a non-peripheral vowel [ə, ɪ] is added to the inventory and accommodation of new non-peripheral vowels is guaranteed only by horizontal expansion of the vowel space periphery. Evidence suggests, when the inventory has just one non-peripheral vowel, it occupies a distinct vowel height evenly contrasting with peripheral vowels [i, e, a, o, u] both horizontally and vertically. Therefore, a single non-peripheral vowel in the inventory does not require excessive horizontal or vertical expansion of the periphery. However, by the addition of a new non-peripheral vowel in the inventory they are both forced to level their heights with the front and the back peripheral vowels. Also, when non-peripheral vowels shift for adjustment, their spacing remains vertically even, but, their spacing becomes horizontally uneven. As a result, the two non-peripheral vowels achieve greater contrast horizontally than vertically. However, the maximum contrast is still maintained in order to differentiate between the two vowels. Hence, the theory proposes that based on the number of vowels in the inventory and on the relation between individual vowels in the inventory, vowel segments maximally disperse from each other and the periphery of acoustic-perceptual vowel space is also widened.

In this way, consistency in operation of Adaptive Dispersion theory assures the correlation between the number of vowels in the inventory and its overall vowel acoustic space. Also, the theory mainly aims to model vowel inventory typologies but does not guide individual languages by its principles. This implies that dispersion models only evaluate vowel inventories as appropriate or inappropriate based on the arrangement of their vowel segments. Therefore, contrary to Distinctive feature theory, which demands that vowel quality should be precise, Adaptive Dispersion theory recognizes the relation between vowels, and maintains the adaptive nature of vowel segments.

Thus, some of the principles of Adaptive Dispersion theory proposed by Becker-Kristal (2010) include a) vowels should be maximally perceptually dispersed from one another; b) vowel qualities may adapt to structural changes causing lesser dispersion in the inventory and ultimately result in vowel shift for maximization of dispersion; and c) maximization of dispersion is achieved via equal spacing between vowel segments. Therefore, based on these principles Adaptive Dispersion theory proposes that inventories with more vowels should cover a larger acoustic space than inventories with fewer vowels. Thereby, in bigger inventories the phonetic realizations should be precise, whereas smaller inventories can have the freedom of variability. Also the theory proposes that cross-linguistically vowel inventories should have an upper limit on the number of possible contrasts in order to maintain minimal distance between vowels in the  $n$  the acoustic-perceptual vowel space.

### **4.2.3 Dispersion-Focalization theory**

According Dispersion-Focalization theory (Schwartz et al., 1997a) vowel systems are examined on the basis of their energy concentration. It is argued that the energy level of a vowel system depends on the type of vowels a vowel system has and on the distribution of individual vowels in that vowel system. Thus, similar to Quantal theory and Adaptive Dispersion theory, Dispersion-Focalization theory also aims to find the most optimal vowel system based on the energy concentration of a vowel system. According to Dispersion-Focalization theory, total energy of a vowel system can be estimated by considering together (a) intrinsic vowel qualities - focal vowels (quantal nature of vowel segments) and (b) acoustic-perceptual distance between vowels (dispersion nature of vowel segments). Thus, based on the total energy of a vowel system, derived from the summation of (a) and (b), Dispersion-Focalization theory predicts that optimal vowel systems have the lowest energy. In other words, it can be stated that vowel systems which include the quantal vowels that are sufficiently

dispersed from each other have lower energy and are the most optimal vowel systems. Whereas, vowel systems that include non-quantal vowels that are also not sufficiently dispersed from each other will have higher energy concentration and are considered to be less optimal vowel systems. Hence, Dispersion-Focalization theory predicts that vowel systems which have peripheral vowels are more optimal vowel systems whereas, vowel systems with central vowels are less optimal vowel systems.

In summary, it is observed that, Quantal theory (Stevens, 1972, 1989) and Adaptive Dispersion theory (Liljencrants & Lindblom, 1972; Lindblom, 1986) mainly try to explain why particular vowel segments and vowel systems are preferred over other vowel segments and vowel systems across languages. However, difference between the two theories lies in the predictions that the two theories make. The Quantal theory predicts that languages usually select their point vowels /i, a, u/ at acoustic regions that are more stable. Quantal theory considers the point vowels as 'quantal' vowels and argues that the quantal vowels are independent of the vowel system size. On the contrary, Adaptive Dispersion theory proposes that sufficient contrast between point vowels in a vowel system leads to 'adaptive dispersion' between the point vowels. Hence, Adaptive Dispersion theory predicts that vowel systems with a large number of vowels will have larger dispersion between point vowels resulting in larger acoustic space. Whereas, vowel systems with smaller number of vowels will have lesser dispersion between the point vowels resulting in smaller acoustic space. On the other hand, Dispersion-Focalization theory (Schwartz et al., 1997a) basically integrates the Quantal theory and the Adaptive Dispersion theory since it considers both the quantal nature and the dispersion nature of vowel segments to predict the optimal vowel system. Hence, the general idea drawn from the three vowel system theories is that organization and structure of vowel systems is not random. It is evident that vowel systems are formed when speakers and listeners are able to produce and perceive every vowel distinctly in a vowel system. Also, the vowel theories reveal

that acoustic characteristics of vowels significantly contribute to understanding how vowels are shaped and structured in a vowel system. Regarding vowel system typologies, it is understood that stable and maximally dispersed vowels are most preferred in languages of the world. Moreover, the theories propose that stable vowels with maximal dispersion are achieved mainly in smaller inventories and therefore, small vowel systems are the most optimal vowel systems found in languages of the world. On the other hand, larger vowel inventories are likely to have unstable vowels that are not maximally dispersed and therefore, large vowel systems are less optimal and only some languages have unusually large vowel systems. Thus, considering these assumptions regarding the shape and structure of vowel systems and considering the salience of acoustic characteristics of vowels, acoustic phonetic characteristics of vowels in Assam Sora are described in the subsequent sections.

### **4.3 Interpretation of Assam Sora vowel formants**

Vowel sounds are produced as a result of vibration in the vocal fold that is caused by air particles passing through the larynx. The principal source of vowel sound is the glottis and adjustments in the configuration of the tongue and the lips produces the distinctive qualities of different vowel sounds (Stevens, 2000). These events in the vocal tract during the production of a vowel sound can be represented in the form of acoustic outputs commonly referred to as formant frequencies. Acoustic characterization of vowel sounds are generally based on their formant frequencies. Formant frequencies relate to the articulatory specification of vowel production and hence formant frequencies are frequently used to represent vowel inventory plots of different languages. Languages such as English, German and Dutch have been subjected to extensive acoustic analysis of their vowel phonemes based on the formant frequencies. Therefore, in the description of vowel phonemes it is important to discuss the

implications of formant frequencies in order to understand the relation between an articulatory configuration and its acoustic output. Traditionally, vowel phonemes have been represented in relation to their first three formant frequencies such that, first formant (F1) is associated with the height feature of the vowel, second formant (F2) is associated with the frontness or backness feature of the vowel and third formant (F3) is associated with the roundedness feature of the vowel (Cox, 1996). While first formant is inversely related to the vowel height, second formant is directly related to vowel frontness or vowel backness and third formant is also inversely related to lip rounding but, F3 is also dependent on the location of the vowel constriction. Also, the F3 value is dependent on language types whereby, only when a language has apparent an contrast between rounded and unrounded vowels such as [u] vs [ʊ], the F3 measure provides the relevant evidences. Therefore, in languages that do not have such roundedness contrasts, only F1 and F2 frequencies are generally used. Thus, this work also mainly examines F1 and F2 frequencies of the vowels in Assam Sora. Hence, it is well accepted that formant frequency measurements provide evidences for phonetic categorization of vowel phonemes and that they also provide the cues for describing the distinctive nature of different vowel phonemes. Therefore, formant frequency measurements are examined here to distinguish the vowel phonemes in Assam Sora. Moreover, it is noticed that in case of a well documented language, such as English, acoustic analysis of the vowel phonemes is often based on a large database representing different time periods. On the other hand, in case of lesser described languages, a large database is often not available. Acoustic analysis of Assam Sora vowel phonemes (Horo & Sarmah, 2015) is also based on a small database, that is, actually the first speech data of the language recorded from Assam Sora speakers living in Singrijhan village only. The analysis adequately supported the finding that Assam Sora has six vowel phonemes /i, e, a, o, u, ə/ by categorically distinguishing the six vowel phonemes based on their formant frequency measurements. Addition-

ally, formant frequency measurement of Assam Sora vowels (Horo & Sarmah, 2015) also revealed that, in disyllabic words, the canonical vowel space in first syllable is different from the vowel space in second syllable. Thus, the first speech database of Assam Sora provided significant information about the vowel phonemes of Assam Sora and formant frequency measurements provided the relevant evidences. Therefore, this chapter further examines Assam Sora vowel phonemes by incorporation of a larger database collected from native Assam Sora speakers from three additional regions namely Sessa, Koilamari and Lamabari. Thus, Assam Sora vowel data from Singrijhan and the other three areas are subjected to the analysis of formant frequency measurements. Also vowel for the present study, the vowel samples of first and second syllable are combined together. Whereas, in the subsequent section (§4.5) the vowel samples are separated as first and second syllable of disyllables. Additionally, in order to show that the six vowels have distinct vowel heights, formant frequency of vowel pairs such as /i/ and /e/; /ə/ and /a/ and /o/ and /u/ are examined below. This will provide evidence that the six vowels have distinct vowel heights since, vowel frontness and vowel backness are generally well separated in a vowel system. Additionally, comparison of vowel formants of Assam Sora vowels of all four regions will also confirm whether Assam Sora has any regional variation on the basis of its vowel formants or not. Moreover, vowel formants reported here are extracted from vowel mid point of steady state vowel formants and formant frequency is converted from Hertz to Mel scale in order to show perceptual difference between the vowels (Lindblom, 1986). Thus, the following sections details the acoustic phonetic properties of Assam Sora vowels on the basis of their formant frequency measurements.

### 4.3.1 Formant frequency of front vowels /i/ and /e/ in Assam Sora

The vowel /i/ in Assam Sora is an unrounded high front vowel and the vowel /e/ in Assam Sora is an unrounded mid front vowel. Minimal pairs showing contrast between the two vowel phonemes in Assam Sora is presented in Table 4.1.

Table 4.1: Contrasting pairs for vowels /i/ and /e/

	Sora	English	Sora	English
1	<i>ida</i>	cut	<i>eda</i>	cry
2	<i>ide</i>	glue	<i>ede</i>	roll
3	<i>ɕi?i</i>	tooth	<i>ɕe?e</i>	red

The stricture of the lips while producing /i/ and /e/ in Assam Sora is similar, such that both the vowels are produced with spreaded lips. Also, both the two vowels are produced by moving the tongue towards the front of the mouth, but the height of the tongue body is different during the production of the two vowels. While the tongue body is in a raised position in the production of /i/, it is lowered to an intermediate position in the production of /e/. Acoustic consequence of the tongue movement towards the front position during the production of /i/ and /e/ is seen in the F2 frequency, whereby F2 frequency of both vowels are increased and are closer to the F3 frequency. On the other hand, the F1 frequency is relatively lower for the the vowel /i/ than for the vowel /e/, which provides evidence that the two vowels have distinct vowel heights in Assam Sora. Region-wise averages of first three formant frequencies of the two vowels is presented in Table 4.2 and standard deviation for the same is given in parenthesis.

From Table 4.2 it is evident that, the vowels /i/ and /e/ are categorically distinct from each other particularly with respect to their vowel height characteristics. While average F2 frequency of the two vowels indicate that they are both front vowels, average F1 frequency of the two vowels indicate that /i/ is a high front vowel and

Table 4.2: Average formant frequency of Assam Sora vowels /i/ and /e/

	Singrijhan		Sessa		Koilamari		Lamabari	
	i	e	i	e	i	e	i	e
F1	273.44 (36.7)	358.38 (47.82)	276.00 (34.45)	351.01 (32.28)	272.62 (33.82)	367.29 (55.55)	281.39 (29.86)	378.37 (31.62)
F2	923.18 (54.89)	868.35 (47.14)	885.28 (78.59)	824.06 (49.62)	927.26 (48.56)	858.55 (41.02)	890.35 (75.20)	850.11 (63.17)
F3	1040.24 (35.58)	994.49 (38.06)	1010.89 (54.06)	974.53 (44.84)	1041.37 (40.38)	994.49 (35.71)	1016.24 (56.99)	982.19 (56.31)

/e/ is a mid front vowel in Assam Sora. Also, distinction between the two vowels are found to be similar in Assam Sora speech data recorded from all four regions. It is found that average F1 frequency of /i/ is between 272 Mel to 281 Mel and average F1 frequency of /e/ is between 351 Mel to 387 Mel in the vowel data of the four regions. Therefore, the average difference in F1 frequency of the two vowels is between 75 Mel to 97 Mel in Assam Sora speech data of all four regions. Moreover, it is observed that the vowel /i/ is often produced higher than its corresponding back high vowel /u/. It can be seen that average F1 frequency of the high back vowel /u/ is between 295 Mel to 327 Mel in Assam Sora speech data of the four regions and therefore, there is a difference of 18 Mel to 54 Mel in F1 frequencies of /i/ and /u/ in Assam Sora speech data of the four regions. Additionally, average F2 frequency of the two vowels indicate that while both /i/ and /e/ are front vowels, /i/ usually has a tendency to be more front than /e/. It is seen that average F2 frequency of /i/ in all four regions is between 885 Mel to 923 Mel and average F2 frequency of /e/ in the four regions is between 824 Mel to 868 Mel. Therefore, a difference of 40 Mel to 69 Mel is found in the F2 frequency of /i/ and /e/ in Assam Sora speech data of the four regions. Hence, F1 and F2 frequencies of two vowels /i/ and /e/ in Assam Sora provide adequate evidence that the two vowels are categorically distinct from each other. Also, the formant frequencies reveal that although both /i/ and /e/ are front vowels, the high vowel /i/ is generally fronter than the mid front vowel /e/ in Assam

Sora. Likewise, although /i/ and /u/ are both high vowels, the front high vowel has a tendency to be produced higher than the back high vowel in Assm Sora vowel system.

### 4.3.2 Formant frequency of central vowels /ə/ and /a/ in Assam Sora

The vowel /ə/ is a mid central vowel and the vowel /a/ is a low central vowel in Assam Sora. Minimal pairs showing contrast between the two vowels in Assam Sora is presented in Table 4.3.

Table 4.3: Contrasting pairs for vowels /ə/ and /a/

	Sora	English	Sora	English
1	<i>əda</i>	fan	<i>ada</i>	curry
2	<i>təda</i>	throw	<i>tada</i>	stumble
3	<i>urəŋ</i>	bamboo	<i>araŋ</i>	stone

The vowels /ə/ and /a/ in Assam Sora are both produced by moving the tongue to a central position, but the height to the tongue body at the time of producing /ə/ is different than at the time of producing /a/. While the tongue body is raised to an intermediate position during the production of /ə/, it is taken down to a lower position during the production of /a/. Evidence for /ə/ and /a/ being central vowels is found from F2 frequency, whereby F2 frequency of both the vowels are increased, but are lesser than the F2 frequency of the front vowels /i/ and /e/. This indicating that /ə/ and /a/ are more centralized vowels in Assam Sora. Similarly, articulatory distinction of vowel heights during the production of the two vowels is also observed in their acoustic outputs. The F1 frequency shows that /a/ has greater F1 frequency than /ə/ providing evidence that while /a/ is a low central vowel, /ə/ is a mid central vowel and that they both have a distinctive vowel height in Assam Sora. Region-wise average of first three formant frequencies of the two vowels is presented in Table 4.4.

Table 4.4 reveals that /ə/ and /a/ are two distinct vowels in Assam Sora and their

Table 4.4: Average formant frequency of Assam Sora vowels /ə/ and /a/

	Singrijhan		Sessa		Koilamari		Lamabari	
	ə	a	ə	a	ə	a	ə	a
F1	333.71 (46.13)	468.59 (70.09)	328.50 (43.18)	442.79 (58.55)	304.26 (54.84)	462.27 (73.38)	321.37 (45.05)	459.62 (59.64)
F2	786.13 (70.42)	743.42 (61.72)	764.35 (52.72)	718.18 (46.93)	797.30 (52.19)	752.01 (53.71)	772.31 (40.29)	753.64 (54.78)
F3	980.36 (42.38)	970.48 (41.91)	946.41 (51.42)	979.93 (51.07)	989.82 (35.85)	981.77 (53.80)	970.47 (54.46)	969.62 (66.85)

corresponding formant frequencies provide evidence that the two vowels differ mainly due to different vowel height characteristics. As indicated by their average F1 and F2 frequencies, it is found that while /ə/ is a mid central vowel, /a/ is a low central vowel in Assam Sora. Also, formant frequency measurements showing distinction between the two vowels are found to be similar in Assam Sora speech data of all four regions. For instance, while average F1 frequency of /ə/ is between 304 Mel to 333 Mel in Assam Sora speech data of the four regions, average F1 frequency of /a/ is between 442 Mel to 468 Mel. Therefore, the difference in F1 frequency of the two vowels is between 114 Mel to 158 Mel in the speech data of the four regions. This indicates that there is a significant vowel height distinction between the two vowels in Assam Sora in all four regions.

Moreover, it is observed that the mid central vowel /ə/ is produced higher than the mid front vowel /e/ and the mid back vowel /o/. This can be seen in the difference of their F1 frequencies, whereby average F1 frequency of the front and back mid vowels are between 351 Mel to 408 Mel in Assam Sora speech varieties of the four regions whereas, average F1 frequency of /ə/ in the speech varieties of the four regions is between 442 Mel to 468 Mel. Therefore, an difference of 23 Mel to 63 Mel is seen between F1 frequencies of /ə/ and /e/ and a difference of 49 Mel to 104 Mel is seen between the F1 frequencies of /ə/ and /o/ in Assam Sora speech data of the four regions. This also indicates that while the mid central vowel is higher than both front

and back mid vowels, the difference is greater for the back vowel /o/ than for the front vowel /e/. However, although the mid central vowel is produced higher than the front and back mid vowels, vowel height of the mid central vowel is not equalized with the high vowels /i/ and /u/.

Additionally, F2 frequency of the two vowels indicate that, sometimes the vowel /ə/ is produced fronter than the vowel /a/ in Assam Sora. It is observed that in Assam Sora data representing Singrijhan, Sessa and Koilamari speakers, the mid central vowel /ə/ tends to be fronter than the low central vowel /a/. Therefore, an average difference of 43 Mel to 46 Mel is seen in F2 frequency of the two vowels indicating that the two vowels are not aligned to each other as mid central and low central vowels. Whereas, Assam Sora data representing Lamabari speakers shows an average difference of only 19 Mel between the F2 frequencies of the two vowels indicating that the two vowels are fairly aligned as central vowels in Assam Sora. Thus, besides categorically distinguishing the vowels /ə/ and /a/ as mid central and low central vowels in Assam, formant frequency measurements also reveal that the mid central /ə/ is often produced fronter than the low central vowel /a/ in Assam Sora vowel system.

### 4.3.3 Formant frequency of back vowels /u/ and /o/ in Assam Sora

The vowel /u/ in Assam Sora is a rounded high back vowel and the vowel /o/ in Assam Sora is a rounded mid back vowel. Minimal pairs showing the contrast between the two vowels is presented in Table 4.5.

Table 4.5: Contrasting pairs for vowels /u/ and /o/

	Sora	English	Sora	English
1	<i>buɕa</i>	suck	<i>boɕa</i>	sew
2	<i>guda</i>	scratch	<i>goda</i>	clean
3	<i>muʔu</i>	nose	<i>moʔo</i>	eye

The vowels /u/ and /o/ are both produced with a rounded lip stricture. Also, the two vowels are produced by retracting the tongue towards the back of the mouth, but the height of the tongue during the production of the two vowels is different. While /u/ is produced by raising the tongue body to a higher position, /o/ is produced by lowering the tongue body to an intermediate position. Acoustic output captured during the production of these two vowels also reflect their articulatory distinction. Thus, evidence for tongue body retraction towards the back of the mouth is found in F2 frequency of both the vowels, whereby F2 of both the vowels are relatively smaller than the central vowels /ə/ and /a/ indicating that both /u/ and /o/ are back vowels in Assam Sora. Similarly, F1 frequency shows that while /u/ has a lower F1 frequency /o/ has relatively higher F1 frequency, providing evidence that /u/ is a high vowel and /o/ is a mid vowel and that they have distinct vowel heights in Assam Sora. Region-wise average of the first three formant frequencies of the two vowels is presented in Table 4.6.

Table 4.6: Average formant frequency of Assam Sora vowels /u/ and /o/

	Singrijhan		Sessa		Koilamari		Lamabari	
	u	o	u	o	u	o	u	o
F1	327.25 (45.15)	403.34 (49.25)	295.49 (30.63)	377.03 (31.294)	301.11 (35.61)	408.17 (42.25)	299.88 (28.35)	402.08 (38.92)
F2	596.61 (94.98)	614.34 (67.14)	624.22 (69.97)	644.93 (67.30)	587.06 (91.71)	601.31 (63.46)	634.25 (69.20)	649.08 (60.43)
F3	987.91 (41.78)	988.68 (41.20)	956.73 (47.30)	947.43 (50.26)	977.20 (34.65)	980.39 (46.72)	966.35 (50.23)	972.60 (50.75)

From Table 4.6 it is evident that the vowel /u/ is acoustically distinct from the vowel /o/ in Assam Sora and the two vowels differ particularly due to their difference in vowel height characteristics. While average F2 frequency of the two vowels imply that they are both back vowels, average F1 frequency implies that /u/ is high back vowel and /o/ is a mid back vowel in Assam Sora. Also, similar F1 frequency distinction between the two vowels is uniformly present in Assam Sora data recorded from

all four regions. It is observed that, while average F1 frequency of /u/ is between 295 Mel to 327 Mel in Assam Sora speech data of the four regions, average F1 frequency of /o/ is between 377 Mel to 408 Mel. Therefore, average difference in F1 frequencies of the two vowels is between 76 Mel to 107 Mel. This indicates that the two vowels have distinct vowel heights in Assam Sora speech varieties of all four regions whereby, /u/ is a high back vowel and /o/ is a mid back vowel. Moreover, unlike front and central vowels, F2 frequency of the back vowels reveal that both /u/ and /o/ are fairly aligned to each other as high and mid back vowels in Assam Sora. Therefore, it is found that difference between F2 frequency of the two vowels in the four regions is only between 14 Mel to 20 Mel. Thus, formant frequency measurements of /u/ and /o/ provide adequate evidence that the two vowels are categorically distinct from each other in Assam Sora varieties of all four regions.

Hence, formant frequency measurements of the vowels /i, e, a, o, u, ə/ in Assam Sora provides evidence that the six vowels are categorically distinct vowels in Assam Sora. It is evident that formant frequency measurements are equally distinct for Assam Sora speech varieties of the four regions including Singrijhan, Sessa, Koilamari and Lamabari. Additionally, it is observed that some vowels such as the high front vowel /i/ and the mid central vowel /ə/ have a tendency to move fronter and higher than their corresponding vowels. This indicates that the two vowels have a dynamic nature in Assam Sora whereby, the high front vowel /i/ is often produced fronter than mid front vowel /e/ and higher than back high vowel /u/. Similarly, the mid central vowel /ə/ in Assam Sora is produced fronter than the low central vowel /a/ and higher than the mid front and back vowels /e/ and /o/. Consequently, these dynamic vowel characteristics are represented in the shape of the over all vowel system of Assam Sora and can be observed by plotting the vowel formant frequencies across F1 and F2 axis. Thus, the following section presents Assam Sora vowel plots based on their F1 and F2 frequencies separately for the four regions namely, Singrijhan, Sessa, Koilamari

and Lamabari. Also, in order to reduce speaker effect on the production of vowels, F1 and F2 frequencies of all vowels are normalized separately for different regions using vowel normalization method proposed by (Lobonov, 1971). Hence, vowel plots illustrated in the following section represent speaker normalized formant frequencies of all the six vowels in Assam Sora.

#### 4.4 Assam Sora vowel plots

Formant frequency of vowel sounds discussed so far represent the typical speech of Assam Sora speakers living in four different regions of Assam. This includes the speech of both men and women of different age groups and different bilingual abilities. Therefore, the speech data also includes some variation caused due to different physiological characteristics of different speakers. Size of the mouth and length of the vocal tract of people are usually different and this causes the formant frequencies to differ. Therefore, in order to control the variation in formant frequencies caused due to the speaker's physiological differences it is crucial to normalize the formant frequencies in the description of vowel sounds. In this regard Adank et al. (2004) has shown that vowel normalization method proposed by Lobonov (1971) effectively removes the speaker's physiological influence, and preserves the phonemic and sociolinguistic information contained in the formant frequency of vowel sounds. Therefore, this work also adopts the vowel normalization method proposed by (Lobonov, 1971) to reduce speaker effect on the six Assam Sora vowels /i, e, a, o, u, ə/ using equation (4.1).

$$z = \frac{x - \mu}{\sigma} \quad (4.1)$$

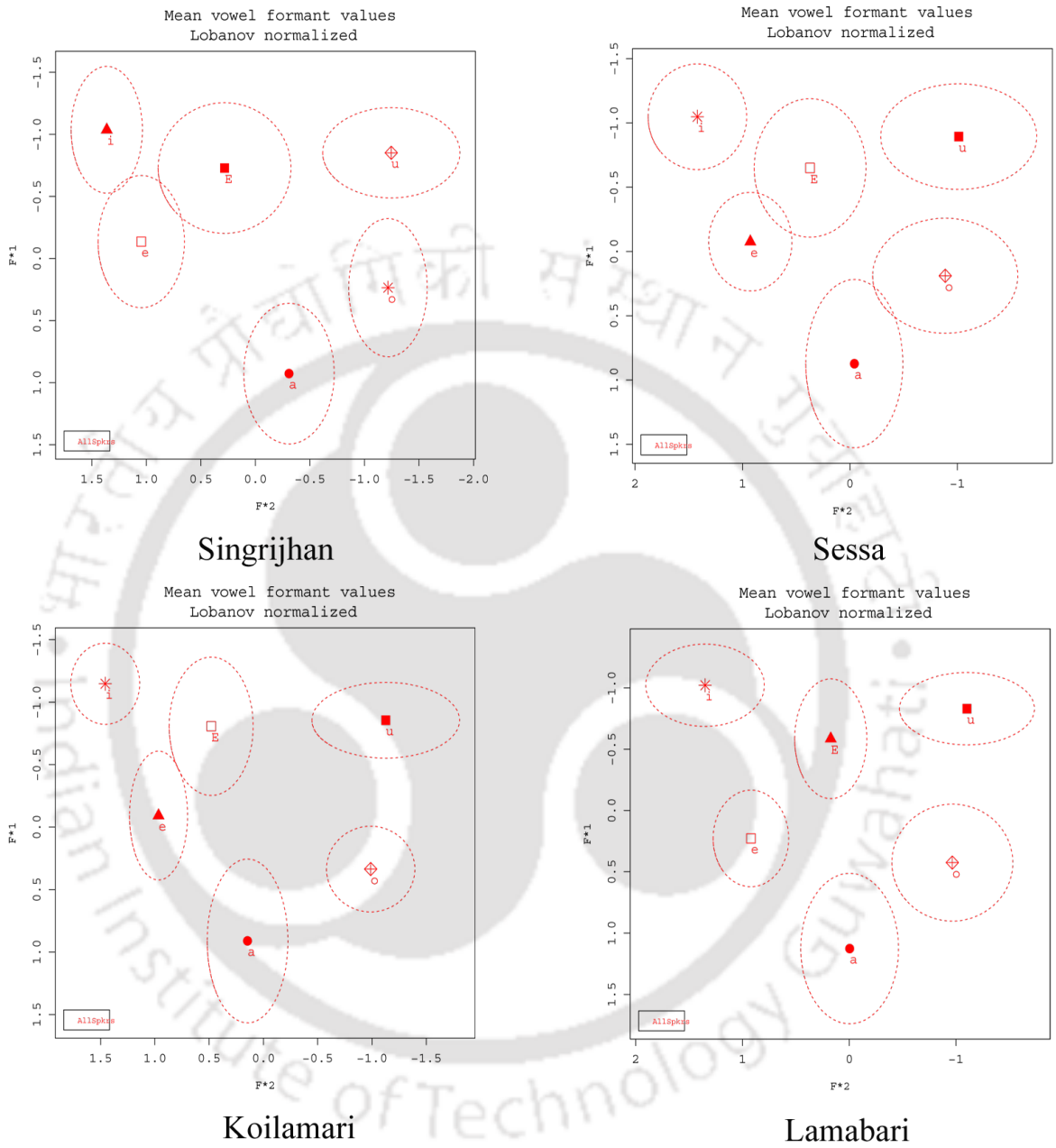
In equation 4.1  $z$  represents the normalized formant frequency calculated for the individual formant value  $x$ ;  $\mu$  represents the mean value of individual speaker's formant frequency values and  $\sigma$  represents the standard deviation in individual speaker's

formant frequency values. Thus, with the help equation 4.1 the six vowel sounds are normalized separately for every speaker and vowel data from the four regions are also treated separately. Subsequently, normalized formant frequencies of all the six vowels are plotted across F1 and F2 axis. A vowel plot across F1 and F2 axis shows the overall shape and structure of the vowel system and therefore, vowel plot based on F1 and F2 frequencies of vowels is an effective way to observe the vowel system of a language. Thus, for this reason, vowel descriptions are often based on vowel plots which also have normalized F1 and F2 frequencies. Thus, in order to visualize the shape of Assam Sora vowel inventory Assam Sora vowel formants are also plotted across F1 and F2 axis separately for the four Assam Sora speaking regions in Assam. Figure 4.1 presents the speaker normalized vowel plots representing Assam Sora speakers living in four different regions of Assam.

Figure 4.1 shows Assam Sora vowel plots for Assam Sora speakers living in Singrihan, Sessa, Koilamari and Lamabari. The vowel plots are based on F1 and F2 frequencies of the six vowels where the central point of every ellipse represents the average of normalized F1 and F2 frequencies of individual vowels and the ellipses represent area containing all F1 and F2 frequencies of individual vowels within one standard deviation of the average. The vowel plots show the shape of Assam Sora vowel inventory including five peripheral vowels /i, e, e, o, u/ and one non-peripheral vowel /ə/. From the vowel plots it is revealed that the six vowels are produced contrastively by Assam Sora speakers of all four regions. Also, it is revealed that the shape of Assam Sora vowel inventory is similar for Assam Sora speakers of all four regions. Thus, the formant frequency analyses here adequately classifies the six vowel system of Assam Sora and evidences suggest that there are no regional variations of Assam Sora on the basis of its vowel phoneme inventory.

Additionally, the study by (Horo & Sarmah, 2015) propose that Assam Sora vowel properties (height and frontness/backness) usually change as a function of syllable

Figure 4.1: Region wise Assam Sora vowel plots



Vowels: i=/i/, e=/e/, a=/a/, o=/o/, u=/u/, E=/ə/

position. Evidences suggest that, Assam Sora vowel property changes across syllables mainly due to the difference of prominence between first and second syllable of disyllabic words. Consequently, overall vowel space of Assam Sora vowel system

also differs across syllables in disyllabic words. Thus, there are evidences that Assam Sora vowel properties are effected as a function of syllable position. Therefore, the following section examines the dynamic nature of Assam Sora vowel formants in first and second syllable of disyllabic words.

## 4.5 Assam Sora vowel formants in individual syllables of disyllables

So far, Assam Sora vowel characteristics have been described by considering vowel samples from both first and second syllable of disyllabic words. However, the acoustic analysis by Horo & Sarmah (2015) provides evidence that Assam Sora vowel property changes across first and second syllable of disyllabic words. Evidences suggest that all six vowels have different F1 and F2 frequencies in first and second syllable of disyllabic words. As a result, similar vowels (/i/ and /i/, /e/ and /e/ etc.) are relatively higher and fronter in one of the two syllables. Also, it is evident that perceptual distance between adjacent vowels change as a function of syllables position. Thus, there are evidences that perceptual distance between adjacent vowels is always greater in one of the two syllables in Assam Sora disyllables. Consequently, due to these effects of syllable position, the shape and structure of Assam Sora vowel system also changes in first and second syllable of disyllabic words. Thus, it is observed that Assam Sora vowel system has different vowel area spaces in first and second syllable of disyllables. However, this dynamic nature of Assam Sora vowel system is mainly reported for Assam Sora speakers living in Singrijhan. Therefore, in order to examine whether Assam Sora vowel properties consistently change across syllables, data from Sessa, Koilamari and Lamabari are also included and are examined here. Hence, considering the fact that Assam Sora vowel formants normally differ in first and second syllable, the following sections mainly examine a) how vowel formants change

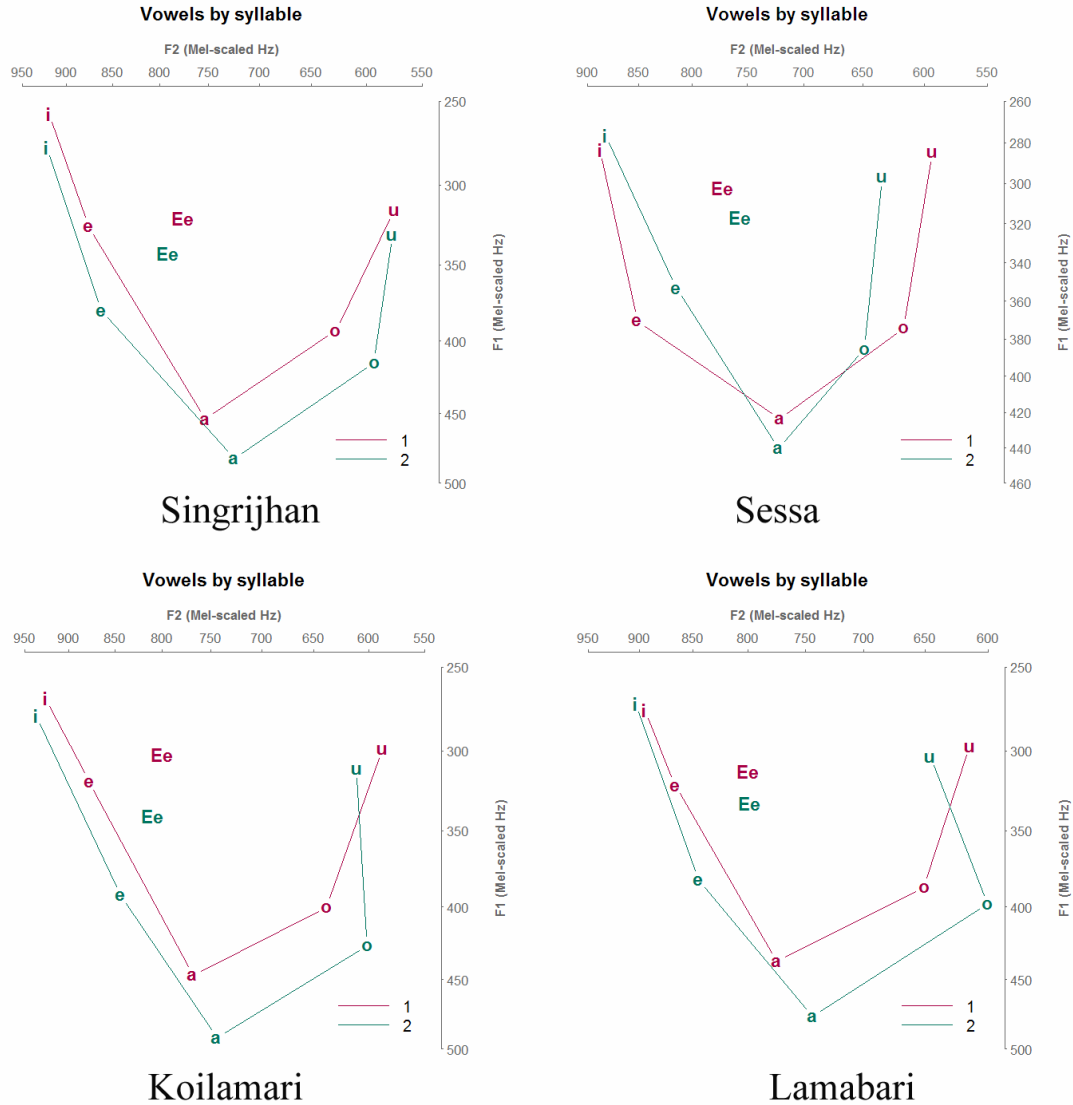
in first and second syllable of Assam Sora disyllables b) how perceptual distance between adjacent vowels differ in first and second syllable of Assam Sora disyllables and c) how the overall vowel space is effected by the syllable position in Assam Sora disyllables. Also, these investigations are done with the same data that has been described in the previous section. However, unlike previous section where vowel samples of both syllables are combined together, in this section the vowel samples are separated on the basis of first and second syllable of Assam Sora disyllables for all disyllable combinations of Assam Sora described in Section 3.2 of Chapter 3 of this work.

#### **4.5.1 F1 and F2 of similar vowels in individual syllables of disyllables**

In order to examine how syllable position effects Assam Sora vowel formants, F1 and F2 frequencies of the six vowels in first and second syllable are plotted next to each other on the same vowel plot. This helps in visualizing the effect of syllable position on Assam Sora vowel formants. Thus, region-wise Assam sora vowel plots across first and second syllable are presented in Figure 4.2.

From Figure 4.2 it is evident that Assam Sora vowels are positioned differently in first and second syllable of disyllabic words. When Assam Sora vowels in first and second syllable are plotted next to each other, it is observed that all the vowels are either raised or fronted in one of the two syllables. For instance, the mid central vowel /ə/ is always produced higher in first syllable than in second syllable. Similarly, the mid front vowel /e/ and the low central vowel /a/ are usually positioned higher in first syllable than in second syllable, except in the vowel plot representing Sessa speakers. Likewise, the mid back vowel /o/ is usually positioned fronter in first syllable than in second syllable in all the vowel plots, except in the vowel plot representing Sessa speakers. Therefore, it is observed that similar vowels in first and second syllable are

Figure 4.2: Assam Sora vowels in 1st and 2nd syllable



Vowels: i=/i/, e=/e/, a=/a/, o=/o/, u=/u/, Ee=/ə/

distant from each other to a certain extent. Thus, by examining Assam Sora vowels separately in first and second syllable of disyllabic words it is revealed that Assam Sora vowels exhibit dynamic nature across first and second syllable of disyllabic words. It is found that, similar vowels in first and second syllable are distant from each other in terms of their F1 and F2 frequencies. Also, this dynamic nature of Assam Sora vowel system is observed regularly in Assam Sora vowel data recorded in all four regions.

However, an exception is found in the Sessa data where the patterns are not similar to the patterns observed in rest of the regions. In this regard, the present work could not find an explainable result, but it is noticeable that despite the exceptions, the primary argument is not effected.

Thus, in order to similar vowels have some degree of perceptual distance, euclidean distance between similar vowels in first and second syllable is calculated for Assam Sora vowel system. Euclidean distance is a mathematical measurement used for calculating distance between any two points in a given space. In case of vowel descriptions, euclidean distance measurement is commonly used to show the perceptual distance between two vowels in the F1 and F2 plane of a vowel space (Lindblom, 1986). In other words, euclidean distance basically measures the distance between the points which represent the position of two vowels in a vowel plot. Also, since vowel points represent a combination of F1 and F2 frequencies, the distance considers both F1 and F2 frequencies of the two vowels for which the euclidean distance is calculated. Therefore, euclidean distance between two vowels represent the combined effect of F1 and F2 frequencies of the two vowels. Thus, in order to estimate the perceptual distance, euclidean distance between similar vowels in first and second syllable of Assam Sora disyllables is calculated using equation 4.2.

$$eud = \sqrt{(F1_x - F1_y)^2 + (F2_x - F2_y)^2} \quad (4.2)$$

In equation 4.2 *eud* represents the euclidean distance between two vowels based on their F1 and F2 frequencies.  $F1_x$ ,  $F1_y$  and  $F2_x$ ,  $F2_y$  represent the F1 and F2 frequencies of the two vowels between which the euclidean distance is calculated. Table 4.7 presents the region-wise euclidean distance estimation between similar vowels across first and second syllable in Assam Sora disyllables.

The euclidean distance estimation in Table 4.7 provides evidence that Assam Sora vowel formants differ across first and second syllable of disyllabic words. Table 4.7

Table 4.7: Euclidean distance between similar vowels across syllables

	Singrijhan	Sessa	Koilamari	Lamabari
i-i	18.07	13.91	15.14	4.95
e-e	56.63	42.54	80.95	65
a-a	42.48	15.36	50.32	48.53
o-o	43.33	35.18	46.84	53.87
u-u	33.09	46.62	28.54	32.03
ə-ə	28.74	21.75	42.64	20.25

reveals that, considerable perceptual distances between similar vowels across first and second syllable of disyllables exists in all six vowels of Assam Sora. As a result, Assam Sora vowel properties such as vowel height, vowel frontness/backness are often changed as a function of syllable position as it is seen in Figure 4.2. This indicates that syllable position has a significant role in vowel characteristics of Assam Sora. Also, it is observed that perceptual distances between similar vowels across first and second syllable of disyllables exists in Assam Sora speech variety of all four regions including Singrijhan, Sessa, Koilamari and Lamabari. This confirms that that changing of vowel formants to certain degree across first and second syllable of disyllabic words is a general tendency in Assam Sora vowel system and the analysis here provides adequate evidences for this phenomenon.

Additionally, it is observed that variation in either F1 or F2 frequency also causes some of the vowels to move away from their canonical positions and to come closer to their adjacent vowels. This can be seen in Figure 4.2 where mid front vowel /e/ is often closer to the high front vowel /i/ in first syllable but not in the second syllable. This indicates that, in addition to changing the vowel quality of individual vowels, syllable position also effects the perceptual distance between adjacent vowels within individual syllables. Thus, in order to examine how syllable position effects the perceptual distance between adjacent vowels within individual syllables, vowel formants of adjacent vowels in first and second syllable are examined in the following section.

#### 4.5.2 F1 and F2 of adjacent vowels in individual syllables of disyllables

By examining the formant frequencies of Assam Sora vowels separately in first and second syllable, it is confirmed that the formant frequencies change as a function of syllable position. In this regard, it is observed that while both F1 and F2 frequencies of all vowels change in first and second syllable, in case of back vowels, greater change is usually observed in F2 frequency, whereas in case of front and central vowels greater change is usually observed in F1 frequency. This indicates that Assam Sora back vowels are often fronted or retracted, and front vowels are often raised or lowered depending on their syllable positions in disyllabic words. This kind of movement is also evident in Figure 4.2 which shows the rising or lowering of front and central vowels and the fronting or retraction of back vowels in Assam Sora vowel system depending on their syllable positions. Thus, it is likely that perceptual distance between front vowels, between central vowels and between back and central vowel is also effected in Assam Sora vowel system as a function of syllable position. Therefore, this section mainly examines the perceptual distance between vowel pairs such as /e/ and /i/, /a/ and /ə/, /o/ and /ə/ and /u/ and /ə/ in the first syllable and the second syllable of Assam Sora disyllables. Moreover, since vowel positions in a vowel system represents both F1 and F2 frequencies, variation in perceptual distance between adjacent vowels in first and second syllable have to be examined by considering both the frequencies. Thus, perceptual distances between the above mentioned vowel pairs are examined by calculating euclidean distance between the vowels using equation 4.2 described earlier.

Regarding front vowels in Assam Sora it is observed that the two front vowels /e/ and /i/ are usually closer to each other in first syllable than in second syllable. This difference in distance between the two front vowels in the first and second syllable can be estimated by calculating euclidean distance between two vowels. Euclidean

distance estimation between the two front vowels /e/ and /i/ in the first and second syllable in all four regions is presented in Table 4.8.

Table 4.8: Euclidean distance between front vowels in first and second syllable

	Singrijhan	Sessa	Koilamari	Lamabari
e-i ( $\sigma$ -1)	81.48	102.54	70.74	54.55
e-i ( $\sigma$ -2)	119.48	104.49	147.07	125.93

From Table 4.8 it is observed that euclidean distance between the front vowels /e/ and /i/ is always lesser in first syllable than in second syllable. This indicates that, front vowels in Assam Sora are often raised or lowered as a function of syllable position. Thus, it is evident that perceptual distance between front vowels /e/ and /i/ in Assam Sora are also effected by the syllable position, whereby distance between the two vowels is often minimized in first syllable and maximized in second syllable.

Similarly, it is observed that low central vowels /a/ and /ə/ are often raised or lowered as a function of syllable position. As a result, the distance between /a/ and /ə/ is often reduced or expanded in Assam Sora vowel system. This variation in distance between /a/ and /ə/ in first and second syllable can be examined from the euclidean distance between the two vowels in first and second syllable of disyllabic words. Region-wise euclidean distance between central vowels /a/ and /ə/ in first and second syllable is presented in Table 4.9.

Table 4.9: Euclidean distance between central vowels in first and second syllable

	Singrijhan	Sessa	Koilamari	Lamabari
a-ə ( $\sigma$ -1)	137.46	103.83	149.63	129.11
a-ə ( $\sigma$ -2)	155.12	114.42	164.92	154.73

From Table 4.9 it is evident that euclidean distance between /a/ and /ə/ is usually lesser in first syllable than in second syllable. This indicates that similar to front vowels, the central vowels are also raised or lowered in Assam Sora vowel system as a function of syllable position in disyllabic words. Thus, it is evident that perceptual distance between the two central vowels is also minimized or maximized in Assam

Sora vowel system depending on their occurrence in the first or second syllable of disyllabic words. Thereby, distance between the two vowels is usually minimized in first syllable but, maximized in second syllable of disyllabic words.

Likewise, in case of back vowels it is observed that the back vowels /o/ and /u/ are often fronted or retracted in Assam Sora vowel system due to the influence of syllable position in disyllabic words. As a result, perceptual distance between the central and back vowels is also effected in Assam Sora vowel system. Thus, different perceptual distances between back and central vowels in first and second syllables can be seen from the euclidean distance between the vowels. Euclidean distance between back and central vowels in Assam Sora vowel system of all four regions is presented in Table 4.10.

Table 4.10: Euclidean distance between back and central vowels in first and second syllable

	Singrijhan	Sessa	Koilamari	Lamabari
o-ə ( $\sigma$ -1)	165.59	146.52	187.98	166.72
o-ə ( $\sigma$ -2)	212.21	112.62	225.76	209.43
u-ə ( $\sigma$ -1)	200.42	178.97	210.71	185.3
u-ə ( $\sigma$ -2)	215.58	121.63	197.97	152.74

From Table 4.10 it is displayed that euclidean distance between /o/ and /ə/ is always lesser in first syllable except in vowel data recorded from Sessa and euclidean distance between /u/ and /ə/ is always lesser in second syllable except in vowel data recorded from Singrijhan. Thus, it is evident that perceptual distance between central and back vowels changes as a function of syllable position. As a result, distance between high back vowel and mid central vowel is usually minimized in the second syllable but is maximized in the first syllable and distance between mid back vowel and mid central vowels is minimized in the first syllable, but maximized in second syllable. However, this work could not find an explainable reason for the exceptions found in Sessa and Singrijhan data reported here, but it is apprehend that the primary argument is not effected by the exceptions.

Hence, by examining the euclidean distance between some of the adjacent vowels in Assam Sora vowels system it is revealed that perceptual distance between adjacent vowels is often minimized or maximized as a function of syllable position. It is evident that front vowels are relatively close to each other in first syllable, but separated further in second syllable. Similarly, the central vowels are relatively closer to each other in the first syllable than in the second syllable. Likewise, the back vowels also have different distances from the mid central vowel depending on their syllable position in disyllabic words. Moreover, in addition to these vowels, it is confirmed that perceptual distance between all adjacent vowels in Assam Sora vowels system is reduced or expanded to a certain degree as a function of syllable position. Hence, as per the predictions of Adaptive Dispersion theory that vowel system may restructure depending on the maximization and minimization of perceptual contrast between vowels, it is likely that the overall shape and structure of Assam Sora vowel system will also differ across syllables. Therefore, in order to examine the overall impact of syllable position on Assam Sora vowel system, Assam Sora vowel area space in disyllables is analyzed in the following section.

### **4.5.3 Assam Sora vowel space in individual syllables of disyllables**

Vowel space basically corresponds to the articulatory space within which all the distinctive vowels of a particular language are produced. Thus, vowel space provides information about the extent of tongue body movement inside the mouth while producing all the distinctive vowels of the language. Moreover, since formant frequencies are acoustic consequences of the articulatory movements, the F1 and F2 space within which all the vowels are located represents the acoustic vowel space of the language. Also, since vowel systems are often described on the basis of perceptual contrast between the distinct vowel sounds, vowel space also represents the perceptual space

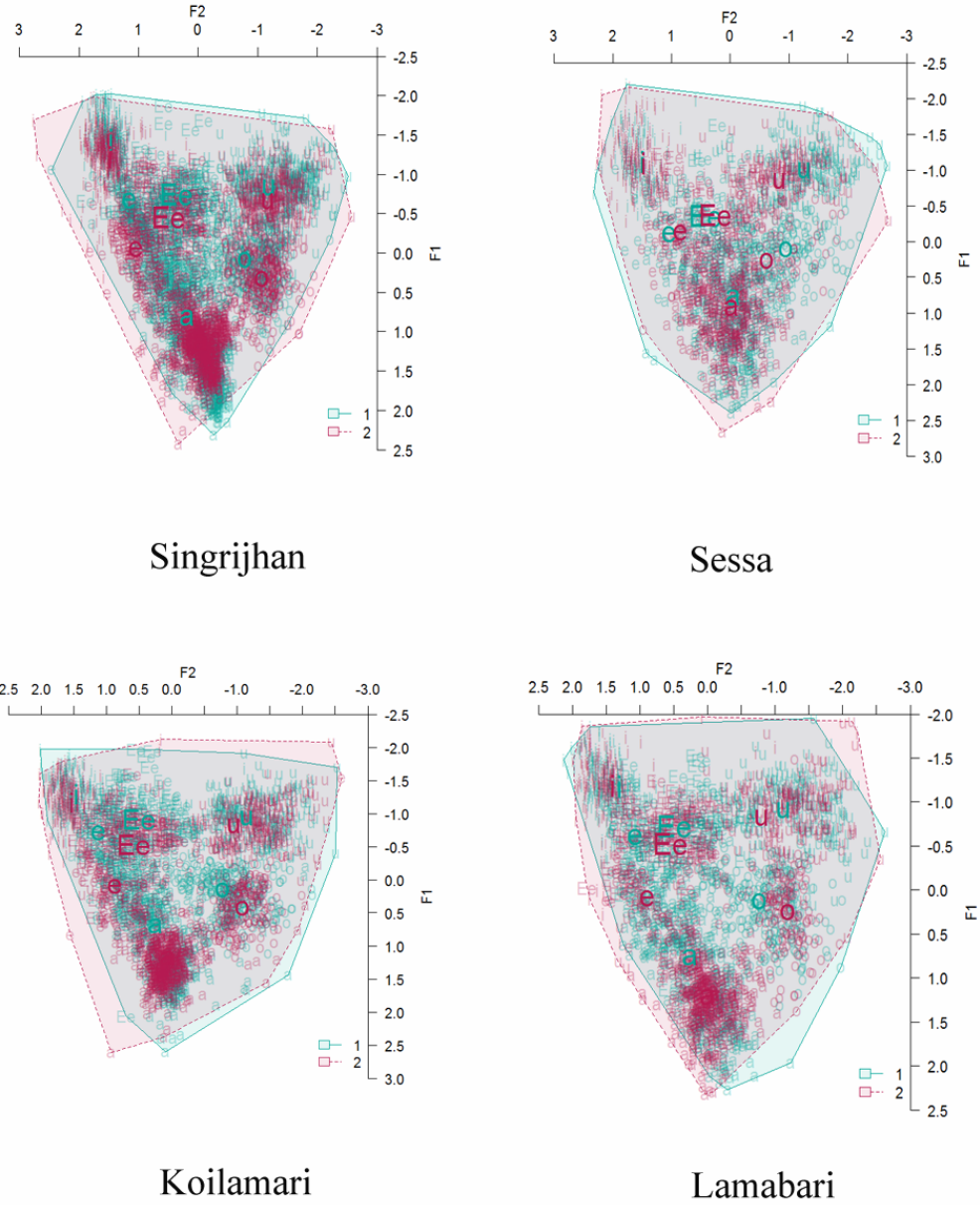
within which all the vowels of a vowel system are contrastively perceived in a language. Therefore, in the description of vowel systems, vowel space is often analyzed in order to to examine the difference in vowel system of two dialects (Jacewicz et al., 2007) or to examine gender variations in the vowel system of a particular language (Lengeris, 2016). Thus, analysis of vowel space provides information about the overall shape and structure of vowel systems. Also, comparison between vowel spaces of two vowel systems provide evidence regarding similarity or dissimilarity in the overall shape and structure of the vowels systems. Hence, considering the fact that Assam Sora vowel formants change across syllables, it is relevant to examine whether the overall shape and structure of the vowel system also changes as a function of syllable position. Therefore, this section examines the vowel space of Assam Sora vowel system in first and second syllable of disyllabic words.

In order to visualize the overall vowel space of Assam Sora vowel system in first and second syllable all the vowel points in each of the two syllables are plotted next to each other with closed circumferences. The vowel plots include all vowel tokens in first and second syllable and all vowel formants are also normalized for speaker effect using vowel normalization method proposed by Lobonov (1971). Thus, the region-wise vowel plots presented in Figure 4.3 represent the overall vowel space plots for Assam Sora vowel system in first and second syllable of disyllabic words.

From Figure 4.3 it is evident that the overall vowel spaces of Assam Sora vowel system in first and second syllable are fairly similar. However, in order to compare the actual vowel space in first and second syllable, vowel area space of Assam Sora vowel system in the two syllables is estimated. For this purpose, the vowel syatem is divided into three parts representing three vowel triangles such as /i-e-a/, /i-a-u/ and /u-o-a/. Then area of the three vowel triangles is calculated using equation 4.3.

$$area = \sqrt{s(s-a)(s-b)(s-c)} \quad (4.3)$$

Figure 4.3: Assam Sora vowel space in first and second syllable



Vowels: i=/i/, e=/e/, a=/a/, o=/o/, u=/u/, Ee=/ə/

In equation 4.3 ‘*a*’, ‘*b*’, ‘*c*’ represent lengths of the three sides of a vowel triangle and ‘*s*’ represents the summation of *a*, *b* and *c* or ( $a+b+c$ ). Also, lengths of the three sides of every vowel triangle basically represents the euclidean distance between two vowel points. For example, in the the vowel triangle /i-e-a/, ‘*a*’ represents the

euclidean distance between /i/ and /e/, 'b' represents the euclidean distance between /e/ and /a/ and 'c' represents the euclidean distance between /a/ and /i/ in the vowel triangle. Subsequently, after calculating the area of the three vowel triangles in a vowel system, overall vowel area space of the vowel system is estimated by adding the areas of the three vowel triangles. Estimated vowel area space of Assam Sora vowel system in first and second syllable of disyllabic words is presented in Table 4.11.

Table 4.11: Vowel area estimation in first and second syllable

	Singrijhan	Sessa	Koilamari	Lamabari
$\sigma$ -1	33765.48	30979.57	32858.72	26708.66
$\sigma$ -2	39110.32	22235.24	41111.6	35219.97

Estimation of vowel area space presented in Table 4.11 reveals that vowel area space of Assam Sora vowel system is essentially different in the first and second syllable of disyllabic words. This indicates that, the vowel space encompassing all the vowel sounds in Assam Sora vowel system is normally effected as a function of syllable position. Consequently, it is observed that vowel area space is usually larger in second syllable than in first syllable except in case of Assam Sora vowel data recorded from Sessa where a reverse relation is observed. In this regard, it has been observed that Assam Sora speakers living in Sessa frequently use an Indo-Aryan language namely Sadri and it is likely that their bilingual ability is influencing their native language. However, more evidences are required to investigate this speech behaviour, and since the vowel properties discussed so far have not shown any significant difference, it is argued that there are no regional variations of Assam Sora based on its vowel properties. On the other hand, a usually larger vowel space in second syllable can be convincingly related to the the phonetic prominence of second syllable over first syllable. As it is shown in Section 3.4 of Chapter 3, the second syllable is phonetically more prominent than first syllable in Assam Sora disyllables. It is proposed that the phonetic prominence of second syllable is influencing the vowel area space of Assam

Sora vowel system. Thus, except in vowel data of Sessa speakers, vowel space of Assam Sora vowel system is always larger in the second syllable and smaller in the first syllable. However, despite the exception in Sessa data, the general analysis of Assam Sora vowel formants in this work adequately confirms that syllable position in disyllabic words is an essential factor effecting the vowel system of Assam Sora. Hence, it is clearly evident that syllable position not only effects the formant frequencies of individual vowels in the two syllables, but it also effects the overall shape and structure of Assam Sora vowel system.

## **4.6 Distinctiveness of Assam Sora vowels**

Finally, in order to see the perceptual distinctiveness among Assam Sora vowels, a series of statistical tests are conducted and the euclidean distance is again calculated among all the six vowels in Assam Sora. Moreover, considering the fact that there are no major regional variations of Assam Sora vowel system, the statistical tests and the euclidean distance measurement are conducted by combining together Assam Sora data of all four regions recorded in this work. Also, in order to see a general distinction between the six vowels in Assam Sora, the vowel data of both first and second syllable of Assam Sora disyllables are combined together. Thus, the following section reports the results drawn from the euclidean distance measurements and statistical tests conducted for confirming the distinctiveness of the six vowels in Assam Sora.

### **4.6.1 Euclidean distance and statistical distinctiveness**

In order to examine the categorical distinction between the six vowel in Assam Sora in terms of their perceptual contrast, euclidean distance between the six vowels in Assam Sora is calculated. For this purpose, vowel data from all four regions are combined together and euclidean distance between every six vowel is calculated. Thus, Table

4.12 presents the euclidean distance between the six vowel in Assam Sora including both first and second syllable produced by Assam Sora speakers of all four regions included in this work.

Table 4.12: Vowel euclidean distance between Assam Sora vowels

	a	i	u	e	o	ə
i	252.79	0	–	–	–	–
u	205.9	310.03	0	–	–	–
e	148.38	105.99	253.34	0	–	–
o	136.19	317.63	92.92	234.52	0	–
ə	143.62	130.94	189.01	68.99	186.69	0

Table 4.12 provides evidence that maximum perceptual contrast exists between all six vowel in Assam Sora. It is found that the maximum perceptual contrast always exists between vowels that are on opposite sides of the vowel periphery. Whereas, perceptual contrast between neighbouring vowels or between the non-peripheral and the peripheral is relatively less. Thus, the maximum perceptual contrast is found between the high front vowel /i/ and the mid back vowel and the minimum perceptual contrast is found between the mid central vowel /ə/ and the mid front vowel /e/. This indicates that while the perceptual contrast between peripheral vowels of opposite sides are maximally distinct from each other, the distinction between the non-peripheral and the peripheral vowels are only sufficiently distinct from each other. Therefore, in order to further confirm the significance of their distinctiveness, a one-way ANOVA test was conducted considering F1 and F2 frequencies of all vowels as the dependent variable and the six vowels as factors. Thus, the statistical tests reveal that both F1 and F2 frequencies of all six vowels in Assam Sora are significantly distinct from each other (For F1  $F(5,12696) = 48442.146$ ,  $p < 0.001$ ; For F2  $F(5, 12396) = 5195.808$ ,  $p < 0.001$ ). Additionally, a post-hoc Bonferroni test further confirmed that the average difference in F1 and F2 frequencies of every vowel is significantly different. Hence, the statistical tests substantiate the acoustic analysis which proposes that Assam Sora has six vowels and every vowel is categorically distinct from each other.

## 4.7 Conclusions

This chapter provides a detailed description of the six vowel phonemes /i, e, a, o, u, ə/ of Assam Sora vowel system. Description of the six vowel phonemes in Assam Sora is based on their phonetic properties and evidence for their distinctive characteristics is substantiated with acoustic analysis. Acoustic analysis of vowels in this chapter is primarily based on formant frequency measurements of the vowels in Assam Sora. Formant frequency analysis in this chapter has mainly examined the two lower formant frequencies (F1 and F2) which provide information about vowel height and vowel frontness/backness. Thus, from the interpretation of F1 and F2 frequencies of Assam Sora vowel system, it is revealed that the six vowels are categorically distinct from each other. Also, the analysis here confirms that formant frequency characteristics of the six vowels are similar in Assam Sora vowel data of all four regions included in this chapter namely, Singrijhan, Sessa, Koilamari and Lamabari.

Additionally, besides categorically distinguishing the six vowels in Assam Sora, formant frequency measurements in this chapter has also revealed some additional characteristics of Assam Sora vowel system. The analysis in this chapter reveals that F1 and F2 frequencies of Assam Sora vowel formants change in first and second syllable of disyllabic words. As a result, it is seen that Assam Sora vowel system has different formant frequencies in first and second syllable of Assam Sora disyllables. Specifically, it is shown that similar vowels in first and second syllable have different vowel formants and adjacent vowels also have different perceptual distances depending of their position of occurrence in first or second syllable of disyllabic words. Consequently, overall shape and structure of Assam Sora vowel system also changes across first and second syllable of disyllabic words. In this regard, evidence has shown that Assam Sora vowel system has different vowel area spaces in first and second syllables. Thereby, Assam Sora vowel system usually has a larger vowel area space in the second syllable in comparison to the vowel area space in the first syllable in disyllabic words.

Hence, this chapter concludes that the six vowels are categorically distinct in Assam Sora across four regions in Assam and syllable position plays an important role in Assam Sora vowel system. Also, since the second syllable is more prominent than the first syllable, the vowel area space is usually larger in the second syllable than in the first syllable. Likewise, perceptual distance between adjacent vowel is also maximum in second syllable than in first syllable. Additionally, from the discussion of vowel system theories in this chapter it is understood that shape and structure of vowel systems depend on the type of vowels in the vowel system as well as on the relation between individual vowels in the vowel system. Thereby, it is realized that vowel systems having stable and maximally dispersed vowels are the most optimal vowel systems in languages of the world. Therefore, small vowel systems are more optimal and crosslinguistically more common than large vowel systems. In this regard, the Assam Sora vowel system with six vowel phonemes, adequately resembles an optimal vowel system and the analysis in this chapter provides the adequate evidences.

## Chapter 5

# Acoustic phonetic properties of stops and nasals in Assam Sora

### 5.1 Introduction

This chapter describes two types of consonants in Assam Sora consonant inventory, namely, stops and nasals with the help of acoustic analysis. This includes describing the acoustic properties correlating to the articulatory characteristics of the two consonant categories in Assam Sora. It is known that articulatory characteristics of consonant sounds are based on their place of articulation, manner of articulation, and voicing contrast. Therefore, acoustic properties of consonant sounds provide the cues for identifying articulatory characteristics of the consonant sounds. Consonant sounds include a specific amount of oral constriction prior to the production of the sound, a period of oral constriction, and a moment of releasing the oral constriction during which the sound is produced. Thus, based on the degree of articulatory constriction, consonants are commonly categorized as sonorants and obstruents. While sonorants include consonant sounds that do not have any articulatory constriction to impede the air flow, obstruents include consonant sounds that have a constriction to impede the air flow, creating turbulence during the release (Ohala, 1997). Thus, acoustic analysis describing the manner of articulation examines the degree of oral constriction involved and identifies its acoustic consequences on the resulting conso-

nant sound. For instance, while a rapid transition of formant frequency at the point of consonant release and vowel start indicates that the consonant is a stop consonant, a smoother rate of formant transition in the same region indicates that the consonant is an approximant consonant (Raphael, 2005). Evidences suggest, depending on the place of oral constriction in the vocal tract, the resulting consonants may have different acoustic properties. Thus, acoustic features such as Voice Onset Time (VOT), transition of higher formants are some of the acoustic correlates of the place of articulation feature of stop consonants. Moreover, during the production of some consonant sounds a buzz like sound can be heard. This is caused by the vibration of vocal cords taking place simultaneously during the production of the consonant sounds. Consequently, some consonants sounds have a voicing feature in addition to their other two features, namely place and manner of articulation. In this regard, the acoustic feature particularly VOT, is proposed to be an important acoustic cue for distinguishing voiced and voiceless stop consonants crosslinguistically. Hence, acoustic studies have substantiated the articulatory descriptions of consonant sounds and as a result consonant sounds are identified by their specific acoustic characteristics.

Regarding the consonant inventory of Assam Sora it is known from Section 2.4 of Chapter 2, that Assam Sora has nine obstruent consonants and nine sonorant consonants. The nine obstruents include three manners of articulations, namely, stops, fricative and affricate. This includes seven stops, one fricative and one affricate consonant. The seven stop consonants /p, b, t, d, k, g, ʔ/ include five places of articulations, namely labial, dental, alveolar, velar and glottal. The voiceless fricative /s/ is an alveolar consonant and the voiced affricate /ɟ/ is a post-alveolar consonant. Moreover, the obstruent consonants except the glottal stop /ʔ/ also have voicing contrasts, but as mentioned in Section 2.4 of Chapter 2, there is an asymmetry of voicing contrast in the case of dental and alveolar stop consonants. Thereby, while the dental stop /t/ has a voiceless segment, the voiced segment is absent and the

alveolar stop /d/ has a voiced segment and the voiceless segment is absent in Assam Sora. Among sonorant consonants, there are five manners of articulations, including nasal, lateral, trill, tap and approximant. The nasal consonants /m, n, ɲ, ŋ/ include four places of articulation, namely labial, alveolar, palatal and velar. The lateral and the trill consonant /l/ and /r/ include one place of articulation, namely, alveolar. The tap consonant /ɽ/ is a retroflex sound and the approximant consonants /w, j/ include two places of articulation, namely labial and palatal.

On the other hand, regarding Orissa Sora consonant inventory, the literature review in Section 2.2.5 of Chapter 2 has shown that different studies have proposed different consonant inventories for the same language. Thus, at least three consonant inventories are proposed including Ramamurti (1931, 1938), Stampe (1965) and A. R. Zide (1982). Additionally, a provisional inventory of the consonant sounds in Orissa Sora is presented by Anderson & Harrison (2008b). However, although these studies propose different consonant sounds in Orissa Sora consonant inventory, it is noticed that at least 15 consonants in Orissa Sora are consistently present in all the consonant inventories proposed. The common consonants include seven obstruents and eight sonorants. The obstruent category includes the six stop consonants /p, b, t, k, g, ʔ/ representing three places of articulations, namely labial, dental, velar and glottal and one fricative consonant /s/ in the alveolar region. Among sonorants there are four nasal consonants /m, n, ɲ, ŋ/ representing four places of articulations, namely, labial, alveolar, palatal and velar; one trill and one lateral consonant /r, l/ in the alveolar region; one retroflex tap consonant /ɽ/ and one approximant consonant /j/ representing the palatal place of articulation.

Significantly, the 15 common consonants in different consonant inventories of Orissa Sora are also found in Assam Sora. Also, the remaining consonants in Assam Sora, except the voiced post-alveolar affricate /ɟʃ/, are included in the individual studies on Orissa Sora mentioned above. For instance, the voiced alveolar stop /d/ is

included in (Ramamurti, 1931), (A. R. Zide, 1982) and (Anderson & Harrison, 2008b). Similarly, the labial approximant /w/ is included in (Stampe, 1965). It is only the voiced post-alveolar affricate /ɖ/ in Assam Sora that is not reported in any of the existing descriptions of Orissa Sora. However, palatal sounds reported in Orissa Sora might have a resemblance with the voiced affricate consonant in Assam Sora. It is reported that, the voiced palatal sounds /j/ (Ramamurti, 1931) and /ɟ/ (A. R. Zide, 1982) in Orissa Sora have [z] as a free variation and according to A. R. Zide (1982) the palatal sound /ɟ/ is also produced as [ɖ] in some instance. Thus, there are some evidences that a voiced post-alveolar affricate consonant is observed in Orissa Sora. Moreover, the voiced post-alveolar affricate consonant is a typical consonant in Munda languages and is not an innovation of the Assam Sora consonant inventory. Hence, it is observed that Assam Sora consonant inventory consists of the consonant phonemes that are also proposed for Orissa Sora by different scholars.

Moreover, a typical characteristics of Orissa Sora stop consonants, namely, the asymmetric relation between voiceless dental stop /t/ and voiced alveolar stop /d/, is also maintained in Assam Sora. Hence, it is found that Orissa Sora consonant phonemes proposed by various scholars are largely similar to the consonant inventory of Assam Sora. Also, Assam Sora consonants are typologically similar to Munda consonant system, particularly the presence of consonant features such as /t/ and /d/ voicing asymmetry and the presence of a tap retroflex consonant /ɽ/. Hence, the consonant system of Assam Sora substantiates the fact that consonant system of Munda languages are consistent.

Thus, considering that there are no major differences between the consonants of Assam Sora and Orissa Sora, this work only examines and documents the phonetic characteristics of two consonant categories in Assam Sora. This includes the stop consonants and the nasal consonants in Assam Sora. In this regard, description of nasal consonants include the four nasal sounds /m, n, ɲ, ŋ/, and description of stop

consonants include the six oral stop consonants /p, b, t, d, k, g/ and a glottal stop /p/. Thus, the following section (§5.2) describes the phonetic properties of Assam Sora nasal consonants with the help of acoustic analysis, then §5.3 describes the six oral stops with particular reference to the characteristics of voicing contrast among the six oral stops in Assam Sora. Subsequently, §5.4 describes the phonetic realization of the glottal stop in Assam Sora, and §5.5 concludes the chapter by summarizing the major findings of this chapter.

## 5.2 Nasal consonants in Assam Sora

Assam Sora has four nasal consonants representing four places of articulations, namely, labial, alveolar, palatal and velar. Of these four nasal consonants, only the labial nasal /m/ and the alveolar nasal /n/ occurs in all three phonemic environments in Assam Sora. Whereas, while the palatal nasal /ɲ/ occurs only word initially and word medially, the velar nasal /ŋ/ occurs only word medially and word finally. This section mainly classifies the phonetic characteristics of the word final nasal in Assam Sora with the help of acoustic analysis.

### 5.2.1 Acoustic characteristics of nasal consonants in Assam Sora

Nasal consonants, being sonorants, have very similar spectral characteristics to vowels (Stevens, 2000). This implies that nasal sounds can also be classified with the help of formant frequencies that are generally used for classifying vowels. However, nasal sounds have some additional formant frequencies. This is because, in the production of nasals, while the nasal cavity is the primary resonator, the oral cavity, closed by the lips (in case of bilabial nasals) and closed by the tongue (in case of alveolar and velar nasals), also act as the side cavity. As a result, while nasal formant is observed

consistently at around 200 Hz for all nasals, other formants also occur for the nasals (Stevens, 2000). Moreover, apart from the formant frequencies, anti-formants are also observed in nasals. Prominent anti-formants arise in nasals mainly due to an oral coupling with nasal sounds. For example, in case of a bilabial nasal, the closed oral cavity acts as a side-branch and the side branch takes out the energy of the main tube. This results in creating of anti-formants in the bilabial nasal. Significantly, due to these anti-formants amplitude of the nasal formant frequency is lowered in nasal sounds. Additionally, anti-formant characteristics also vary with place of articulation of nasal sounds. For instance, anti-formant characteristics of alveolar and velar nasals are much lower than the bilabial nasal, since in alveolar and velar nasals the side oral cavity is much smaller than the side oral cavity of the bilabial nasal. Thus, prominence of nasal formants and anti-formants are generally taken into consideration for the classification of nasal consonants. Therefore, the classification of the nasals in Assam Sora is done by examining the nasal formants and the anti-formants.

In order to examine the acoustic characteristic of nasal sounds in Assam Sora, a data set consisting words with nasal codas including /m/, /n/ and /ŋ/ is considered. For /m/, 320 tokens; for /n/, 79 tokens and for /ŋ/, 638 are extracted from the speech database collected from all four regions of Assam Sora. From the nasal data, spectrum is extracted for each nasal at the middle of the total duration of the nasal sounds in the coda position only using a Praat script. In order to do this, each sound file containing nasal sounds is transformed to a spectrogram using Praat (Boersma & Weenink, 1992) with a Gaussian window of 5 ms window, 2 ms time steps, and 20 Hz of frequency step with a maximum frequency of 5000 Hz. Following this, spectrum at the mid point of the total duration of every nasal is calculated and power of the spectrum is derived using Equation 5.1, where  $r$  and  $i$  are the real and imaginary components, respectively, of the Discrete Fourier Transform of the speech signal. Moreover, power of the spectrum is calculated in the 0 – 5000 Hz frequency range at

115 equidistant points, resulting in an interval of 43.5 Hz. Subsequently, the averaged spectrum across all nasal tokens is computed and displayed as seen in Figure 5.1.

$$pwr = 10 \times (\log_{10}(r^2 + i^2)) \quad (5.1)$$

Figure 5.1: Average nasal spectra of Assam Sora with formants indicated with up arrows and anti-formants indicated with down arrows

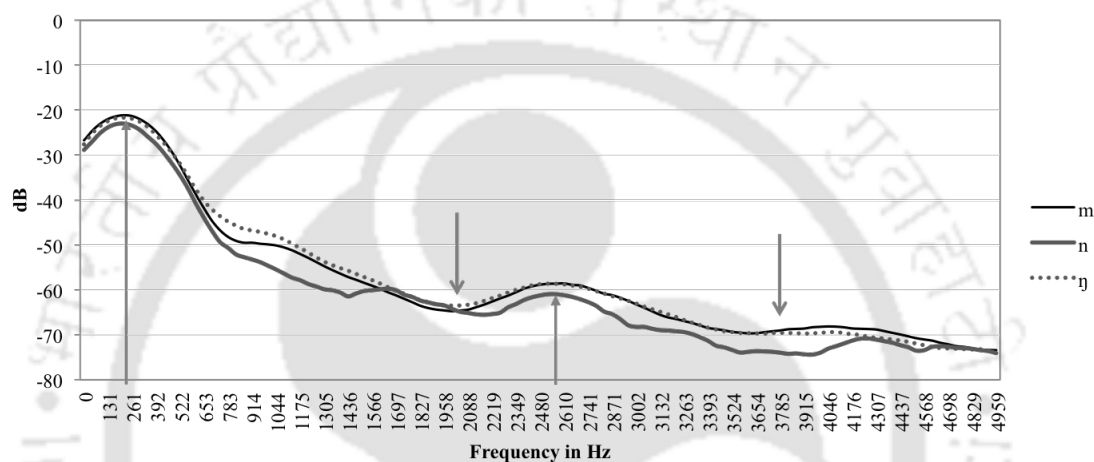
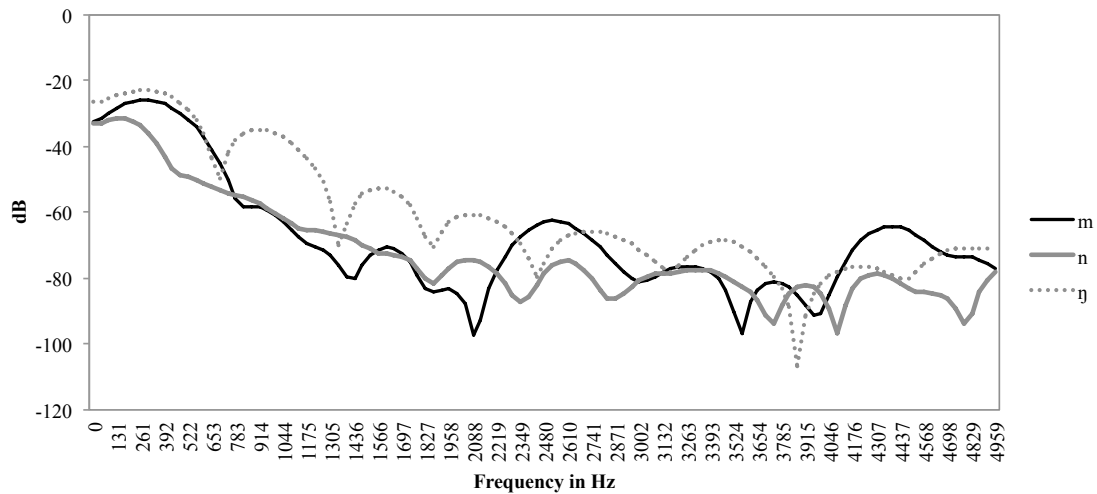


Figure 5.1 shows the occurrence of a prominent nasal formant at around 250 Hz, as indicated by an up arrow. Similarly, another prominent formant is noticed at around 2500 Hz. It is noteworthy that there is a considerable difference between the amplitude of the first formant and the second formant. This difference is caused by the presence of anti-formants just preceding the second formant as indicated with a down arrow. Similarly, another anti-formant realized at about 3700 Hz has dampened the following spectrum almost flat. While the spectral peaks are clearly visible in the averaged spectrum of /m/, /n/ and /ŋ/, the side-branch induced differences among the three nasals are not noticeable. The reason for this maybe due to the averaging of the nasal spectrum that has smoothed off any place of articulation specific (or side-branch specific) differences. Therefore, in order to see the characteristics of the nasal spectrum, without spectral characteristics being smoothed out, single spectra of a female speaker producing three nasal sounds /m/, /n/ and /ŋ/ is plotted in Figure

5.2.

Figure 5.2: Nasal spectra of Assam Sora as produced by a single speaker



From Figure 5.2 it is evident that, all three nasals show a formant peak at about 200 Hz. The bilabial consonant /m/ shows prominent spectral peaks at 260 Hz, 2500 Hz and 4300 Hz. Another peak at about 1500 Hz is dampened by anti-formants approximately at 1300 Hz and 2000 Hz. In case of the alveolar nasal /n/, only the first nasal peak is the prominent one whereas, potential nasal peak at about 2500 Hz is dampened by anti-formants. The spectrum for the velar nasal /ŋ/ shows a clear nasal formant at around 250 Hz and its harmonics are also quite prominent. However, no other prominent nasal peaks are noticed in the spectrum apart from the harmonic peaks. In case of the bilabial nasal /m/, there are two noticeable anti-formants at about 2000 Hz and at about 3500 Hz. Similarly, in case of the alveolar nasal /n/ a prominent anti-formant is noticed at about 4000 Hz. In case of velar nasal /ŋ/ also, an anti-formant around 4000 Hz is prominent. Hence, considering, Figure 5.1 and Figure 5.2, it is can concluded that in case of Assam Sora nasals, nasal peaks are predominantly notice between 200 – 250 Hz and another nasal peak is noticed between 2500 – 3000 Hz. Similarly, a prominent nasal anti-formant is also noticed between 3500 – 4000 Hz. hence, from these nasal peaks and prominent anti-formants,

nasal consonants in the word final position can be identified in a continuous speech in Assam Sora.

### 5.3 Voicing contrast in Assam Sora stop consonants

Voicing in stop consonants is a phonemic feature in Assam Sora. Based on the voicing contrast, there are two categories of oral stops in Assam Sora namely voiced stops including /b, d, g/ and voiceless stops including /p, t, k/. Also, a typical voicing feature of the stop consonants in Assam Sora is the presence of voicing asymmetry between the voiceless dental stop /t̪/ and the voiced alveolar stop /d/. Speech data recorded in this work clearly shows that while /t̪/ is a voiceless dental stop, /d/ is a voiced alveolar stop in Assam Sora. However, the voicing asymmetry is sometimes neutralized, particularly during the production of the voiced counterpart. Some Assam Sora speakers tend to produce a voiced dental stop [d̪] in place of the voiced alveolar stop /d/. This alternation is also reported in Orissa Sora. In this regard, Ramamurti (1931) suggests that in Orissa Sora, the alternation between voiced dental stop [d̪] and voiced alveolar stop /d/ is phonologically conditioned by homophonic sounds in the environment of the stop consonant. However, in case of Assam Sora it is observed that the alternation is basically a free variation, and such instances are not very frequent. Assam Sora speakers only occasionally tend to produce a voiced dental stop [d̪] in place of the voiced alveolar stop /d/. Moreover, it is observed that production of the voiceless dental stop /t̪/ is never altered by any of the Assam Sora speakers recorded in this work. Additionally, it has been observed that only a few Assam Sora speakers occasionally produce an aspirated velar stop consonant [k<sup>h</sup>] in place of its unaspirated counterpart /k/, particularly in the word initial position. In this regard, it is apparent that the aspirated consonant is borrowed from an

Indo-Aryan language namely Sadri, which is predominantly spoken in the vicinity of Assam Sora speakers in all four regions recorded in this work. However, adaptation of aspirated stop consonants by Assam Sora speakers is still in its early stage. It is observed, only a few fluent bilingual Assam Sora speakers occasionally produce the voiceless aspirated velar stop [k<sup>h</sup>] in place of the voiceless unaspirated velar stop /k/. Whereas, the dental stop /t/ and the labial stop /p/ is never found to be aspirated by any of the speakers. Similarly, the voiced stops /b, d, g/ are also never aspirated by any Assam Sora speakers, recorded in this work. Thus, besides a two way voicing contrast and a voicing asymmetry between /t/ and /d/, there are also instances of some Assam Sora speakers occasionally producing a voiced dental stop [d̥] as a free variation, and a voiceless aspirated stop [k<sup>h</sup>] as a borrowed sound. Hence, in order to show these voicing features of Assam Sora stop consonants, the acoustic correlates of voicing contrast in stop consonants, Voice Onset Time (VOT), is examined for the six stop consonants in Assam Sora in the subsequent section. However, prior to this the following section briefly describes the process of examining the acoustic characteristics of voicing contrast in stop consonants.

### **5.3.1 Acoustic characteristics of voicing contrast in stop consonants**

Phonetically stop consonants are categorized as voiced or voiceless on the basis of their difference in articulation at the source of the sound. Voiced stops are normally produced with a glottal buzz. The glottal buzz is caused by the vibration of vocal cords at the time of releasing voiced stops which is absent during the production of voiceless stops. Thus, action at the larynx provides the basic articulatory cue for distinguishing between voiced and voiceless stop consonants. Laryngeal activity provides the acoustic cue for distinguishing between voiced and voiceless stops in languages of the world. Generally, voiced stop segments exhibit low frequency harmonics in the

spectrogram which are absent in voiceless stop segments. Thus, presence versus absence of the low frequency region represents the basic acoustic cue for distinguishing voiced and voiceless stop consonants. However, acoustic study of stop consonants has considered a few other cues which claim to distinguish voiced and voiceless stop consonants in a better way, in terms of their production and perception features. The studies have examined both naturally produced speech signals, as well as syntactically modified speech stimulus. Two commonly known acoustic cues that can be used to identify and discriminate voiced and voiceless stop consonants are formant transition (Liberman et al., 1954), Halle et al. (1957), (Liberman et al., 1958), Harris et al. (1958) and Voice Onset Time (VOT) (Lisker & Abramson, 1964), (Lisker & Abramson, 1965), (Abramson & Lisker, 1965), (Lisker et al., 1977) and (Lisker, 1978). Evidences suggest that formant transition represents the rapid change in vowel formants when a stop consonant is adjacent to a vowel sound. Therefore, formant transition is believed to capture the articulatory movement made while articulators move from a consonant sound to the adjacent vowel sound. And Voice Onset Time (VOT) represents the time interval between the release of an oral stop and the onset of the glottal pulse of the following vowel. Acoustic studies on stop consonants have employed both these methods to distinguish voicing contrast, as well as to predict place of articulation of stop consonants crosslinguistically.

Formant transition cues are mostly used to define the perceptual quality of stop consonants. The focus is primarily on the listener's ability to discriminate voiced and voiceless stop consonants. Studies have explored formant transition cue in first, second and third formant and it is assumed that usually voiced stops have a transition into the following vowel which is lacking in voiceless stops. Studies examining formant transition cues in stop consonants are often based on syntactically modified speech stimuli, and are mostly tested on speakers of a well documented language such as, English. Remarkably, such studies have provided satisfactory results and shown that

listeners were able to discriminate between voiced and voiceless stops on the basis of formant transition cues only. However, Kewley-Port (1982) found results using natural speech, which actually do not support the findings. His research reveals that natural speech has different formant transitions, and listeners are not able to discriminate the stop categories as accurately as reported in the earlier studies. Thus, although formant transition is used to discriminate voiced and voiceless stops and to predict consonantal place of articulation, it has limitations in the analysis of naturally produced stop consonants.

The other acoustic cue used for discriminating voiced and voiceless stop consonants is voice onset time or VOT. As mentioned earlier, VOT represents the time interval between the release of a stop consonant and the onset of glottal pulse of the following vowel sound. Similar to formant transition cue, studies examining the VOT cue are also based on natural as well as syntactically modified speech stimuli. Studies based on syntactically modified speech stimuli show that voiced stops have shorter VOT and voiceless stops have longer VOT. Also, evidences suggest that there is a sharp perceptual contrast between the VOT of voiced and voiceless stop consonants. Likewise, natural speech data also shows that VOT of stops can be broadly classified into just two clusters, but intermediate VOT is rare. Thus, unlike formant transition cue, VOT measurement has shown that discrimination of voiced and voiceless stop consonants is equally effective for synthetic stimuli and natural speech. Also, VOT measurement has helped in generalizing the phonetic universal for distinguishing voiced and voiceless stop consonants crosslinguistically. Hence, the following section examines the VOT characteristics of the six stop consonants in Assam Sora.

### **5.3.2 VOT of Assam Sora stop consonants**

Acoustic categorization of voiced and voiceless stops in Assam Sora is accomplished by examining their VOT features. The analysis is based on naturally produced stop

consonants in prevocalic position and the speech data represents Assam Sora speakers of four geographical regions including Singrija, Sessa, Koilamari and Lambari. The data set consists of initial CV and CVC syllable of disyllabic words where all six stop consonants precede the low central vowel /a/ of Assam Sora vowel system. A set of near minimal pairs showing voicing contrast in word initial stop consonants is presented in Table 5.1.

Table 5.1: Word initial voicing contrast in Assam Sora stop consonants

	Sora	English	Sora	English	Sora	English
/p/	pata	crack	palu	white	paʒa	bring
/b/	batar	bad taste	bali	feather	baʒsa	good
/t/	tagi	hot	taŋga	already eaten food	taʔta	slap
/d/	daga	sun	daʒa	cover from heat	daʔa	water
/k/	kata	story	kani	this	kaʒip	long sword
/g/	gatar	hit	ganai	tuber	gaʒiŋ	alliance

For examining VOT difference in the voiced and voiceless stop consonants, the speech data is manually annotated for VOT interval between release of the stop consonants and the beginning of glottalic pulse of the following vowel. This interval marks the voice onset time and provides information about the voicing contrast in stop consonants. In order to visualize the VOT period in the speech data, Figure 5.3, 5.4 and 5.5 presents the visual display of wave forms of the six stop consonants captured in words such as *pata* meaning crack; *batar* meaning bad taste; *taʔta* meaning slap; *daʔa* meaning water; *kata* meaning story and *gatar* meaning hit.

Subsequently, VOT values of the six stop consonants are extracted. Also, VOT of the voiced stops /b, d, g/ are assigned negative values since it is observed that the glottalic pulse in voiced stops, indicating voicing, begins much earlier before the oral closure is released. Whereas, in voiceless stop, the glottalic pulse begins only after the release of the stop consonants. This indicates that the voiced stops in Assam Sora have a voicing lead whereas voiceless stops have a voicing lag (Lisker & Abramson, 1964). Thus, the pre-voicing in voiced stops is also visible in the wave

Figure 5.3: Wave forms of word initial /p/ and /b/

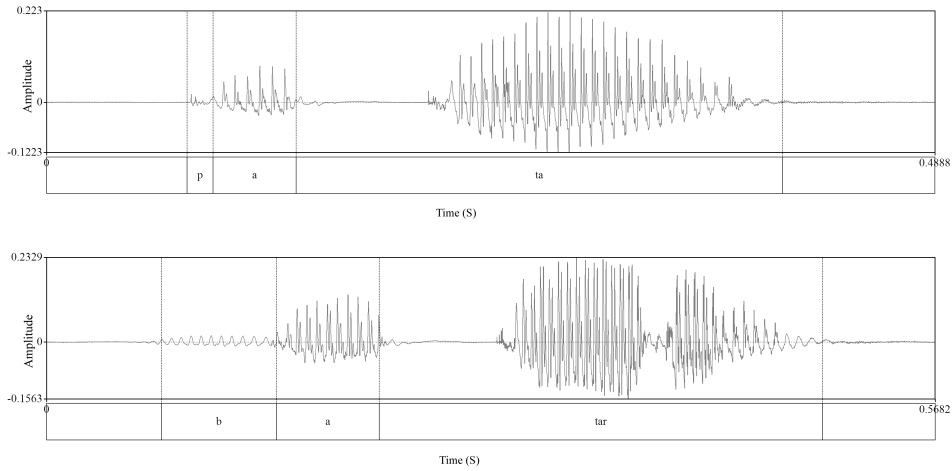
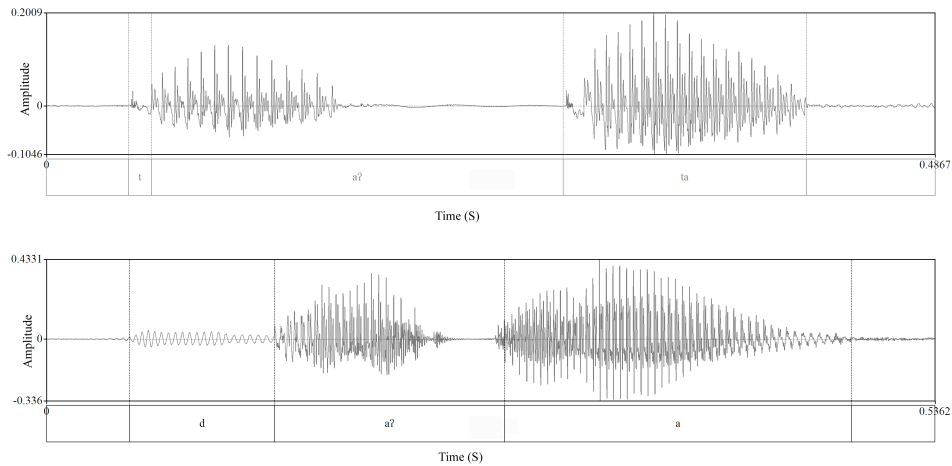


Figure 5.4: Wave forms of word initial /t/ and /d/



forms presented above. In order to represent the pre-voicing period of the voiced stop, their VOT values are assigned negative values. Table 5.2 and 5.3 present the average VOT of /p, t, k/ and /b, d, g/ in the word initial position as produced by Assam Sora speakers of different geographical locations. Also, standard deviation for the same is given in parenthesis.

Table 5.2: Average VOT of voiced stops

	Singrijhan	Sessa	Koilamari	Lamabari
b	-54.68 (21.25)	-56.32 (19.09)	-57.96 (16.37)	-62 (13.66)
d	-56.13 (18.80)	-59.77 (19.41)	-55.83 (17.17)	-56.87 (17.17)
g	-57.36 (14.92)	-57.70 (18.93)	-50.51 (16.13)	-59.55 (17.20)

Figure 5.5: Wave forms of word initial /k/ and /g/

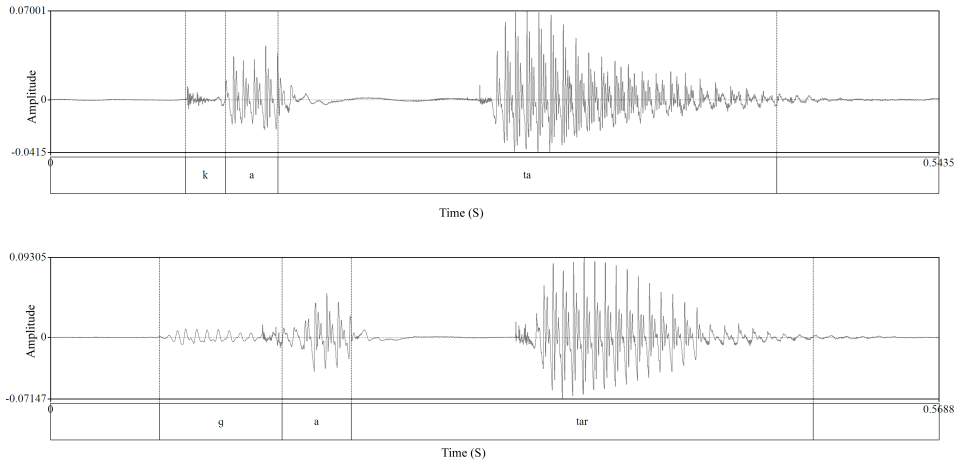


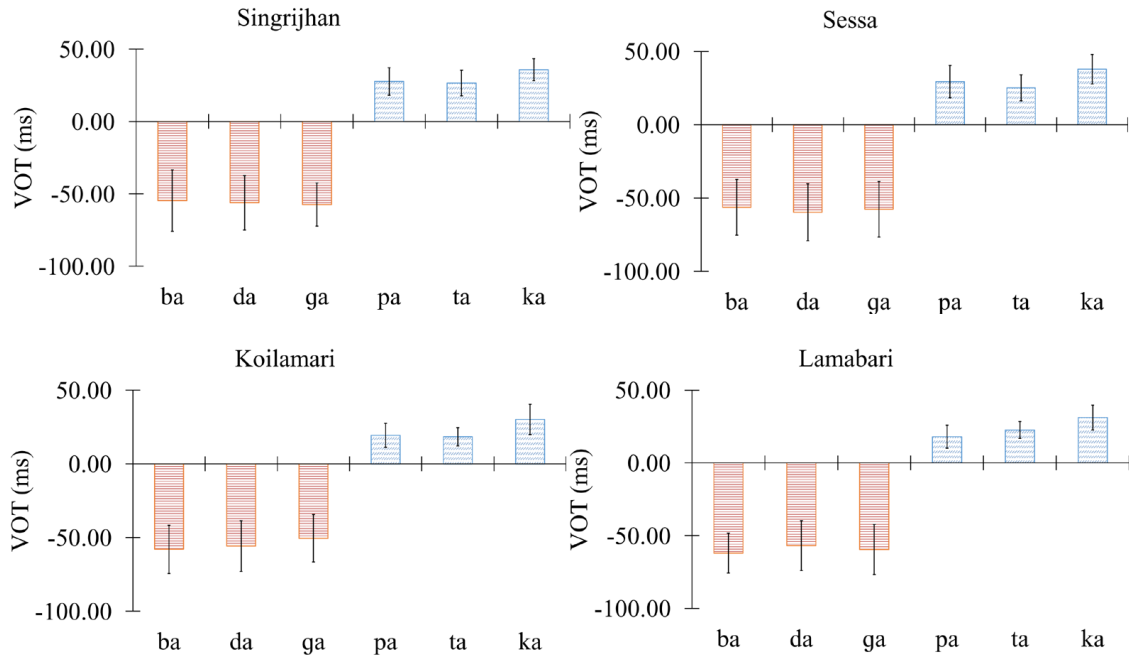
Table 5.3: Average VOT of voiceless stops

	Singrijhan	Sessa	Koilamari	Lamabari
p	27.65 (9.50)	29.29 (11.04)	19.35 (8.09)	17.95 (7.92)
t	26.61 (8.86)	25.10 (8.93)	18.33 (6.20)	22.60 (5.76)
k	35.75 (7.58)	37.82 (10.09)	30.12 (10.26)	31.14 (8.52)

From Table 5.2 and 5.3, the acoustic analysis reveals that VOT boundary between the two stop categories in Assam Sora is clearly demarcated. It is evident that while the average VOT of voiceless stops /p, t, k/ ranges between 17 to 37 milliseconds, the average VOT of voiced stops /b, d, g/ ranges between -50 milliseconds to -62 milliseconds. Also, the VOT difference between voiced and voiceless stops in Assam Sora is fairly similar to Assam Sora speakers of all four regions including Singrijhan, Sessa, Koilamari and Lamabari. This indicates that the voicing contrast in Assam Sora does not overlap and voicing contrast between voiced and voiceless stop consonants is categorically distinct. Thus, in order to visualize the distinction Figure 5.6 presents a graphical representation of the average VOT values with error bars representing standard deviations.

Figure 5.6 shows that voiced and voiceless stops in Assam Sora can be effectively categorized on the basis of their VOT values. Moreover, in order to further confirm

Figure 5.6: VOT of voiced and voiceless stops in Assam Sora



the results, a one-way ANOVA was conducted considering VOT as the dependent variable and voicing contrast as factor. Consequently, it is found that VOT of voiced and voiceless stop consonants are significantly different in word initial position for Assam Sora speakers of all four regions recorded in this work (Singrijhan  $F(1,232) = 109.366, p < 0.001$ ); (Sessa  $F(1,318) = 247.510, p < 0.001$ ); (Koilamari  $F(1,236) = 257.621, p < 0.001$ ); (Lamabari  $F(1,185) = 256.664, p < 0.001$ ). Hence, acoustic analysis of the six stop consonants in Assam Sora confirms that the VOT of voiced and voiceless stops categorically distinguishes their voicing contrasts, and the difference is statistically significant.

Additionally, regarding the occasional production of a voiced dental stop [d̪] and a voiceless velar aspirated stop [k<sup>h</sup>], it is found that their VOT distinctions are also categorically distinct from the six stop consonants described above. Table 5.4 presents the average VOT of the voiced dental [d̪] stop and the voiceless velar aspirated stop

occasionally produced by a few Assam Sora speakers recorded in this work. The standard deviation is given in parenthesis and data sample size is also given in the form of percentage.

Table 5.4: VOT of dental [d̪] and aspirated [kʰ]

	VOT of [d̪]	% of samples	VOT of [kʰ]	% of samples
Singrijhan	-82.13 (32)	0.84	–	0
Sessa	-86.37 (42.31)	0.92	62.07 (9.95)	1.23
Koilamari	-68.36 (12.73)	2.57	38.46 (3.79)	1.24
Lamabari	-86.42 (28.42)	5.07	40.41 (0.96)	1.05

From Table 5.4 it is primarily observed that while the voiced dental stop [d̪] occurs less than 5% in Assam Sora speech data, the voiceless velar aspirated stop [kʰ] occurs less than 1.5% in Assam Sora speech data. Secondly, it is observed that average VOT of both the sounds are distinct from voiced and voiceless stop categories described above. It is observed that while the average voicing lead of the three voiced stops /b, d, g/ is between -50 milliseconds to -62 milliseconds, for the voiced dental stop [d̪] it is between -68 milliseconds to -86 milliseconds. This indicates that the pre-voicing period for the voiced dental stop [d̪] is usually longer than the three voiced stops in Assam Sora. Similarly, it is observed that while the VOT of the voiceless stops in Assam Sora are between 17 milliseconds to 37 milliseconds, for the voiceless aspirated stop it is above 38 milliseconds. This indicates that the voicing lag time for the voiceless aspirated sounds is longer than that of the voiceless stops in Assam Sora. Thus, it is evident that the free variant [d̪] and the borrowed sound [kʰ] also have distinguishable VOT differences. However, since, the two sounds are only occasionally produced by a few speakers, it is difficult to make generalizations about these two sounds in Assam Sora. The analysis here only provides evidence which shows, although the two sounds are not completely adapted in Assam Sora, their phonetic features are apparently distinct from the existing stop consonants in the language. If the two sounds are eventually adapted into Assam Sora, they may

maintain distinct characteristics or merge with the characteristics of the existing stop consonants in Assam Sora.

## 5.4 Glottal stop /ʔ/ in Assam Sora

Glottal constriction or glottalization is a frequently occurring sound phenomena in the Munda language, as well as in other branches of the Austroasiatic Language Family. In these languages, a glottal constriction is generally realized as a phonation type or as a phonemic sound segment. Commonly occurring phonation types in Austroasiatic languages include glottalized consonants, glottalized vowels and pre-glottalized obstruents. Most Munda languages have a glottal stop /ʔ/ or/and a pre-glottalized obstruent usually known as checked consonants in Munda literature. Thus, pre-glottalized obstruents reported in Munda languages include the following sounds /<sup>ʔ</sup>p, <sup>ʔ</sup>b, <sup>ʔ</sup>t, <sup>ʔ</sup>d, <sup>ʔ</sup>k, <sup>ʔ</sup>g, <sup>ʔ</sup>c, <sup>ʔ</sup>j/. However, the review of Munda segmental properties in Section 2.2.2 of Chapter 2 suggests that pre-glottalized obstruents in Munda languages are better represented as allophones of the non-glottalized stop consonants, since they occur only in word final position in place of the non-glottalized stop consonants. For instance in Remo, Bhattacharya (1965) mentions that there is a glottal stop as well as two pre-glottalized obstruents, /<sup>ʔ</sup>k/ and /<sup>ʔ</sup>p/. However, distribution of the glottal elements in Remo suggests that while the pre-glottalized consonants occur only in word final position, the glottal stop occurs in final, medial as well as in word initial position. Moreover, evidences suggest that the pre-glottalized consonants in Remo are also phonologically conditioned whereby, [<sup>ʔ</sup>k] and [<sup>ʔ</sup>p] are changed of voiced stops /g/ and /b/ when followed by a vowel sound. Similarly, although Ramamurti (1931) proposes that Orissa Sora has a glottal stop /ʔ/ and some pre-glottalized consonants including [<sup>ʔ</sup>b, <sup>ʔ</sup>d, <sup>ʔ</sup>r, <sup>ʔ</sup>n, <sup>ʔ</sup>ɲ, <sup>ʔ</sup>ɳ, <sup>ʔ</sup>j, <sup>ʔ</sup>z, <sup>ʔ</sup>l], his descriptions suggest, similar to other Munda languages, Orissa Sora also has only one glottal element being the

glottal stop /ʔ/. The pre-glottalized consonants are only allophonic variation of the non-glottalized consonants which occur in the word final position only. Thus, it is more likely that a glottal stop /ʔ/ is the actual glottal phoneme in Munda languages, whereas the pre-glottalized consonants are only allophonic variation of the non-glottalized stop consonants in Munda languages.

Additionally, other important characteristics of the glottal constriction in Munda languages is reported in a South Munda language namely Gorum. Rau (2011) argues, glottal constriction in Gorum is realized as three distinct glottal phonemes including a glottal stop /ʔ/, a pre-glottalized voiced obstruent /ʔ<sup>o</sup>/ and creaky vowels /ʋ/. Evidences suggest, while the glottal stop and the creaky vowels are clear glottal phenomena differing only in the degree of glottal constriction, the pre-glottalized onstruents which occur in the boundary between a vowel and a non-glottalized voiced stop appear as complex sounds. Also, Rau (2011) proposes that the three glottal sounds are essentially consonant phonemes in Gorum, and are related to the syllable structure and prosodic structure of the language. Evidences suggest that only one glottal element is allowed in a syllable, and it is confined to the syllable rhyme of the word. Likewise, syllable weight is also attributed to the syllable which consists of the glottal element. Thus, it is evident that glottal constriction is a significant phenomena in Gorum, and evidences suggest that unlike other Munda languages Gorum exhibits an extensive phonologization of the three glottal elements.

In case of Assam Sora, the overview of the consonant phonemes, in Section 2.2.5 of Chapter 2 reveals that, similar to Orissa Sora, Assam Sora also has only one glottal element, namely the glottal stop /ʔ/. Also, evidences suggest that the glottal stop in Assam Sora mainly occurs word medially and word finally. Additionally, as part of this research work, acoustic phonetic characteristics of the glottal in Assam Sora are also presented in Kalita et al. (2016). This study primarily classifies the glottal stop in Assam Sora on the basis of its phonetic properties. The study propose acoustic

features which helps in identifying the glottal stop in Assam Sora speech data. Thus, some of the major findings reported in the study are described in the following section.

#### 5.4.1 Phonetic realization of /ʔ/ in Assam Sora

The study by Kalita et al. (2016) mainly examines intervocalic glottal stops in Assam Sora. This includes disyllabic words such as *uʔu* meaning hair, *iʔi* meaning louse, *daʔa* meaning water etc. The study reveals that intervocalically, the glottal stop in Assam Sora has three different phonetic realizations including a complete voiceless glottal stop, a voiced glottal stop and a creaky voiced glottal stop. A wave form captured while producing the word *iʔi*, in Assam Sora, illustrates the presence of a complete voiceless glottal stop in Assam Sora in Figure 5.7.

Figure 5.7: Wave form showing complete glottal stop in the word *iʔi*

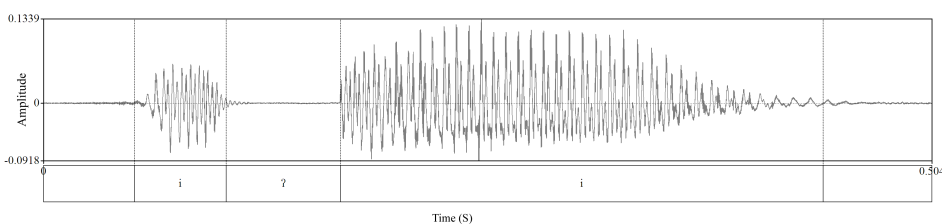


Figure 5.7 shows that periodic wave forms, representing vowel sounds, are completely ceased during the production of the glottal stop in the word *iʔi*. This indicates that the glottal stop is able to interrupt the voice source by completely closing the vocal folds for a certain period of time. Following this when the glottal constriction is released, the vocal folds start vibrating again resulting in resuming of the periodic waves. Thus, as indicated in Figure 5.7, a complete voiceless glottal stop in Assam Sora is characterized by the absence of acoustic energy, caused by an occlusion at the source of the sound namely, the glottis. Therefore, a complete voiceless glottal stop in intervocalic position can be clearly discriminated from the adjacent vowel sounds in Assam Sora.

The second type of glottal stop in Assam Sora is a voiced glottal stop. The voiced glottal stop is distinct from the complete voiceless glottal stop because unlike the complete voiceless glottal stop, the voiced glottal stop does not completely interrupt the voice source. Rather, in voiced glottal stops continuous voicing over the glottal stop region is observed. This realization of the glottal stop in Assam Sora can be seen in Figure 5.8, which shows the wave form of the voiced glottal stop captured during production of the word *iʔi* in Assam Sora.

Figure 5.8: Wave form showing voiced glottal stop in the word *iʔi*

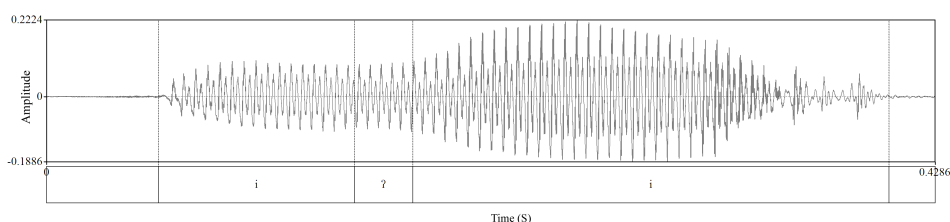


Figure 5.8 shows that glottal occlusion during the production of voiced glottal stop is not as prominent as it is in case of the complete voiceless glottal stop. This indicates, in voiced glottal stops the vocal cords are not completely occluded, and therefore they continue to vibrate. As a result, glottalic cycles of the preceding vowel are carried over the glottal stop and continued onto the following vowel sound. However, due to a pressure of creating a glottalic occlusion, amplitude of the speech signal where the glottal stop occurs is relatively lower than the adjacent vowel segments. Thus, besides having the voiced characteristics of the adjacent vowels, voiced glottal stop in intervocalic position is also characterized with a low amplitude in Assam Sora.

The third type of glottal stop in Assam Sora is a creaky voiced glottal stop. The creaky voiced glottal stop is similar to the voiced glottal stop in Assam Sora. The creaky voiced glottal stop also has continuous voicing, but unlike the voiced glottal stop, it is characterized with abrupt changes in the voice source. The abrupt change in the voice source during the production of creaky voiced glottal stops distinguishes it from the voiced glottal stop. A wave form of the creaky voiced glottal stop captured

while producing the word *iʔi*, in Assam Sora, is presented in Figure 5.9.

Figure 5.9: Wave form showing creaky voiced glottal stop in the word *iʔi*

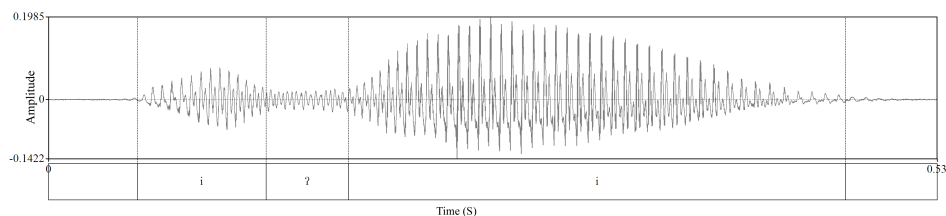
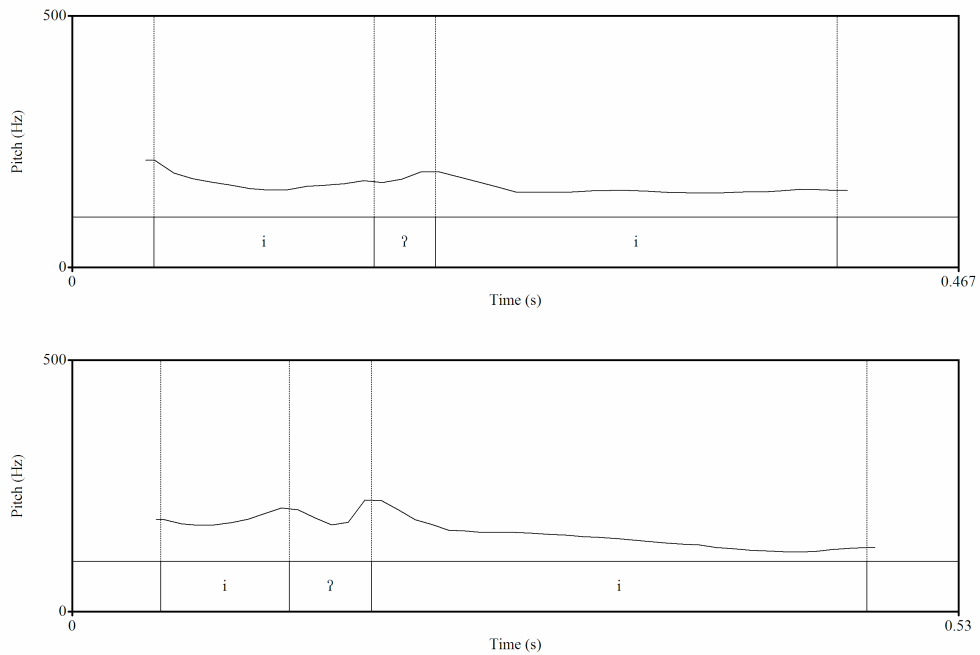


Figure 5.9 shows that some interruption is caused during the production of a creaky voiced glottal stop, but the interruption is not as significant as it is seen in case of the complete glottal stop. Creaky voiced glottal stop allows continuation of vocal cords vibration from the preceding vowel to following vowel. However, the pressure to make the glottal occlusion is so strong that the acoustic energy is significantly lowered at the point where the creaky voiced glottal stop occurs intervocalically. Consequently, the voice source abruptly changes at the point where the creaky voiced glottal stop occurs. Additionally, creaky voiced glottal stop in Assam Sora also exhibits a variation in its pitch pattern. Figure 5.10 shows the pitch extracts of the creaky voiced glottal stop captured while producing the word *iʔi* in Assam Sora.

The top panel in Figure 5.10 shows that the creaky voiced glottal stop has a rising pitch. And the bottom panel shows that the creaky voiced glottal stop has a dipped pitch. This indicates that abrupt changes in voice source during the production of creaky voiced glottal stop also perturbs pitch variation. In this regard, (Hillenbrand & Houde, 1996) proposes that pitch perturbation is natural phenomena in intervocalic glottal stops. Also, evidences suggest that a dip pitch is perceptually more prominent than a rising rising pitch in identifying intervocalic glottal stops in continuous speech. Thus, perturbation of pitch pattern during the production of creaky voiced glottal stop is an important phonetic characteristic which helps in identifying intervocalic glottal stops in Assam Sora.

Figure 5.10: Two types of pitch patterns in creaky voiced glottal stop



## 5.5 Conclusions

This chapter has described the stops and nasals of Assam Sora consonant inventory with the help of acoustic analysis. Description of the nasal consonants in this chapter has mainly shown the acoustic characteristics of word final nasals in Assam Sora. Assam Sora has four nasal consonants including a bilabial nasal, an alveolar nasal, a palatal nasal and a velar nasal. Except the palatal nasal, all three nasals can occur in the word final position in Assam Sora. From the descriptions in this chapter, it is confirmed that the nasals, being sonorant sounds, can be classified with the help of prominent nasal formants and anti-formants. Thus, for the three word final nasals in Assam Sora, including /m, n, ŋ/, it is confirmed that prominent nasal formants are noticeable between 200 - 250 Hz and between 2500 - 4000 Hz. Also, anti-formant, that is created due to an oral coupling during the production of nasal consonants, is noticeable between 3500 - 4000 Hz for the three word final nasal consonants in Assam Sora. Hence, based on these characteristics, this chapter concludes that the nasal sounds in Assam Sora can be adequately described with the help of spectral

characteristics. Also, it is proposed that taking these characteristics of the nasal sounds, word final nasal consonants can be identified in a continuous speech in Assam Sora.

For the oral stop in Assam Sora, this chapter has mainly shown the acoustic characteristics of voicing contrast among the stop consonants in the word initial position. Assam Sora has six oral stops including four places of articulations, namely, labial, dental, alveolar and velar and all the six stop consonants can occur word initially. The six oral stops also have voicing contrasts in Assam Sora. However, there is a voicing asymmetry between the dental and the alveolar stop consonants in Assam Sora. While Assam Sora has a voiceless dental stop, the voiced counter part is an alveolar stop. Thus, Assam Sora has neither a voiced dental stop nor a voiceless alveolar stop. Only occasionally some Assam Sora speakers tend to produce a voiced dental stop in place of the voiced alveolar stop. In this regard, this chapter has concluded that the voiced dental stop occurs as a free variation of the voiced dental stop in Assam Sora. Also, the data in this chapter reveals that the voiced dental stop has occurred in less than 5% of the total data used in the study of Assam Sora stop consonants. This chapter has also revealed that Assam Sora speakers tend to produce a voiceless aspirated velar stop in place of the voiceless unaspirated velar stop. In this regard, this chapter proposes that an aspirated consonant is not native to Assam Sora consonant inventory and only some fluent bilingual Assam Sora speakers produce an aspirated voiceless velar stop in place of the voiceless unaspirated velar stop. Thus, this chapter has concluded that the aspirated consonant is borrowed from the Indo-Aryan lingua franca Sadri, which is predominantly spoken in the vicinity of the Assam Sora speakers in Assam. Moreover, since the aspirated consonant occurs in less than 1.25% of the data used for the study of Assam Sora stop consonants, this chapter concludes that the aspirated consonant is not completely adopted in Assam Sora. Additionally, regarding the acoustic analysis of voicing contrast in Assam Sora stop consonants,

this chapter concludes that the Voice Onset Time (VOT) feature clearly differentiates the voiced stop from the voiceless stop in Assam Sora, particularly in the word initial position. Analysis in this chapter has shown, in the word initial position the voiced stops have a period of prevoicing, usually between -50 to -62 milliseconds. As a result, the voiced stops have a voicing lead even before the oral closure is released in those consonants. On the other hand, in voiceless stops, the voicing in the following vowel begins only after oral closure of the stop consonant is released, usually after 17 to 37 milliseconds. As a result, the voiceless stops are characterized with a voicing in Assam Sora. Hence, the analysis in this chapter concludes that the VOT feature adequately discriminates the voiced and the voiceless consonants in Assam Sora in the word initial position.

Description of the glottal stop in Assam Sora in this chapter is reported from a study presented in Kalita et al. (2016), which was carried out as a part of this research. Thus, the descriptions reported in this chapter show that intervocalic glottal stop in Assam Sora has three types of phonetic realizations. They include a complete voiceless glottal stop, a voiced glottal stop and a creaky voiced glottal stop. In this regard, it is shown that the complete voiceless glottal stop is characterized by a complete interruption of the voice source where the glottal stop occurs intervocalically. On the other hand, the voiced glottal stop is characterized by continuation of vocal cords vibration from the preceding vowel onto the following vowel. Also, a relatively low amplitude at the point where the intervocalic glottal stop occurs is an important characteristic of the voiced glottal stop in Assam Sora. The third type of glottal stop in Assam Sora, the creaky voiced glottal stop, is similar to the voiced glottal stop as there is continuous voicing even in the creaky voiced glottal stop. However, the pressure to make an glottal occlusion in the creaky voiced glottal stop is stronger than in the voiced glottal stop, but less significant than the complete voiceless glottal stop. As a result, there are abrupt changes in voice source at the point where the

creaky voiced glottal stop occurs intervocalically. Consequently, the abrupt changes in the voice source causes perturbation of pitch at the point where the creaky voiced glottal stop occurs. The resulting pitch pattern includes a rising pitch as well as a dipped pitch. Thus, creaky voiced glottal stop in Assam Sora is characterized by a low amplitude as well as a perturbed pitch patterns in the the intervocalic position. Hence, based on the findings presented in (Kalita et al., 2016), this chapter concludes that phonetic properties contribute significantly in identifying an intervocalic glottal stop in Assam Sora.



## Chapter 6

# Synchronic comparison between Assam Sora and Orissa Sora

### 6.1 Introduction

This chapter describes the phonetic relation between Assam Sora and Orissa Sora by making a synchronic comparison between their vowel systems, phonetic realizations of word stress and voicing contrast in stop consonants. Chapter 1 has established the fact that Assam Sora is a transplanted variety of Orissa Sora which emerged due to the migration of Sora people from Orissa to Assam as indentured tea labourers in the 19<sup>th</sup> century. The aim of this chapter is to examine how the two Sora varieties spoken in two distant geographical locations relate to each other at this point of time.

Comparison between the the two Sora varieties is purposely based on the three phonetic features, since two of the features are frequently occurring speech segments in all natural language and one of the features is very relevant to the Austroasiatic language phylum. For instance, vowel sounds are basic to every language and stop consonants also commonly occur in all consonant inventories. On the other hand, although word stress is not common crosslinguistically, it is of particular interest in this work because the unique stress pattern ‘sesquisyllabic’, is an a typical segmental feature of Austroasiatic languages. Its presence versus absence in various Austroasiatic languages divides the language phylum into South and Southeast Asian Austroasi-

atic languages. Moreover, the three features have convincingly described the relation between language varieties of well documented languages such as English, German, Dutch, and Korean. Therefore, it is also believed that the three phonetic features can adequately describe the phonetic relation between the two Sora varieties discussed in this chapter. Hence, from the comparison it will be known if Assam Sora has shifted away from the synchronic of Orissa Sora or not.

Additionally, a synchronic comparison between the two Sora varieties will also help in understanding the impact of migration on Assam Sora. By reviewing the relevant literature it is known that human migration is an important extra-linguistic factor of language change (Kerswill, 2006). Studies have revealed that changes in languages which migrate are caused by factors such as language contact and dialect contact. In this regard, it is known that Sora groups in Assam have been living with speakers of various language groups for several years, and Sora itself is speculated to have different dialects. Therefore, it is very likely that migration has had a significant effect on Assam Sora. Thus, the phonetic relation between Assam Sora and Orissa Sora will confirm if migration of Sora people has actually affected their language or not.

Comparison between the Assam Sora and Orissa Sora is made with the help of acoustic analysis based on synchronic speech data recorded from both Sora varieties. Assam Sora data in this chapter represents the speech variety of Assam Sora speakers (five males and five females from each village) living in four villages of Assam namely Singrijhan, Sessa, Koilamari and Lamabari. Orissa Sora data represents the speech variety of ten Orissa Sora speakers (five males and five females) living in Raiguda village of Southern Orissa (See Figure 1.1 in Chapter 1). Thus, in the following section (§6.2) phonetic properties of the vowel system of Assam Sora is compared with the vowel system of Orissa Sora. Then §6.3 compares Assam Sora and Orissa Sora vowel characteristics in individual syllables of disyllabic words. Subsequently,

§6.4 compares the phonetic realization of word stress in the two Sora varieties, and §6.5 presents the comparison of voicing contrast in stop consonants in Assam Sora and Orissa Sora. Finally, §6.6 concludes the chapter through summarizing the findings and by giving concluding remarks.

## 6.2 Vowels in Assam Sora and Orissa Sora

Analysis of Assam Sora vowel data by Horo & Sarmah (2015) reveals, Assam Sora has six vowel phonemes including /i, e, o, a, u, ə/. Also, the analysis in Chapter 4 has shown that Assam Sora speakers, living in four different, geographical locations have the same vowel inventory. On the other hand, the literature review in Chapter 2 reveals, Orissa Sora vowels are described differently by different scholars. In this regard, while consensus regarding a nine vowel system of Orissa Sora proposed by Stampe (1965) and Donegan (1981) is found in the works of A. R. Zide (1982) and Donegan (1993), the works by Ramamurti (1931) and Ramamurti (1938) are often criticized for inadequate data evidences. On the other hand, the description by P. Mohanty (1997) and Anderson & Harrison (2008b) are not discussed by other scholars yet. Thus, considering the general consensus of among scholars, this work also assumes that Orissa Sora has nine vowels. Therefore, prior to comparing Assam Sora data with synchronic Orissa Sora data, data evidences of two existing studies on Orissa Sora are evaluated in this research. For this purpose, the data list presented by Anderson & Harrison (2011) and Donegan & Stampe (2002) are elicited from Assam Sora and Orissa Sora participants recorded in this dissertation.

In this regard, by eliciting the data evidences of two existing studies on Orissa Sora by Anderson & Harrison (2011) and Donegan & Stampe (2002), it is found that some of the vowels in Orissa Sora are produced differently in Assam Sora. For instance, while Anderson & Harrison (2011) suggests, [i] is a distinct vowel sound in Orissa

Sora, in Assam Sora it is produced as either [i], [e] or [ə]. Similarly, while Donegan & Stampe (2002) proposes, /i/ is a distinct phoneme in Orissa Sora, in Assam Sora it is produced as a mid central vowel [ə]. In addition to these vowel alterations, Table 6.1 and 6.2 show how more Orissa Sora vowels are produced differently in Assam Sora.

Table 6.1: Orissa Sora /i/ and /u/ (Anderson & Harrison, 2011) compared with Assam Sora

Vowel	Orissa Sora	Assam Sora	English
i ~ i	iɟa adil	iʔɟa adil	never dirty
i ~ ə	aɲimnəm ɟumbɪrna	aɲəmnəm ɟumbərna	name steal
i ~ e	anindun pilisida	anenduɲ pilesida	roam uproot
u ~ u	kɔduple	kuduple	all
u ~ a	bɔŋsa	baŋsa	fine

Table 6.2: Orissa Sora vowel minimal set (Donegan & Stampe, 2002) compared with Assam Sora

Vowel	Orissa Sora	Assam Sora	English
i	iɟ-	id-	scratch
ə	ə d-	əd-	prop
u	-luɟ-	luʔ	ear
a	a d-	ad-	drive
i ~ ə	i d-	əʔ-	fan
e ~ i	-eɟ-	iʔ-	thorn
ɛ ~ e	ɛ d-	ed-	roll
ɔ ~ o	ɔ d-	od-	knead
o ~ u	-loɟ-	lud	cord

Thus, evidences seen in Table 6.1 and 6.2 suggests, Orissa Sora vowels such as [i], [u], [ɛ], [ɔ], [e] and [o] are produced as different vowels in Assam Sora. Strikingly, in some cases the same vowel is produced as three different vowels or two different vowels in Assam Sora. From these evidences, it may be postulated that Assam Sora vowel inventory has been restructured and reduced in size after the migration of Sora speakers from Orissa to Assam. It appears from Table 6.2 that front and back vowels are either raised or lowered and the central vowel is lowered. This appears to be

a systematic change because the changing vowels are simply assimilating with the neighbouring vowel sounds. However, the changes seen in Table 6.2 do not appear as systematic because the same vowels are seen assimilating into two or three different vowels, which is not likely to occur in any language.

Moreover, by eliciting the data in 6.1 and 6.2 from Orissa Sora participants recorded in this dissertation reveals that these vowel variations are also observed in synchronic Orissa Sora. Consequently, synchronic speech data of Orissa Sora reveals that Orissa Sora speakers living in Raiguda village of Southern Orissa also produce the six vowels /a, e, i, o, u, ə/ similar to Assam Sora speakers. Therefore, it is apparent that synchronically Assam Sora and Orissa Sora have similar vowel systems with one non-peripheral vowel /ə/ and five peripheral vowels /i, e, a, o, u/. This, is significant considering the existing literature on Orissa Sora vowel inventory since although, most studies agree that Orissa Sora has nine vowels, data from the present field work suggests that Orissa Sora has six vowels similar to Assam Sora. Therefore, in order to further examine the vowel correspondence between Assam Sora and Orissa, their acoustic characteristics are compared in this section. Hence, the following subsection discusses the acoustic correlates of the vowel sounds in Assam Sora and Orissa Sora.

### **6.2.1 Vowel formants in Assam Sora and Orissa Sora**

Comparison of acoustic properties of the vowels in Assam Sora and Orissa Sora in this section is based on synchronic speech data of the two Sora varieties. Assam Sora speech data represents four geographical locations in Assam including Singrijhan, Sessa, Koilamari, and Lamabari, and Orissa Sora speech data represents one geographical location in Orissa namely, Raiguda. The data set consists of vowel minimal pairs taken from Horo & Sarmah (2015) and word lists generated from Ramamurti (1938) and Anderson & Harrison (2011). Participants recorded while saying these data sets include both male and female (five each) speakers between 20 to 60 years of

age in all the regions of Assam Sora and Orissa Sora. Also, for the present analysis vowels samples include vowel sounds of both first and second syllable considered together. Whereas, in the next section (§6.3) vowel samples are separated as first and second syllables of disyllables.

Acoustic comparison of synchronic vowel data of both Sora varieties is based on formant frequency measurements, because significant information about Assam Sora vowel system is achieved from the examination of vowel formants. Analysis of vowel formants particularly includes F1 and F2 frequencies. The analysis of Assam Sora F1 and F2 frequencies have adequately shown that the six vowels /a, e, i, o, u ə/ are categorically distinct from each other, in Assam Sora vowel system. Therefore, in order to compare Assam Sora and Orissa Sora vowel formants, formant frequency measurements of the six vowels /i, e, a, o, u, ə/ produced by Orissa Sora speakers are compared with the formant frequency measurements of the same six vowels produced by Assam Sora speakers. Table 6.3 presents the region wise average formant frequency of F1, F2 and F3 for all six vowels with standard deviations in parenthesis.

Table 6.3 reveals the average F1, F2 and F3 frequencies of the six vowels /i, e, a, o, u, ə/ in Orissa Sora are fairly similar to the average F1, F2 and F3 frequencies of the same six vowels in Assam Sora. Also, the average formant frequencies suggest, similar to Assam Sora, the six vowels are also categorically distinct in Orissa Sora. For instance, the F1 frequency reveals, vowel pairs such as /i/ and /e/, /a/ and /ə/ and /o/ and /u/ have distinct vowel heights. Likewise, the F2 frequency distinguishes vowels /i/ and /e/, /a/ and /ə/ and /o/ and /u/ as front, central and back vowels in Orissa Sora. Thus, the vowel formants of the six vowels in Assam Sora and Orissa Sora provide evidence that the six vowels are fairly similar in both the Sora varieties.

Additionally, Table 6.3 reveals, some of the typical characteristics of Assam Sora vowel system are also found in the vowels of Orissa Sora. For instance, in Chapter 4 it is shown that in Assam Sora vowel system, the high front vowel /i/ and the mid

Table 6.3: Average vowel formants of Assam Sora and Orissa Sora

		F1	F2	F3
ə	Singrijhan	333.71 (46.13)	786.13 (70.42)	980.36 (42.38)
	Sessa	328.49 (43.17)	764.34 (52.71)	946.4 (51.42)
	Koilamari	304.25 (54.84)	797.29 (52.19)	989.81 (35.85)
	Lamabari	321.36 (45.05)	772.31 (40.29)	970.47 (54.45)
	Raiguda	353.14 (40.82)	785.51 (69.88)	982.7 (60.01)
i	Singrijhan	273.44 (36.7)	923.18 (54.89)	1040.24 (35.58)
	Sessa	276 (34.44)	885.28 (78.59)	1010.88 (54.06)
	Koilamari	272.61 (33.82)	927.25 (48.56)	1041.37 (40.37)
	Lamabari	281.39 (29.86)	890.34 (75.19)	1016.23 (56.98)
	Raiguda	311.12 (44.79)	905.08 (64.12)	1010.72 (47.78)
e	Singrijhan	358.38 (47.82)	868.35 (47.14)	994.49 (38.06)
	Sessa	351.01 (32.28)	824.06 (49.62)	974.53 (44.84)
	Koilamari	367.28 (55.54)	858.54 (41.01)	994.48 (35.71)
	Lamabari	378.36 (31.61)	850.11 (63.17)	982.19 (56.31)
	Raiguda	389.2 (41.84)	826.97 (60.34)	969.85 (50.59)
a	Singrijhan	468.59 (70.09)	743.42 (61.72)	970.48 (41.91)
	Sessa	442.79 (58.55)	718.17 (46.92)	979.92 (51.06)
	Koilamari	462.26 (73.37)	752.01 (53.71)	981.77 (53.8)
	Lamabari	459.61 (59.63)	753.64 (54.77)	969.62 (66.84)
	Raiguda	477.67 (57.49)	755.3 (54.71)	960.94 (55.71)
o	Singrijhan	403.34 (49.25)	614.34 (67.14)	988.68 (41.2)
	Sessa	377.03 (31.29)	644.93 (67.30)	947.43 (50.26)
	Koilamari	408.16 (42.24)	601.3 (63.46)	980.38 (46.72)
	Lamabari	402.08 (38.91)	649.08 (60.42)	972.6 (50.75)
	Raiguda	426.33 (38.05)	635.65 (64.2)	973.02 (53.28)
u	Singrijhan	327.25 (45.15)	596.61 (94.98)	987.91 (41.78)
	Sessa	295.49 (30.63)	624.22 (69.96)	956.72 (47.3)
	Koilamari	301.11 (35.6)	587.06 (91.7)	977.19 (34.64)
	Lamabari	299.88 (28.34)	634.24 (69.19)	966.34 (50.22)
	Raiguda	331.04 (38.7)	605.08 (99.92)	984.76 (50.24)

central vowel /ə/ are fronter than their corresponding vowels namely the the mid front vowel /e/ and the low central vowel /a/. These vowel characteristics are also found in Orissa Sora. The average F2 frequency of /i/ is recorded as 905.08 Mel and average F2 frequency of /e/ is recorded as 826.97 Mel. This shows a difference of

78 Mel in F2 frequencies of the two front vowels, indicating similar to Assam Sora, /i/ is also fronter than /e/ in Orissa Sora. Since a higher F2 frequency correlates to fronter vowels in a vowel system. Similarly, average F2 frequency of /ə/ in Orissa Sora is recorded as 785.51 Mel and average F2 frequency of /a/ is recorded as 755.3 Mel. This shows a difference of 30 Mel in F2 frequencies of the two central vowels and indicates that similar to Assam Sora, /ə/ is also fronter than /a/ in Orissa Sora.

However, unlike the Assam Sora vowel system where back vowels /o/ and /u/ are fairly aligned to each other with less than 20 Mel difference in their F2 frequencies, in Orissa Sora the F2 frequency difference between /o/ and /u/ is 30 Mel. This indicates that similar to Assam Sora, /o/ is fronter than /u/ in Orissa Sora. Significantly, these kind of uniformities indicate that besides having fairly similar vowel formants, individual vowels also interact in similar ways in both Assam Sora and Orissa Sora.

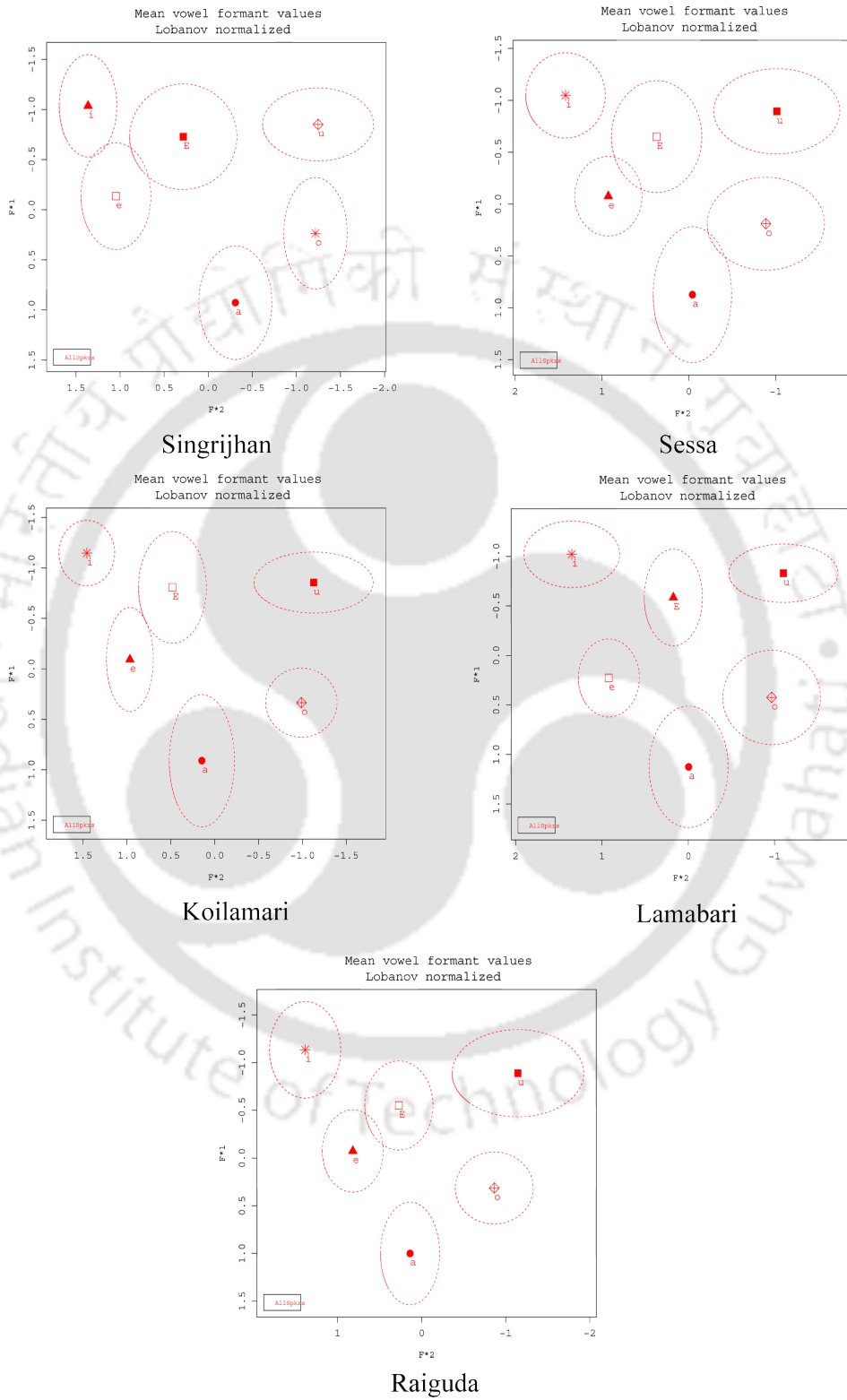
Another similarity between vowel characteristics of Assam Sora and Orissa Sora can be seen in the vowel height difference between the high vowels /i/ and /u/ and between the mid vowels /e/, /ə/ and /o/. Chapter 4 in this work has shown that in the Assam Sora vowel system the high front vowel /i/ is usually higher than the high back vowel /u/. Also, the mid central vowel /ə/ is usually higher than the mid front vowel /e/, and the mid back vowel /o/. The difference in height between /ə/ and /o/ is greater than the difference in height between /ə/ and /e/.

Table 6.3 shows that these vowel characteristics are also found in Orissa Sora. The average F1 frequency of /i/ is 311.12 Mel, and the average F1 frequency of /u/ is 331.04 Mel, indicating that /i/ is higher than /u/ in Orissa Sora. Since, lower F1 frequency correlates to higher vowel in a vowel system. Also, vowel height distinction between the three mid vowels can be seen in the vowels of Orissa Sora. While average F1 frequency of /ə/ is 353.14 Mel, average F1 frequency of /e/ is 389.2 Mel, and average F1 frequency of /o/ is 426.33 Mel. This indicates that the mid central vowel /ə/ is higher than the mid front vowel /e/ and the back front vowel /o/. Also, the

difference in F1 frequencies of the three mid vowels indicate that the difference is greater between mid central vowel /ə/ and mid back vowel /o/.

Thus, the formant frequency measurements primarily reveal that the six vowels /i, e, a, o, u, ə/ are categorically distinct in Orissa Sora as they are in Assam Sora vowel system. Secondly, it is evident that besides having fairly similar formant frequencies, the six vowels in Orissa Sora also exhibit the typical vowel characteristics of Assam Sora vowel system. This indicates that the vowel systems of Assam Sora and synchronic Orissa Sora are identical to each other. In order to substantiate this fact, visual evidences of Assam Sora and Orissa Sora vowel formants are plotted on a F1 and F2 axis with speaker normalized formant frequencies attained through the vowel normalization method proposed by Lobonov (1971). Figure 6.1 presents speaker normalized vowel plots for Assam Sora and Orissa Sora.

Figure 6.1: Assam Sora and Orissa Sora vowel plots



Vowels: i=/i/, e=/e/, a=/a/, o=/o/, u=/u/, E=/ə/

In Figure 6.1 the four vowel plots including Singrijhan, Sessa, Koilamari and Lamabari represent Assam Sora vowel plots and one vowel plot, namely Raiguda represents the vowel plot of Orissa Sora. From Figure 6.1 it is evident, the vowel plots of Assam Sora are similar to the vowel plots of Orissa Sora. This provides evidence showing that besides individual vowel sounds, the shape and structure of Assam Sora and Orissa Sora vowel systems are also similar to each other. Thus, from the analysis of vowel formants in Assam Sora and Orissa Sora it is evident that synchronically both Assam Sora and Orissa Sora have the same vowel system including five peripheral vowels /i, e, a, o, u/ and one non-peripheral vowel /ə/. Also, it is evident that vowel characteristics such vowel frontness and vowel height differences between adjacent vowels in the two Sora varieties are similar to each other.

Additionally, the analysis of Assam Sora vowel formants in Chapter 4 reveal that syllable structure has an important effect on Assam Sora vowel system. Evidences suggest, Assam Sora vowel formants usually change as a function of syllable position in disyllabic words. Thus, considering the fact that vowels formants are fairly similar in Assam Sora and Orissa Sora, it is relevant to investigate the influence of syllable position in Orissa Sora vowel system. As a result, the findings will provide additional evidences regarding the synchronic relation between Assam Sora and Orissa Sora vowel systems. Therefore, in order to examine the relation of Assam Sora and Orissa Sora vowel systems, particularly with respect to syllable position, the following section examines Assam Sora and Orissa Sora vowel formants in individual syllables of disyllables.

### 6.3 Assam Sora and Orissa Sora vowel formants in individual syllables of disyllables

In the previous section, comparison of the Assam Sora and Orissa Sora vowel systems was based on formant frequencies drawn from vowel samples of both first and second syllable of disyllables. However, analysis of Assam Sora vowel formants by (Horo & Sarmah, 2015) reveals the dynamic nature of vowels in Assam Sora. It is found that Assam Sora vowel formants change as a function of syllable position whereby, both F1 and F2 frequencies of Assam Sora vowels are different in first and second syllable of disyllabic words. This indicates that Assam Sora vowels have different vowel heights and vowel frontness/backness depending on their syllable position in disyllabic words. Moreover, perceptual distance between adjacent vowels in Assam Sora vowel system are observed to change, as a function of syllable position. Consequently, the vowel area space also changes across first and second syllable of Assam Sora disyllables, whereby vowel area space is usually larger in second syllable than in first syllable.

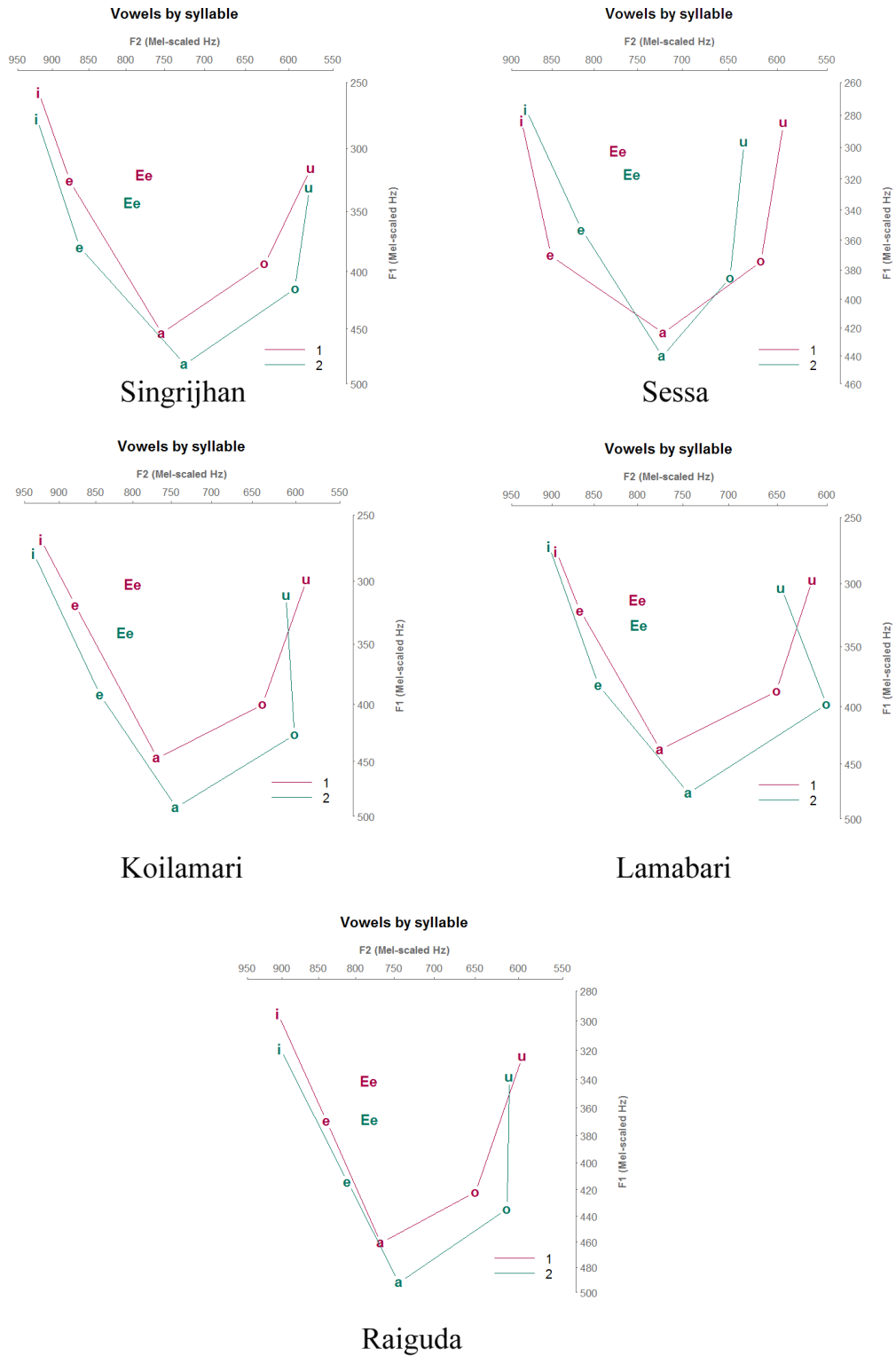
Thus, considering that Assam Sora and Orissa Sora have similar vowel systems, it is likely that this feature also exists in Orissa Sora. In order to examine the effect of syllable position on Orissa Sora vowel system the vowel formants are extracted and examined separately in first and second syllable of Orissa Sora disyllables. This mainly includes an examination of the vowel formants in individual syllables of disyllables, an examination of perceptual distance between adjacent vowels in first and second syllable of disyllables, and finally an examination of the vowel area space in first and second syllable of disyllables. Subsequently, the results are compared with the existing results of Assam Sora in order to examine the synchronic relation between the two Sora varieties. These examinations are done by taking the same set of data which is described in the previous section. However, while in the previous section the vowel samples of first and second syllables are combined together, in this section they

are examined separately. Also, the disyllables in both Assam Sora and Orissa Sora represent all the disyllable combinations in Assam Sora that are described in Chapter 3 of this work.

### **6.3.1 F1 and F2 of similar vowels in individual syllables of Assam Sora and Orissa Sora disyllables**

In order to examine if vowel formants dynamically change in first and second syllable of Orissa Sora disyllables, F1 and F2 frequencies of the six Orissa Sora vowels in first and second syllable are plotted next to each other on the same vowel plot. Figure 6.2 presents Orissa Sora vowel plots showing the difference in vowel formants in first and second syllable of disyllabic words. Also, Assam Sora vowel plots showing difference in first and second syllable of disyllabic words are presented alongside Orissa Sora vowel plots for the purpose of comparison between the two Sora varieties.

Figure 6.2: Assam Sora and Orissa Sora vowel plots in individual syllables of disyllables



Vowels: i=/i/, e=/e/, a=/a/, o=/o/, u=/u/, Ee=/ə/

Significantly, Figure 6.2 shows the F1 and F2 frequencies of all six vowels in Orissa Sora change, in first and second syllable of disyllabic words. This indicates, similar to Assam Sora, Orissa Sora vowels have different vowel heights and vowel frontness/backness, depending on the position where the vowels occur in disyllabic words. As a result, it is evident from Figure 6.2 that similar vowels in first and second syllable are distant from each other on the vowel plots. For instance it is observed, while the mid front vowel /e/ is higher in first syllable, it is lower in second syllable. Likewise, the mid back vowel /o/ is fronter in first syllable, but is retracted in second syllable. Thus, it is found similar to Assam Sora, Orissa Sora vowel formants change as a function of syllable position. Also, the vowel plots indicate, there can be perceptual distance between similar vowels in first and second syllable of Orissa Sora disyllables, as it was observed in Assam Sora. In order to show perceptual distance exists between similar vowels in first and second syllable of Orissa Sora disyllables, euclidean distance between similar vowels in first and second syllable is estimated for Orissa Sora vowel system by using equation 4.2 described in Chapter 4.

Table 6.4 presents the euclidean distance between similar vowels in first and second syllable of disyllabic words for the Orissa Sora vowel system. Also, euclidean distance estimation for the Assam Sora vowel system is also given alongside for the purpose of comparison.

Table 6.4: Euclidean distance between similar vowels across syllables

	Singrijhan	Sessa	Koilamari	Lamabari	Raiguda
i-i	18.07	13.91	15.14	4.95	24.05
e-e	56.63	42.54	80.95	65	52.58
a-a	42.48	15.36	50.32	48.53	38.56
o-o	43.33	35.18	46.84	53.87	40.40
u-u	33.09	46.62	28.54	32.03	21.88
ə-ə	28.74	21.75	42.64	20.25	27.18

From Table 6.4, it is evident the vowel formants of all six vowels in Orissa Sora change in first and second syllable of disyllables. The euclidean distance estimations

provide evidence that perceptual distance exists between similar vowels in first and second syllable of disyllables. Consequently, Orissa Sora vowel properties such as, vowel height and vowel frontness/backness change to a certain degree, depending on its syllable position in disyllables. This is evident in Figure 6.2.

Hence, it is revealed syllable position in disyllables plays an important role in Orissa Sora vowel system. The analysis here confirms that similar to Assam Sora vowel system, Orissa Sora vowel formants also change as a function of syllable position. Additionally, Assam Sora vowel system shows perceptual distance between adjacent vowels within individual syllables also change as a function of syllable position. This implies that perceptual distance between adjacent vowels such as /i/ and /e/, /a/ and /ə/ etc. is different in first and second syllables of disyllables. Consequently, adjacent vowels are often nearer or further from each other as a function of syllable position, so their perceptual contrasts are also effected by the syllable position. Therefore, since Orissa Sora vowel formants change as a function of syllable position, it is relevant to examine this particular characteristic in Orissa Sora vowel system. Hence, the following section examines the perceptual distance between adjacent vowels in the Orissa Sora vowel system and compares the findings with Assam Sora vowel system.

### **6.3.2 F1 and F2 of adjacent vowels in individual syllables of Assam Sora and Orissa Sora disyllables**

Regarding changes in perceptual distance between adjacent vowels in first and second syllable of Assam Sora disyllables it has been shown in Chapter 4 that usually front and central vowels are raised or lowered and back vowels are fronted or retracted depending on their occurrence in first or second syllable of disyllables. Consequently, evidences suggest perceptual distance between front vowels /e/ and /i/ and between central vowels /a/ and /ə/ change by the influence of syllable position. Similarly, evi-

dences suggest perceptual distance between central and back vowels /ə/ and /o/ and /ə/ and /u/ also changes as a function of syllable position. As a result, it is observed from the descriptions in Chapter 4 that some vowels have better perceptual contrast in either of the two syllables, and usually the adjacent vowels have better perceptual contrast in the second syllable. Therefore, in order to examine this phenomena in the Orissa Sora vowel system, perceptual distance between the above mentioned adjacent vowel pairs in individual syllables of Orissa Sora disyllables are examined for the Orissa Sora vowel system with the help of euclidean distance estimation. Euclidean distance estimation between front vowels /e/ and /i/, between central vowels /a/ and /ə/ and between back and central vowels /o/ and /ə/ and between /u/ and /ə/ are presented in Table 6.5, 6.6 and 6.7. Also, euclidean distance estimation for the same in Assam Sora vowel system is presented alongside for the purpose of comparison.

Table 6.5: Euclidean distance between front vowels in first and second syllable

	Singrijhan	Sessa	Koilamari	Lamabari	Raiguda
e-i ( $\sigma$ -1)	81.48	102.54	70.74	54.55	101.22
e-i ( $\sigma$ -2)	119.48	104.49	147.07	125.93	133.51

Table 6.6: Euclidean distance between central vowels in first and second syllable

	Singrijhan	Sessa	Koilamari	Lamabari	Raiguda
a-ə ( $\sigma$ -1)	137.46	103.83	149.63	129.11	121.73
a-ə ( $\sigma$ -2)	155.12	114.42	164.92	154.73	130.1

Table 6.7: Euclidean distance between back and central vowels in first and second syllable

	Singrijhan	Sessa	Koilamari	Lamabari	Raiguda
o-ə ( $\sigma$ -1)	165.59	146.52	187.98	166.72	154.54
o-ə ( $\sigma$ -2)	212.21	112.62	225.76	209.43	182.14
u-ə ( $\sigma$ -1)	200.42	178.97	210.71	185.3	186.14
u-ə ( $\sigma$ -2)	215.58	121.63	197.97	152.74	171.23

From Table 6.5, 6.6 and 6.7 it is clearly evident that perceptual distance between adjacent vowels in the Orissa Sora vowel system changes due to the influence

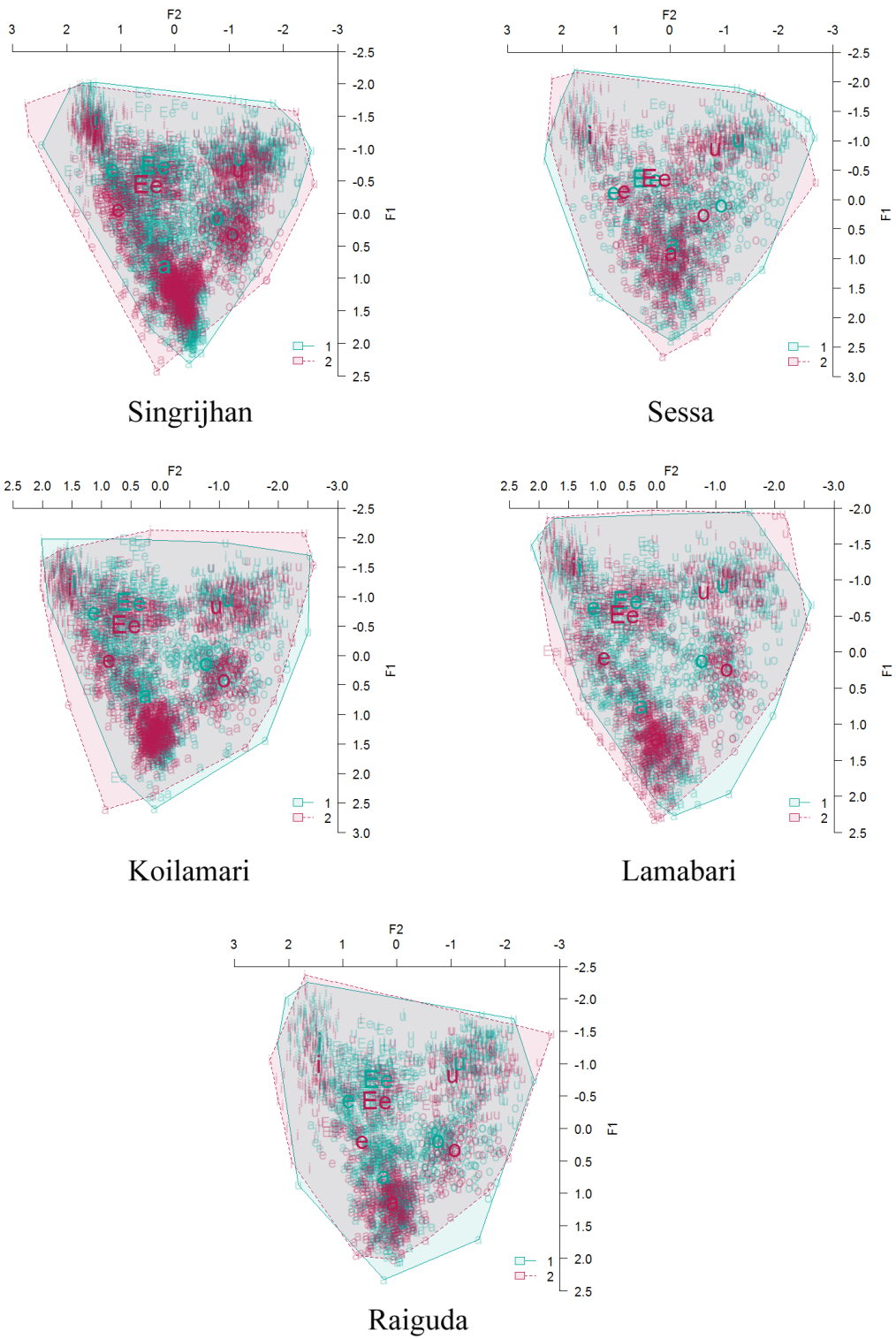
of syllable position in disyllables. Perceptual distance between front vowels /i/ and /e/ is greater in second syllable than in first syllable. Also, perceptual distance between central vowels /a/ and /ə/ is found to be greater in second syllable than in first syllable. Likewise, it is observed that while the perceptual distance between mid back vowel /o/ and mid central vowel /ə/ is greater in second syllable, perceptual distance between high back vowel /u/ and mid central vowel /ə/ is greater in first syllable than in second syllable. These characteristics imply that in addition to changing the vowel formants of individual vowels, syllable position also influences the internal relation between adjacent vowels in Orissa Sora vowel system. Specifically, the perceptual distances are effected, whereby perceptual distance between adjacent vowels are maximized in second syllable rather than in first syllable. The analysis confirms that perceptual distance between adjacent vowels in Orissa Sora vowel system is effected by the influence of syllable position in disyllables. It is evident, these perceptual distance variations in Orissa Sora are similar to the perceptual distance variations in Assam Sora. However, the distance between /o/ and /ə/ in Sessa data, and the distance between /u/ and /ə/ in Singrijhan data differs from the Orissa Sora data. It is not clear why this discrepancy exists, but it is clear the Orissa Sora data patterns, are very similar to Assam Sora data patterns of the other two regions. Hence, by examining the effect of syllable position on perceptual distance between adjacent vowels, it is confirmed that the internal relation between adjacent vowels across syllables is similar in Assam Sora and Orissa Sora vowel systems.

The next section examines the influence of syllable position on the overall shape and structure of Orissa Sora vowel system since it is known that syllable position in disyllables adequately effects the overall shape and structure of Assam Sora vowel system.

### 6.3.3 Assam Sora and Orissa Sora vowel space in individual syllables of disyllables

It is known from Chapter 4 in this work that the Assam Sora vowel system has different vowel area spaces in first and second syllable of disyllables, due to the dynamic nature of vowel formants in the two syllables. Also, evidences suggest, the vowel area space in second syllable is usually larger than the vowel area space in first syllable. Thus, considering that syllable position also effects Orissa Sora vowel formants and perceptual distance between adjacent vowels in first and second syllable are different, it is likely that vowel space of the Orissa Sora vowel system also differs in first and second syllable of disyllables. Therefore, in order to visually examine the vowel space of Orissa Sora vowel system in individual syllables of disyllables, all vowel samples in first and second syllable are plotted in a F1 and F2 axis within closed circumferences. Figure 6.3 presents the vowel space plots in first and second syllable of Orissa Sora disyllables with speaker normalized formant frequencies attained by using the vowel normalization method proposed by (Lobonov, 1971). Also, similar plots for Assam Sora vowel system are presented alongside for the purpose of comparison.

Figure 6.3: Assam Sora and Orissa Sora vowel space in disyllables



Vowels: i=/i/, e=/e/, a=/a/, o=/o/, u=/u/, Ee=/ə/

From Figure 6.3 it is observed, vowel space of the Orissa Sora vowel system in first and second syllable of disyllabic words are fairly similar to each other. This indicates that, all vowel samples are produced within fairly similar vowel spaces in first and second syllable of Orissa Sora disyllables. However, vowel area estimation provides evidence that suggests Assam Sora has different vowel area spaces in first and second syllable of disyllables. Also, it appears vowel space is usually larger in second syllable than in first syllable. Therefore, in order to compare the actual vowel spaces of Orissa Sora vowel system in first and second syllable of disyllables, vowel area has been estimated for Orissa Sora vowel system by using equation 4.3 described in Chapter 4. Table 6.8 presents the vowel area estimation of Orissa Sora vowel system in first and second syllable of disyllables. Also, vowel area estimation of Assam Sora vowel system in first and second syllable is presented alongside in order to make a comparison between Assam Sora and Orissa Sora.

Table 6.8: Vowel area estimation in first and second syllable of Assam Sora and Orissa Sora

	Singrijhan	Sessa	Koilamari	Lamabari	Raiguda
$\sigma$ -1	33765.48	30979.57	32858.72	26708.66	27902.53
$\sigma$ -2	39110.32	22235.24	41111.6	35219.97	29464.21

From Table 6.8 it is evident Orissa Sora vowel system has different vowel area spaces in first and second syllable of disyllables. Also, it is evident that vowel area space is larger in second syllable than in first syllable. This provides evidence in addition to changing vowel formants and effecting the perceptual distance between adjacent vowels, that syllable position also effects the over all vowel space of Orissa Sora vowel system. Thus, the analysis here confirms that syllable position plays an important role in Orissa Sora vowel system. Additionally, it is observed that vowel space difference in Orissa Sora vowel system across syllables is essentially similar to Assam Sora vowel system. As mentioned earlier, Assam Sora vowel system also has different vowel spaces in first and second syllable of disyllables and usually the vowel

space is larger in second syllable except in case of Assam Sora data representing Sessa speakers. In this regard, it has been reported in Chapter 4 that the exception in Sessa data may be due to the influence of a second language namely Sadri on Assam Sora speakers living in Sessa. Thus, by excluding the exception found in Sessa data, it is observed the pattern is largely similar with rest of the Assam Sora data. Also, evidences suggest, the exception in Sessa data is not void of the fact that Assam Sora vowel spaces are different in first and second syllable of disyllables. Hence, the analysis of formant frequencies in individual syllables of disyllables provides adequate evidence to prove syllable position is an important factor influencing the vowel system of Assam Sora as well as Orissa Sora. Also, it is confirms the following are similiar in Assam Sora and synchronic Orissa Sora: a) variation in formant frequencies across syllables, b) different perceptual distances between adjacent vowels across syllables and c) variation in vowel area space across syllables.

Additionally, the study by (Horo & Sarmah, 2015) and Chapter 3 of this work have examined the phonetic realization of word stress in Assam Sora and the descriptions propose that Assam Sora has an iambic stress pattern. Thus, evidences suggest that second syllable is phonetically more prominent than the first syllable in Assam Sora disyllables. Therefore, seeing that syllable position has similar influences on vowel systems of Assam Sora and Orissa Sora, it is the interest of this work to examine the possibilities of word stress in synchronic Orissa Sora and to compare the findings with Assam Sora. Also, the review of Austroasiatic and Munda languages reveal that one of the typical characteristics of Austroasiatic languages is the sesquisyllabic word prosody, that is reported to be lacking in Munda languages. However, some studies propose that an iambic word stress in some Munda languages may actually resemble the sesquisyllabic word prosody of Austroasiatic languages. Thus, the presence of an iambic stress in Assam Sora substantiates the hypothesis and therefore, it is the interest of this work to compare the phonetic realization of word stress in Orissa

Sora and compare it with the findings of Assam Sora. Hence, the following section investigates word stress in synchronic Orissa Sora and subsequently compares the results with the existing discoveries of syllable prominence in Assam Sora.

## 6.4 Word stress in Assam Sora and Orissa Sora

As mentioned earlier, analysis of word stress in Assam Sora reveals that in disyllabic words the second syllable is more prominent than the first syllable. The study by Horo & Sarmah (2015) proposes Assam Sora has an iambic stress pattern in disyllabic words. While this finding is mainly based on Assam Sora speech data recorded from speakers living in Singrijhan only, the descriptions in Chapter 3 provides evidence that the pattern is similar in Assam Sora speech variety of three additional regions in Assam including Sessa, Koilamari and Lamabari. Thus, from the speech data recorded from all these areas it is observed, stress in second syllable is realized by greater vowel duration, greater fundamental frequency and greater vowel intensity. Hence, the claim that Assam Sora has iambic stress is also substantiated with the help of acoustic analysis.

Regarding Orissa Sora, the word stress pattern is undescribed. However, evidences suggest, Munda languages such as Santali Ghosh (2008) and Plains Remo Anderson (2008a) have stress on the second syllable. Additionally, Austroasiatic languages are associated with the presence of a typical word prosody called the ‘sesquisyllables’. Sesquisyllabic words in Southeast Asian Austroasiatic languages usually have a first minor syllable followed by a second major syllable. Also, while the major syllable is phonologically and phonetically heavier, the minor syllable is generally restricted to short and centralized vowels only. This kind of word prosody is typical to Austroasiatic languages of Southeast Asia, but is believed to be lacking in Munda languages. Studies such as (Donegan & Stampe, 1983) and (Donegan & Stampe,

2004) propose, the typical sesquisyllabic word prosody of Austroasiatic languages is lacking in Munda languages. Also, it is argued that Munda languages lack the sesquisyllabic word prosody mainly due to the influence of South Asian non-Munda languages. However, Anderson (2015) argues, the weak strong word prosody of many Munda languages including Santali and Plains Remo is in fact equivalent to the word prosody of sesquisyllabic words. The only difference lies in the inventory of the minor syllable, as the weak syllable in Munda languages can include the full vowel inventory, the minor syllable in sesquisyllabic words include only short and centralized vowels. This analogy of iambic stress in Munda languages is applicable to Assam Sora and there are phonetic evidences to prove the hypothesis.

However, since the properties of word stress in Orissa Sora are unknown, it is not clear whether Orissa Sora can also be described along similar lines or not. Therefore, in the following subsections word stress in Orissa Sora is examined with the with the help of acoustic analysis. Subsequently, in order to find the synchronic relation between the two Sora varieties, results of Orissa Sora are also compared to the existing results of Assam Sora. Moreover, while the Assam Sora data represents speech data recorded from Assam Sora speakers living in four region of Assam namely Singrijhan, Sessam Koilamari and Lamabari, Orissa Sora speech data represents Orissa Sora speakers living in Raiguda village of Southern Orissa. Also, the comparison is mainly based on the data set presented by (Horo & Sarmah, 2015) since, the data set in this work has primarily revealed the presence of iambic stress in Assam Sora with the help of acoustic analysis. Three acoustic correlates of word stress namely vowel duration, fundamental frequency, and vowel intensity are examined for Orissa Sora in the following sections, since these cues provide adequate evidences regarding the presence of iambic stress in Assam Sora.

### 6.4.1 Vowel duration across syllables in Assam Sora and Orissa Sora

In Chapter 4, vowel duration was found to be one of the most important characteristics of word stress in Assam Sora. Evidences suggest, vowel duration in second syllable is always longer than the vowel duration in first syllable. Vowel duration distinction in first and second syllable is observed in all types of disyllabic words, and in all four regions of Assam Sora. This confirms that longer vowel duration in second syllable of Assam Sora disyllables is an indicator of syllable prominence or word stress in Assam Sora. Therefore, in order to examine word stress in Orissa Sora, total vowel duration in terms of time interval between vowel start and vowel end is calculated for every vowel in first and second syllable of Orissa Sora disyllables. Table 6.9 presents the average vowel duration in first and second syllable of disyllabic words as produced by the ten Orissa Sora speakers recorded in this chapter and standard deviation for the same is given in parenthesis. Also, average vowel duration in first and second syllable of Assam Sora disyllables is presented in the same table for the purpose of comparison.

Table 6.9: Average vowel duration in first and second syllable

	Singrijhan	Sessa	Koilamari	Lamabari	Raiguda
Syllable 1	83.72 (38.14)	77.13 (31.06)	84.6 (41.95)	74.27 (30.83)	95.3 (47.75)
Syllable 2	153.62 (58.47)	105.74 (44.99)	159.42 (70.37)	135.92 (49.34)	147.66 (49.35)

From Table 6.9 it is observed that average vowel duration in Orissa Sora disyllabic words is longer in second syllable than in first syllable. It is observed, while the average vowel duration in second syllable is 147.66 milliseconds, the average vowel duration in first syllable is only 95.3 milliseconds. This indicates that longer vowels are produced in second syllable whereas, the vowels are relatively shorter in first syllable of Orissa Sora disyllables. Thus, it is evident similar to Assam Sora, Orissa Sora also distinguishes

vowel duration in first and second syllable of disyllables. Moreover, similar to Assam Sora, Orissa Sora also has longer vowels in second syllable of disyllables. Hence, in order to illustrate the difference, Figure 6.4 presents a graphical representation of vowel duration distinction between first and second syllable in both Assam Sora and Orissa Sora. Also, the illustration bellow represents speaker normalized vowel durations attained by using the *z-score* equation for both Assam Sora and Orissa Sora disyllables.

Figure 6.4: Speakers normalized vowel duration in first and second syllable

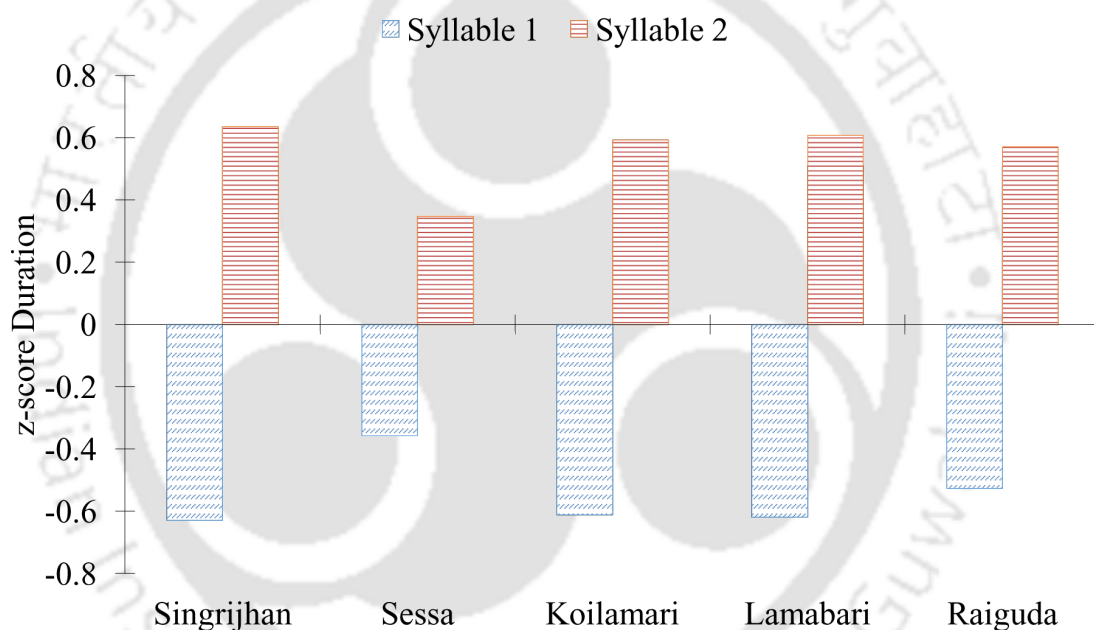


Figure 6.4 illustrates the vowel duration difference in first and second syllable of disyllables in both Assam Sora and Orissa Sora. The illustration confirms that vowel duration is usually longer in second syllable than in first syllable for both Assam Sora and Orissa Sora disyllables. Additionally, while the statistical significance of longer vowel duration in second syllable of Assam Sora disyllables is already reported in Chapter 3, a one-way ANOVA test considering normalized vowel duration as dependent variable and syllable position as factor reveals, vowel duration is significantly longer in second syllable of Orissa Sora disyllables as well ( $F(1,2382) = 1029.040$ ,

$p < 0.001$ ). Thus, the analysis here confirms that disyllabic words in both Assam Sora and Orissa Sora have different vowel durations across first and second syllable, whereby vowels are always longer in the second syllable. This provides evidence that longer vowels in second syllable effectively contributes to the stressing of second syllable in disyllables for both Assam Sora and Orissa Sora. Hence, this work provides evidence that phonetic realization of word stress through longer vowel duration in second syllable is adequately similar in the two synchronic varieties of the Sora language.

#### **6.4.2 Fundamental frequency across syllables in Assam Sora and Orissa Sora**

Fundamental frequency, the acoustic correlate of pitch variation, is an important cue for identifying word stress in Assam Sora disyllables. Evidences discussed in Chapter 4 of this work suggest, fundamental frequency is usually higher in the vowel nuclei of the second syllable in comparison to that of the first syllable in Assam Sora disyllables. Also, fundamental frequency difference in first and second syllable of Assam Sora disyllables is observed in average fundamental frequency as well as in maximum fundamental frequency of the vowel nuclei in the two syllables. Thus, both average fundamental frequency and maximum fundamental frequency of the vowel nuclei in second syllable is higher than that of the vowel nuclei in the first syllable of Assam Sora disyllables. Moreover, while this pattern is regularly seen in Assam Sora speech data, there is an exception with regard to the speech data recorded from Assam Sora speakers living in Lamabari region. As reported earlier, the Lamabari data shows a completely uniform pitch in first and second syllable of disyllables, therefore a conclusive reason for the exception could not be derived in this work. However, since a large portion of Assam Sora speech data significantly shows the importance of fundamental frequency in identifying word stress in Assam Sora

disyllables, fundamental frequency in the vowel nuclei of Orissa Sora disyllables is also examined.

Both average and maximum fundamental frequency is calculated between vowel start and vowel end for every vowel in first and second syllable of Orissa Sora disyllables. The average fundamental frequencies and maximum fundamental frequencies of first and second syllable of Orissa Sora disyllables are presented in Table 6.10 and 6.11, with standard deviation in parenthesis. Also, fundamental frequency measurements for the same in Assam Sora disyllables is given alongside for the purpose of comparison.

Table 6.10: Average  $f_0$  in first and second syllable

	Singrijhan	Sessa	Koilamari	Lamabari	Raiguda
Syllable 1	172.4 (51.09)	142.17 (34.15)	167.79 (59.15)	162.42 (51.63)	178.72 (40.63)
Syllable 2	181.61 (50.42)	153.27 (32.01)	181.27 (53.47)	161.94 (39.21)	197.37 (47.63)

Table 6.11: Maximum  $f_0$  in first and second syllable

	Singrijhan	Sessa	Koilamari	Lamabari	Raiguda
Syllable 1	182.85 (61.18)	152.28 (43.78)	180.48 (67.24)	174.24 (61.41)	187.95 (45.92)
Syllable 2	198.43 (62.81)	164.68 (36.81)	196.4 (63.50)	173.58 (48.47)	211.76 (57.33)

From Table 6.10 and 6.11 it is observed that both average and maximum frequencies of the vowel nuclei in second syllable is higher than that of the first syllable in Orissa Sora disyllables. This indicates that the second syllable in Orissa Sora disyllable has higher pitch variation than the first syllable. In order to reduce the speaker effect on the pitch variation in first and second syllable of Orissa Sora disyllables, both average and maximum fundamental frequencies are normalized using the  $z$ -score equation. Figure 6.5 shows the comparison between speakers normalized average fundamental frequency in first and second syllable and Figure 6.6 shows the comparison between

speakers normalized maximum fundamental frequency in first and second syllable of disyllables. Also, fundamental frequency distinction for first and second syllable of Assam Sora disyllables is presented alongside in order to compare between the two Sora varieties.

Figure 6.5: Speakers normalized average  $f_0$  in first and second syllable

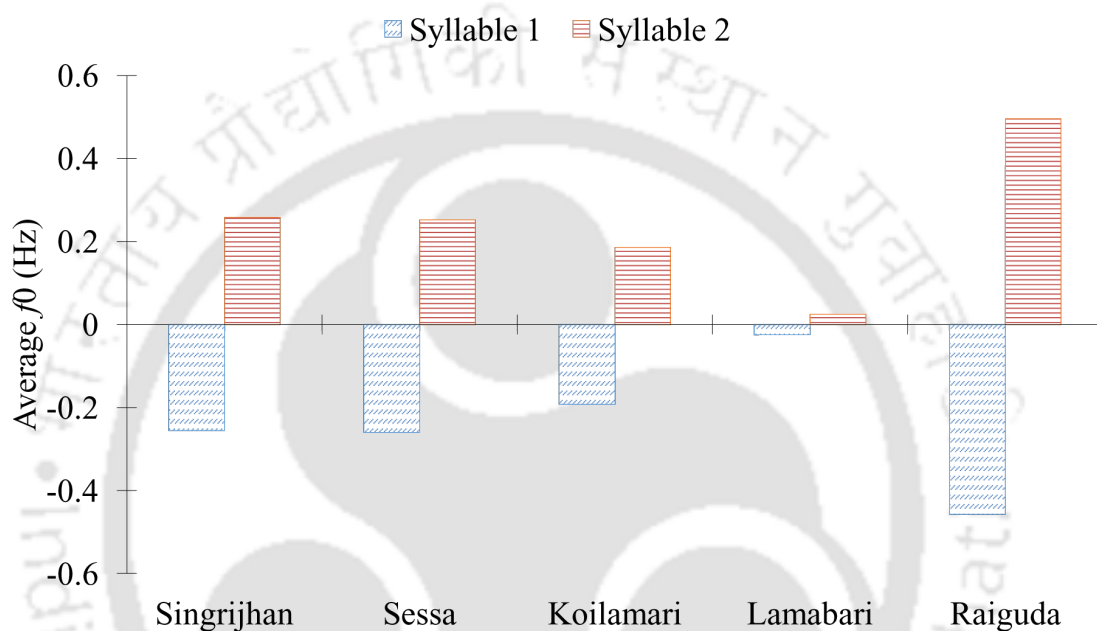
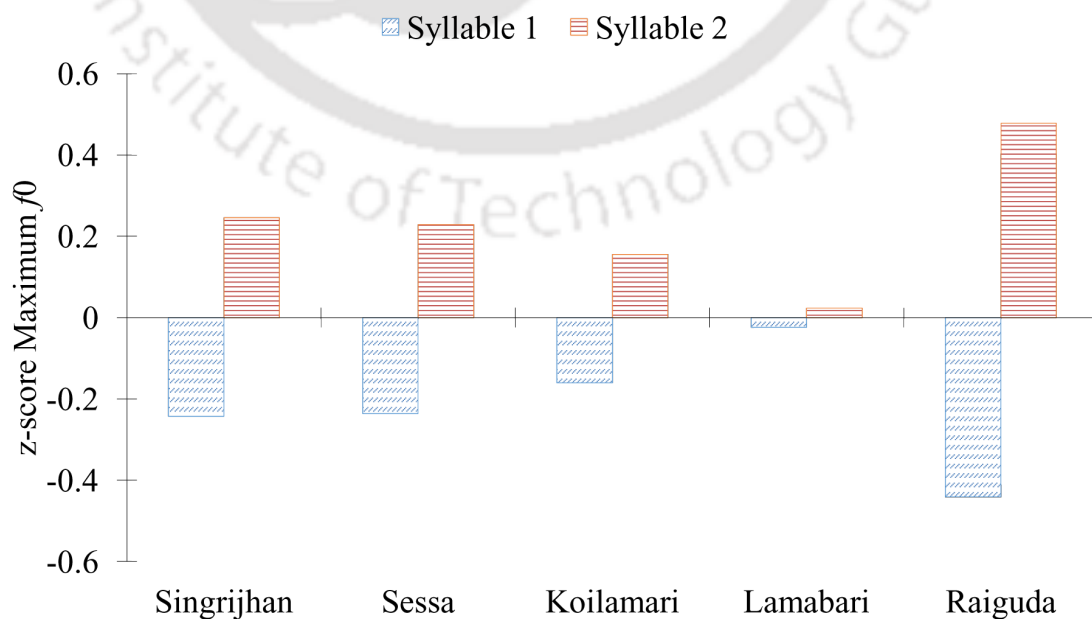


Figure 6.6: Speakers normalized maximum  $f_0$  in first and second syllable



From Figures 6.5 and 6.6 it is evident that the vowel nuclei in first and second syllable of Orissa Sora disyllables have different average and maximum fundamental frequencies. It is also observed, similar to Assam Sora, both average and maximum fundamental frequency in the vowel nuclei of the second syllable is higher than that of the first syllable. This provides evidence that fundamental frequency adequately differentiates the first and second syllable of Orissa Sora disyllables.

Moreover, a one-way ANOVA test considering normalized average fundamental frequency as the dependent variable and the syllable position as factor, also reveals the average fundamental frequency difference between first and second syllable of Orissa Sora disyllables is statistically significant ( $F(1,2382) = 618.880, p < 0.001$ ). Similarly, a one-way ANOVA test considering normalized maximum fundamental frequency as the dependent variable and the syllable position as factor reveals, the maximum fundamental frequency difference between first and second syllable of Orissa Sora disyllables is also statistically significant ( $F(1,2382) = 533.350, p < 0.001$ ). Thus, the analysis here confirms that higher fundamental frequency in the second syllable of Orissa Sora disyllables contributes to the syllable prominence of the second syllable. Consequently, this work provides evidence for the phonetic realization of word stress through higher fundamental frequency in the second syllable of disyllables is largely similar between the synchronic varieties of Assam Sora and Orissa Sora, except for one exception found in Assam Sora speech data recorded from Lamabari region.

### **6.4.3 Vowel intensity across syllables in Assam Sora and Orissa Sora**

Vowel intensity signifies the acoustic energy in vowel sounds, the auditory equivalent of intensity is amplitude or loudness. Vowel intensity difference in first and second syllable is also an evidence that second syllable is more prominent than the first syllable in Assam Sora disyllables. Thus, it is observed that Assam Sora speakers

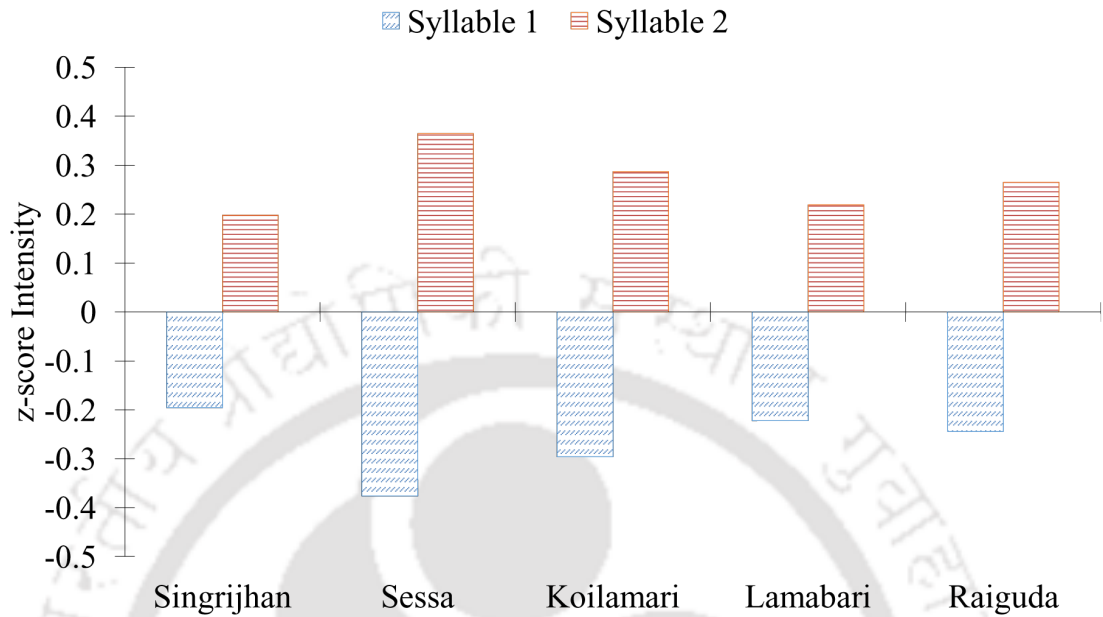
normally produce louder vowels in second syllable than in first syllable. Therefore, in order to examine whether Orissa Sora has similar characteristics or not, vowel intensity in first and second syllable of Orissa Sora disyllables is examined. For this purpose, average vowel intensity is calculated between vowel start and vowel end for every vowel in first and second syllable of Orissa Sora disyllables. Table 6.12 presents the average of average vowel intensities in first and second syllable of Orissa Sora disyllables, with standard deviation in parenthesis. Also, average vowel intensity variation between first and second syllable of Assam Sora disyllables is given in the same table in order to compare between the two Sora varieties.

Table 6.12: Average vowel intensity in first and second syllable

	Singrijhan	Sessa	Koilamari	Lamabari	Raiguda
Syllable 1	72.86 (4.16)	74.1 (5.54)	61.28 (4.46)	58.28 (4.39)	62.16 (4.45)
Syllable 2	74.24 (3.17)	77.23 (5.56)	63.41 (4.63)	59.91 (4.83)	64.28 (4.32)

Significantly, Table 6.12 reveals, first and second syllable of Orissa Sora disyllables have different average vowel intensities. Also, it is evident that similar to Assam Sora, Orissa Sora has greater vowel intensity in the second syllable, whereas the average vowel intensity in first syllable is relatively smaller. This indicates similar to Assam Sora, Orissa Sora vowel nuclei in the second syllable can be perceived as louder than the vowel nuclei of the first syllable. However, considering the fact that there can be individual speaker's effect on the vowel intensity of Orissa Sora disyllables, vowel intensity measurement in first and second syllable is also been normalized using the *z-score* equation. Thus, Figure 6.7 is a graphical representation of the normalized vowel intensity distinction between first and second syllable of Orissa Sora disyllables. Moreover, graphs showing normalized vowel intensity distinction in first and second syllable of Assam Sora disyllables are presented alongside for the purpose of comparison.

Figure 6.7: Speakers normalized average vowel intensity in first and second syllable



From Figure 6.7 it is evident similar to Assam Sora disyllables, Orissa Sora disyllables also have greater vowel intensity in the second syllable. This indicates that vowels are relatively louder in second syllable in comparison to the vowels in the first in both Assam Sora and Orissa Sora disyllables. Additionally, a one-way ANOVA test considering normalized average vowel intensity as the dependent variable and syllable position as factor also reveals that average vowel intensity is significantly greater in second syllable than in first syllable of Orissa Sora disyllables ( $F(1,2382) = 165.761, p < 0.001$ ). This provides evidence that similar to Assam Sora, greater vowel intensity in the second syllable can effectively contribute in the realization of iambic stress in Orissa Sora disyllables. Thus, the analysis that confirms greater vowel intensity is an indicator of word stress in Orissa Sora. Also, this work provides evidence showing the phonetic property of word stress namely vowel intensity to be similar in Assam Sora and Orissa Sora disyllables. In which, word stress is generally indicated by greater vowel intensity in the second syllable of disyllables in the two synchronic varieties of the Sora language.

Thus, by examining the three phonetic cues of word stress in Orissa Sora, it is confirmed that the second syllable in disyllables is more prominent than the first syllable. In this regard, the discussion so far has provided evidence that the second syllable is characterized with longer vowel duration, higher fundamental frequency and greater vowel intensity. Therefore, these phonetic characteristics in Orissa Sora are similar to the those of Assam Sora proposed in Horo & Sarmah (2015) as well as in Chapter 3 of this work. Thus, this work proposes similar to Assam Sora, synchronic Orissa Sora also has an iambic stress pattern in the disyllabic words. Also, this work has provided adequate evidences to prove the phonetic realization of iambic stress in both the Sora varieties are similar to each other. Additionally the research undertaken has shown, two important segmental characteristics of Assam Sora and Orissa Sora namely, vowels and word stress are similar to each other. Therefore, the next section compares a third segmental characteristic, voicing contrast in stop consonants, to further verify the phonetic relation between Assam Sora and Orissa Sora.

## **6.5 Stop consonants in Assam Sora and Orissa Sora**

Scholars in the past have agreed that Orissa Sora has six oral stop consonants /p, b, t, d, k, g/ including four places of articulation, namely, labial, dental, alveolar and velar with two-way voicing contrasts, namely, voiced and voiceless. It is also agreed that /t/ and /d/ have voicing asymmetries such that, while /t/ is a voiceless dental stop, /d/ is a voiced alveolar stop in Orissa Sora. Thus, there is neither a voiceless alveolar stop nor a voiced dental stop in Orissa Sora consonant inventory. Likewise, the overview of Assam Sora consonants in Chapter 2 reveals, all the six stop consonants that are found in Assam Sora and their distributions are similar to the stop consonants in Orissa Sora. However, it is observed that while the /t/ and /d/ voicing asymmetry is largely maintained in Assam Sora, it is often subjected to a free

variation. Evidences show a voiced dental stop [d̪] is often produced instead of the voiced alveolar stop /d/ in Assam Sora. This phenomena is also reported in Orissa Sora. In this regard, Ramamurti (1931) proposes, the alternation is conditioned by homophonic sounds in the environment, whereby a dental stop occurs next to a dental consonant and an alveolar stop occurs next to an alveolar consonant in Orissa Sora. However, in Assam Sora the alternation is observed to be a clear case of free variation. It is found that the same word is produced with a voiced alveolar stop /d/ as well as a voiced dental stop [d̪] irrespective of consonantal environments.

However, analysis of Assam Sora speech data in Chapter 5 reveals, the sound occurs in less than 5% of the speech data accounted for the study of stop consonants in Assam Sora. This confirms that the voiced dental stop only occasionally occurs as a free variation of the voiced alveolar stop in Assam Sora, and therefore it is not completely incorporated in Assam Sora. Moreover, the voiceless dental stop /t/ does not show any such variation. Additionally, Assam Sora speech data described in Chapter 5 reveals a few Assam Sora speakers occasionally produce a voiceless velar aspirated stop [k<sup>h</sup>] in place of its unaspirated counterpart /k/. For this sound, it is proposed that the aspirated consonant is borrowed from the Indo-Aryan language, Sadri. However, similar to the voiced dental stop, the aspirated velar stop also occurs in less than 1.25% of the speech data accounted for the analysis of stop consonants in Assam Sora. Thus, it is observed only a few fluent bilingual Assam Sora speaker occasionally produce the aspirated consonant in Assam Sora, therefore the sound is not completely adopted in Assam Sora. The presence of an aspirated stop consonant is not mentioned in any of the Orissa Sora literature reviewed in this work.

Thus, in general it is found that the six stop consonants have similar voicing contrasts in Assam Sora and Orissa Sora. However, it is not know whether the phonetic characteristics of voicing contrast in stop consonants of Assam Sora and Orissa Sora are similar or not. Therefore, it is the interest of this work to compare the phonetic

characteristics of voicing contrasts in the two Sora varieties. Also, studies have shown that stop consonants are basic to the consonant inventory of every language. Studies such as Cho & Ladefoged (1999) propose, individual languages can actually be distinguished depending on the ways in which the languages differentiate between voiced and voiceless stop consonants. Similarly, studies such as Choi (2002) have shown that voicing contrast in stop consonants can be used to distinguish between two Korean dialects, Chonnam and Seoul. Thus, considering the phonetic relevance of studying voicing contrast in stop consonants of related and related languages, this work examines the phonetic characteristics of voicing contrast of stop consonants in Assam Sora and Orissa Sora. To accomplish this task, the acoustic correlates of voicing contrast in stop consonants, Voice Onset Time (VOT), is examined for the stop consonants in Assam Sora and Orissa Sora in the following section.

### **6.5.1 VOT of stop consonants in Assam Sora and Orissa Sora**

Acoustic analysis of Assam Sora stop consonants in Chapter 5 of this work reveals, the VOT feature adequately categorizes the six stop consonants as voiced and voiceless stops in Assam Sora. It is found that while the VOT of voiceless stop ranges between 17 to 37 milliseconds, the VOT of voiced stops ranges between -50 to -62 milliseconds. Thus, the analysis confirms that there are no voicing overlaps and there are no intermediate voicing categories between the voiced and voiceless stop consonants in Assam Sora. Additionally, the VOT measurements show the two infrequent stop consonants, including the voiced dental stop [d̪] and the voiceless velar aspirated stop [kʰ], also have distinct VOT in comparison to the VOT of the voiced and voiceless stops in Assam Sora. It is found, while the average VOT of the voiced dental stop [d̪] is usually between -68 millisecond to -86 milliseconds, the VOT of the voiceless velar aspirated stop is between 38 milliseconds to 62 milliseconds. This indicates that although the two sounds are infrequent in Assam Sora, they are still distinct

from the six regular stop in Assa Sora in the speech of fluent bilingual Assam Sora speakers. Thus, acoustic analysis of stop consonants in Assam Sora has provides as adequate categorization of the two stop categories, and also helps in understanding the phonetic characteristics of voicing contrast in the six stop consonants in Assam Sora.

On the other hand, phonetic characteristics of voicing contrast in the stop consonants of Orissa Sora are not yet known. Therefore, this work has recorded Orissa Sora speech data from speaker living in Raiguda village of southern Orissa. Similar to Assam Sora, the voicing contrast in Orissa Sora stop consonants is mainly examined for VOT variation in word initial stop consonants where all six stops follow a low central vowel /a/ of Orissa Sora vowel system. Calculation of VOT in the speech data includes calculating the time interval between the release of oral occlusion of the stop consonant and beginning of glottalic pulses of the following vowel sound. Moreover, VOT values of the voiced stops are assigned negative values since it is observed that the voiced stops have a period of prevoicing whereby, the glottalic pulses begin even before the oral closure is released. Table 6.13 and 6.14 presents the average VOT of the six voiced and voiceless stop consonants in Orissa Sora with standard deviation in parenthesis. Also, the VOT of Assam Sora stop consonants are given alongside for the purpose of comparison.

Table 6.13: Average VOT of voiced stops in Assam Sora and Orissa Sora

	Singrijhan	Sessa	Koilamari	Lamabari	Raiguda
b	-54.68 (21.25)	-56.32 (19.09)	-57.96 (16.37)	-62 (13.66)	-65.30 (22.88)
d	-56.13 (18.80)	-59.77 (19.41)	-55.83 (17.17)	-56.87 (17.17)	-71.43 (17.87)
g	-57.36 (14.92)	-57.70 (18.93)	-50.51 (16.13)	-59.55 (17.20)	-53.98 (14.28)

From Table 6.13 and 6.14 it is observed that the average VOT of voiced and voiceless stops in Orissa Sora are clearly demarcated from each other. It is observed,

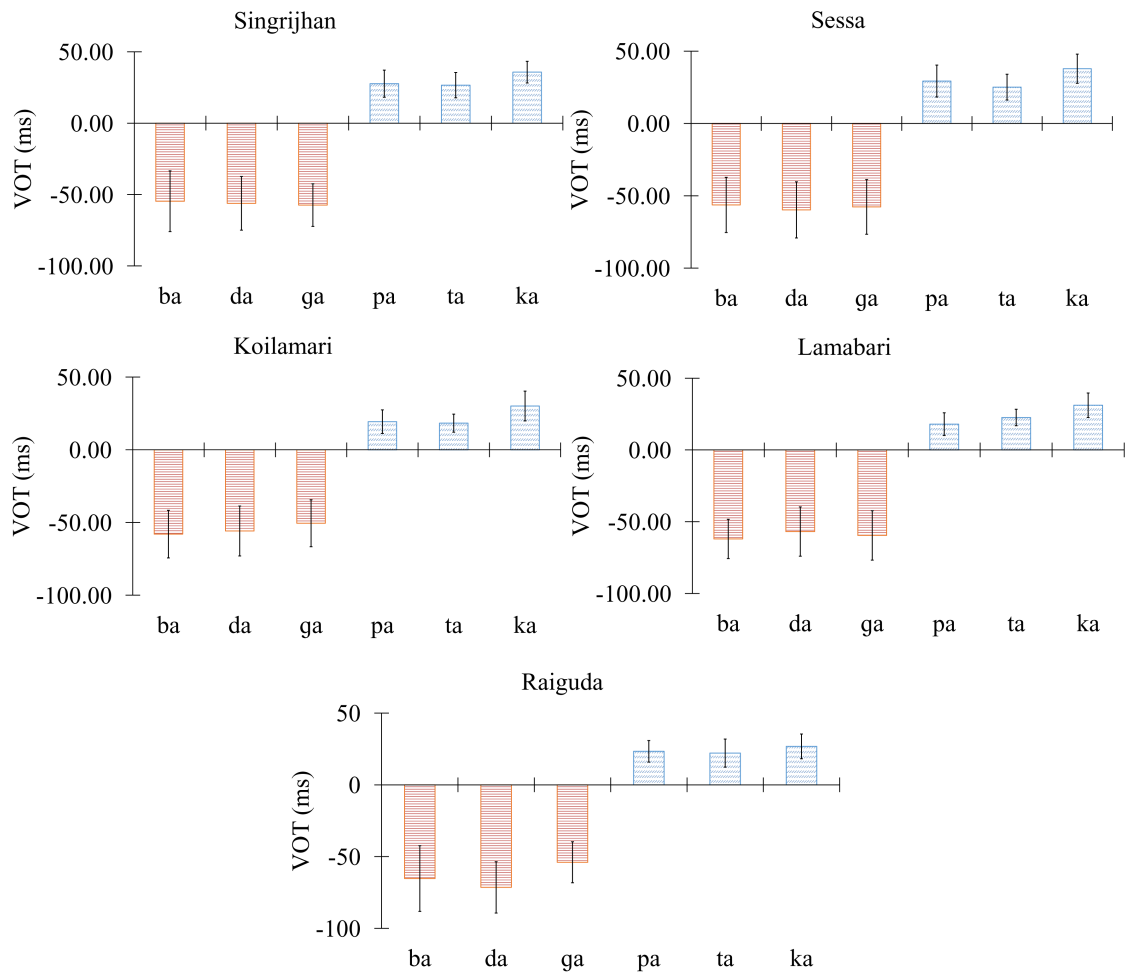
Table 6.14: Average VOT of voiceless stops in Assam Sora and Orissa Sora

	Singrijhan	Sessa	Koilamari	Lamabari	Raiguda
p	27.65 (9.50)	29.29 (11.04)	19.35 (8.09)	17.95 (7.92)	23.34 (7.48)
t	26.61 (8.86)	25.10 (8.93)	18.33 (6.20)	22.60 (5.76)	22.12 (9.77)
k	35.75 (7.58)	37.82 (10.09)	30.12 (10.26)	31.14 (8.52)	26.78 (8.62)

while the average VOT of the voiceless stop ranges between 22 milliseconds to 26 milliseconds, the average VOT of voiced stops ranges between -53 milliseconds to -71 milliseconds. Moreover, it is evident the average VOT of voiced and voiceless stops in Assam Sora and Orissa Sora are fairly similar to each other. Word initial voiced stops /b, d, g/ in both the Sora varieties have nearly the same amount of prevoicing period, which accounts for voicing lead in the voiced consonants. Similarly, voicing is almost equally delayed in the voiceless stops /p, t, k/, resulting in voicing lag in the voiceless stop consonants of both Sora varieties. In order to illustrate the similarity, Figure 6.8 presents a graphical representation of the average VOT difference between the six stop consonants in Orissa Sora with error bars showing standard deviations. Also, the graphs for average VOT difference between voiced and voiceless stops in Assam Sora are presented alongside for the purpose of comparison.

From Figure 6.8 it is clear that both Assam Sora and Orissa Sora have only two way voicing contrasts and the VOT distinction clearly demarcates voiced and voiceless stops in both the Sora varieties. Moreover, while the VOT distinction between voiced and voiceless stops in Assam Sora is shown to be statistically significant, a one-way ANOVA test with Orissa Sora data considering the VOT distinction as dependent variable and voicing contrast as factor reveals, voicing contrast in stop consonants of Orissa Sora is also statistically significant ( $F(1,138) = 274.905, p < 0.001$ ). Thus, the analysis confirms similar to Assam Sora stop consonants, synchronic Orissa Sora also does not have voicing overlaps between the voiced and the voiceless stops and

Figure 6.8: VOT of oral stops in Assam Sora and Orissa Sora



there are no intermediate stops between the two stop categories in both Sora varieties. Additionally, it is found that the voiced dental stop [d̪], occasionally found in Assam Sora, is not present in the synchronic speech data of Orissa Sora recorded in this work. However, a voiceless velar aspirated stop [k<sup>h</sup>] is found in less than 3% of synchronic Orissa Sora speech data accounted for the study of stop consonants. In this regard, it is seen that only one Orissa Sora speaker occasionally produces a voiceless aspirated velar stop in place of its unaspirated counterpart /k/. Thus, it is likely that the speaker has borrowed the sound from Oriya which is predominantly spoken in the vicinity of Sora speakers living in Orissa. Significantly, average VOT of the voiceless aspirated stop in Orissa Sora is found to be 95.63 milliseconds which also appears

distinct from the VOT of both the voiced and voiceless stop consonants in Orissa Sora. However, as the sound is not very regular in Orissa Sora, it is difficult to make any generalization at this point of time. Therefore, the primary findings of this work suggest that VOT distinction clearly demarcates voiced and voiceless stops in both Assam Sora and Orissa Sora. Also, the VOT measure shows that voicing lead in voiced stops is accompanied by a period of prevoicing, and voicing lag in voiceless stops is accompanied by a delay in inception of glottalic pulses after the release of oral occlusion. Hence, from the analysis of VOT in Assam Sora and Orissa Sora it is confirmed that the phonetic characteristics of voicing contrast in word initial stop consonants in Assam Sora is similar to the word initial stop consonants in synchronic Orissa Sora.

## 6.6 Conclusion

This chapter has compared the phonetic properties of two synchronic Sora varieties, namely, Assam Sora and Orissa Sora which is one of the important objectives of this dissertation. Earlier this work has shown, Assam Sora is a transplanted variety of Orissa Sora which emerged mainly due to the migration of Sora speakers from Orissa to Assam in the 19<sup>th</sup> century. Thus, by comparing the phonetic properties of synchronic Assam Sora and synchronic Orissa Sora, this chapter has shown how the two Sora varieties relate to each other after being spoken in two different geographical locations for such a long period. Also, by comparing synchronic Assam Sora and Orissa, this chapter has shown the effect of migration of a language community from one place to another, since studies in the past have proposed that transplanted languages have a general tendency to change. Also, the phonetic features that are compared between Assam Sora and Orissa Sora are typologically relevant since various studies in the past have used those features to compare language varieties of related as well

as unrelated languages. Thus, comparison of phonetic properties in Assam Sora and Orissa Sora includes comparison of the vowel system, comparison of the phonetic realization of word stress and comparison of voicing contrast in the stop consonants. Moreover, the phonetic comparisons are based on speech data collected from four geographical locations in Assam including Singrijhan, Sessa, Koilamari and Lamabari and one geographical location in Orissa namely Raiguda.

Comparison of the vowel system in Assam Sora and Orissa Sora is based on the formant frequency measurements, particularly F1 and F2, which provide information about vowel height and vowel frontness/backness in a vowel system. From the comparison of vowel formants in Assam Sora and Orissa Sora it is confirmed that synchronically the vowel systems of Assam Sora and Orissa Sora are similar to each other. Evidences suggest, both Assam Sora and synchronic Orissa Sora have a six vowel system and all six vowels are categorically distinct from each other in both the Sora varieties. Also, the analysis of vowel formants provide evidence that some of the typical vowel characteristics such as vowel frontness/backness difference between /i/ and /e/ and between /ə/ and /a/ and vowel height difference between /i/ and /u/ and between /ə/, /o/ and /e/ are similar in the vowel systems of both the Sora varieties. Likewise, the analysis of vowel formants in individual syllables of disyllables provide evidence that syllable position plays an important role in the vowel systems of both Assam Sora and Orissa Sora. Evidences show vowel formants change as a function of syllable position, perceptual distance between adjacent vowels are effected as a function of syllable position and overall vowel space also changes as a function of syllable position. Also, it is confirmed that these characteristics are essentially similar in both Assam Sora and Orissa Sora. Thus, it is concluded that besides having similar vowel systems, Assam Sora and synchronic Orissa Sora also have similar vowel characteristics within the vowel system.

Similarly, the comparison of word stress in Assam Sora and Orissa Sora in this

chapter also reveals that both the Sora varieties have an iambic word stress in disyllables. In order to come to this conclusion three phonetic properties of word stress, namely, vowel length, pitch variation and vowel amplitude are examined in this chapter. Consequently, the acoustic correlates of these three phonetic properties, namely, vowel duration, fundamental frequency and vowel intensity provide evidence that second syllable is more prominent than the first syllable. Thus, evidence suggest that the vowel nuclei of the second syllable is usually longer, pitched higher and perceived louder than the vowel nuclei of first syllable. Also, evidences suggest that these characteristics of word stress representing an iambic stress pattern are largely similar in Assam Sora and Orissa Sora.

Finally, the comparison of voicing contrast in stop consonants of Assam Sora and Orissa Sora, also reveals that the VOT characteristic of stop consonants clearly demarcates voiced and voiceless stops in both Assam Sora and Orissa Sora. Thus, evidences show, while a prevoicing period gives a voicing lead in voiced stops, a delayed initiation of glottalic pulses after the release of oral occlusion gives causes a voicing lag in the voiceless stops. Also, the effects are essentially similar in both Assam Sora and Orissa Sora.

In summary, this chapter concludes that even after being separated for more than hundred years, and being spoken in two remotely located geographical locations, Assam Sora and Orissa Sora still exhibit significant similarities in the three important phonetic segments; and the analysis in this chapter provides the relevant evidences for this claim. Moreover, although there are a few exceptions in some cases where the results do not match the regular pattern, the patterns are largely similar across a majority of the compared data; and therefore, the exceptions do not effect the general conclusions of this chapter.

# Chapter 7

## Conclusions and future research prospects

### 7.1 General conclusions

The primary objective of this work was to provide a detailed description of the phonetic properties of Assam Sora. Accordingly, this work has provided a description of the phoneme inventory and syllable structure of Assam Sora as spoken in four villages of Assam, namely, Singrijhan, Sessa, Koilamari and Lamabari. The descriptions are supported by acoustic phonetic analysis. This work also investigated if there are significant differences arising due to geographical variation, among the speech of the speakers of four varieties included in this work. Apart from that, in order to understand the similarity of Assam Sora and Sora spoken in Orissa, data was also collected from Raiguda village in southern Orissa and was subsequently analyzed and compared with the Assam Sora data.

Considering phonetic descriptions reported in the literature being highly inconsistent in terms of the number and types of distinctive segments in Sora, this work also arrived at a conclusion regarding the number of segments and their acoustic characteristics in both Assam Sora and Orissa Sora.

Finally, the study investigated if Assam Sora fits the characteristics of a transplanted language in the sense of Kerswill (2006) or not.

This work concludes that Assam Sora as spoken in four villages of Assam, namely, Singrijhan, Sessa, Koilamari and Lamabari, demonstrates distinctive characteristics of the Munda languages in terms of syllable structure, segmental properties etc. Secondly, it also concludes that Assam Sora spoken in the four villages of Assam is mostly homogenous. Additionally, Assam Sora has strong resemblance to the Sora spoken in Orissa. Thirdly, this work concludes that Assam Sora does not fit into the description of transplanted languages in the sense of Kerswill (2006).

This chapter concludes by providing a note on the segmental properties of Assam Sora in §7.2. A comparison of Assam Sora and Orissa Sora is provided in §7.3. Closing remarks and future research prospects are provided in §7.4.

## **7.2 Segmental properties of Assam Sora**

Prior to the description of Assam Sora segmental properties, this research reviewed the existing literature on the Sora language. It was found that existing studies on the Sora language are mainly based on Sora spoken in Orissa or Andhra Pradesh. There are very few studies based on Sora spoken in Assam. Additionally, this research found different scholars have proposed different segmental properties for the Sora language, as it is spoken in Orissa. Thus, this work concludes that existing descriptions of Orissa Sora segmental properties are inconsistent. Assam Sora segmental features were found to be only partly described. For this reason, this study has conducted linguistic field work to describe the segmental properties of Assam Sora. Thus, segmental properties of Assam Sora reported in this work are based on Assam Sora speech data recorded in four regions of Assam, namely Singrijhan, Sessa, Koilamari and Lamabari. To describe the segmental properties, the phonetic features of syllable structure and phoneme inventory of Assam Sora has been investigated. Syllable structure and phoneme inventory of Assam Sora presented in this work are found uniform in the

speech data of all four regions recorded in this dissertation. Therefore, the segmental properties described show that there are no regional variations of Assam Sora.

### 7.2.1 Assam Sora syllable structure

Data in this work reveals the basic syllable structure of Assam Sora is (C)V(C). Description of Assam Sora syllable structure in this work has also revealed that monosyllabic words are rarely found in Assam Sora. Therefore, this work concludes that disyllabic words are the minimal word template in Assam Sora. In this regard, evidences suggest although Proto-Munda was mainly monosyllabic, a bimoraic constraint in many modern Munda languages and necessitates the formation of longer words (Anderson & Zide, 2002). Thus, similar to Assam Sora, Gtaʔ, a south Munda language spoken in Eastern India, also has disyllabic words as the minimal word template. Moreover, besides identifying the basic syllable template and the basic word template, this work also presents the phonotactic constraints in Assam Sora syllable structure. This work presents evidence showing, both first and second syllable have the same basic syllable template (C)V(C). Likewise, phonemic distribution of all vowels and consonants are similar in both syllable positions, except the retroflex tap /ɽ/ and the labial approxinant /w/. This work concludes that while the retroflex tap occurs only in the onset of the second syllable, the labial approxinant occurs only in the onset of the first syllable in Assam Sora syllable structure.

Additionally, from the acoustic analysis of word stress, this study confirms the claim by Horo & Sarmah (2015) that Assam Sora has an iambic stress pattern in disyllables. By incorporating a larger database, this work displays in Assam Sora disyllables, the second syllable is more prominent than the first syllable. Prominence of the second syllable was evaluated by examining the three phonetic correlates of word stress namely vowel duration, fundamental frequency, and vowel intensity. By examining these three features, the present work provides evidence showing the vowel

nuclei of the second syllable is generally longer, pitched higher and sounds louder than the vowel nuclei of the first syllable. Hence, based on these three phonetic properties, this work concludes Assam Sora has an iambic stress pattern in disyllabic words. However, it is observed that while the vowel duration is invariably longer in the second syllable, fundamental frequency and vowel intensity are sometimes subjected to exceptions. Therefore, this work concludes, while vowel duration is a robust cue for identifying iambic stress in Assam Sora disyllables, fundamental frequency, and vowel intensity mostly provide additional information about iambic stress in Assam Sora disyllables.

## 7.2.2 Assam Sora vowel phonemes

By incorporating a larger database, this work substantiates the claim by Horo & Sarmah (2015) that Assam Sora vowel system has six vowels as shown in Table 7.1.

Table 7.1: Assam Sora vowel phonemes

	Front	Central	Back
High	i		u
Mid	e	ə	o
Low		a	

Evidence regarding the six vowel system in Assam Sora is presented in this work with the help of acoustic analysis. Acoustic analysis of the vowel sounds in this study is based on formant frequencies, particularly with the help of F1 and F2 frequencies. F1 and F2 frequencies provide evidence that the six vowels have distinct vowel properties including height and vowel frontness/backness. Thus, based on their F1 and F2 frequencies, this work provides evidence showing all six vowels are categorically distinct from each other. Additionally, acoustic analysis of the vowel sounds in this work also reveals Assam Sora vowel properties, including vowel height and vowel frontness/backness, generally change as a function of syllable position. Thus,

by examining the F1 and F2 frequencies of Assam Sora vowels in individual syllables of disyllables, this work confirms perceptual distance between adjacent vowels change as a function of syllable position. Consequently, this work has shown perceptual distance between the vowel sounds is generally maximized in the second syllable, whereas perceptual distance between vowels is minimized in the first syllable. Likewise, this work presents evidence showing the overall vowel space of Assam Sora vowel system changes as a function of syllables position. By examining the vowel area space, this work has shown that vowel area space is normally bigger in second syllable than in first syllable. Thus, considering these vowel changes across syllables, this work proposes the changes in vowel properties of Assam Sora vowel system is mainly motivated by the presence of iambic stress in disyllables. Hence, this work concludes that syllable position has an important role in Assam vowel system and prominence of second syllable in disyllables is also reflected on Assam Sora vowel system.

### 7.2.3 Assam Sora consonant phonemes

This work proposes that Assam Sora has eighteen consonant phonemes. The consonant phonemes of Assam Sora proposed in this work are presented in Table 7.2.

Table 7.2: Assam Sora consonant phonemes

	Labial	Dent	Alveo	Post-Alveo	Retro	Palat	Velar	Glott
Plosive	p	b	t	d			k	ʔ
Fricative			s					
Nasal		m		n		ɲ	ŋ	
Lateral				l				
Trill				r				
Tap						ɽ		
Approx		w				j		
Affricate					ɟ			

The eighteen consonants in Assam Sora represent eight places of articulation, namely Bilabial, Dental, Alveolar, Post-Alveolar, Retroflex, Palatal, Velar and Glott-

tal and eight manners of articulations, namely Plosives (stops), Fricatives, Nasals, Lateral, Trill, Tap and Approximants and Affricate. The plosives also have voicing contrasts in Assam Sora such that while /p, t, k/ are voiceless plosives, /b, d, g/ are voiced plosives in Assam Sora. Moreover, in the plosives category, this work has found a voicing asymmetry between /t/ and /d/ whereby, /t/ is a voiceless dental plosive, and /d/ is a voiced alveolar plosive in Assam Sora. Thus, it is proposed that there is neither a voiced dental plosive nor a voiceless alveolar plosive in Assam Sora. Additionally, from the acoustic analysis of stop consonants, this work concludes, the voice onset time feature clearly separates the voiceless stops from the voiced stops in Assam Sora. This work provides evidence showing voiced stops are characterized with a period of pre-voicing that results in voicing lead in the voiced stops. Which indicates voicing in voiced stops starts much earlier even before the oral occlusion is released. On the other hand, the voiceless stops are characterized by a voicing lag which is caused by a delay in starting of the glottalic pulses after the release of the stop consonant. This difference in the voicing contrast between voiced and voiceless stops, is captured by the voice onset time feature of the stop consonants in Assam Sora. For the nasal consonant, it is concluded that, due to the sonority feature in nasal sounds, Assam Sora nasal sounds can be adequately classified with the help of nasal formants. Moreover, since the production of nasal consonants include an oral coupling, it is concluded that Assam Sora nasal consonants exhibit prominent anti-formants which also helps in identifying the nasal sounds in Assam Sora. Finally, for the glottal stop in Assam Sora, this work reports that intervocalic glottal stop in Assam Sora has three phonetic realizations including a complete voiceless glottal stop, a voiced glottal stop and a creaky voiced glottal stop. Also, the creaky voiced glottal stop exhibits pitch perturbation in Assam Sora.

### 7.3 Comparing Assam Sora and Orissa Sora

Synchronic comparison between Assam Sora and Orissa Sora in this work has revealed that segmental and suprasegmental features including word stress, vowels and stop consonants have similar phonetic realizations in the two Sora varieties. The acoustic correlates of these segmental features provide convincing evidence for the synchronic similarity between Assam Sora and Orissa Sora. Hence, the synchronic speech data recorded from Assam Sora and Orissa Sora provide evidence that both the Sora varieties have six distinct vowels including /i, e, a, o, u, ə/, in disyllabic words the second syllable is always prominent and word initial voiced and voiceless stops are separated by a difference in voice onset time of the stop consonants. Additionally, this work has also shown, among the three cues of word stress the vowel duration cue is the most robust cue in both the Sora varieties. Consequently, the vowel duration is found to be invariably long in second syllable rather than in first syllable. Therefore, vowel duration alone is able to attribute prominence to the second syllable in disyllabic words of both the Sora varieties. Also, this research has shown the vowel quality in both Sora varieties change as a function syllable position. As a result, vowel properties including vowel height and vowel frontness/backness change across first and second syllable of disyllables in both Assam Sora and Orissa Sora. Similarly, the difference in perceptual distance between adjacent vowels is seen to be maximized in second syllable and minimized in first syllable in both Assam Sora and Orissa Sora. As a result, vowel space in both Sora varieties appear to be larger in the second syllable than in the first syllable. Thus, through a series of examination of the vowel properties, it is confirmed that syllable position is an important factor effecting the vowel system of both Assam Sora and Orissa Sora. Also, evidences pointed out in this work has shown that, prominence of second syllable is clearly reflected on the shape and structure of the vowel system in both Assam Sora and Orissa Sora.

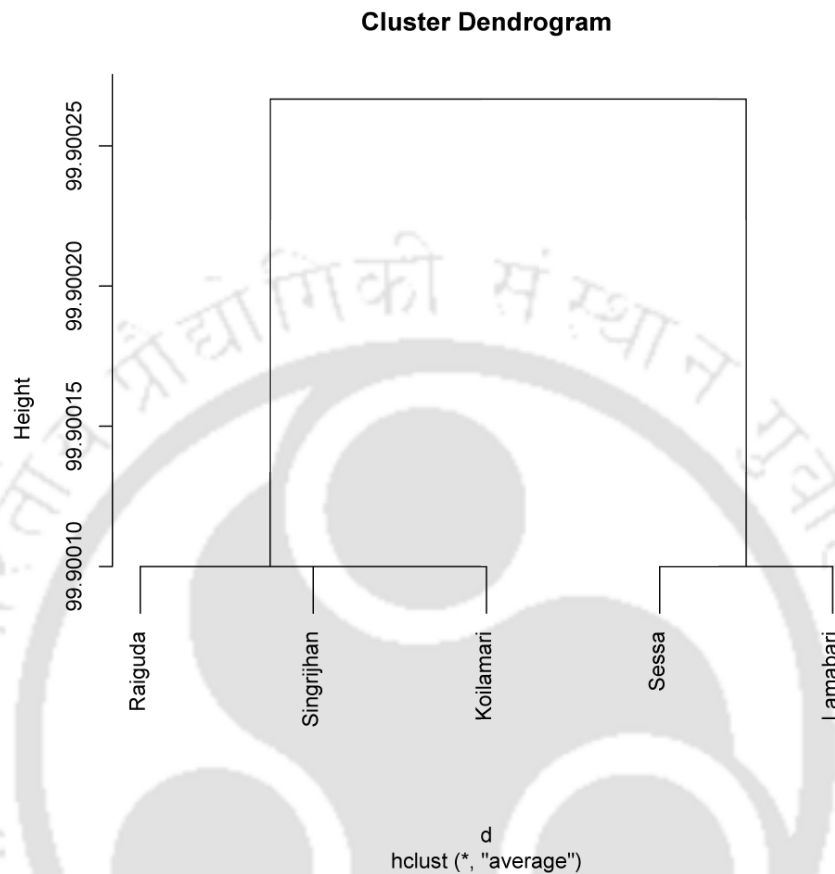
Finally, in order to visualize the correlation between Assam Sora and Orissa

Sora on the basis their phonetic properties, a cluster analysis is conducted using the ‘Weighted Pair Group Method of Arithmetic averages’ (WPGMA) (Johnson, 2008). The cluster analysis is based on similarity measures that uses equation 7.1. to create a cluster dendrogram in order to show correlation between groups that are compared.

$$distance_{c1,c4} = 1/2(distance_{c1,c2} + distance_{c1,c3}) \quad (7.1)$$

Thus, in order to visualize the correlation between Assam Sora and Orissa Sora, WPGMA cluster analysis is conducted on two phonetic features, namely, vowel formants and VOT of stop consonants. For this purpose, average F1 and F2 values for all vowels across all the areas considered for this work were correlated using Pearsons correlation and individual similarity values were obtained and used as inputs to derive cluster dendrograms using the WPGMA method. Thus, the cluster dendrogram in Figure 7.1 shows the correlation between Assam Sora on the basis of their vowel formant features.

Figure 7.1: Vowel correlation between Assam Sora and Orissa Sora



From Figure 7.1 it is observed that the vowel features of Orissa Sora clusters together with vowel features of Assam Sora of two regions, namely, Singrijhan and Koilamari whereas, Sessa and Lamabari regions of Assam Sora form a separate cluster. While there are separate clusters in the dendrogram if the distances are noticed on the y-axis, it is evident that there are no significant differences among the formant frequencies of the vowels across Sora speakers in various regions.

Figure 7.2: Stop consonants correlation between Assam Sora and Orissa Sora

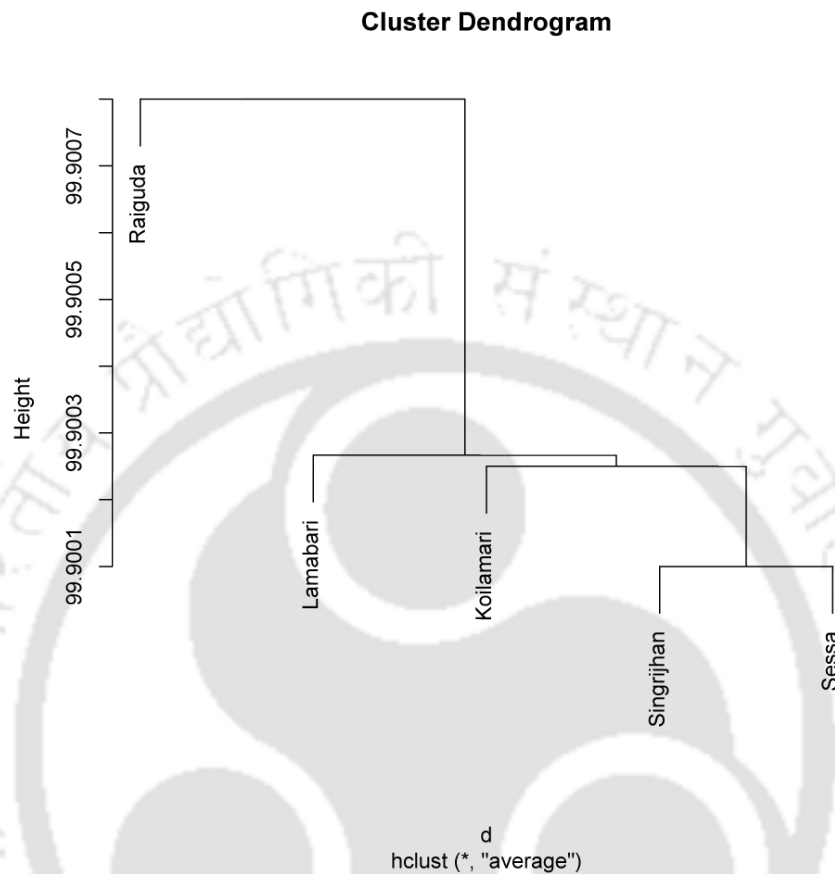


Figure 7.2 shows the correlation between Assam Sora and Orissa Sora on the basis of the VOT values of their stops. As far as VOT is concerned, Figure 7.2 shows that Raiguda stops are marginally different in terms of VOT, from the stops produced by Assam Sora speakers. However, as mentioned in Chapter 6, the difference of VOT among the Sora speakers is not statistically significant. The values on the y-axis clearly show a difference of only 0.0004 between Orissa and Assam Sora VOT measures.

## 7.4 Closing remarks and future research prospects

This work shows that Assam Sora emerged due to the migration of Sora people from Orissa to Assam in the 19<sup>th</sup> century. The Sora people migrated to Assam as indentured tea labourers through various labour recruitment processes. Although the migration was motivated by poor economic conditions in Orissa, evidences highlighted in this work argue that some Soras were deceitfully migrated by the tea planters of Assam. Prior to the migration of Soras, and other indentured labourers from central India, the tea industry in Assam was facing labour scarcity. When the first group of indentured tea labourers arrived, of which the Soras were an integral part, the tea planters viewed them as the most suitable labour force. This was due to the indentured tea labourers' tolerant attitude that they did not complain under the harsh working conditions in the tea industry. Eventually, a large number of indentured tea labourers, including the Soras, were procured and transported to Assam under inhuman conditions, and this continued for several years.

Subsequently, after their arrival in Assam the Soras were inducted in the tea industry, and strategically secluded from local communities. The Soras were confined inside a closed territory and were also subjugated under strict authority of the tea planters. Consequently, the Soras did not get opportunity to interact with the local communities of Assam for many years. The local communities were also prevented from entering into tea plantation territory. Eventually, the social gap between the tea labourers and the local Assamese population increased and contact with the tea garden labourers was socially stigmatized in many cases. This resulted in minimal linguistic contact of the tea garden labourers to other languages spoken in their vicinity. This indirectly helped in preserving the linguistic salience of many tea garden languages. Hence, while Assam Sora was transplanted, there were not many linguistic influences that could change the phonetic and lexical properties of the language.

After completion of the indentured period, the Sora people could not return to

their homeland due to various reasons. Their lack of ability to speak other languages than their mother tongue also made it difficult for them to undertake their journey back to their places of origin.

Significantly, although the Soras have been living in Assam for more than a hundred years, very little is known about their existence in the region. The official population records indicate that there are only 406 Soras in Assam (Registrar General of India, 2001). However, field studies conducted during this research revealed that there are more than 2000 Soras living in just four villages of Assam. The community estimates that there are nearly 40,000 Soras currently living in Assam. This clearly indicates that the Soras are underrepresented in Assam. Similarly, very little is known about the language of the Soras in Assam.

This work is a small step towards understanding the phonetic properties of Sora, a Munda language spoken in Assam, the north eastern part of India. This work has shown that migration of Soras from Orissa to Assam did not affect the phonetic properties of their language. Additionally, the lexical database collected in this work demonstrated that lexical loans in the language is very limited.

In this regard, this work argues that the social isolation of the Soras in Assam has helped them in preserving the phonetic and lexical characteristics of their language. The social isolation of the Soras in Assam is also evident from the linguistic and educational background information collected from the participants in this study. As shown in Table A.1 and Table A.2, only 40% of the Assam Sora population are able to speak Assamese, in spite of being born and brought up in Assam in areas surrounded by Assamese speakers. Apart from that, it is also evident that their social isolation has negatively affected their education. Among all the Sora participants in Assam, only 50% had any institutional education and only 13% had education of class 10 level.

While it is argued in this work that Assam Sora has not changed its phonetic

properties in spite of being a transplanted language, evidently the argument is based on only the phonetic properties of Assam Sora. It is not known whether other linguistic characteristics of Assam Sora, such as morphology or syntax have resemblance to Orissa Sora or not. Therefore, although the phonetic properties of Assam Sora adequately resembles synchronic Orissa Sora, it cannot be generalized that the two Sora varieties are similar to each other in every respect. Therefore, in order to claim that Assam Sora and Orissa Sora are completely similar to each other, it is essential that other linguistic characteristics of the two Sora varieties are analyzed and compared in future.

Additionally, Munda languages spoken among the tea garden labourers of Assam are generally less studied. Studies in India, in the past have mostly focused on the Munda languages of eastern and central India. Thus, it is expected that the findings of this work will motivate future works on other Munda languages outside the Munda heartland.

# Appendix A

## Background information of participants

Table A.1: Information of female participants

#	Name	Age	Sex	Education	Languages Known	Village
1	SS	50	F	None	Sora, Sadri, Assamese	Koilamari
2	DS	40	F	None	Sora, Sadri, Assamese, Hindi	Koilamari
3	PS	45	F	None	Sora, Sadri, Assamese, Hindi	Koilamari
4	JS	30	F	None	Sora, Sadri, Assamese	Koilamari
5	RS	23	F	10	Sora, Sadri, Assamese, Hindi	Koilamari
6	PS	55	F	None	Sora, Sadri	Lamabari
7	GS	60	F	None	Sora, Sadri	Lamabari
8	RS	50	F	None	Sora, Sadri	Lamabari
9	ZS	46	F	None	Sora, Sadri	Lamabari
10	MS	58	F	None	Sora, Sadri	Lamabari
11	BS	65	F	4	Sora, Sadri	Singrijhan
12	MS	60	F	None	Sora, Sadri	Singrijhan
13	CS	62	F	None	Sora, Sadri	Singrijhan
14	LS	47	F	10	Sora, Sadri, Assamese, Hindi	Singrijhan
15	MS	45	F	None	Sora, Sadri	Singrijhan
16	MS	58	F	2	Sora, Sadri	Sessa
17	SS	62	F	1	Sora, Sadri	Sessa
18	MK	64	F	4	Sora, Sadri, Hindi	Sessa
19	AK	50	F	4	Sora, Sadri	Sessa
20	LS	72	F	None	Sora, Sadri	Sessa
21	SS	44	F	None	Sora, Oriya	Raiguda
22	JS	46	F	None	Sora, Oriya	Raiguda
23	BS	36	F	None	Sora, Oriya	Raiguda
24	PS	47	F	None	Sora, Oriya	Raiguda
25	CS	32	F	None	Sora, Oriya, Telugu	Raiguda

Table A.2: Information of male participants

#	Name	Age	Sex	Education	Languages Known	Village
26	BS	30	M	12	Sora, Sadri, Assamese, Hindi	Koilamari
27	SS	57	M	None	Sora, Sadri, Assamese, Hindi	Koilamari
28	RS	40	M	10	Sora, Sadri, Assamese, Hindi	Koilamari
29	GS	29	M	6	Sora, Sadri, Assamese, Hindi	Koilamari
30	KS	49	M	5	Sora, Sadri, Assamese, Hindi	Koilamari
31	RS	55	M	None	Sora, Sadri	Lamabari
32	KS	46	M	None	Sora, Sadri	Lamabari
33	DS	45	M	None	Sora, Sadri	Lamabari
34	PS	70	M	None	Sora, Sadri	Lamabari
35	RS	37	M	8	Sora, Sadri	Lamabari
36	SS	64	M	2	Sora, Sadri	Singrijhan
37	MS	50	M	7	Sora, Sadri, Assamese, Hindi	Singrijhan
38	MS	46	M	8	Sora, Sadri, Assamese, Hindi	Singrijhan
39	BS	53	M	9	Sora, Sadri, Assamese, Hindi	Singrijhan
40	KS	57	M	10	Sora, Sadri, Assamese, Hindi	Singrijhan
41	AS	81	M	None	Sora, Sadri	Sessa
42	AS	69	M	None	Sora, Sadri	Sessa
43	BS	47	M	4	Sora, Sadri, Assamese, Hindi	Sessa
44	BS	49	M	7	Sora, Sadri, Assamese, Hindi	Sessa
45	BS	66	M	None	Sora, Sadri, Hindi, Assamese	Sessa
46	JG	49	M	12	Sora, Oriya	Raiguda
47	SD	48	M	5	Sora, Oriya, Telugu	Raiguda
48	LS	25	M	None	Sora, Oriya	Raiguda
49	RS	36	M	12	Sora, Oriya, Telugu	Raiguda
50	GS	39	M	10	Sora, Oriya, Telugu, Hindi	Raiguda

## Appendix B

### Data sets collected for the dissertation

Table B.1: Assam Sora consonant minimal sets

#	Sora	English	#	Sora	English
1	bənum	ant hill	23	alin	separate
2	dənum	hatching	24	aliŋ	liquor
3	gənum	dreaming	25	səkur	rough
4	pənum	warm	26	ɕəkur	form into lump
5	tənum	collecting	27	asaŋ	bitter
6	kənum	handful	28	aɕaŋ	bone
7	iʔi	louse	29	ɕuɕu	grand father
8	iti	what	30	juju	wrinkle
9	idi	cut	31	kuru	hair
10	aniʔ	written	32	kuɕu	rice
11	anib	root	33	karu	monkey
12	uʔba	brother	34	kaɕu	blue
13	uba	pull out hand	35	jaŋ	mother
14	mama	get away!	36	raŋa	deny
15	nama	this	37	laŋga	good
16	ɲama	touch	38	waŋ	father
17	ajim	to remember	39	laʔa	wide
18	ajin	he	40	raʔa	elephant
19	ajiŋ	sister	41	jap	squeeze
20	ɲama	touch	42	wap	leafy vegetable
21	ɲaŋa	bow	43	ara	tree
22	ɲanaŋ	learn	44	ala	tail

Table B.2: Assam Sora vowel minimal sets (Horo &amp; Sarmah, 2015)

#	Sora	English	#	Sora	English
1	iʔi	louse	34	ɕile	over boiled rice
2	uʔu	hair	35	ɕelu	meat
3	daʔa	water	36	alaŋ	tongue
4	doʔoŋ	body	37	aluŋ	in
5	luʔu	ear	38	ala	tail
6	moʔo	eye	39	kaʔu	blind
7	muʔu	nose	40	kuʔa	earthen oven
8	raʔa	elephant	41	aʔer	ice
9	siʔi	hand	42	oʔar	earth
10	soʔo	rotten foul smell	43	aʔəm	name
11	toʔo	mouth	44	aʔum	urine
12	ɕaʔa	snake	45	kakiŋ	elder sister
13	ɕeʔe	red	46	kakuŋ	elder brother
14	ɕiʔi	tooth	47	məʔam	blood
15	ɕoʔo	fruit	48	pəʔum	warm
16	boɕa	sew	49	aŋgi	battle axe
17	buɕa	suck finger	50	eŋsi	finger ring
18	baɕa	spit out	51	iŋsa	plough shaft
19	bəsə	salt	52	oŋro	kind of honey
20	galɕi	ten	53	əʔ-na	to fan
21	gulɕi	seven	54	uŋru	heating
22	gada	cut like stab	55	araŋ	sour
23	goda	rub/clean	56	areŋ	stone
24	guda	scratch	57	gariŋ	alliance
25	kani	this	58	iroŋ-or	myna
26	kuni	that	59	urəŋ	bamboo
27	kəna	tiger	60	uruŋ-a	bring
28	kina	sing	61	baŋsa	good
29	ajo	fish	62	sese	choose
30	saju	cold	63	tusi	push
31	tagi	hot	64	kisso	dog
32	togi	fire	65	bəsə	salt
33	ɕali	long	66	asu	sick

Table B.3: Assam Sora disyllables (1)

#	Sora	English	Syllable	#	Sora	English	Syllable
1	bada	stick	CV.CV	41	juju	grandmother/ wrinkle	CV.CV
2	baṛa	gun	CV.CV	42	kadi	wooden	CV.CV
3	baru	hill	CV.CV	43	kadu	bangle	CV.CV
4	bati	mushroom	CV.CV	44	kani	this	CV.CV
5	bubu	foam	CV.CV	45	karu	monkey	CV.CV
6	bura	fox	CV.CV	46	kaṛu	blue	CV.CV
7	daga	sun ray	CV.CV	47	kasi	tender fruit	CV.CV
8	daku	keep	CV.CV	48	kəna	tiger	CV.CV
9	daṛa	hard bamboo	CV.CV	49	kina	sing	CV.CV
10	daḍa	climb	CV.CV	50	kora	scold	CV.CV
11	deba	knock softly	CV.CV	51	kota	there	CV.CV
12	deda	move aside	CV.CV	52	kude	gave birth	CV.CV
13	deli	it worked	CV.CV	53	kuni	that	CV.CV
14	deti	not wanted	CV.CV	54	kura	bite	CV.CV
15	didi	behind	CV.CV	55	kuṛa	earth oven	CV.CV
16	dija	waist	CV.CV	56	kuru	body hair	CV.CV
17	dina	rise up	CV.CV	57	laba	slough of snake	CV.CV
18	dira	lean	CV.CV	58	labo	soil	CV.CV
19	disa	country	CV.CV	59	lagu	left out staff	CV.CV
20	doja	hang	CV.CV	60	lebu	money	CV.CV
21	dola	hit	CV.CV	61	luda	sleep	CV.CV
22	dusa	blame/accuse	CV.CV	62	luma	do	CV.CV
23	gada	cut	CV.CV	63	maṛu	dead body	CV.CV
24	galu	fall	CV.CV	64	mola	roast	CV.CV
25	ganu	kind of tuber	CV.CV	65	moṛa	fog	CV.CV
26	gare	liver	CV.CV	66	mudu	blunt knife	CV.CV
27	garu	tiger roaring	CV.CV	67	munu	kind of pulse	CV.CV
28	gate	drink	CV.CV	68	mure	wound/ulcer	CV.CV
29	gati	steer	CV.CV	69	muḍa	open cloth	CV.CV
30	giḍa	look	CV.CV	70	ṇama	hold/touch	CV.CV
31	goba	sit	CV.CV	71	nida	pick and collect	CV.CV
32	goda	clean	CV.CV	72	pada	lock the door	CV.CV
33	guma	sorting rice by throwing up	CV.CV	73	pajo	dent	CV.CV
34	guri	draw line	CV.CV	74	pala	break pot	CV.CV
35	guṛi	proved accused	CV.CV	75	palu	white	CV.CV
36	jaga	black	CV.CV	76	papu	kind of pulse	CV.CV
37	jaji	sister	CV.CV	77	pata	crack	CV.CV
38	jida	cut	CV.CV	78	patu	hole	CV.CV
39	jira	go/run/jump	CV.CV	79	pele	leak from hole	CV.CV
40	jojo	press firmly	CV.CV	80	pete	leaking	CV.CV

Table B.4: Assam Sora disyllables (2)

#	Sora	English	Syllable	#	Sora	English	Syllable
81	peti	leaking	CV.CV	121	tuṛu	friend	CV.CV
82	pila	wooden box	CV.CV	122	tusi	push	CV.CV
83	rama	chew	CV.CV	123	ḍaji	small red pulse	CV.CV
84	rami	indian myna	CV.CV	124	ḍala	lick	CV.CV
85	reja	loan	CV.CV	125	ḍali	long	CV.CV
86	rejam	son in law	CV.CV	126	ḍami	guava	CV.CV
87	rere	cane	CV.CV	127	ḍani	glue	CV.CV
88	reti	went	CV.CV	128	ḍate	catch	CV.CV
89	rida	tear apart	CV.CV	129	ḍati	sesame	CV.CV
90	ridi	bronze	CV.CV	130	ḍaḍa	loan	CV.CV
91	riṛe	numb tooth	CV.CV	131	ḍelu	meat	CV.CV
92	rodu	hard	CV.CV	132	ḍida	start enmity	CV.CV
93	rogo	kind of pulse	CV.CV	133	ḍine	frozen	CV.CV
94	ruda	spread water	CV.CV	134	ḍiḍi	glue	CV.CV
95	ruga	chicken pox	CV.CV	135	ḍona	paste it yourself	CV.CV
96	ruku	gather	CV.CV	136	ḍoṛa	big drain	CV.CV
97	sadi	open loose	CV.CV	137	ḍoro	pouring	CV.CV
98	saju	cold	CV.CV	138	ḍuma	eat	CV.CV
99	sara	squirt grain	CV.CV	139	ḍure	kind of net	CV.CV
100	sare	kind of pulse	CV.CV	140	balaj	loose of knot	CV.CVC
101	saru	straight	CV.CV	141	baliʔ	feather	CV.CVC
102	seda	throw	CV.CV	142	baseṅ	bottom of hill	CV.CVC
103	sida	throw dispose	CV.CV	143	bənum	ant hill	CV.CVC
104	soja	flame the fire	CV.CV	144	biḍol	spit	CV.CVC
105	sona	squint	CV.CV	145	bokuj	one bunch	CV.CVC
106	suda	pain	CV.CV	146	bomaṅ	stubborn	CV.CVC
107	suṛa	big	CV.CV	147	bujaṅ	brother	CV.CVC
108	suruj	curled	CV.CV	148	bukaj	say lie	CV.CVC
109	susu	start	CV.CV	149	buṛaṅ	new budding	CV.CVC
110	tada	stick something forcefully	CV.CV	150	buṛar	bamboo to carry on shoulder	CV.CVC
111	tagi	hot	CV.CV	151	dajdoṅ	climb	CV.CVC
112	təda	beat hit	CV.CV	152	dajer	turn	CV.CVC
113	tətə	beat/hit	CV.CV	153	danir	support	CV.CVC
114	tije	gave	CV.CV	154	danoj	hanging	CV.CVC
115	tiki	at last	CV.CV	155	datul	trap	CV.CVC
116	tiri	finished	CV.CV	156	deliṅ	lose	CV.CVC
117	toba	knot	CV.CV	157	dəṛum	thunder	CV.CVC
118	togi	fire	CV.CV	158	deḍiṅ	how much	CV.CVC
119	tuda	cut	CV.CV	159	didiṅ	preserve	CV.CVC
120	tula	support	CV.CV	160	diṛir	embankment	CV.CVC

Table B.5: Assam Sora disyllables (3)

#	Sora	English	Syllable	#	Sora	English	Syllable
161	dojal	visit relatives	CV.CVC	191	kupal	kind of mat	CV.CVC
162	gajur	turn around half	CV.CVC	192	kurab	bushes	CV.CVC
163	gamaŋ	rich	CV.CVC	193	kuɽup	making circle	CV.CVC
164	ganaj	eatable tuber	CV.CVC	194	lagam	greedy	CV.CVC
165	gənur	rain	CV.CVC	195	lagaʔ	console	CV.CVC
166	gəɽiŋ	alliance	CV.CVC	196	lakəj	sand	CV.CVC
167	gəruŋ	excessive	CV.CVC	197	lataŋ	melt	CV.CVC
168	gətər	hit	CV.CVC	198	latum	cotton	CV.CVC
169	gobaj	sit	CV.CVC	199	lemer	melt	CV.CVC
170	gurat	unripe	CV.CVC	200	mabar	go	CV.CVC
171	gusaj	cover chest with cloth	CV.CVC	201	madoj	drowsy	CV.CVC
172	jiraj	come	CV.CVC	202	majoŋ	breast	CV.CVC
173	jogab	eat in one gulp	CV.CVC	203	manəŋ	sidewise	CV.CVC
174	jutin	shaking body	CV.CVC	204	marəŋ	growing	CV.CVC
175	kanin	this specific	CV.CVC	205	mariʔ	clear	CV.CVC
176	kaniŋ	wound	CV.CVC	206	metaŋ	slow breeze of air	CV.CVC
177	kaɽib	long sword	CV.CVC	207	mijam	blood	CV.CVC
178	kaɽub	throb	CV.CVC	208	mujeʔ	day before yesterday	CV.CVC
179	kaɽuŋ	some pattern in white	CV.CVC	209	mujeʔ	that day	CV.CVC
180	kaɽup	hazy	CV.CVC	210	ɲadap	at once/ decidedly	CV.CVC
181	kənam	wood dust	CV.CVC	211	namiʔ	just now	CV.CVC
182	kituŋ	god	CV.CVC	212	ɲanaŋ	learn	CV.CVC
183	kolin	that side	CV.CVC	213	ɲatiŋ	heating	CV.CVC
184	kotan	that side	CV.CVC	214	paniʔ	offering	CV.CVC
185	kotin	there	CV.CVC	215	pasiʔ	child	CV.CVC
186	kudup	together	CV.CVC	216	pəɲum	warm	CV.CVC
187	kulam	spirit	CV.CVC	217	pitar	box	CV.CVC
188	kulin	there	CV.CVC	218	radur	hard bamboo	CV.CVC
189	kulma	ghost devil soul	CV.CVC	219	ratul	land slide	CV.CVC
190	kumap	funeral ashes	CV.CVC	220	regam	medicine	CV.CVC

Table B.6: Assam Sora disyllables (4)

#	Sora	English	Syllable	#	Sora	English	Syllable
221	roḍol	straight	CV.CVC	261	baʔdo	curse	CVC.CV
222	rubən	yesterday	CV.CVC	262	birda	meeting	CVC.CV
223	ruḍum	offer as food	CV.CVC	263	birna	talk or speak	CVC.CV
224	sabaj	inconvenient	CV.CVC	264	buʔtu	gram pulse	CVC.CV
225	sajim	make quite	CV.CVC	265	daṅtu	cover from fire	CVC.CV
226	sakab	crowded	CV.CVC	266	deʔna	move aside	CVC.CV
227	sanaṅ	door	CV.CVC	267	diṅli	late	CVC.CV
228	saṅaṅ	swollen	CV.CVC	268	diṅlu	pull one another	CVC.CV
229	sariʔ	come fast	CV.CVC	269	diṅte	delay	CVC.CV
230	satar	nerves	CV.CVC	270	dirga	wait	CVC.CV
231	sərum	good smell	CV.CVC	271	dirna	agreed	CVC.CV
232	sogar	foul smell	CV.CVC	272	diʔne	showing height	CVC.CV
233	somiṅ	hen dung	CV.CVC	273	dumba	non Sora	CVC.CV
234	sonum	ghost	CV.CVC	274	dumle	squint	CVC.CV
235	sopa	mat	CV.CVC	275	dumta	guard at night	CVC.CV
236	tadaj	drunk	CV.CVC	276	dundu	swing baby	CVC.CV
237	tadar	tuft of cock	CV.CVC	277	duṅna	take out	CVC.CV
238	taniṅ	stuff to carry load on head	CV.CVC	278	gabba	cut	CVC.CV
239	tanol	rope	CV.CVC	279	galle	placed across	CVC.CV
240	tanul	support	CV.CVC	280	galḍi	ten	CVC.CV
241	teram	kind of honey	CV.CVC	281	gojte	knows	CVC.CV
242	teṅer	making circle	CV.CVC	282	gulti	back of head	CVC.CV
243	tulab	forest	CV.CVC	283	gulḍi	seven	CVC.CV
244	tunal	python	CV.CVC	284	guṅle	fallen	CVC.CV
245	tunol	set as support	CV.CVC	285	guṅli	fell down	CVC.CV
246	turub	cloud	CV.CVC	286	guṅti	will fall	CVC.CV
247	wabaṅ	tea firewood	CV.CVC	287	gupna	take something on waist	CVC.CV
248	ḍabeʔ	tasteless food	CV.CVC	288	gurri	mature ripe	CVC.CV
249	ḍadoṅ	did not catch	CV.CVC	289	janti	in what way	CVC.CV
250	ḍalem	smooth	CV.CVC	290	jarme	day after tomorrow	CVC.CV
251	ḍaṅaṅ	matured	CV.CVC	291	jarna	roam	CVC.CV
252	ḍəgət	sticky	CV.CVC	292	jerte	going	CVC.CV
253	ḍiraṅ	terrace on tree	CV.CVC	293	jette	will go	CVC.CV
254	ḍitiʔ	little/some	CV.CVC	294	jimna	cover oneself	CVC.CV
255	ḍodoṅ	do not swipe	CV.CVC	295	jirba	go together	CVC.CV
256	ḍoraṅ	goosebumps	CV.CVC	296	jirna	run	CVC.CV
257	ḍuḍuṅ	grandfather	CV.CVC	297	jirte	will go/leaving	CVC.CV
258	banti	divide	CVC.CV	298	kalla	stupid man	CVC.CV
259	baṅke	ankle	CVC.CV	299	kambu	bear	CVC.CV
260	baṅsa	good	CVC.CV	300	kansi	small flower	CVC.CV

Table B.7: Assam Sora disyllables (5)

#	Sora	English	Syllable	#	Sora	English	Syllable
301	kapɾa	shoulder	CVC.CV	331	papdi	fish fins	CVC.CV
302	kaʔdi	wood	CVC.CV	332	paska	twist neck	CVC.CV
303	kaʔpa	wings	CVC.CV	333	paʔde	wrench	CVC.CV
304	kəmme	goat	CVC.CV	334	piŋda	float	CVC.CV
305	kinlu	stubborn	CVC.CV	335	piŋli	float	CVC.CV
306	kisso	dog	CVC.CV	336	poʔde	twist	CVC.CV
307	kujtu	bunch	CVC.CV	337	rajna	meeting	CVC.CV
308	kulba	ghost/ dead body	CVC.CV	338	raŋgi	wind	CVC.CV
309	kullu	fox	CVC.CV	339	riŋra	loose	CVC.CV
310	kunde	there	CVC.CV	340	rojle	weak person	CVC.CV
311	kunde	there that	CVC.CV	341	rondu	lazy	CVC.CV
312	kundi	knife	CVC.CV	342	ruʔma	pour	CVC.CV
313	kundu	embrace	CVC.CV	343	sajte	scare someone to beat	CVC.CV
314	kuple	fitted perfectly	CVC.CV	344	sandi	boundary	CVC.CV
315	kuʔli	give birth	CVC.CV	345	sanna	small	CVC.CV
316	kuʔte	labor pain	CVC.CV	346	sapa	mat	CVC.CV
317	laŋga	good	CVC.CV	347	saʔpa	ask	CVC.CV
318	laʔgu	save preserve	CVC.CV	348	silla	spear	CVC.CV
319	laʔlu	save preserve	CVC.CV	349	sindi	date fruit	CVC.CV
320	lojle	take something secretly	CVC.CV	350	sipe	hand finger	CVC.CV
321	luŋga	drain	CVC.CV	351	sirsa	doesn't do anything	CVC.CV
322	mandaj	lame man	CVC.CV	352	siʔda	leave	CVC.CV
323	manni	goddess	CVC.CV	353	siʔli	lost	CVC.CV
324	maʔne	kill	CVC.CV	354	sojtu	set fire	CVC.CV
325	monna	like/ attract to mind	CVC.CV	355	soʔpa	bamboo mat	CVC.CV
326	naŋte	scare someone to beat	CVC.CV	356	soʔte	leaping	CVC.CV
327	nappe	immediately	CVC.CV	357	tajta	flesh on the back	CVC.CV
328	ɲarme	yesterday	CVC.CV	358	tamba	day	CVC.CV
329	pabba	stumble	CVC.CV	359	tamba	stumble	CVC.CV
330	paŋɕa	bring bone	CVC.CV	360	taŋga	taboo leftover food	CVC.CV

Table B.8: Assam Sora disyllables (6)

#	Sora	English	Syllable	#	Sora	English	Syllable
361	taŋka	money	CVC.CV	401	dirtyn	persuade	CVC.CVC
362	taŋli	cow	CVC.CV	402	gajgaj	bore hole	CVC.CVC
363	taple	bring out	CVC.CV	403	gajlaj	tuber	CVC.CVC
364	tapre	lightning	CVC.CV	404	galɕiŋ	way to sit	CVC.CVC
365	tarba	flower	CVC.CV	405	gannai	tuber	CVC.CVC
366	taʔle	ceased	CVC.CV	406	gaʔgab	cut	CVC.CVC
367	taʔlu	tell lie	CVC.CV	407	gaʔtaj	kind of knife	CVC.CVC
368	taʔmu	make sleep	CVC.CV	408	giʔdəm	see	CVC.CVC
369	taʔre	lightning	CVC.CV	409	giʔlaj	saw	CVC.CVC
370	teʔna	swing	CVC.CV	410	giʔtam	will see you	CVC.CVC
371	teʔte	swing	CVC.CV	411	gojbir	joking	CVC.CVC
372	teʔtin	swing	CVC.CV	412	gorser	dead	CVC.CVC
373	toʔle	left out	CVC.CV	413	gumbaj	matured	CVC.CVC
374	tubda	beat cloth to wash	CVC.CV	414	jabjab	crowded	CVC.CVC
375	tumbe	carry on back	CVC.CV	415	jaggar	completely	CVC.CVC
376	turru	ant house	CVC.CV	416	jamdoŋ	do like this	CVC.CVC
377	ɕabda	chew	CVC.CV	417	jamlaj	stretch hand	CVC.CVC
378	ɕajta	down	CVC.CV	418	jaŋboj	goddess	CVC.CVC
379	ɕilka	guava	CVC.CV	419	jaŋgam	how	CVC.CVC
380	ɕindo	why	CVC.CV	420	jarnaj	turn	CVC.CVC
381	ɕiŋe	wash legs	CVC.CV	421	jartin	returning	CVC.CVC
382	ɕiŋli	frozen	CVC.CV	422	jettaj	coming	CVC.CVC
383	ɕiŋŋe	leg finger	CVC.CV	423	jiŋjiŋ	puff of air	CVC.CVC
384	ɕirga	vulture	CVC.CV	424	jirtaj	will come	CVC.CVC
385	ɕonni	ladies finger	CVC.CV	425	jirɕaŋ	leap/jump	CVC.CVC
386	ɕoʔba	muddy	CVC.CV	426	jubjub	fire flame	CVC.CVC
387	ɕoʔna	put on skin	CVC.CV	427	jurpuʔ	shake cloth	CVC.CVC
388	ɕumdo	do not eat	CVC.CV	428	kajkaj	scold	CVC.CVC
389	bajtam	you will be joined	CVC.CVC	429	kambam	sal tree	CVC.CVC
390	bandan	pond	CVC.CVC	430	kampuŋ	stomach	CVC.CVC
391	banrab	angry	CVC.CVC	431	kandar	branch	CVC.CVC
392	bantoŋ	scared	CVC.CVC	432	kanduŋ	back of body	CVC.CVC
393	barrap	scold	CVC.CVC	433	kantuj	owl	CVC.CVC
394	bəbŋen	of head	CVC.CVC	434	kardul	very hungry	CVC.CVC
395	bəʔsuŋ	mud plaster	CVC.CVC	435	karkar	shake in fear	CVC.CVC
396	bojboj	very	CVC.CVC	436	kassim	chicken	CVC.CVC
397	boŋtel	buffalo	CVC.CVC	437	kəntal	wall	CVC.CVC
398	bubjuŋ	sun set	CVC.CVC	438	kəntəŋ	stubborn	CVC.CVC
399	bujbuj	leak	CVC.CVC	439	kinlaj	pumpkin	CVC.CVC
400	dirdir	lean	CVC.CVC	440	kinlaj	singing	CVC.CVC

Table B.9: Assam Sora disyllables (7)

#	Sora	English	Syllable	#	Sora	English	Syllable
441	kinloŋ	man turns into tiger	CVC.CVC	471	pasɕəŋ	child	CVC.CVC
442	kinsar	small deer	CVC.CVC	472	paʔduŋ	chickenpox	CVC.CVC
443	kinɕaŋ	back of body	CVC.CVC	473	paʔɕaŋ	neck bone twisted	CVC.CVC
444	kiʔɕiŋ	stubborn	CVC.CVC	474	puŋtar	punch hole	CVC.CVC
445	kumgam	back of the neck	CVC.CVC	475	purban	ancient	CVC.CVC
446	kundup	tight fist	CVC.CVC	476	purpur	worship	CVC.CVC
447	kuntuŋ	dry gourd	CVC.CVC	477	rammaŋ	separate	CVC.CVC
448	kuntuʔ	bottle guard	CVC.CVC	478	raŋtal	tree terrace	CVC.CVC
449	kuŋkam	vermilion	CVC.CVC	479	raŋɕiŋ	way priests sit	CVC.CVC
450	kuʔbar	given birth	CVC.CVC	480	raʔtaj	way of cutting	CVC.CVC
451	kuʔdub	all	CVC.CVC	481	riʔɕaŋ	openly directly	CVC.CVC
452	lambaʔ	cover with cloth	CVC.CVC	482	ruŋtab	unmarried lady	CVC.CVC
453	laʔlad	press firmly	CVC.CVC	483	sajlaj	which one	CVC.CVC
454	liblib	fire flames	CVC.CVC	484	sajsaj	cut slices	CVC.CVC
455	limtam	greeting	CVC.CVC	485	salpam	sour flavor added to rice	CVC.CVC
456	londeŋ	collect	CVC.CVC	486	sarpiʔ	soft rain	CVC.CVC
457	lublub	shade	CVC.CVC	487	sirsar	healing	CVC.CVC
458	mannam	my eyes	CVC.CVC	488	siʔruŋ	mango terrace for wedding	CVC.CVC
459	maʔmab	fall when pushed	CVC.CVC	489	sojsoj	flatterer	CVC.CVC
460	maʔmar	insect	CVC.CVC	490	sublaj	related to fence	CVC.CVC
461	maʔmaʔ	fell facing the ground	CVC.CVC	491	subsub	husking rise	CVC.CVC
462	maʔnen	your eyes	CVC.CVC	492	subsub	falsely	CVC.CVC
463	miʔmir	cockroach	CVC.CVC	493	sumsum	touch	CVC.CVC
464	miʔtal	half cut bamboo	CVC.CVC	494	sundaŋ	bamboo pole	CVC.CVC
465	ŋamdoŋ	don't touch	CVC.CVC	495	sunruj	bamboo basket	CVC.CVC
466	ŋamlaj	catch/touch	CVC.CVC	496	surbaŋ	kind of bee	CVC.CVC
467	ŋaʔtaj	sun heat	CVC.CVC	497	taklam	crazy man	CVC.CVC
468	palpal	divide	CVC.CVC	498	tallam	rescue	CVC.CVC
469	paŋkuj	elope	CVC.CVC	499	tampal	push to make fall	CVC.CVC
470	pardaj	spark	CVC.CVC	500	taŋdiŋ	sorting rice	CVC.CVC

Table B.10: Assam Sora disyllables (8)

#	Sora	English	Syllable	#	Sora	English	Syllable
501	taŋgun	related to fence	CVC.CVC	541	ɕaʔa	snake	CVC.V
502	taŋjiŋ	loan interest	CVC.CVC	542	ɕiʔi	leg	CVC.V
503	taŋkub	cut	CVC.CVC	543	ɕoʔo	fruit	CVC.V
504	taplaj	extract honey	CVC.CVC	544	boʔor	kind of bee	CVC.VC
505	taʔtaʔ	slap	CVC.CVC	545	doʔoŋ	body	CVC.VC
506	teŋdəm	abundant	CVC.CVC	546	jaʔaŋ	mother	CVC.VC
507	timpuŋ	half not filled	CVC.CVC	547	kiʔiŋ	wound	CVC.VC
508	tiŋkir	kind of bee	CVC.CVC	548	loŋjir	lines to plant potato	CVC.VC
509	tiʔtin	there	CVC.CVC	549	luŋuʔ	dark	CVC.VC
510	tujtuj	star	CVC.CVC	550	luʔmu	shade	CVC.VC
511	tumbat	stick forcibly	CVC.CVC	551	maʔaŋ	laughing	CVC.VC
512	tunʂal	punch	CVC.CVC	552	meʔeŋ	alive	CVC.VC
513	tunʂtuŋ	punch	CVC.CVC	553	moʔoj	excitement	CVC.VC
514	tunʂɕaŋ	speak directly	CVC.CVC	554	saʔaŋ	tea firewood	CVC.VC
515	turŋam	way of falling	CVC.CVC	555	suʔur	boil water	CVC.VC
516	turɕab	jump	CVC.CVC	556	taŋor	road	CVC.VC
517	ɕablaj	touch	CVC.CVC	557	tundu	tuft of cock	CVC.VC
518	ɕantam	appoint	CVC.CVC	558	tunʂal	inflammation	CVC.VC
519	ɕarɕar	not smooth	CVC.CVC	559	tunʂar	halve	CVC.VC
520	ɕaʔtob	drop of liquid	CVC.CVC	560	wabol	tea leaf	CVC.VC
521	ɕaʔɕab	not smooth	CVC.CVC	561	waʔaŋ	father	CVC.VC
522	ɕiʔdaŋ	little	CVC.CVC	562	aba	madhuca indica	V.CV
523	ɕolɕol	say lie	CVC.CVC	563	abe	jump to cross	V.CV
524	ɕoŋɕoŋ	equalize	CVC.CVC	564	abu	red ant	V.CV
525	ɕumbur	stealing	CVC.CVC	565	ada	curry	V.CV
526	buʔu	hot object	CVC.V	566	adi	cotton	V.CV
527	daŋa	cover from heating	CVC.V	567	adu	two palms together	V.CV
528	daʔa	water	CVC.V	568	aga	sharpener	V.CV
529	dəŋa	pull	CVC.V	569	aji	sister	V.CV
530	jaʔa	mother	CVC.V	570	ajo	fish	V.CV
531	laʔa	wide	CVC.V	571	aku	sugarcane	V.CV
532	moʔo	eyes	CVC.V	572	ala	tail	V.CV
533	ŋaŋa	bow	CVC.V	573	alu	ear	V.CV
534	paŋa	bring object	CVC.V	574	amu	nose	V.CV
535	raŋa	deny	CVC.V	575	ara	tree/wood	V.CV
536	siʔi	hand	CVC.V	576	are	egg	V.CV
537	soʔo	foul smell	CVC.V	577	aʂu	arrive	V.CV
538	tortor	fence	CVC.V	578	asu	sick	V.CV
539	toʔo	mouth	CVC.V	579	aɕi	root	V.CV
540	tuŋa	punch	CVC.V	580	əda	fan verb	V.CV

Table B.11: Assam Sora disyllables (9)

#	Sora	English	Syllable	#	Sora	English	Syllable
581	ede	curl/roll	V.CV	621	ajim	think of	V.CVC
582	ædi	fan verb	V.CV	622	akar	nail finger	V.CVC
583	æli	rice beer new	V.CV	623	aki?	dead	V.CVC
584	ida	cut	V.CV	624	akol	hook together	V.CVC
585	ide	sticky substance	V.CV	625	akub	close firmly	V.CVC
586	idi	write	V.CV	626	alam	hay	V.CVC
587	igu	carry on shoulder	V.CV	627	alaŋ	tongue	V.CVC
588	ila	peel	V.CV	628	alib	deer	V.CVC
589	ira	pour	V.CV	629	alin	separate grain with feet	V.CVC
590	isa	quarrel	V.CV	630	aluŋ	inside	V.CVC
591	ite	tipsy/drunken	V.CV	631	amad	eyes	V.CVC
592	oba	watchman	V.CV	632	amam	skim over the surface	V.CVC
593	oda	mix	V.CV	633	aman	you	V.CVC
594	ola	leaf	V.CV	634	ameŋ	unripe	V.CVC
595	ole	saliva	V.CV	635	anam	fasting	V.CVC
596	oŋa	water canal	V.CV	636	anaŋ	sling	V.CVC
597	oŋe	approve	V.CV	637	anar	match stick	V.CVC
598	oŋa	peel	V.CV	638	aŋəm	name	V.CVC
599	uba	put hand inside	V.CV	639	anəŋ	netted basket	V.CVC
600	uma	vapor	V.CV	640	anib	bottom of tree	V.CVC
601	uŋa	mango	V.CV	641	anin	he	V.CVC
602	uraj	open something	V.CV	642	ani?	written	V.CVC
603	usi	mischievous	V.CV	643	aŋum	urine	V.CVC
604	uŋa	throw water with hand	V.CV	644	anuŋ	netted basket	V.CVC
605	abaj	seed	V.CVC	645	aŋur	smell	V.CVC
606	abar	substitute	V.CVC	646	apiŋ	float	V.CVC
607	aboj	one	V.CVC	647	aput	rebound	V.CVC
608	adab	kind of tuber for pig	V.CVC	648	apu?	broken bangle	V.CVC
609	adaj	promote	V.CVC	649	aram	gather	V.CVC
610	adam	near	V.CVC	650	araŋ	sour	V.CVC
611	adaŋ	honey	V.CVC	651	aŋaŋ	kite	V.CVC
612	adil	dirt from body	V.CVC	652	arel	hail storm	V.CVC
613	adup	two palms together	V.CVC	653	areŋ	weighing stone	V.CVC
614	adur	harvest festival	V.CVC	654	aŋer	stone	V.CVC
615	agab	tall grass	V.CVC	655	aroj	fly insect	V.CVC
616	aguj	impatience	V.CVC	656	aŋup	milk	V.CVC
617	agur	ripe fruit	V.CVC	657	aŋur	set free	V.CVC
618	ajam	long time ago	V.CVC	658	asaj	charcoal	V.CVC
619	ajaŋ	unripe	V.CVC	659	asaŋ	bitter	V.CVC
620	ajer	change	V.CVC	660	asar	dry	V.CVC

Table B.12: Assam Sora disyllables (10)

#	Sora	English	Syllable	#	Sora	English	Syllable
661	asin	boiled	V.CVC	701	osoj	direct kept here	V.CVC
662	asuj	little	V.CVC	702	otaʔ	this side	V.CVC
663	asup	sharp	V.CVC	703	oteʔ	stumble	V.CVC
664	ataj	fart	V.CVC	704	oɕoj	cut to small	V.CVC
665	atal	palm	V.CVC	705	udiʔ	beetle	V.CVC
666	ataŋ	brain	V.CVC	706	udoŋ	perhaps	V.CVC
667	ateŋ	excessive	V.CVC	707	ugaj	shocked	V.CVC
668	atuŋ	pumpkin	V.CVC	708	ugar	chest	V.CVC
669	aɕaŋ	bones	V.CVC	709	ujoŋ	fall off	V.CVC
670	aɕiŋ	grass	V.CVC	710	ujul	hot	V.CVC
671	aɕiʔ	little	V.CVC	711	ujuŋ	sun/day	V.CVC
672	edal	dog sleeping	V.CVC	712	ulam	slightly evening	V.CVC
673	egej	appear	V.CVC	713	ulim	cock crowing	V.CVC
674	ekam	a/an/ the	V.CVC	714	umaŋ	quick	V.CVC
675	eliʔ	pull stomach inside	V.CVC	715	umeŋ	fast	V.CVC
676	eraŋ	wait	V.CVC	716	umiŋ	quick	V.CVC
677	əriŋ	loose of object	V.CVC	717	upul	sweat	V.CVC
678	idaj	soul of dead people	V.CVC	718	upər	small feather	V.CVC
679	idaŋ	cut firewood	V.CVC	719	uraŋ	tasteless uneaten food	V.CVC
680	idol	written/pen	V.CVC	720	uɕar	mango tree	V.CVC
681	ijaŋ	mother	V.CVC	721	urəŋ	bamboo	V.CVC
682	imad	make signs with eyes	V.CVC	722	uɕiʔ	roll like ball	V.CVC
683	irab	interest effort	V.CVC	723	uruŋ	take person	V.CVC
684	itin	what/which	V.CVC	724	usaj	boil cook	V.CVC
685	odin	move aside	V.CVC	725	usuŋ	falling	V.CVC
686	odul	taunt/insult	V.CVC	726	adde	not wanted	VC.CV
687	ojan	inflamm fire	V.CVC	727	addi	lift up	VC.CV
688	ojer	winnowing basket	V.CVC	728	akka	stuck	VC.CV
689	okiʔ	reserve food	V.CVC	729	alti	soft bran	VC.CV
690	olaj	exchange	V.CVC	730	ansa	separate	VC.CV
691	olaŋ	call/enquire	V.CVC	731	anti	that	VC.CV
692	oloj	kind of tree	V.CVC	732	aŋɕi	childless woman	VC.CV
693	onal	wooden mortar	V.CVC	733	aŋgi	battle axe	VC.CV
694	open	my child	V.CVC	734	aptu	hole	VC.CV
695	oniʔ	write fast	V.CVC	735	arbu	ant eater	VC.CV
696	orab	fill by pressing	V.CVC	736	arda	strain water	VC.CV
697	oraj	kind of pulse	V.CVC	737	arra	unfruitful	VC.CV
698	orol	guard at daytime	V.CVC	738	arre	parrot	VC.CV
699	orub	evening	V.CVC	739	arsi	long tail monkey	VC.CV
700	oseŋ	less	V.CVC	740	arti	deal	VC.CV

Table B.13: Assam Sora disyllables (11)

#	Sora	English	Syllable	#	Sora	English	Syllable
741	aʔda	one forth	VC.CV	781	uble	leftover food	VC.CV
742	aʔpa	thrash rice	VC.CV	782	ujda	sprinkle water	VC.CV
743	eŋda	swim	VC.CV	783	ulda	marrow	VC.CV
744	eŋga	how	VC.CV	784	uŋle	astonished	VC.CV
745	eŋlu	watermelon	VC.CV	785	uŋru	heating/humid	VC.CV
746	eŋra	cucumber	VC.CV	786	uŋte	glitter	VC.CV
747	eŋre	garden herb	VC.CV	787	urpa	unstring a bow	VC.CV
748	eŋsi	finger ring hand	VC.CV	788	uʔba	younger brother	VC.CV
749	eŋte	fly/swim	VC.CV	789	uʔka	habituated	VC.CV
750	ərda	sprinkle water	VC.CV	790	ablaj	something to eat	VC.CVC
751	ərri	winner	VC.CV	791	aldəŋ	broken	VC.CVC
752	erɕa	raise the fire	VC.CV	792	aldub	broken	VC.CVC
753	əʔna	fan verb	VC.CV	793	alkaj	quarrel	VC.CVC
754	eʔne	such as this	VC.CV	794	ambin	you PL	VC.CVC
755	ilda	evil spirit	VC.CV	795	amriʔ	not clear	VC.CVC
756	inɕa	no	VC.CV	796	andəm	very much	VC.CVC
757	iŋda	send	VC.CV	797	ansoj	alone	VC.CVC
758	iŋsa	handle of ploughing tool	VC.CV	798	andzəm	plenty	VC.CVC
759	iŋte	will go together	VC.CV	799	andziŋ	trapped legs	VC.CVC
760	iʔka	amusement	VC.CV	800	aŋgaj	month	VC.CVC
761	irse	taboo	VC.CV	801	aŋgaj	moon	VC.CVC
762	iʔba	thorn	VC.CV	802	aŋguŋ	falling	VC.CVC
763	iʔle	write	VC.CV	803	aŋgur	sort rice	VC.CVC
764	iʔto	will cut	VC.CV	804	aŋtəʔ	throw	VC.CVC
765	iʔɕa	no	VC.CV	805	appuŋ	say speak out	VC.CVC
766	obsa	remove	VC.CV	806	ardam	close by	VC.CVC
767	ojda	mix	VC.CV	807	argal	thirst	VC.CVC
768	omda	made free	VC.CV	808	arpaʔ	spread legs	VC.CVC
769	omdi	wait	VC.CV	809	arraŋ	mango pickle	VC.CVC
770	omro	kind of honey	VC.CV	810	arɕab	crowded	VC.CVC
771	onde	which one	VC.CV	811	aʔgaj	won't eat	VC.CVC
772	onri	husking rod	VC.CV	812	aʔgem	ignite fire	VC.CVC
773	onti	bird	VC.CV	813	aʔjil	press firmly	VC.CVC
774	oŋro	variety of honey	VC.CV	814	aʔmaŋ	before	VC.CVC
775	orda	ridge thrown while ploughing	VC.CV	815	aʔnap	chase cow	VC.CVC
776	orro	plough	VC.CV	816	aʔsun	let it be	VC.CVC
777	oʔda	line to plant	VC.CV	817	aʔɕaŋ	stuck in throat	VC.CVC
778	oʔla	ban dog	VC.CV	818	endziŋ	way of sitting	VC.CVC
779	oʔle	mix things	VC.CV	819	endzum	axe	VC.CVC
780	oʔɕo	make gesture	VC.CV	820	eŋɕiŋ	finger ring feet	VC.CVC

Table B.14: Assam Sora disyllables (12)

#	Sora	English	Syllable	#	Sora	English	Syllable
821	ərtuʔ	shake cloth	VC.CVC	861	oskaj	get ready	VC.CVC
822	eʔtaj	cut	VC.CVC	862	oʔnəj	not wanted	VC.CVC
823	əʔtin	shaking	VC.CVC	863	oʔseŋ	that side	VC.CVC
824	iddəb	tie on waist	VC.CVC	864	oʔtin	there	VC.CVC
825	ilɕab	hybrid	VC.CVC	865	ubdal	leftover food	VC.CVC
826	imtaj	remember	VC.CVC	866	ulley	grandson	VC.CVC
827	indir	lean	VC.CVC	867	umpeʔ	puff of air	VC.CVC
828	iŋtaj	i will go	VC.CVC	868	umreŋ	made free	VC.CVC
829	iŋtin	leaving	VC.CVC	869	umu	smoke	VC.CVC
830	irtub	angle of ploughing tool	VC.CVC	870	unruj	up root	VC.CVC
831	irɕab	jump	VC.CVC	871	uŋbaj	vomit	VC.CVC
832	iʔbəb	beheaded	VC.CVC	872	uŋlaj	stunned	VC.CVC
833	iʔdub	take something on waist	VC.CVC	873	uŋpəl	hot wind	VC.CVC
834	iʔlaj	contracted belly	VC.CVC	874	uŋtaj	astonished	VC.CVC
835	iʔmaʔ	winking	VC.CVC	875	uro	sneeze loudly	VC.CVC
836	iʔraŋ	wait	VC.CVC	876	urpiŋ	smell of goat	VC.CVC
837	iʔsal	reap spadix of liquor tree	VC.CVC	877	urraŋ	beak skin	VC.CVC
838	iʔsum	desire willing	VC.CVC	878	uʔdaj	climb	VC.CVC
839	iʔtaj	cut	VC.CVC	879	uʔmul	shade	VC.CVC
840	iʔtal	written on wall	VC.CVC	880	uʔpul	shadow	VC.CVC
841	iʔtam	frighten someone	VC.CVC	881	uʔsal	skin	VC.CVC
842	iʔtin	i am leaving	VC.CVC	882	uʔsun	while ago	VC.CVC
843	oblay	turn upside down	VC.CVC	883	eŋa	fly verb	VC.V
844	obsaʔ	leave it	VC.CVC	884	iŋa	go	VC.V
845	obseŋ	turn	VC.CVC	885	iʔi	louse	VC.V
846	obɕoŋ	measure	VC.CVC	886	uʔu	hair	VC.V
847	oɕlay	put fire	VC.CVC	887	aŋal	firewood	VC.VC
848	ombuŋ	lie down	VC.CVC	888	aŋan	when	VC.VC
849	omraŋ	horizontal	VC.CVC	889	aŋol	witchcraft	VC.VC
850	omtar	in that way	VC.CVC	890	aŋur	foul smell	VC.VC
851	ondeŋ	indian date	VC.CVC	891	aʔaj	deny	VC.VC
852	onder	finger/off shoot	VC.CVC	892	aʔam	arrow	VC.VC
853	onsir	help	VC.CVC	893	aʔeŋ	bird stomach	VC.VC
854	ontum	finished	VC.CVC	894	eŋer	gooseberry	VC.VC
855	ontuŋ	completed	VC.CVC	895	oŋal	evening	VC.VC
856	ontup	natural medicine	VC.CVC	896	oŋan	where	VC.VC
857	oŋɕər	fig tree	VC.CVC	897	oŋol	dusky	VC.VC
858	oŋger	male	VC.CVC	898	oŋoŋ	for worship	VC.VC
859	oŋgun	big rat	VC.CVC	899	oʔon	child	VC.VC
860	oŋkum	kind of rat	VC.CVC	900	uŋun	feeling hot	VC.VC

# Glossary

**All Assam Saora Samaj** The literary society of Assam Sora community. 13

**Arakati** A process of tea labour recruitment. 18

**Chalan** An emigration certificate given to indentured tea labourers of Assam. 24

**Koineisation** A process of language change reported in transplanted languages. 33

**Labour line** Labour quarters pattern in the tea gardens. 28

**Sadri** An Indo-Aryan pidgin used by the tea labourers in Assam as lingua-franca.  
41

**Sardar** A tea labour recruiting agent. 19

**Sardari** A process of tea labour recruitment. 18

**Segmental** Referring to phonetic segments like vowels, consonants and syllables. 11

**Telenga** A tea labour recruiting contractor. 25

**Transplanted language** A language that has migrated from one place to another.  
30

## References

- Abramson, A. S., & Lisker, L. (1965). Voice Onset Time in stop consonants: Acoustic analysis and synthesis. In *Proceedings of the 5th international congress of acoustics* (Vol. 51).
- Adank, P., Smits, R., & Van Hout, R. (2004). A comparison of vowel normalization procedures for language variation research. *The Journal of the Acoustical Society of America*, 116(5), 3099–3107.
- Anderson, G. D. (1999). A new classification of the Munda languages: Evidence from comparative verb morphology. In *209th meeting of the american oriental society, baltimore, md.*
- Anderson, G. D. (2001). A new classification of South Munda: Evidence from comparative verb morphology. *Indian linguistics*, 62(1-4), 21–36.
- Anderson, G. D. (2008a). Gta?. In G. D. Anderson (Ed.), *The munda languages* (pp. 682–763). London and New York: Routledge.
- Anderson, G. D. (Ed.). (2008b). *The Munda languages*. London and New York: Routledge.
- Anderson, G. D. (2014). Overview of the Munda languages. In J. Mathias & P. Sidwell (Eds.), *The handbook of austroasiatic languages* (Vol. 1, pp. 364–414). Leiden and Boston: Brill.
- Anderson, G. D. (2015). Prosody, phonological domains and the structure of roots, stems and words in the Munda languages in a comparative/historical light. *Journal of South Asian Languages and Linguistics*, 2(2), 163–183.
- Anderson, G. D., & Harrison, K. D. (2008a). Remo (Bonda). In G. D. Anderson (Ed.), *The munda languages* (pp. 557–632). London and New York: Routledge.
- Anderson, G. D., & Harrison, K. D. (2008b). Sora. In G. D. Anderson (Ed.), *The munda languages* (pp. 299–380). London and New York: Routledge.
- Anderson, G. D., & Harrison, K. D. (2011). *Sora talking dictionary*. Living Tongues Institute for Endangered Languages. Retrieved 2017-01-06, from <http://sora.talkingdictionary.org>

- Anderson, G. D., & Rau, F. (2008). Gorum. In G. D. Anderson (Ed.), *The munda languages* (pp. 381–433). London and New York: Routledge.
- Anderson, G. D., & Zide, N. H. (2002). Issues in Proto-Munda and Proto-Austroasiatic nominal derivation: The bimoraic constraint. *Papers from the Tenth Annual Meeting of the Southeast Asian Linguistics Society*, 55–74.
- Arsenault, P. (2017). Retroflexion in South Asia: Typological, genetic, and areal patterns. *Journal of South Asian Languages and Linguistics*, 4(1), 1–53.
- Avram, A. A. (2014). Immigrant workers and language formation: Gulf Pidgin Arabic. *Lengua y Migración*, 6(2), 7–40.
- Barz, R. K., & Siegel, J. (Eds.). (1988). *Language Transplanted: The development of overseas Hindi* (Vol. 1). Wiesbaden: Otto Harrassowitz Verlag.
- Becker-Kristal, R. (2010). *Acoustic typology of vowel inventories and dispersion theory: Insights from a large cross-linguistic corpus* (Unpublished doctoral dissertation). University of California, Los Angeles.
- Behal, R. P. (1985). Forms of labour protest in Assam valley tea plantations, 1900–1930. *Economic and Political Weekly*, 20(4), 19–26.
- Behal, R. P. (2014). *One hundred years of servitude: Political economy of tea plantations in colonial Assam*. New Delhi: Tulika Books.
- Bhatia, T. (1981). Transplanted South Asian languages: An overview in dimensions of South Asian linguistics. *Studies in the Linguistic Sciences Urbana, Ill.*, 11(2), 129–134.
- Bhattacharya, S. (1965). Glottal stop and checked consonants in Bonda. *Indo-Iranian Journal*, 9(1), 69–71.
- Bhattacharya, S. (1975). *Studies in comparative Munda linguistics*. Simla: Indian Institute of Advanced Study.
- Boersma, P., & Weenink, D. (1992). *Praat: A system for doing phonetics by computer*.
- Chatterjee, S., & Gupta, R. D. (1981). Tea-labour in Assam: Recruitment and government policy, 1840–80. *Economic and Political Weekly*, 16(44–46), 1861–1868.
- Cho, T., & Ladefoged, P. (1999). Variation and universals in VOT: Evidence from 18 languages. *Journal of phonetics*, 27(2), 207–229.
- Choi, H. (2002). Acoustic cues for the Korean stop contrast: Dialectal variation. In *Zas papers in linguistics* (p. 1–12).

- Coleman, D. A. (1997). Origins of multi-cultural societies and problems of their management under democracy. In *Proceedings of the 23rd international population conference* (Vol. 3, pp. 1457–1496). Beijing, China: Liege Belgium International Union for the Scientific Study of Population [IUSSP].
- Cox, F. (1996). *An acoustic study of vowel variation in Australian English* (Unpublished doctoral dissertation). Macquarie University, Australia.
- Dash, A. (2014). Tribal origin of the cult of the Jagannath. *Odisha Review*, 1–6.
- Davenport, M., Davenport, M., & Hannahs, S. (2010). *Introducing phonetics and phonology*. London and New York: Routledge.
- Diffloth, G. (2005). The contribution of linguistic palaeontology to the homeland of Austroasiatic. In L. Sagart, R. Blench, & A. Sanchez-Mazas (Eds.), *The peopling of east asia: Putting together archaeology, linguistics and genetics* (pp. 77–81). London and New York: Routledge Curzon.
- Diffloth, G., & Zide, N. (1992). Austroasiatic languages. *International Encyclopedia of Linguistics*, 1, 137–142.
- Domingue, N. (1981). Internal change in a transplanted language in dimensions of South Asian linguistics. *Studies in the Linguistic Sciences Urbana, Ill.*, 11(2), 151–159.
- Donegan, P. J. (1980). *Sora prosody*. Osmania University, Hyderabad, India. (Second International Conference on Austroasiatic Languages and Linguistics)
- Donegan, P. J. (1981). *The phonology of Sora vowels*. University of Chicago, Chicago, United States. (Conference on Munda Languages and Cultures)
- Donegan, P. J. (1993). Rhythm and vocalic drift in Munda and Mon-Khmer. *Linguistics of the Tibeto-Burman Area*, 16(1), 1–43.
- Donegan, P. J. (2009). *Essentials of sora phonology*. Mahidol University, Salaya, Thailand. (Fourth International Conference on Austroasiatic Linguistics)
- Donegan, P. J., & Stampe, D. (1983). Rhythm and the holistic organization of language structure. In J. F. Richardson, M. L. Marks, & A. Chukerman (Eds.), *Papers from the parasession on the interplay of phonology, morphology and syntax* (pp. 337–353). Chicago: Chicago Linguistic Society.
- Donegan, P. J., & Stampe, D. (2002). South-East Asian features in the Munda languages: Evidence for the analytic-to-synthetic drift of Munda. In *Annual meeting of the berkeley linguistics society* (Vol. 28, pp. 111–120).
- Donegan, P. J., & Stampe, D. (2004). Rhythm and the synthetic drift of Munda. In R. Singh (Ed.), *The yearbook of south asian languages and linguistics* (p. 3–36). Berlin, Germany: Muoton de Gruyter.

- Duarah, D. (2005). *Saora bhaxar aabhas*. Pahi Offset Press, Moranhat.
- Everson, M. (2009). *Proposal for encoding the Sora Sompeng script in the unicode standard*. Writing systems, computers and people. Retrieved 2017-01-06, from [http://scriptsource.org/cms/scripts/page.php?item\\_id=entry\\_detail&uid=r4c4bsfy7v](http://scriptsource.org/cms/scripts/page.php?item_id=entry_detail&uid=r4c4bsfy7v)
- Fry, D. B. (1955). Duration and intensity as physical correlates of linguistic stress. *The Journal of the Acoustical Society of America*, 27(4), 765–768.
- Gambhir, S. K. (1981). *The East Indian speech community in Guyana: A sociolinguistic study with special reference to koine formation* (Unpublished doctoral dissertation). University of Pennsylvania, Philadelphia.
- Ghosh, A. (2008). Santali. In G. D. Anderson (Ed.), *The munda languages* (pp. 11–98). London and New York: Routledge.
- Gogoi, J. K., & Goswami, H. (2004). Migration and its consequences for Assam with special reference to the illegal immigration from Bangladesh. *Assam Economic Journal*, 17, 1–10.
- Goswami, P. (1999). *Assam in the nineteenth century: Industrialisation & colonial penetration*. Guwahati: Spectrum.
- Grierson, G. A. (1906). *Linguistic survey of India* (Vol. 4). India: Office of the superintendent of government printing.
- Griffiths, A. (2008). Gutob. In G. D. Anderson (Ed.), *The munda languages* (pp. 633–81). London and New York: Routledge.
- Guha, A. (1977). *Planter-Raj to Swaraj: Freedom struggle and electoral politics in Assam 1826-1947*. Delhi: Indian Council of Historical Research.
- Gupta, H. L. (1959). An unknown factor in the annexation of Assam. In *Proceedings of the indian history congress* (Vol. 22, pp. 412–418).
- Halle, M., Hughes, G. W., & Radley, J.-P. (1957). Acoustic properties of stop consonants. *The Journal of the Acoustical Society of America*, 29(1), 107–116.
- Harris, K. S., Hoffman, H. S., Liberman, A. M., Delattre, P. C., & Cooper, F. S. (1958). Effect of third-formant transitions on the perception of the voiced stop consonants. *The Journal of the Acoustical Society of America*, 30(2), 122–126.
- Hillenbrand, J. M., & Houde, R. A. (1996). Role of f0 and amplitude in the perception of intervocalic glottal stops. *Journal of Speech, Language, and Hearing Research*, 39(6), 1182–1190.
- Horo, L., & Sarmah, P. (2015). Acoustic analysis of vowels in Assam Sora. *North East Indian Linguistics*, 7, 69–88.

- Jacewicz, E., Fox, R. A., & Salmons, J. (2007). Vowel space areas across dialects and gender. In *International congress of phonetic sciences* (Vol. 16, pp. 1465–1468).
- Jenny, M., Weber, T., & Weymuth, R. (2014). The Austroasiatic languages: A typological overview. In J. Mathias & P. Sidwell (Eds.), *The handbook of austroasiatic languages* (Vol. 1, pp. 13–143). Leiden and Boston: Brill.
- Johnson, K. (2008). *Quantitative methods in linguistics*. Malden, MA and Oxford: Blackwell.
- Kager, R. (2007). Feet and metrical stress. In P. d. Lacy (Ed.), *The cambridge handbook of phonology* (pp. 195–228). Cambridge: Cambridge University Press.
- Kalita, S., Horo, L., Sarmah, P., Prasanna, S. M., & Dandapat, S. (2016). Analysis of glottal stop in Assam Sora language. In *Interspeech* (pp. 1049–1053).
- Kalita, S., Lalhminghlui, W., Horo, L., Sarmah, P., Prasanna, S. M., & Dandapat, S. (2017). Acoustic characterization of word-final glottal stops in Mizo and Assam Sora. *Proc. Interspeech 2017*, 1039–1043.
- Kar, R. K. (1979). A migrant tribe in a tea plantation in India: Economic profile. *Anthropos Freiburg*, 74(5-6), 770–784.
- Kar, R. K. (1981). *Savaras of Mancotta*. New Delhi: Cosmo Publications.
- Kerswill, P. (2002). Koineization and accommodation. In J. K. Chambers, P. Trudgill, & N. Schilling-Estes (Eds.), *The handbook of language variation and change* (pp. 669–702). Oxford: Blackwell.
- Kerswill, P. (2006). Migration and language. *Sociolinguistics/Soziolinguistik. An international handbook of the science of language and society*, 3, 1–27.
- Kewley-Port, D. (1982). Measurement of formant transitions in naturally produced stop consonant–vowel syllables. *The Journal of the Acoustical Society of America*, 72(2), 379–389.
- Konow, S. (1908). Notes on the Munda family of speech in India. *Anthropos*, 3, 68–82.
- Lakshmi Narayana, K. (2013). *A study on the health care practices with emphasis on the maternal and child care among the Savara: A most vulnerable tribe of Andhra Pradesh* (Unpublished doctoral dissertation). Andhra University.
- Lee, H. Y. H. (2014). Losing Chinese as the first language in Thailand. *Asian Social Science*, 10(6), 176.
- Lengeris, A. (2016). Comparison of perception-production vowel spaces for speakers of Standard Modern Greek and two regional dialects. *The Journal of the Acoustical Society of America*, 140(4), 314–319.

- Liberman, A. M., Delattre, P. C., & Cooper, F. S. (1958). Some cues for the distinction between voiced and voiceless stops in initial position. *Language and speech*, 1(3), 153–167.
- Liberman, A. M., Delattre, P. C., Cooper, F. S., & Gerstman, L. J. (1954). The role of consonant-vowel transitions in the perception of the stop and nasal consonants. *Psychological Monographs: General and Applied*, 68(8), 1.
- Liljencrants, J., & Lindblom, B. (1972). Numerical simulation of vowel quality systems: The role of perceptual contrast. *Language*, 48(4), 839–862.
- Lindblom, B. (1986). Phonetic universals in vowel systems. In J. J. Ohala & J. J. Jaeger (Eds.), *Experimental phonology* (pp. 13–44). Orlando: FL: Academic Press.
- Lisker, L. (1978). In qualified defense of VOT. *Language and Speech*, 21(4), 375–383.
- Lisker, L., & Abramson, A. S. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *Word*, 20(3), 384–422.
- Lisker, L., & Abramson, A. S. (1965). Stop categorization and Voice Onset Time. In *5<sup>th</sup> international congress of phonetic science, munster* (pp. 389–391). Basel: Karger.
- Lisker, L., Liberman, A. M., Erickson, D. M., Dechovitz, D., & Mandler, R. (1977). On pushing the Voice-Onset-Time (VOT) boundary about. *Language and Speech*, 20(3), 209–216.
- Lobonov, B. (1971). Classification of Russian vowels spoken by different listeners. *Journal of the Acoustic Society of America*, 49(2B), 606–608.
- Mannell, R., Cox, F., & Harrington, J. (2009). *An introduction to phonetics and phonology*. Sydney: Macquarie University. Retrieved 2017-01-06, from <http://clas.mq.edu.au/speech/phonetics/index.html>
- Matisoff, J. A. (1973). Tonogenesis in Southeast Asia. In L. Hyman (Ed.), *Consonant types and tone* (Vol. 1, pp. 71–96). Southern California: Occasional Papers in Linguistics.
- Mesthrie, R. (1995). Language change, survival, decline: Indian languages in South Africa. In R. Mesthrie (Ed.), *Language and social history: Studies in south african sociolinguistics* (pp. 116–128). Cape Town: David Philip.
- Mohanty, A. (1996). Saora. In Sachchidananda & R. R. Prasad (Eds.), *Encyclopaedic profile of indian tribes* (Vol. 4, pp. 893–895). Delhi: Discovery Publishing House Pvt. Ltd.
- Mohanty, P. (1997). Loss of /o/ in Kui, Sora, and Oriya: A clue for a sub-linguistic area. *Languages of Tribal and Indigenous Peoples of India: The Ethnic Space*, 8, 109–119.

- Mufwene, S. S. (2007). Population movements and contacts in language evolution. *Journal of language contact*, 1(1), 63–92.
- Munda, R. D. (1968). Proto-Kherwarian phonemic system. *Unpublished MA thesis. Chicago University.*
- Munda, R. D. (1969). *Proto-Kherwarian phonology* (Unpublished doctoral dissertation). University of Chicago, Chicago.
- Obed, M. (2015). *Sora idolan kinolan do danisaman aharangna kambol*. Mariani: Concept Online.
- Ohala, J. J. (1997). Emergent stops. In *Proceedings of the 4th seoul international conference on linguistics* (pp. 84–91). Seoul: Linguistic Society of Korea.
- Osada, T. (1996). Notes on the Proto-Kherwarian vowel system. *Indo-Iranian Journal*, 39(3), 245–258.
- Pattanayak, D. (1985). Languages for the masses and Chinese languages in India. *Anthropological Linguistics*, 27(2), 222–227.
- Pinnow, H.-J., & Kuiper. (1963). The position of the Munda languages within the Austroasiatic language family. In H. L. Shorto (Ed.), *Linguistic comparison in southeast asia and the pacific* (pp. 140–152). London: School of Oriental and African Studies.
- Priyadarshini, S. S. (2016). “idital”-digitization and revival of tribal script, Sora Sompeng. In *Typoday* (Vol. 9, pp. 1–17). Bangalore: Srishti Institute of Art, Design and Technology.
- Ramamurti, R. S. G. V. (1914). *Telugu-Savara Dictionary*. Superintendent, Government Press, Madras: Published under the authority of the Director of Public Instruction.
- Ramamurti, R. S. G. V. (1931). *A manual of the Sora (or Savara) language*. Superintendent, Government Press, Madras: Published under the authority of the Director of Public Instruction.
- Ramamurti, R. S. G. V. (1933). *English-Sora Dictionary*. Superintendent, Government Press, Madras: Published under the authority of the Director of Public Instruction.
- Ramamurti, R. S. G. V. (1938). *Sora-English Dictionary*. Delhi: Mittal Publication.
- Ramyead, L. P. (1988). Hindi in Mauritius: A perspective. *Language Transplanted: The Development of Overseas Hindi*, 1, 23.
- Rao, M. V. K. (1996). Savara. In Sachchidananda & R. R. Prasad (Eds.), *Encyclopaedic profile of indian tribes* (Vol. 4, pp. 903–908). Delhi: Discovery Publishing House Pvt. Ltd.

- Raphael, L. J. (2005). Acoustic cues to the perception of segmental phonemes. In D. B. Pisoni & R. E. Remez (Eds.), *The handbook of speech perception* (pp. 182–206). Malden, MA: Blackwell Publishing Ltd.
- Rau, F. (2011). Notes on glottal constriction in Gorum. *Austroasiatic Studies, Papers from ICAAL 4, Mon-Khmer Studies Journal, Special Issue(2)*, 174–183.
- Registrar General of India. (2001). *Distribution of the 100 Non-Scheduled Languages – India, States & Union Territories – 2001 Census*. New Delhi: Office of the Registrar General and Census Commissioner, India.
- Reinecke, J. E. (1969). *Language and dialect in Hawaii, A sociolinguistic history to 1935*. Honolulu: University Press of Hawaii.
- Schmidt, P. W. (1906). *Die Mon-Khmer-Völker ein Bindeglied zwischen Völkern Zentralasiens und Australasiens*. Braunschweig: Friedrich Vieweg und Sohn.
- Schwartz, J.-L., Boë, L.-J., Vallée, N., & Abry, C. (1997a). The Dispersion-Focalization Theory of vowel systems. *Journal of phonetics*, 25(3), 255–286.
- Schwartz, J.-L., Boë, L.-J., Vallée, N., & Abry, C. (1997b). Major trends in vowel system inventories. *Journal of phonetics*, 25(3), 233–253.
- Sharma, B. (1992). *Forma education among the Savara: An ethnographic study* (Unpublished doctoral dissertation). University of Hyderabad, Hyderabad.
- Sharma, J. (2006). British science, Chinese skill and Assam tea making empire's garden. *Indian Economic & Social History Review*, 43(4), 429–455.
- Sharma, J. (2009). 'Lazy' natives, coolie labour, and the Assam tea industry. *Modern Asian Studies*, 43(06), 1287–1324.
- Sidwell, P. (2013). Issues in Austroasiatic classification. *Language and Linguistics Compass*, 7(8), 437–457.
- Sidwell, P. (2015). The Austroasiatic language phylum: A typology of phonological restructuring. In C. Bowern & B. Evans (Eds.), *Routledge handbook of historical linguistics* (p. 675-703). Abingdon: Routledge.
- Sidwell, P., & Rau, F. (2014). Austroasiatic comparative-historical reconstruction: An overview. In J. Mathias & P. Sidwell (Eds.), *The handbook of austroasiatic languages* (Vol. 1, pp. 221–363). Leiden and Boston: Brill.
- Siegel, J. (1985). Koines and Koineization. *Language in society*, 14(03), 357–378.
- Siegel, J. (1988). The development of Fiji Hindustani. In R. K. Barz & J. Siegel (Eds.), *Language transplanted: The development of overseas hindi* (pp. 121–49). Wiesbaden: Otto Harrassowitz.

- Siegel, J. (1998). Indian languages in Fiji: Past, present and future. *South Asia: Journal of South Asian Studies*, 21(s1), 181–214.
- Stampe, D. (1963). *Proto-Sora-Parengi Phonology* (Unpublished master's thesis). University of Chicago, Chicago.
- Stampe, D. (1965). Recent work in Munda linguistics I. *International journal of American linguistics*, 31(4), 332–341.
- Starosta, S. (1967). *Sora syntax: A generative approach to a Munda language* (Unpublished doctoral dissertation). University of Wisconsin, Madison.
- Starosta, S. (1976). Case forms and case relations in Sora. In P. Jenner, L. Thompson, & S. Starosta (Eds.), *Austroasiatic studies* (pp. 1069–1108). Honolulu: University Press of Hawaii.
- Starosta, S. (1992). Sora combining forms and pseudo-compounding. *Mon-Khmer Studies*, 18–19, 78–105.
- Stevens, K. N. (1972). The Quantal nature of speech: Evidence from articulatory-acoustic data. In E. E. David & D. P. B (Eds.), (pp. 51–66). New York: McGraw-Hill.
- Stevens, K. N. (1989). On the Quantal nature of speech. *Journal of phonetics*, 17(1), 3–45.
- Stevens, K. N. (2000). *Acoustic Phonetics*. London: MIT press.
- Trudgill, P. (1986). *Dialects in contact*. Oxford: Basil Blackwell.
- Trudgill, P. (2004). Linguistic and social typology. In J. K. Chambers, P. Trudgill, & N. Schilling-Estes (Eds.), *The handbook of language variation and change* (pp. 707–728). Oxford: Blackwell.
- Vitebsky, P. (1980). Birth, entity and responsibility: The spirit of the sun in Sora cosmology. *L'Homme*, 20(1), 47–70.
- Vitebsky, P. (1993). *Dialogues with the dead: The discussion of mortality among the Sora of eastern India*. Cambridge: Cambridge University Press.
- Vitebsky, P. (1998). A farewell to ancestors? Deforestation and the changing spiritual environment of the Sora. In R. Grove, V. Damodaran, & S. Sangwan (Eds.), *Nature and the orient* (p. 967-82). Delhi: Oxford University Press.
- Zide, A. R. (1976). Nominal combining forms in Sora and Gorum. In P. N. Jenner, L. C. Thompson, & S. Starosta (Eds.), *Austroasiatic studies* (p. 1259-1294). Honolulu: The University Press of Hawaii.
- Zide, A. R. (1982). *A reconstruction of Proto-Sora-Juray-Gorum Phonology* (Unpublished doctoral dissertation). University of Chicago, Chicago.

- Zide, N. (1999). Three Munda scripts. *Linguistics of the Tibeto-Burman Area*, 22(2), 199-232.
- Zide, N. H. (1965). Gutob-Remo vocalism and glottalised vowels in Proto-Munda. *Lingua*, 14, 43-53.
- Zide, N. H. (1966). Korku low tone and the Proto-Korku-Kherwarian vowel system. In N. H. Zide (Ed.), *Studies in comparative austroasiatic linguistics* (p. 214-1000). London: Mouton —& Co.
- Zide, N. H. (1969). Munda and Non-Munda Austroasiatic languages. *Current trends in Linguistics*, 5, 411-430.

