10th International Conference on Speech Prosody 2020
25-28 May 2020, Tokyo, Japan

Rhythm in Sora Trilingual Readers

Luke Horo[1], *Priyankoo Sarmah*[2,3]

[1]Living Tongues Institute for Endangered Languages
[2]Department of Humanities and Social Sciences
[3]Center for Linguistic Science and Technology
Indian Institute of Technology Guwahati

luke.horo@livingtongues.org, priyankoo@iitg.ac.in

Abstract

This work investigates the rhythm of read speech as produced by Sora speakers in Assam. Sora is an Austro-Asiatic language and not much is known about the speech rhythm of these languages. Many of the Sora speakers in Assam are trilingual, as they speak Sora, Sadri and Assamese. While Assamese is the dominant language in the area, Sadri is a lingua-franca used by Sora speakers to communicate with speakers of other languages. In this work, we investigated the difference in rhythm when the speakers of Sora read Assamese, Sadri and Sora texts. Conventional rhythm measures, such as %V, nPVI, r-PVI, varco-V , varco-C, ΔV and ΔC are calculated for the read speech. The results showed that read Sora speech tends to be more mora-timed. However, when the same speakers read Assamese text, their rhythm properties are neither Assamese-like nor Sora-like. Similarly, the rhythm in the Sora speakers speaking Sadri also is very distinct and does not reflect any L1 influence.

Index Terms: Sora, Sadri, Assamese, Rhythm, Austro-Asiatic, Munda

1. Introduction

Sora is a Munda language of the Austroasiatic language family spoken by approximately four hundred thousand Sora individuals in India [1]. While most Soras reside in Odisha of eastern India, a diasporic Sora population is found in Assam of north-east India. This work is based on the variety of Sora spoken in Assam only. The Soras in Assam are generally bi-lingual or tri-lingual. Their speech repertoire includes Sora as their first language (L1) and two Indo-Aryan languages namely, Sadri and Assamese as their second (L2) and third language (L3). Sadri is a language of wider communication among tea garden workers of Assam of which the Soras are an integral part. The Soras use Sadri to communicate with non-Sora tea garden workers residing in their neighbourhood. On the other hand, Assamese is the official language of Assam and the Soras can speak in Assamese only if they have a formal education. This complex linguistic situation has resulted in many Soras being trilingual. Hence, it is interesting to investigate how Sora trilingual speakers negotiate their prosodic features, such as rhythm, while speaking the three languages they are proficient in.

Traditionally speech rhythm has been categorized as stress-timed, syllable-timed and mora-timed [2, 3, 4, 5, 6]. At the same time studies based on acoustic evidence have claimed that rhythm classes are not categorical but rather a continuum [7, 8, 9]. The instrumental studies on rhythm have considered isochrony as one of the basis for measuring rhythm in languages. It is claimed that the duration of every syllable is equal in a syllable-timed language. This also assumes that there will be less variability in segment duration in these languages and specifically vowel duration need to be adjusted to fit in a particular stress type. In case of the mora-timed languages, the duration of every mora is claimed to be equal (mora-timed). On the other hand, in case of a stress-timed language, the temporal duration between two stressed syllables is to be kept equal, which requires the vowel durations in syllables to be adjust to accommodate a particular stress type. The stress-timed languages have greater variety of syllable types than syllable-timed languages and exhibit vowel reduction in unstressed syllables [8].

Based on these observations, temporal measures like ΔC (standard deviation of consonantal intervals), ΔV (standard deviation of vocalic intervals) and %V (percentage of vocalic intervals in an utterance) are measured to capture rhythm type derived differences in languages [10]. However, it was also noticed that speaking rate can affect measures like %V, ΔV and ΔC in varying degrees and hence, a rhythm measure based on Pairwise Variability Index (PVI) was proposed to minimize the effect by measuring the difference in duration between each pair of successive vocalic or intervocalic intervals [11]. Similarly, VarcoΔC and VarcoΔV measures were proposed as they minimize the effect of speech rate on ΔC and ΔC, respectively [12].

It is argued that stress-timed languages allow a greater variety of consonant clusters, they were found to have a higher ΔC, whereas %V is lower due to vowel reduction. However, ΔV can not be interpreted clearly because it can be affected by a variety of language specific or contextual factors. Among these three absolute temporal measures, the combination of %V and ΔC was found to be the best way of distinguishing rhythm classes [10]. Though it was found that ΔC is highly affected by speech rate and %V remains rather stable, the results still corresponded to the claim that languages of the two different rhythm classes are distinguishable by %V and ΔC [10]. It is also to be noted that there are also claims that the rhythm metrics proposed so far in the literature cannot satisfactorily classify languages into separate rhythmic classes [13].

Notwithstanding with the criticisms meted towards rhythm metrics, there is a body of literature that have also investigated rhythm measures in second languages (L2). While these studies have noticed first language (L1) influence on L2 rhythm, most of the times rhythm is noticed to be intermediate between the speakers' L1 and L2 [14]. Considering this, the current study is important as it attempts to study rhythm in speech of the native speakers of Sora while reading texts written in their L1 (Sora), L2 (Sadri) and L3 (Assamese). The remainder of this paper is organized as follows: Section 2 describes the participants, material and measurements considered in this study. Section 3 reports the results for the rate of articulation, correlation between rate of articulation, utterance length and rhythm measures. Finally, Section 4 discusses the results and concludes the study.

10.21437/SpeechProsody.2020-119

2. Methodology

2.1. Participants

Five Sora men living in Singrijhan Tea Estate of Sonitpur District in Assam participated in the study. Their average age is twenty four years and standard deviation of their age is two years. They all have completed high school and they speak Sora as their first language, Sadri as second language and Assamese as third language. The participants regularly use the three languages in different domains and are also proficient in reading texts written in Assamese script.

2.2. Material

Data used in this work includes read speech of the passage 'North Wind and the Sun' [15] translated to Sora, Sadri and Assamese. The translations are done by native speakers of each language. The Assamese translation is done in the standard Assamese script and due to the lack of a standardized orthography, Sora and Sadri translations are also done in Assamese script. Prior to conducting the reading tasks, the participants were given time to familiarize themselves with the passage and were also allowed to rehearse a few times. The details of the paragraphs are provided in Table1.

2.3. Measurements

After data recording was complete, the speech data was annotated at phoneme level by the first author using Praat and saved as a TextGrid file [16]. Using Praat, a script was run to extract the phoneme labels and the time at the phoneme boundaries from the TextGrid file. Duration for each phoneme was saved to a spreadsheet. Subsequently, a Perl script was run to identify successive vocalic and intervocalic units in the spreadsheet and rhythm measures, namely, %V, rPVI-C, nPVI-V, ΔC, ΔV, VarcoΔC and VarcoΔC were automatically calculated and saved in another file as comma separated values (csv). Descriptive and investigative statistics were run on the csv file using R [17]. While plots are made with *ggplot2* package, statistical analyses were conducted using *car, emmeans* and *lme4* packages [18, 19, 20].

3. Results

3.1. Rate of articulation in reading

It has been noticed that several rhythm measures are highly correlated with the rate of articulation. Hence, we decided to measure the rate of articulation of the the Assamese, Sadri and Sora read speech of the speakers in this study. Rate of articulation was calculated by dividing the duration of a breath group by the number of phonemes produced in the same breath group. As seen in Figure 1, the rate of articulation for Sora speakers in reading Sora passages is slightly higher than their reading Assamese and Sadri passages.

In order to confirm these findings we modelled rate of articulation with a Linear Mixed Effects model where languages of the read passages were considered fixed effect and speaker was considered random effect. The models were built on R with the *lme4* package[17, 20]. The models were subjected to a Wald χ^2 test for analysis of deviance using the *car* package [18]. Additionally the significance of contrast in articulation rate between pair of language was estimated by using a post-hoc Bonferroni test using the *emmeans* package on R[19]. The Wald χ^2 test conducted on the LME model showed a significant effect of language on rate of articulation [$\chi^2 = 41.2, p < 0.001$]. A post-hoc Bonferroni test showed that, the rate of articulation in reading Sora was significantly greater than reading Assamese and Sadri as seen in Table 2.

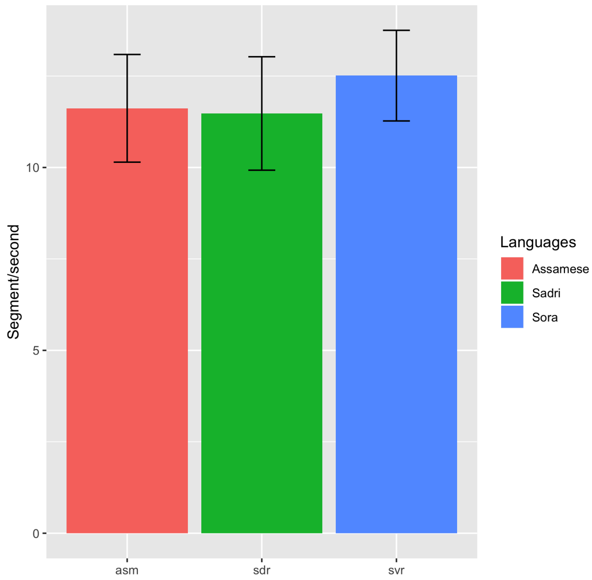

Figure 1: *Rate of articulation of Sora speakers*

Table 2: *Contrasts of articulation rate*

Contrasts	Standard Error	df	p-value
Assamese - Sadri	0.170	267	0.3670
Assamese - Sora	0.170	267	< 0.001
Sadri - Sora	0.158	267	< 0.0001

3.2. Correlation of rhythm measures

It has been shown in the literature that several rhythm measures are either directly or indirectly proportionate to the rate of speaking. Similarly, utterance length is also said to be one of the factors rhythm measures are sensitive to. Apart from that it has also been shown that such effects on rhythm measures vary according to the language investigated. Hence, Pearson correlation was measured for each of the passages separately as seen in Figure 2, Figure 3 and Figure 4. As seen in the figures, the correlation for the measures are different according to the language. In case of their L1, ΔV, ΔC and rPVI-C are negatively affected by the rate of articulation as measured by segment/second. On the other hand, length of utterance has a strong negative correlation with Varco-ΔC. In case of L2, Sadri, rate of articulation has strong negative correlation to ΔV, ΔC and rPVI-C. Apart from that in case of Sadri rate of articulation shows a strong negative correlation with %V. Varco-ΔC shows a strong negative correlation with the length of utterance. In case of their L3, Assamese, the effect of rate of articulation is prominent in all seven rhythm measures considered in this study. As seen in Figure 4, segment/second is negatively correlated with all the rhythm measures. However, in case of the effect of the length of utterance, as with L1 and L2, Varco-ΔC shows a negative correlation.

Table 1: *Number of words and syllables with types in the three passages*

Languages	Words	Syllables	1σ	2σ	3σ	4σ	5σ	CV	CVC	CVCC	CCV	V	VC
Sora	85	243	2	31	33	15	4	117	84	0	1	23	18
Sadri	99	222	3	71	23	2	0	108	83	1	0	16	14
Assamese	103	242	10	54	33	5	1	185	34	0	0	12	4

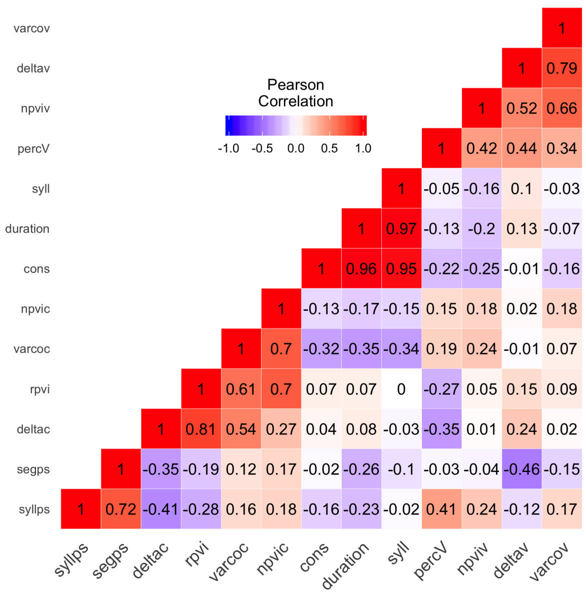

Figure 2: *Pearson correlation matrix of measures for Sora passage*

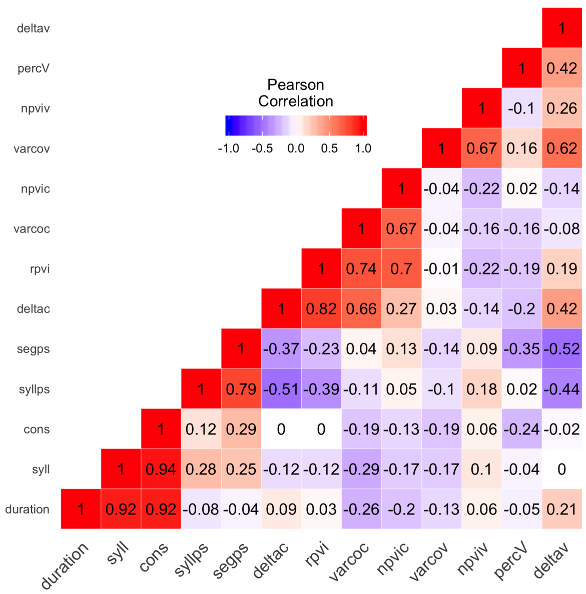

Figure 3: *Pearson correlation matrix of measures for Sadri passage*

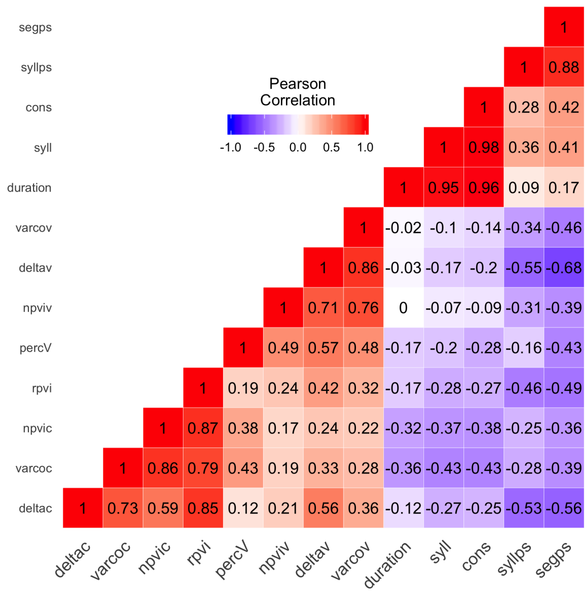

Figure 4: *Pearson correlation matrix of measures for Assamese passage*

3.3. Rhythm measures

The rhythm measures of the Sora speakers are presented in Table 3, along with results for Assamese L1 speakers, British English (stess-timed), Spanish (syllable-timed) and Japanese (mora-timed). As far as %V and nPVI-V values are concerned, Sora L1 values are high on both measures. Hence, it is difficult to categorize native Sora read speech into one of the two categories.These results are also plotted on an x-y plot on Figure 5 and Figure 6. The two figures show that for Sora speakers, the L1 rhythm is between the prototypical stress-timed and syllable-timed languages. While it is close to Japanese, in terms of nPVI it is farther away from Japanese. While Sora speakers' production of Sadri shows more similarity to mora-timed rhythm, their production of Assamese as L3 is more syllable-timed. The %V values for Assamese are even more than the Spanish speakers.

In order to see the reliability of the measures and to see if language has an effect on the rhythm measures obtained for the Sora speakers, we constructed LME models with each of the seven rhythm measures, with language (L1, L2 or L3) as fixed factor, and speaker as random factor. On each model, a Wald χ^2 test was conducted to obtained p-values for significance testing. The results of the χ^2 tests conducted on the LME models are provided in Table 4. Each model was also subjected to comparison by means of a Bonferroni post-hoc test for contrast. Among the seven measures, only %V and ΔC showed significance difference among all three languages.

Table 3: *Rhythm measures of Sora speakers with other L1 speakers*

Languages	%V	nPVI-V	r-PVI	ΔV	ΔC	Varco-V	Varco-C
Sora	50.8	50.6	70.4	55.0	63.5	55.8	64.9
Sadri	44.3	42.8	75.8	53.4	75.8	52.9	59.4
Assamese (Sora)	55.3	47.4	55.1	56.4	51.5	51.7	60.5
Assamese (Jorhat) [21]	48.1	48.9	43.4	41.4	45.9	53.8	54.6
British English [11, 22]	41.1	57.2	64.1	46.6	56.7	64.0	47.0
Spanish [11, 22]	50.8	29.7	57.7	20.7	47.5	41.0	46.0
Japanese [11, 14]	45.5	40.9	62.5	53.0	55.5	56.0	-

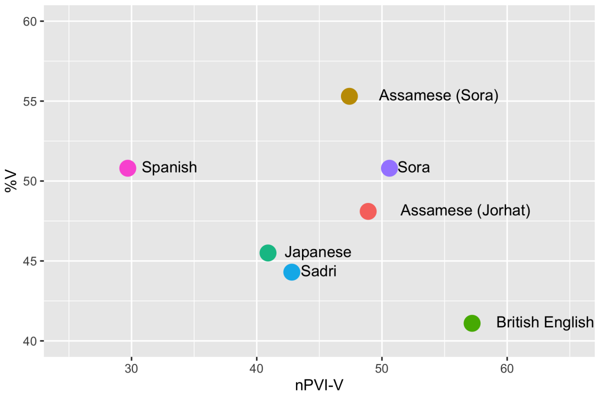

Figure 5: *%V and nPVI-V for Sora L1, L2 and L3 speech with other L1s*

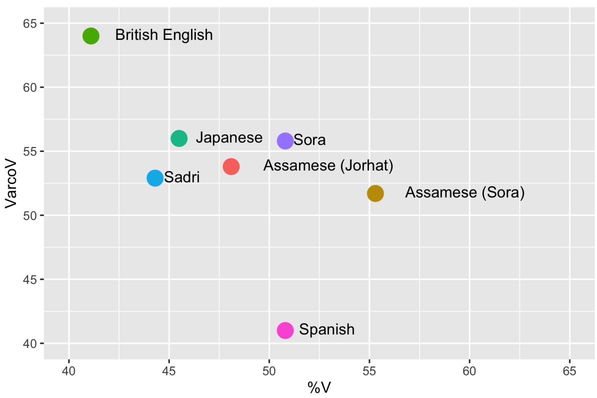

Figure 6: *%V and VarcoΔV for Sora L1, L2 and L3 speech with other L1s*

Table 4: *Wald χ^2 tests on LME models for rhythm measures*

Measure	χ^2	**p-value**
%V	187.85	< 0.001
ΔV	0.51	0.774
ΔC	58.40	< 0.001
Varco-ΔV	6.19	< 0.05
Varco-ΔC	11.36	< 0.01
nPVI-V	12.37	< 0.01
rpvi-C	32.36	< 0.001

4. Conclusions

The results of this study have several implications. Firstly, we concluded that there is a clear difference in the articulation rate in terms of the language of the text provided. While the rate of articulation was fastest and significantly faster for L1 (Sora), there was no difference in the rate of articulation between the Assamese and the Sadri text. However, when asked about their reading habits, the Sora speakers informed that they are more used to reading Assamese than reading Sora or Sadri. Hence, it is clear that in spite of not being used to written Sora text, L1 proficiency helped the Sora speakers to read the text significantly better.

In case of the effect of rate of articulation and length of utterance on the rhythm measures, it is clear that it has an effect on Varco-ΔC, which seems language independent effect. However, in case of the effect of rate of articulation on rhythm measures, we see a distinct pattern for L1 and L2 vs. L3. When Sora speakers read L1 (Sora) and L2 (Sadri) texts, the effect of rate of articulation is comparable. While reading L1 and L2 texts, rate of articulation has negative correlation with ΔV, ΔC and rPVI-C. On the other hand, in case of L3, Assamese, all seven rhythm measures are negatively correlated with the rate of articulation. This correlation may be a result of Sora speakers being too familiar with reading Assamese text. As a result of their familiarity with reading Assamese, the speakers were probably reading Assamese offhand, without any conscious effort. As a result of that rhythm manifested was merely a function of rate of articulation in reading Assamese text. While reading in all three languages did not lack naturalness, L2 and L3 readings did sound highly L1 accented. However, such accent was probably only restricted to segmental influences.

The distribution of %V and nPVI-V measures on an x-y plane shows that none of the L1, L2 or L3 can be categorized into one of the rhythm categories. Even though, Sadri values are very similar to that of Japanese. L1, Sora, tends to be more like a syllable timed language such as Spanish. However, it also has high n-PVI-V values not quite close to syllable-timed languages. It is to be noted that earlier works on Sora has indicated that Sora words are primarily disyllabic and there is a significant prominence on the second syllables of Sora, rendering the initial syllables short with the vowels reduced [23, 24, 25]. As seen in Table 1, the current data collected for Sora has mostly polysyllabic words where there is a possibility of vowel reduction, both in terms of temporal and spectral features. Hence, the higher nPVI-V for Sora is not unexpected. While we have found systematic differences in rhythm of Sora L1, L2 and L3 reading, as in %V and ΔC measures, categorizing them to a strict rhythm category is difficult. More evidence will possibly come from natural speech of the three languages and newer rhythm measurement techniques.

5. References

[1] Registrar General of India, *Distribution of the 100 Non-Scheduled Languages – India, States & Union Territories – 2011 Census*. New Delhi: Office of the Registrar General and Census Commissioner, India, 2011.

[2] K. L. Pike, *The intonation of American English*. ERIC, 1945.

[3] D. Abercrombie *et al.*, *Elements of general phonetics*. Edinburgh University Press Edinburgh, 1967, vol. 203.

[4] M. Han, "The feature of duration in Japanese," *Onsei no kenkyu*, vol. 10, pp. 65–80, 1962.

[5] R. F. Port, S. Al-Ani, and S. Maeda, "Temporal compensation and universal phonetics," *Phonetica*, vol. 37, no. 4, pp. 235–252, 1980.

[6] R. F. Port, J. Dalby, and M. O'Dell, "Evidence for mora timing in Japanese," *The Journal of the Acoustical Society of America*, vol. 81, no. 5, pp. 1574–1585, 1987.

[7] P. Roach, "On the distinction between 'stress-timed'and 'syllable-timed' languages," *Linguistic controversies*, pp. 73–79, 1982.

[8] R. M. Dauer, "Stress-timing and syllable-timing reanalyzed." *Journal of phonetics*, 1983.

[9] R. Hamdi, M. Barkat-Defradas, E. Ferragne, and F. Pellegrino, "Speech timing and rhythmic structure in Arabic dialects: a comparison of two approaches," in *Proceedings of Interspeech*, vol. 4, 2004.

[10] F. Ramus, M. Nespor, and J. Mehler, "Correlates of linguistic rhythm in the speech signal," *Cognition*, vol. 73, no. 3, pp. 265–292, 1999.

[11] E. Grabe and E. L. Low, "Durational variability in speech and the rhythm class hypothesis," *Papers in laboratory phonology*, vol. 7, no. 515-546, 2002.

[12] V. Dellwo, *Language and language-processing)*. Frankfurt am Main: Peter Lang, 2006, ch. Rhythm and Speech Rate: A Variation Coefficient for deltaC, pp. 231–241.

[13] A. Arvaniti, "The usefulness of metrics in the quantification of speech rhythm," *Journal of Phonetics*, vol. 40, no. 3, pp. 351–373, 2012.

[14] I. Grenon and L. White, "Acquiring rhythm: A comparison of l1 and l2 speakers of canadian english and japanese," in *Proceedings of the 32nd Boston University conference on language development*. Citeseer, 2008, pp. 155–166.

[15] I. P. Association, I. P. A. Staff *et al.*, *Handbook of the International Phonetic Association: A guide to the use of the International Phonetic Alphabet*. Cambridge University Press, 1999.

[16] P. Boersma *et al.*, "Praat, a system for doing phonetics by computer," *Glot international*, vol. 5, 2002.

[17] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2019. [Online]. Available: https://www.R-project.org/

[18] J. Fox and S. Weisberg, *An R Companion to Applied Regression*, 3rd ed. Thousand Oaks CA: Sage, 2019. [Online]. Available: https://socialsciences.mcmaster.ca/jfox/Books/Companion/

[19] R. Lenth, *emmeans: Estimated Marginal Means, aka Least-Squares Means*, 2019, r package version 1.4.1. [Online]. Available: https://CRAN.R-project.org/package=emmeans

[20] D. Bates, M. Mächler, B. Bolker, and S. Walker, "Fitting linear mixed-effects models using lme4," *Journal of Statistical Software*, vol. 67, no. 1, pp. 1–48, 2015.

[21] L. Dihingia and P. Sarmah, "Rhythm and speaking rate in Assamese varieties," submitted.

[22] L. White and S. L. Mattys, "Rhythmic typology and variation in first and second languages," *AMSTERDAM STUDIES IN THE THEORY AND HISTORY OF LINGUISTIC SCIENCE SERIES 4*, vol. 282, p. 237, 2007.

[23] L. Horo and P. Sarmah, "Acoustic analysis of vowels in Assam Sora," *North East Indian Linguistics*, vol. 7, pp. 69–88, 2015.

[24] L. Horo, "A phonetic description of assam sora," Ph.D. Dissertation, Indian Institute of Technology Guwahati, Guwahati, Assam, 2017.

[25] L. Horo and P. Sarmah, "A synchronic comparison of orissa sora and assam sora," 2017, international Seminar on Munda Linguistics.