Speech Prosody 2024
2-5 July 2024, Leiden, The Netherlands

Effect of Sociolinguistic Variations on Rate and Rhythm of Hindi L2 Speech

Joyshree Chakraborty[1], *Leena Dihingia*[2], *Priyankoo Sarmah*[1], *and Rohit Sinha*[1]

[1]Centre for Linguistic Science and Technology, Indian Institute of Technology Guwahati
[2]Department of Assamese, Gauhati University

jchakraborty@iitg.ac.in, leena@gauhati.ac.in, priyankoo@iitg.ac.in, rsinha@iitg.ac.in

Abstract

The paper investigates the effect of Assamese L1 in native Assamese speakers speaking Hindi as an L2 in terms of speech rate and rhythm metrics in various sociolinguistic contexts. While Hindi is a syllable-timed language, the rhythm in Assamese varieties is reported to be akin to Japanese, a mora-timed language. We investigate spontaneous Hindi speech produced by 75 native Assamese speakers. A total of 4645 breath groups are analyzed for speech rate and rhythm using measures like syllable per second, segment per second, %V, nPVI, rPVI, VarCo-V, VarCo-C, ΔV, and ΔC. Assamese accented Hindi spontaneous speech data is compared with Hindi and Assamese read speech data and further grouped into age, gender, rural/urban area, and dialect. All these groups appear to show rhythmicity similar to Assamese read speech indicating L1 influence on L2 rhythm. Age and gender effects on rhythm measures are observed.

Index Terms: rhythm, speech-rate, sociolinguistic variation

1. Introduction

Several studies have shown how speakers' native language (L1) can affect the rhythm of second language (L2) produced by them [1, 2, 3, 4, 5]. In the Indian context, a sizeable population (about 14%) speaks Hindi as either a second or third language [6], yet most of the literature on Indian native languages effect deal with English L2. There is a derth of work on the rhythmic influence of L1 on Hindi L2. Hence, the current work examines native Assamese speakers' L1 influence on the rate and rhythm of Hindi L2 spontaneous speech. Assamese, the official language of the northeastern state of Assam, is spoken by approximately 15.3 million individuals in India. According to the 2011 census, out of these speakers, 2.45 million also speak Hindi as their secondary language [6]. Additionally, the paper investigates the effect of sociolinguistic variations on Hindi L2 speech.

Based on how speech units are organized in time, languages have been traditionally categorized into three different classes, namely syllable-timed, stress-timed [7, 8], and mora-timed [9, 10, 11]. Hindi, an Indo-Aryan language, has been reported to be syllable-timed based purely on auditory impression [12]. Assamese, another Indo-Aryan language spoken in India, had been previously reported to be a syllable-timed language [13]. Whereas, recent quantitative and acoustic studies have shown that the rhythm measures of Assamese are comparable to those of the mora-timed languages [14, 15, 16]. Accordingly, we use three different types of metrics to quantify rhythm:

- *Interval measures*: The authors in [17] proposed interval measures like ΔC (standard deviation of consonantal intervals), ΔV (standard deviation of vocalic intervals), and %V (percentage of vocalic intervals in an utterance) to measure durational variability in vowel and consonant durations. Stress-timed languages were found to have higher ΔC since they can allow a greater variety of consonant clusters. On the other hand, due to vowel reduction in these languages, %V is low. However, ΔV can not be interpreted clearly as it can be affected by a variety of language-specific or contextual factors [15, 16]. The combination of %V and ΔC was found to be the best way of distinguishing rhythm classes, among these three absolute temporal measures [17].
- *Pairwise variability index*: It was found that interval measures are strongly influenced by speaking rate [18, 19]. Hence, the pairwise variability index (PVI) of vocalic and consonantal intervals was proposed in which the speech rate influence was reduced [18].
- *Rate-normalized measure*: An additional metric, obtained by normalizing the rate of the vowel intervals (VarcoV: coefficient of variation of vowel interval) and consonant intervals (VarcoC: coefficient of variation of the consonantal interval), which normalize ΔV and ΔC for speaking rates were proposed [20].

Variations in speech rate and rhythm can be attributed to factors such as age, gender, dialect, proficiency, and spontaneity [21, 22, 3, 4]. The variability of consonant and vowel intervals in [4] was affected by different English experiences or proficiency levels among Japanese learners of English. In [22], a study done on the German language showed that speech rate differences were largely accountable for rhythmic differences in interval measures between younger and older adults. Sociolinguistic factors can also change a language's rhythm class when spoken by a non-native speaker [23, 24]. The English (stress-timed) spoken by French natives showed French rhythmicity (syllable-timed) in terms of %V, nPVI-V, and VarcoV [24]. Additionally, studies have shown that rhythm class cannot be assigned based purely on values of durational metrics as speech rhythm can be significantly affected by speaker timing strategies [25, 26] Hence, in this paper, we also investigate the effect of four sociolinguistic factors, viz. age, gender, dialect, and region (urban/rural) on Hindi L2 rhythm.

The rest of the paper is organized as follows: Section 2 describes the database statistics and the methodology used. The results of the experimentations as well as their discussions are presented in Section 3. The salient findings of the study are concluded in Section 4.

2. Methodology

The dataset collected for this study includes ASsamese-accented spontaneous HIndi (ASHI) speech data produced by 75 native Assamese speakers. Hindi being an official language of India is taught in schools from the primary level as a second or third language. However, its fluency depends on individual use and exposure to the language, with the younger gener-

10.21437/SpeechProsody.2024-4

Table 1: *The values of rhythm and speech rate measures of four speech datasets created for this study along with those for Spanish [27], Japanese [28], and Engish [18].*

Measure	Spanish	Japanese	HI_NWS	AS_NWS	AH_NWS	ASHI	English
%V	50.8	45.5	47.2	48.2	48.0	45.3	41.1
nPVI-V	29.7	40.9	46.9	46.0	52.5	54	57.2
ΔV	20.7	53	44.9	38.5	58	65.2	46.6
VarCo-V	41	56	51.9	49.4	55.9	62.4	64
rPVI	57.7	62.5	75.2	37.6	88.6	110.4	64.1
nPVI-C	-	-	62.6	48.2	65.7	65.1	-
VarCo-C	46	-	64.4	49.8	66.3	78.4	47
ΔC	47.5	55.5	72.4	38.8	84.8	115.5	56.7
sylps	-	-	4.8	6.4	4.6	5.6	-
segps	-	-	9.1	13.2	8.8	10.7	-

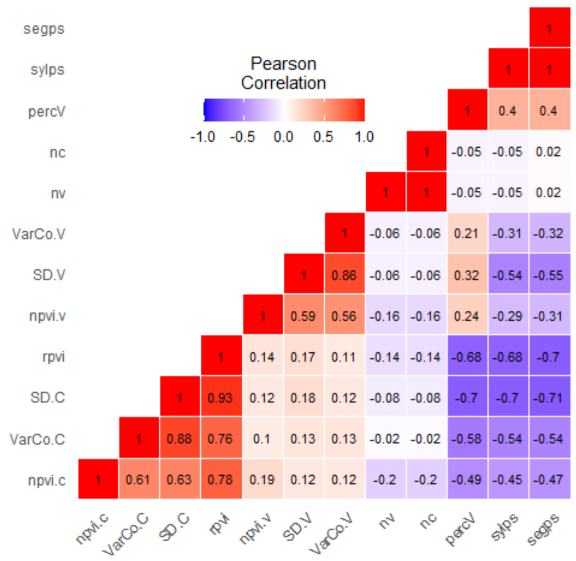

Figure 1: *Pearson's correlation matrix of rhythm measures.*

ation and those from urban regions generally being more fluent in speaking the language. The participants varied in age (61 speakers below 30 years and 14 speakers above 30) and gender (26 male and 49 female). They were also grouped by the geographical dialectal regions they belonged to (19 speakers from Eastern Assam (EA), 49 from Western Assam (WA), and 7 from Central Assam (CA)) and by the region (urban/rural) they came from. The participants were recorded speaking spontaneously for about five minutes on any topic related to their interests. The data collection was conducted remotely over the telephone and all the speech files were later stored in a unified format of 16 kHz sampling frequency, 16-bit encoding, and mono-channel.

Two other datasets were considered to compare ASHI with L1 Assamese and L1 Hindi speech rhythm. The L1 Assamese speech data was obtained from a previous study on Assamese rhythm [15] (AS_NWS), whereas the L1 Hindi speech database (HI_NWS) was collected from native Hindi speakers. For both datasets speakers were recorded reading a version of the fable "The North Wind and the Sun" (NWS) translated into their respective native (Assamese or Hindi) languages.

It is evident from the literature that there is a difference in the rhythmicity between spontaneous and read speech [23]. Therefore, we collected read Assamese accented Hindi speech data referred to as AH_NWS. This database contains Hindi speech recordings of the NWS passage produced by 14 native Assamese speakers (7 male and 7 female). The AH_NWS database comprises speech data collected from a total of 14 individuals (7 male and 7 female) belonging to the three dialectal regions (4 from EA, 5 from WA, and 6 from CA). All the speakers were below 30 years of age.

The speech datasets from all the groups were segmented at the breath group level using the Praat software. The speech files corresponding to the breath groups were subjected to a MATLAB® code for vowel onset point (VOP) and vowel endpoint (VEP) detection [29]. Annotation and analysis were limited to breath groups including seven syllables or more, and without any pause or hesitation. The durations of vowels and consonants were determined from VOPs and VEPs. In a previous study on Assamese rhythm, the speech data was manually annotated for the duration of vowels and consonants [15]. This study utilized that corpus and it was found that the rhythm metrics obtained using both methods are in line.

A Tascam linear PCM recorder, linked via an audio connector to a Shure unidirectional head-worn microphone, was used for the recording of all the speech datasets except ASHI. The NWS passage translated into Hindi was provided to Hindi and Assamese natives. Similar words corresponding to the Assamese text were used during the Hindi translation. The number of words in the Hindi text was kept almost the same as that of the Assamese text. For this study, the number of breath groups obtained by analyzing ASHI, AS_NWS, AH_NWS, and HI_NWS were 4645, 855, 358, and 225, respectively. The breath groups are analyzed for speech rate and rhythm using measures like syllable per second, segment per second, %V, nPVI-V, nPVI-C, rPVI, VarCo-V, VarCo-C, ΔV, and ΔC represented as sylps, sygs, percV, npvi.v, npvi.c, rpvi, VarCo.V, VarCo.C, SD.V, and SD.C, respectively, in Figure1.

In the next section, the results for speaking rhythm and rate for each speech database group are examined. In addition, a co-variation matrix produced by utilizing Pearson's correlation approach to compare speaking rate and rhythm measurements is provided. The findings of a random forest classification, which was done to categorize the various speech groups according to rate and rhythm aspects, are also reported in the section that follows. Lastly, several statistical analyses are carried out to determine how sociolinguistic factors affect the rhythm and articulation rate measurements for ASHI.

3. Results and Discussion

3.1. Correlation Analysis

As previously pointed out, the number of syllables in the analyzed utterances influences the rhythm measures. Speaking rates have also been seen to influence rhythm measures [19, 18]. Hence, a Pearson's correlation test was conducted to investigate the interaction between utterance length, speaking rate, and rhythm. The matrix in Figure1 presents the results of the Pearson's correlation test. ΔC, ΔV, rPVI, Varco-C, and nPVI-C exhibit a strong correlation with the articulation rate, as evidenced by sylps and segps. These measures are lowered as the rate of articulation increases. However, the length of the breath group (as indicated by nv, number of syllables in a breath group) does not show a strong correlation with rhythm measures. Among the rhythm measures, %V followed by nPVI-V are the least affected by speech rate.

3.2. Rhythm Analysis

The rhythm metrics obtained for the four speech groups (ASHI, AH_NWS, AS_NWS, HI_NWS) are provided in Table 1 along with measures of Spanish, British English, and Japanese, for comparison, obtained from earlier observations [18, 28, 27]. To check if the four speech groups are distinguishable by rhythm metrics, we performed a MANOVA test with rhythm measures as dependent variables and group as a factor. Results showed a Pillai score of 0.79 indicating that rhythm measures significantly differ by speech group. Next, we wanted to see if, when two speech groups are compared, which rhythm metrics can distinguish them. For the same, the MANOVA tests are performed under four cases as defined below:

- Case-1: HI_NWS vs. AS_NWS
- Case-2: AS_NWS vs. AH_NWS
- Case-3: AH_NWS vs. HI_NWS
- Case-4: AH_NWS vs. ASHI

Table 2: *Rhythm measure comparisons between sets of two speech groups.*

Measure	Case-1	Case-2	Case-3	Case-4
%V	*	!	!	*
nPVI-V	!	*	*	!
ΔV	*	*	*	*
VarCo-V	*	*	*	*
rPVI	*	*	*	*
nPVI-C	*	*	!	!
VarCo-C	*	*	!	*
ΔC	*	*	*	*

*=significantly different; !=no significant difference

The results of those tests are summarised in Table 2. As observed, for Case-1, all rhythm measures except nPVI-V are found to be significantly different. Similarly, for Case-2, all rhythm measures except %V are found significantly different. In the case of Case-3, all rhythm measures except %V, nPVI-C, and VarCo-C are found significantly different. For Case-4, all rhythm measures except nPVI-V and nPVI-C are found significantly different. We also tried to classify the speech groups according to the cases mentioned above. Except for Case-4, other cases showed fair discrimination between the two speech groups under consideration. The accuracies for Case-1, Case-2, and Case-3 are found to be 88%, 96%, and 73%, respectively.

In Section 3.1, we note that rhythm measures %V and nPVI-V are least affected by speech rate and utterance length. Hence, we plot the measures for the four speech databases along with Japanese (mora-timed), British English (stress-timed), and

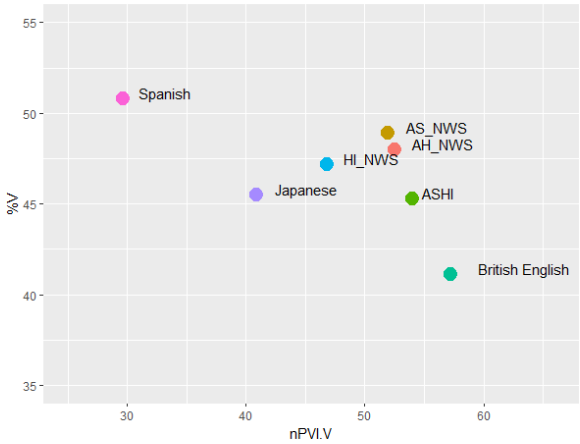

Figure 2: *Plot showing the nPVI-V and %V variation for AS_NWS, HI_NWS, AH_NWS and ASHI.*

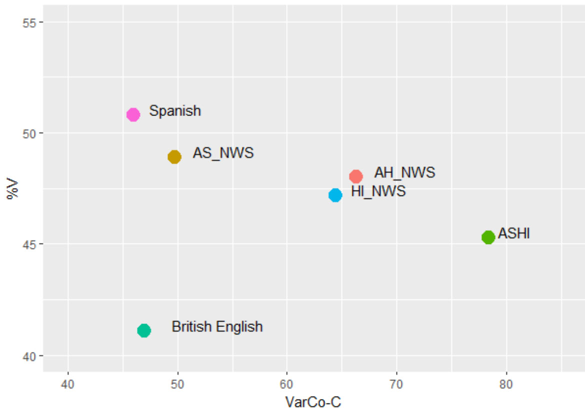

Figure 3: *Plot showing the VarCo-C and %V variation for AS_NWS, HI_NWS, AH_NWS and ASHI.*

Spanish (syllable-timed) for comparison. Figure 2 shows the distribution of speech groups/languages on a 2D plane with %V and nPVI-V as two coordinate axes. Previous acoustic study with AS_NWS speech data showed that Assamese rhythm is akin to mora-timed Japanese [15], as also seen in the figure. As observed, AH_NWS is positioned the closest to AS_NWS indicating a positive L1 effect on Hindi L2 read speech. Similarly, ASHI is also positioned closer to AS_NWS indicating a positive L1 effect on Hindi L2 spontaneous speech. As mentioned earlier, based on impressionistic evidence, Hindi has been reported to be syllable-timed like Spanish [12]. It has been observed that individual timing strategies have a significantly greater influence on rhythm, understood here as durational variability than a language's phonological structure [25]. In our analysis of Hindi read speech, we observed that the language is more akin to mora-timed rhythm which might be attributed to the overall speaker timing strategy, rather than the phonology of Hindi.

The authors in [20] proposed that VarCo-C remains unchanged regardless of speech rates and may be deemed more effective in distinguishing stress-timed languages from syllable-timed languages.. However, in our study, we found that speech rate affects VarCo-C (refer to Figure 1). On further analysis, we found that the VarCo-C value decreases with increasing speech rates. Nevertheless, we decided to plot the %V and VarCo-C measures on a 2D plot as shown in Figure 3. Although Japanese is not plotted here for lack of data, we see that the four groups are positioned between Spanish and English indicating a mora-timed rhythm.

A classification task was performed using random forest (RF) and rhythm measures to determine the separability among four speech databases utilized in this study. Figure 4 shows the confusion matrix for the said classification. From the figure, it can be seen that ASHI and AS_NWS are well-separated. The other groups show confusion with ASHI. This is attributed to a significant imbalance between the number of breath groups present in the read speech groups (AS_NWS, AH_NWS, and HI_NWS) and ASHI. To address that, we attempted to reduce the number of breath groups to 10 per speaker for all the speakers of ASHI to make it comparable to other groups. The RF classification task is repeated and the confusion matrix is shown in Figure 5. It is observed that all four speech groups are well separated except HI_NWS. This might be because HI_NWS has relatively lower breath groups.

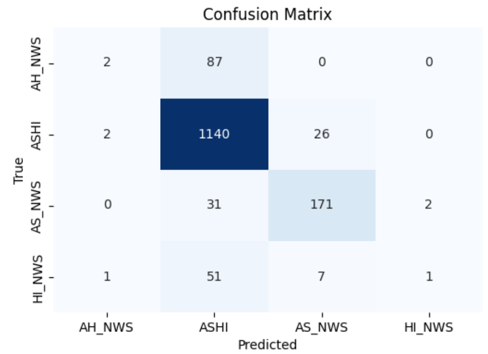

Figure 4: *Confusion matrix obtained from RF classification based on rhythm metrics for ASHI, AS_NWS, AH_NWS, and HI_NWS.*

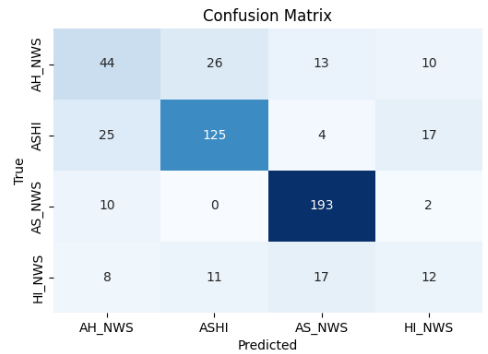

Figure 5: *Confusion matrix for repeat RF classification reported in Figure 4 with balanced ASHI data.*

3.3. Effect of sociolinguistic factors on rate and rhythm of ASHI

To check the effect of sociolinguistic factors on ASHI, separate linear mixed effect (LME) models for each speech rate and rhythm measure were created with age, gender, dialect, and region (urban/rural) as fixed effects, and speaker as a random effect. The LME models were then subjected to an ANOVA test, and the results are presented in Table 3. It can be seen from the table, dialect and region do not seem to have any effect on rhythm measures. However, the effect of gender has been seen on %V, nPVI-V, and ΔV. Similarly, the effect of age is seen on %V, rPVI, VarCo-C, and ΔC. Table 4 shows these measures for male and female speakers of ASHI and in the above and below 30 years age group. For each of the measures, female speakers are found to have significantly higher values than male speakers. Older speakers tend to have significantly higher values for rhythm measures except for %V. We then construct separate LME models with rhythm measures as dependent variables and the interaction between gender and age as fixed effects. It was found that the interaction between gender and age was not significant enough for any of the rhythm measures.

Table 3: *Effect of sociolinguistic factors on speech rate and rhythm measures for ASHI.*

Measure	Gender	Age	Dialect	Region
%V	*	*	!	!
nPVI-V	*	!	!	!
ΔV	*	!	!	!
VarCo-V	!	!	!	!
rPVI	!	*	!	!
nPVI-C	!	!	!	!
VarCo-C	!	*	!	!
ΔC	!	*	!	!
Sylps	!	!	!	!
Segps	!	!	!	!

*=significantly different; !=no significant difference

Table 4: *Gender- and age-wise rhythm measures for ASHI.*

Measure	Female	Male	Below 30	Above 30
rPVI	-	-	106.5	129.9
VarCo-C	-	-	76.7	86.5
ΔC	-	-	110.8	138.4
%V	46.6	42.8	45.9	42.1
nPVI-V	55.5	51.1	-	-
ΔV	68.8	58.4	-	-

4. Conclusions

This study attempted to investigate the effect of Assamese L1 on Hindi L2 speech rate and rhythm. For this purpose, the speech rate and rhythm metrics across various sociolinguistic contexts are compared for L1 Assamese read speech (AS_NWS), L2 Hindi read speech (HI_NWS), L1 Assamese accented L2 Hindi read speech (AH_NWS) and spontaneous Assamese L1 accented L2 Hindi speech (ASHI). MANOVA test reveals that the four speech groups are significantly distinguishable. On the other hand, a comparison done on %V and nPVI-V plot with Spanish, British English, and Japanese shows that all four studied speech groups are positioned between syllable-timed and stress-timed languages similar to Japanese. Interestingly, both AS_NWS and AH_NWS are observed to have higher %V than HI_NWS. Statistical tests showed that %V for AS_NWS and AH_NWS did not differ significantly indicating L1 effect on L2 speech. ASHI is found to be closest to Japanese in terms of %V. Among the different sociolinguistic factors analyzed, age and gender are found to affect the rhythm while they are not found to have any significant effect on the speech rate in the context of ASHI.

5. Acknowledgements

We thank the participants for actively taking part in the data collection process. We also thank Ms. Bhagyasree Talukdar for her contribution to creating a part of the ASHI database.

6. References

[1] U. Gut, "Rhythm in L2 Speech," *Speech and Language Technology*, vol. 14, no. 15, pp. 83–94, 2012.

[2] L. White and S. L. Mattys, "Calibrating Rhythm: First Language and Second Language Studies," *Journal of Phonetics*, vol. 35, no. 4, pp. 501–522, 2007.

[3] W. L. Law, O. Dmitrieva, and A. Francis, "Convergence of L1 and L2 Speech Rhythm in Cantonese-English Bilingual Speakers," in *Proc. of Speech Prosody*, 2020, pp. 547–550.

[4] S. Kawase, J. Kim, and C. Davis, "The Influence of Second Language Experience on Japanese-accented English Rhythm," in *Proc. of Speech Prosody*, 2016, pp. 746–750.

[5] A. Tortel and D. Hirst, "Rhythm Metrics and the Production of English L1/L2," in *Proc. of Speech Prosody*, 2010.

[6] Office of the Registrar & General Census Commissioner, India, "C-17 population by bilingualism and trilingualism," 2011, Accessed: 2024-01-10. [Online]. Available: https://censusindia.gov.in/census.website/data/census-tables

[7] K. L. Pike, *The Intonation of American English*. ERIC, 1945.

[8] D. Abercrombie *et al.*, *Elements of General Phonetics*. Edinburgh University Press Edinburgh, 1967.

[9] M. Han, "The Feature of Duration in Japanese," *Onsei no kenkyu [Studies in Phonetics]*, vol. 10, pp. 65–80, 1962.

[10] R. F. Port, S. Al-Ani, and S. Maeda, "Temporal Compensation and Universal Phonetics," *Phonetica*, vol. 37, no. 4, pp. 235–252, 1980.

[11] 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.

[12] R. M. Dauer, "Stress-timing and syllable-timing reanalyzed," *Journal of Phonetics*, vol. 11, no. 1, pp. 51–62, 1983.

[13] P. S. Ray, "A Reference Grammar of Bengali." Center for Applied Linguistics, Washington DC, Tech. Rep., 1966.

[14] 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.

[15] L. Dihingia and P. Sarmah, "Rhythm and Speaking Rate in Assamese Varieties," *Measurements*, vol. 10, no. 205, p. 33, 2020.

[16] L. Dihingia, "Vowel acoustics of Assamese spoken in five geographical locations," Ph.D. dissertation, Indian Institute of Technology Guwahati, 2020.

[17] F. Ramus, M. Nespor, and J. Mehler, "Correlates of Linguistic Rhythm in the Speech Signal," *Cognition*, vol. 73, no. 3, pp. 265–292, 1999.

[18] 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.

[19] V. Dellwo and P. Wagner, "Relations between Language Rhythm and Speech Rate," in *Proc. of the 15th International Congress of Phonetics Sciences*, 2003, pp. 471–474.

[20] V. Dellwo, P. Karnowski, and I. Szigeti, *Language and Language-Processing*, 2006, ch. Rhythm and Speech Rate: A Variation Coefficient for deltaC, pp. 231–241.

[21] T. Piske, I. R. MacKay, and J. E. Flege, "Factors Affecting Degree of Foreign Accent in an L2: A Review," *Journal of Phonetics*, vol. 29, no. 2, pp. 191–215, 2001.

[22] E. Pellegrino, L. He, and V. Dellwo, "The Effect of Ageing on Speech Rhythm: A Study on Zurich German," in *Proc. of Speech Prosody*, 2018.

[23] O. Maxwell and E. Payne, "Investigating (Rhythm) Variation in Indian English: An Integrated Approach," in *Speech Rhythm in Learner and Second Language Varieties of English*, 2023, pp. 17–57.

[24] S. Kaminskaïa, "Interplay of Sociolinguistic Factors in Rhythmic Variation in a Minority French Dialect," in *Proc. of Speech Prosody*, 2016.

[25] A. Loukina, B. Rosner, G. Kochanski, E. Keane, and C. Shih, "What determines duration-based rhythm measures: text or speaker?" *Laboratory Phonology*, vol. 4, no. 2, pp. 339–382, 2013.

[26] A. Arvaniti, "Rhythm, timing and the timing of rhythm," *Phonetica*, vol. 66, pp. 46–63, 02 2009.

[27] 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.

[28] I. Grenon and L. White, "Acquiring Rhythm: A Comparison of L1 and L2 Speakers of Canadian English and Japanese," in *Proc. of the 32nd Boston University Conference on Language Development*, 2008, pp. 155–166.

[29] P. Gogoi, P. Sarmah, and S. Prasanna, "Automatic Rhythm and Speech Rate Analysis of Mising Spontaneous Speech," in *Proc. of International Conference on Speech and Computer*, 2022, pp. 201–213.