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

Rhythm and Speaking Rate in Assamese Varieties

Leena Dihingia[1], *Priyankoo Sarmah*[1,2]

[1]Department of Humanities and Social Sciences
[2]Center for Linguistic Science and Technology
Indian Institute of Technology Guwahati, INDIA
{l.dihingia,priyankoo}@iitg.ac.in

Abstract

This work investigates the rhythm and speech rate in Assamese read speech as spoken in five geographic regions within the state of Assam in India. These five areas are categorized into three dialectal areas. For rhythm %V, nPVI, r-PVI, varco-V , varco-C, ΔV and ΔC measures are calculated. For rate of speaking syllable per second and segments per second are calculated. The results obtained show that the rhythm measures of Assamese are comparable to that of the mora-timed languages, such as Japanese. The results also showed that the five regions differed significantly from each other in terms of rhythm measures and rate of speaking. A quadratic discriminant analysis showed that the Assamese spoken in the five regions can be discriminated with an accuracy of about 42% with rhythm measures and speaking rate. A tertiary finding also showed that %V and nPVI-V are the least affected measure by rate of speaking.

Index Terms: Assamese, rhythm, rate of speaking, dialects, quadratic discriminant analysis

1. Introduction

Assamese is an Indo-European language spoken by about 15 million people in the province of Assam in India. The language has been broadly divided into three main dialect groups: Eastern Assamese (EA), Central Assamese (CA) and Western Assamese (WA), with each division having several sub-dialects [1, 2, 3]. While Assamese is spoken by a large number of speakers, there are not many works that discuss the phonetics of prosody in Assamese. Hence, there is next to no work on Assamese speakers' rhythm in speaking Assamese. World's languages are said to fall into three rhythm categories, namely, syllable-timed, stress-timed [4, 5] and mora-timed [6, 7, 8], based on how speech units are organized in time. Assamese has been reported to be a syllable timed language, based purely on auditory impression [9]. However, another quantitative study with a single Assamese speaker claims that Assamese rhythm is mora-timed [10].

Studies based on acoustic evidence have claimed that rhythm classes are not categorical but a continuum [11, 12, 13]. However, it was observed that languages perceived as stress-timed, have greater variety of syllable types than syllable-timed languages and exhibit vowel reduction in unstressed syllables [12]. 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 [14]. Since stress-timed languages can 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 [14]. When %V and ΔC are plotted on an x-y plane, the stressed-timed and syllable-timed languages cluster into different groups. It was seen that speaking rate can affect measures like %V, ΔV and ΔC in varying degrees and thus are not always effective in differentiating rhythm classes [15]. Hence, rhythm measures based on Pairwise Variability Index (PVI) was proposed to minimize the effect of speaking rate [15]. This approach categorizes languages based on durational variability of successive units of speech and can represent both rate normalized or raw values. Similarly, another measure, VarcoΔC, was proposed to minimize the effect of speech rate on ΔC, as it measures relative variation of consonantal intervals instead of absolute variation. The effect of speaking rate on the vocalic measure %V and intervocalic measure ΔC was analyzed by [16]. 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 [14].

Studies have shown that successful identification and classification of languages and their varieties, based on rate of speaking and rhythm, is possible. For example, significant speaking rate differences have been observed between varieties of American English [17, 18]; New Zealand English and American English [19] and Dutch spoken in the Netherlands and Belgium [20]. Also, rhythm has been found to be different for different varieties of spoken Arabic [13, 21, 22]. Using these phonetic correlates, dialect identification was also attempted [21]. Even though rhythm has been found to be distinctive for many languages there are claims that it is highly affected by the rate of speaking, depending on the way it is measured [15]. On the other hand there are also claims that the rhythm metrics proposed so far in the literature cannot actually classify languages into separate rhythmic classes [23].

Considering the discussions and limitations pointed out regarding the usefulness of the rhythm metrics, it is important that the measurements are also conducted on languages hitherto unanalyzed for rhythm. Hence, in this work, we provide rhythm metrics for Assamese, as it is spoken in five different geographical regions in the Assam province of India. These five regions are Tinsukia and Jorhat, belonging to EA; Nagaon, belonging to CA; and Nalbari and Barpeta, belonging to WA. We also examine the correlation among rhythm metrics, rate of articulation and the number of syllables in a measured utterance. Finally, we conduct a quadratic discriminant analysis with the rhythm metrics and articulation rate to determine the accuracy in classifying Assamese spoken in five different geographic areas, comprised of three dialect regions.

 10.21437/SpeechProsody.2020-115

2. Methodology

In order to compare the articulation rate and rhythm of Assamese spoken by speakers in five areas of Assam, namely, Tinsukia, Jorhat, Nagaon, Nalbari and Barpeta, fifty native speakers of Assamese were recorded reading a passage of Assamese text in a noiseless environment. The recording was done using a Shure unidirectional head-worn microphone connected to a Tascam linear PCM recorder via xlr jack. The sampling frequency was kept at 44.1 kHz at 24 bit in .wav format. The data recorded was segmented phoneme-wise in Praat 6.1.06[24]. Only breath groups with at least seven syllables and without any pause or hesitation were annotated and analyzed. Initially, the breath groups were manually segmented at phoneme level using Praat[24]. A praat script was used to extract the duration of each phoneme onto a spreadsheet. Speech rate was calculated with this data. A perl script was used to extract the duration of vocalic and non-duration intervals from the spreadsheet and calculate rhythm metrics.

2.1. Speakers

Speech data was collected from fifty native speakers of Assamese from the five areas. From each area, a total of ten (five female and five male, except Nalbari which had four female and six male) speakers were recorded reading an Assamese passage. Before collecting the data it was ensured that the speakers were proficient in reading Assamese. They also had knowledge of Hindi and English in varying degrees.

2.2. Materials

An Assamese translation of the *North Wind and the Sun* passage was provided to the speakers for reading. The passage was in Assamese script and comprised of 103 words, 252 syllables and 447 phonemes. The types of syllables in the passage is listed in Table 1. Out of the 50 speakers, 16 speakers read the passage twice.

Table 1: *Syllable types in the read passage*

V	CV	CVC	VC
10	205	33	4

2.3. Measurements

In order to ensure that pauses and fillers do not affect the measurements, only breath groups with at least seven continuous syllables were considered for analysis. For speaking rate, syllable per second and segments per second measures were calculated [17]. In total, 855 breath groups were analyzed. The total number of syllables considered for measurement is 11151. Region-wise distribution of number of syllables in provided in Table 2.

Table 2: *Total number of syllables analyzed*

Dialect region	Region	No. of syllables
EA	Tinsukia	1689
EA	Jorhat	2687
CA	Nagaon	1896
WA	Nalbari	3088
WA	Barpeta	1791
	Total	11151

To analyze rhythm, four interval measures namely, %V, ΔC, ΔV, VarcoΔV and VarcoΔC were calculated. %V is the percentage of vocalic intervals in an utterance. ΔC is the standard deviation of the duration of intervocalic intervals in a breath group and ΔV is the standard deviation of the duration of vocalic intervals in a breath group [14]. VarcoΔC is defined as the percentage of standard deviation of consonantal interval duration (ΔC) of the average duration of consonantal intervals (mean C) [25]. Similarly, VarcoΔV is calculated from the vocalic intervals.

Two Pairwise Variability Index measures, nPVI-V and rPVI are evaluated following [15]. nPVI-V is the rate normalized measures of the durational variation of two consecutive vocalic intervals and rPVI is the raw measure of the durational variation of two consecutive consonantal intervals.

In the following section, the results obtained for rate of speaking and rhythm for Assamese spoken in five regions are discussed. Apart from that, a co-variation matrix obtained by comparing the rate of speaking and rhythm measures using Pearson's correlation method is presented. The following section also reports the results of a quadratic discriminant analysis, conducted to see the effectiveness of articulation rate and rhythm in discriminating Assamese spoken in five different geographic regions. Finally a series of statistical tests are conducted to estimate the effect of different geographical regions on the articulation rate and rhythm measures obtained.

3. Results

3.1. Correlation matrix

It has been pointed out that rhythm measures are influenced by the number of syllables in the measured utterances. Similarly, it has also been noticed that rate of speaking affects some of the rhythm measures. In order to see how utterance length, speaking rate and rhythm interact with each other, we conducted a Pearson's correlation test. The results obtained are presented in form of a matrix in Figure1. As seen in Figure 1, ΔC, rPVI, ΔV and Varco-ΔV are strongly correlated with the rate of articulation (as indicated by Sylps and Segps). As the rate of speaking increases, these measures are lowered. Length of breath group (as indicated by nv, number of syllables in a breath group) is strongly correlated to Varco-ΔC, ΔC, Varco-ΔV and ΔV. As the length of an utterance increases, Varco-ΔC, ΔC, Varco-ΔV and ΔV decrease. Considering these results we can conclude that, overall, nPVI-V and %V are least affected by both rate of articulation and length of utterance. Interestingly, while Varco-ΔC is least affected by rate of articulation, it is affected by the length of utterance.

3.2. Rate of Articulation

The rate of articulation of speakers of Assamese from five regions of Assam are presented in Figure 2. As seen in the figure, there is a noticeable difference in the rate of articulation by regions. Assuming articulation rate difference by region, we fitted a couple of Linear Mixed Effect (LME) models using the *lme4* package on R [26, 27]. In both models, syllable/second and segment/second were the dependent variables and region was the fixed factor. Speaker and gender were considered random variables. After the models were built, Wald chisquare tests were conducted to see the effect of region on rate of articulation. The Wald chisquare tests conducted on both measures of articulation rate yielded non-significant p-values, (χ^2=6.4, p = 0.17, for segment/ second; χ^2=5.7, p = 0.22, for syllable/ sec-

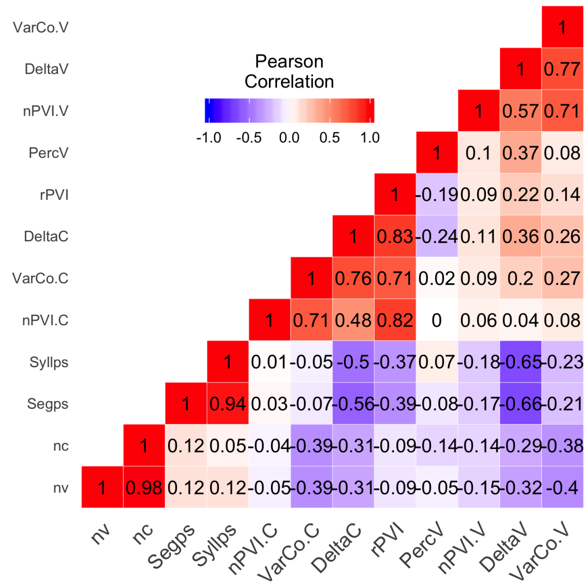

Figure 1: *Pearson's correlation matrix*

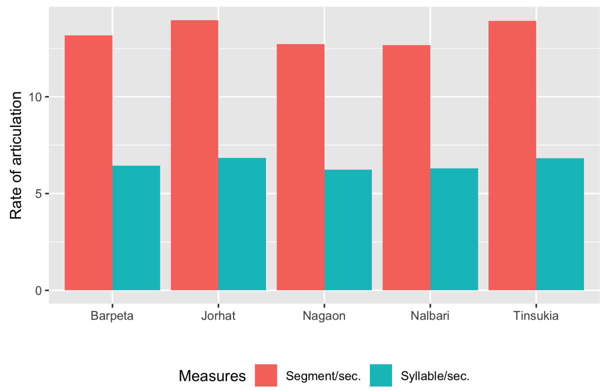

Figure 2: *Rate of articulation in five Assamese regions*

ond) indicating a weak correlation between region and rate of articulation.

Subsequently, the five regions were subsumed under the three dialect areas they are said to belong to, and again a couple of LME models were built to see if dialect areas differ significantly in terms of their rate of articulation. Dialect area was considered fixed factor and speaker and gender were considered random factors. As expected, dialect areas showed no significant interaction with rate of articulation. For dialect areas, Wald chisquare tests yielded non-significant p-values for segment/second (χ^2=5.5, p = 0.06) and syllable/second (χ^2=5.4, p = 0.07).

3.3. Rhythm metrics

The rhythm measurements obtained for Assamese spoken in five regions of Assam are provided in Table 3. The table also presents rhythm measures of Japanese, English and Spanish for comparison, obtained from previous observations [15, 28, 29].

We observed in Section 3.1 that in case of Assamese, all measures, except %V and nPVI-V, are significantly affected by rate of articulation and length of utterance. Hence, we initially plotted the %V and nPVI-V values obtained from Assamese speakers, separated by regions. In order to compare them with prototypical syllable-timed, stress-timed and mora-timed languages, we plotted the measures from Assamese speakers along with that of British English, Spanish and Japanese for comparison, obtained from previous observations [15, 28, 29]. Figure 3 shows the distribution of languages on a %V and nPVI-V plane. As observed from the figure, the Assamese regions get clustered closer to Japanese.

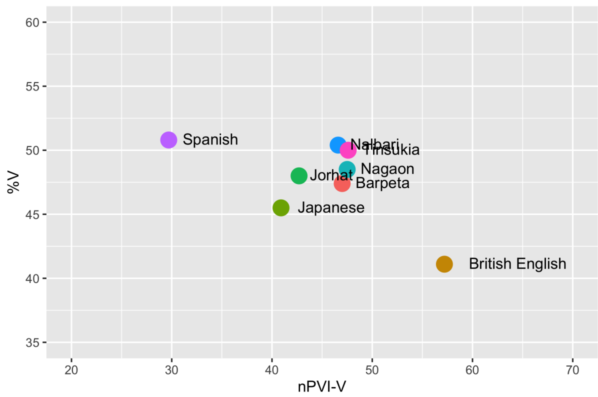

Figure 3: *nPVI-V and %V in Assamese speakers*

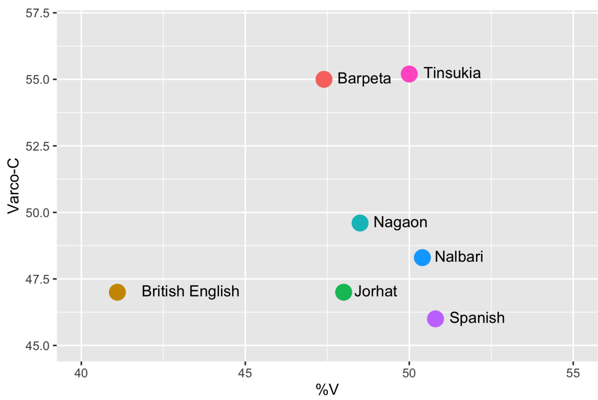

Figure 4: *%V and Varco-C in Assamese speakers*

It has been suggested that Varco-ΔC is not affected by speech rates and is considered better in discriminating stress-timed languages from syllable-timed languages [25]. We have noticed in Section 3.1, that in case of Assamese, even though speech rate does not affect Varco-ΔC, utterance length does. Nevertheless, we decided to plot the two measures on an x-y plot as seen in Figure 4. While data for Japanese is not plotted in this plot, we do not see any coherent pattern of Varco-ΔC distribution of Assamese varieties. Hence, separation between rhythm type is best, with the given data, when we plot %V and nPVI-V on an x-y plot.

In order to see if the effect of region and dialect areas on rhythm measures, we built LME models for each measure where region or dialect area was considered fixed effect and gender and speaker were considered random effects. The models were subjected to Wald χ^2 test and the results obtained are presented in Table 4. Table 4 shows that all the rhythm measures, except ΔC and nPVI-V, show a significant effect of region. On the other hand, for dialect areas, %V, Varco-ΔC and nPVI-V do not show any effect of dialect areas.

Table 3: *Rhythm measure comparisons*

region	%V	nPVI-V	r-PVI	ΔV	ΔC	Varco-V	Varco-C
Barpeta	47.4	51.8	56.5	43.7	53.3	55.2	59.1
Jorhat	48.1	48.9	43.4	41.4	45.9	53.8	54.6
Nagaon	48.6	51.3	49.0	46.2	50.6	54.3	55.2
Nalbari	50.3	55.5	48.2	52.5	51.3	58.4	57.4
Tinsukia	50.0	52.4	46.4	42.6	47.0	54.5	58.3
British English	41.1	57.2	64.1	46.6	56.7	64.0	47.0
Spanish	50.8	29.7	57.7	20.7	47.5	41.0	46.0
Japanese	45.5	40.9	62.5	53.0	55.5	56.0	-

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

	Region		Dialect area	
Measure	χ^2	**p-value**	χ^2	**p-value**
%V	10.08	0.039	0.81	0.66
ΔV	17.32	0.002	7.48	0.02
ΔC	07.23	0.124	6.10	0.04
Varco-ΔV	14.28	0.006	8.48	0.01
Varco-ΔC	12.81	0.012	4.83	0.09
nPVI-V	08.63	0.070	3.67	0.16
rPVI	15.77	0.003	6.84	0.03

3.4. QDA results

In order to see the discriminability of Assamese spoken in different regions or different dialectal areas by means of rhythm and rate of articulation measures, a Quadratic Discriminant Analysis (QDA) was conducted by combining the rhythm measures and rate of articulation measures as features. Table 5 presents the results of QDA. Overall, Assamese spoken in five regions is discriminated with an accuracy of about 42.1% with all the rhythm and speaking rate measures. However, the best result of 42.8% is achieved when ΔC is removed from the analysis. In case of dialect identification, best accuracy of 47.4% is achieved when rPVI is removed from the analysis.

Table 5: *Results of Quadratic Discriminant Analysis*

Sl.	Features	Region ID	Dialect ID
1	Seg/sec.	24.2%	39.7%
2	Syl/sec.	24.0%	39.3%
3	%V	22.2%	33.3%
4	ΔV	25.8%	40.0%
5	Seg/sec.+/sec.	25.9%	41.5%
6	%V+ΔV + VarCo-V+ΔC+ VarCo-C+nPVI-V+rPVI	37.8%	42.6%
7	Seg/sec.+Syl/sec.+ %V+ΔV+VarCo-V+ΔC+ VarCo-C+ nPVI-V+rPVI	42.1%	47.2%
8	7, excluding ΔC	**42.8**%	46.9%
9	7, excluding rPVI	41.5%	**47.4**%

Table 6 and Table 7 provide the accuracy by region and dialect areas for the best discrimination accuracy. As seen in Table 6, in the best case, all areas, except Nagaon, are correctly categorized with an average accuracy of 46%. Similarly, Eastern and Western Assamese dialect areas are correctly categorized with an average accuracy of 63%. However, in case of the Nagaon region, which also is the sole member in the CA category, in this study, the accuracy in the lowest. It is important to note here that CA dialect division which is geographically nestled between EA and WA speaking regions and has been considered as an intermediate dialect [3]. We assume this to be the reason for the lower accuracy in categorizing the Nagaon region.

Table 6: *Region-wise accuracy in QDA for region ID*

Tinsukia	Jorhat	Nagaon	Nalbari	Barpeta
59.5%	38.5%	28.1%	47.2%	40.7%

Table 7: *Dialect-wise accuracy in QDA for Dialect ID*

EA	CA	WA
71.3%	15.4%	55.6%

4. Conclusions

This study provided an analysis of speech rate and rhythm measures in five Assamese speaking regions belonging to three broad dialect areas. Rate of articulation measures indicated no strong correlation between region or dialect with speaking rate. In case of rhythm measures nPVI-V and %V are found to be least affected by speaking rates and length of breath groups. As suggested in the literature, VarcoΔC was not affected by rate of speaking however, it highly correlated with the length of utterance. When plotted on an nPVI-V-%V plane, Assamese varieties clustered close to Japanese which is a mora-timed language. Previous works differ in concluding the rhythm class for Assamese [9, 10]. Considering our analysis of rhythm class in Assamese, we see evidence that Assamese rhythm may be more akin to mora-timed rhythm class. It is noteworthy in Assamese a foot is claimed to be always bimoraic and which contains two moras [30]. Hence, rhythm measures clustering Assamese with mora-timed languages, is not surprising after all. Finally, we also report that rhythm and rate of articulation measures can discriminate Assamese spoken in different regions, at least above the chance level.

5. Acknowledgements

The Assamese data presented in this work comes from a project titled "A broad Sociolinguistic Study of Vowel Variation in Assamese", funded by the Indian Council of Social Science Research (ICSSR), India.

6. References

[1] G. A. Grierson, *Linguistic survey of India*. New Delhi: Motilal Banarasidass, 1927, vol. 5, no. 1,2.

[2] B. Kakati, *Assamese, its formation and development*. Lawyers̀ Book, 1941.

[3] G. C. Goswami and J. Tamuli, "Asamiya," in *The Indo-Aryan languages*, D. Jain and G. Cardona, Eds. Routledge, 2003, pp. 391–443.

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

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

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

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

[8] 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.

[9] P. S. Ray, "A reference grammar of Bengali." Center for Applied Linguistics, Washington DC, Tech. Rep., 1966.

[10] S. R. Savithri, "Speech rhythm in indian languages," in *Dr N. Rathna Oration, presented at the 41st Indian Speech and Hearing Association Conference, Pune*, 2009.

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

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

[13] 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.

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

[15] 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.

[16] V. Dellwo and P. Wagner, "Relations between language rhythm and speech rate," in *Proceedings of the 15th International Congress of Phonetics Sciences*. Citeseer, 2003, pp. 471–474.

[17] E. Jacewicz, R. A. Fox, and L. Wei, "Between-speaker and within-speaker variation in speech tempo of American English," *The Journal of the Acoustical Society of America*, vol. 128, no. 2, pp. 839–850, 2010.

[18] E. Jacewicz, R. A. Fox, C. O'Neill, and J. Salmons, "Articulation rate across dialect, age, and gender," *Language variation and change*, vol. 21, no. 02, pp. 233–256, 2009.

[19] M. P. Robb and G. T. Gillon, "Speech rates of New Zealand English and American English speaking children," *Advances in Speech Language Pathology*, vol. 9, no. 2, pp. 173–180, 2007.

[20] J. Verhoeven, G. De Pauw, and H. Kloots, "Speech rate in a pluricentric language: A comparison between Dutch in Belgium and the Netherlands," *Language and Speech*, vol. 47, no. 3, pp. 297–308, 2004.

[21] F. Biadsy and J. B. Hirschberg, "Using prosody and phonotactics in Arabic dialect identification," in *Proceedings of INTERSPEECH'09*, 2009.

[22] S. Ghazali, R. Hamdi, and M. Barkat, "Speech rhythm variation in arabic dialects," in *Speech Prosody 2002, International Conference*, 2002.

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

[24] P. Boersma and D. Weenink, "Praat: doing phonetics by computer (version 5.4).[logiciel, en ligne]," 2014.

[25] 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.

[26] 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.

[27] 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/

[28] 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.

[29] 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.

[30] S. Mahanta, "Some aspects of prominence in assamese and assamese english," MPhil Thesis, Central Institute of English and Foreign Languages, 2001.