



On Comparing Time- and Frequency-Domain Rhythm Measures in Classifying Assamese Dialects

Joyshree Chakraborty¹, Leena Dihingia², Priyankoo Sarmah¹, Rohit Sinha¹

¹Centre for Linguistic Science and Technology, Indian Institute of Technology Guwahati, India

²Department of Assamese, Gauhati University, India

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

Abstract

The rhythm measures are one of the primary means of differentiating languages and their dialects. They are broadly classified into two groups: time domain and frequency domain measures. The time-domain rhythm measures include temporal metrics like vocalic and non-vocalic durations and their derivatives, etc. and the frequency-domain rhythm measures comprise amplitude-modulated rhythm formant trajectory frequency and magnitude. To conduct this research, we focused on four Assamese regional varieties spoken in four districts of the Indian state of Assam. Data used in this study consists of read speech data of native Assamese speakers reading the Assamese translation of the “North Wind and the Sun” passage. The speech data was obtained from a total of 10 speakers for each of the four varieties. The average accuracy of dialect classification using quadratic discriminant analysis turns out to be 42% and 35%, respectively, for time- and frequency-domain rhythm measures. **Index Terms:** Rhythm categories, Rhythm metrics, Rhythm formant analysis, Assamese dialects

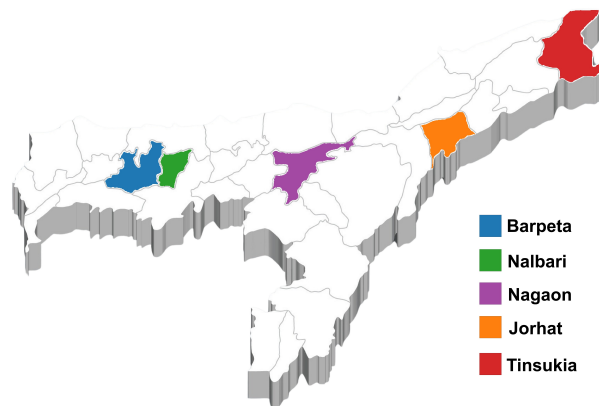


Figure 1: Map of Assam illustrating 5 districts representing the regional varieties of Assamese. Speakers from 4 districts (Barpeta, Nalbari, Jorhat and Tinsukia) were selected for this analysis.

1. Introduction

The Assamese language is spoken in the state of Assam in India. It has been broadly divided into three geographically distinct dialect groups: Central Assamese (CA), Eastern Assamese (EA), and Western Assamese (WA), each comprising several regional varieties. These dialects vary from each other in a wide range of linguistic features, including morphology, phonology, and vocabulary [1, 2, 3]. Although the phonetics and phonology of Assamese dialects have been well-studied, recent works have mostly attempted to classify these dialects using vowel acoustics [4, 5, 6, 7]. To the best of our knowledge, the only work that classified the Assamese regional varieties using rhythm as the prosodic feature used temporal rhythm metrics [8]. It employed a quadratic discriminant analysis (QDA) and reported an average accuracy of 42% in classifying five regional varieties, namely, Jorhat and Tinsukia belonging to EA dialect group, Nalbari and Barpeta belonging to WA dialect group and Nagaon belonging to CA dialect group. The geographical location of those regional varieties of Assamese are shown in Figure 1. In a recent study that used frequency domain rhythm measures in the form of low-frequency amplitude-modulated rhythm formants, different varieties of English and German languages were classified based on a majority vote rule [9]. In this work, we attempt to analyze the rhythms of four regional varieties of Assamese (Barpeta, Jorhat, Nalbari, and Tinsukia) as shown in Figure 1, using both time-domain and frequency-domain rhythm measures. The paper also evaluates dialect classification using QDA for both the rhythm measures.

Dialect identification is the process of identifying and dif-

ferentiating the various dialects of a language spoken in various social or geographic contexts. On account of vulnerabilities due to linguistic variations and geographic restrictions, dialect classification is thought to be difficult [10]. In literature, dialect identification has been carried out by acoustic modeling using MFCC [10], MFCC and higher order SDC features [11] and phonotactic modeling [12]. It is seen that prosodic features like rhythm and pitch, along with speaking rate, when combined with the phonotactic features yield better identification of the dialects [13]. There are many attempts to classify the dialects of Assamese. In one such attempt, mel frequency cepstral coefficient features were used to train Gaussian mixture models to classify four dialects of Assam [14]. In another work, the first four formant frequencies of eight vowels of Assamese are combined to form a 12-dimensional feature vector [15] for dialect identification. An attempt to discriminate 5 dialects corresponding to 3 dialectal areas of Assam, namely EA, WA, and CA was carried out by combining the vocalic and non-vocalic rhythm measures and rate of articulation (ROA) measures as features [8, 16].

Previous research indicates that rhythm is a multifaceted percept that cannot be easily analyzed considering a single aspect of the speech signal [17, 18]. Concentrating solely on interval durations, event timing, or event prominences might be misleading, as perceptions of these aspects have been demonstrated to interact with other variables within the speech stream [17, 19]. To address this, researchers looked into speech amplitude envelop-based metrics for rhythm analysis in various

languages [20]. Language and dialect classification has been carried out based on time and frequency domain rhythm features [21, 22, 9]. As mentioned earlier, several studies of rhythm in languages, have widely used the linguistic-phonetic interface wherein temporal metrics are used to categorize languages into different ‘rhythm classes’. However, to the best of our knowledge, no other study has used the frequency domain features for any Indian language or dialect classification. Therefore, although temporal rhythm metrics have been used previously in Assamese dialect classification, we explore the use of rhythm formants to classify the dialects of this language.

The rest of the paper is organized as follows. Section 2 describes the salient prosodic attributes of Assamese and its major dialects. Following that, the methodology employed for this study, in particular, describing the data, and the time and frequency domain rhythm measures used in Section 3. The experimental results are presented and discussed in Section 4. The paper is concluded in Section 5.

2. Assamese and its dialects

Assamese belongs to the Indo-Aryan language family and is spoken in the Brahmaputra valley of Assam. Over time it has varied enormously across the valley giving rise to several dialects and sub-dialects, with the major divisions being EA, CA, and WA [23, 2, 24]. The EA variety is also considered the standard colloquial Assamese (SCA) variety. While the EA exhibits homogeneity, WA consists of numerous local dialects, each exhibiting unique features compared to the others, differing mostly in phonetic terms. However, EA and WA vary in a broad range of vocabulary, morphology, and phonology [2].

2.1. Prosodic features

Among the prosodic differences, scholars have mostly commented on the stress or prominence of Assamese words as mentioned below. It is worth highlighting that prosodic differences are often related to differences in other segmental phonological processes.

- EA and CA have medial stress, whereas WA has initial stress. The strong initial stress of WA generally shortens words. In the other two dialects the nucleus of the syllable immediately following the stressed one is attenuated in timbre. For example, ‘pumpkin’ is pronounced as *komóra* in EA; *kumúra* in CA, and *kúmra* in WA [2, 25].
- SCA exemplifies a left-to-right trochaic system with an iterative binary rhythm [26]. In the Nalbari variety of Assamese spoken in western Assam, the prominence pattern is more forceful which is augmented by phonological processes like segmental deletion, addition, and diphthongization [27].

Rhythm constitutes a major prosodic feature of languages traditionally classifying the world’s languages into three rhythm categories, namely, stress-timed, syllable-timed [28, 29] and mora-timed [30, 31, 32]. These classifications determined by the organization of speech units over time. Assamese has been previously reported as a syllable-timed language, based solely on auditory impression [33]. However, a quantitative study involving one Assamese speaker [34] and a recent study using temporal rhythm metrics [8] claim that Assamese rhythm is mora-timed.

3. Methodology

This section elaborates on the speech data used and the time-domain and frequency-domain rhythm measures employed in this study.

3.1. Data

The Assamese speech data was obtained from a previous study on Assamese rhythm [8]. This data is collected from 40 voluntary native Assamese speakers. Of these, 10 (6 female and 5 male) speakers belong to Barpeta, 10 (5 female and 5 male) speakers belong to Jorhat, 10 (4 female and 6 male) speakers belong to Nalbari and 10 (5 female and 5 male) speakers belong to the Tinsukia district respectively. In the previous study on Assamese dialect classification using rhythm metrics, the classification accuracy was the lowest for the Nagoan region (CA variety). The authors assumed that this low accuracy was a result of CA being geographically situated between EA and WA and has been considered as an intermediate dialect [8, 16]. Therefore in this study, we have considered only EA and WA dialectal varieties for classification.

The speech data is recorded at a sampling frequency of 44.1 kHz and with 24-bit resolution [8]. The recorded speech files were downsampled to 16 kHz and 16-bit resolution for ease of computation and to support the input requirements of the RFA tool¹. A read-aloud IPA standard story called “The North Wind and the Sun,” translated into Assamese, comprises the speech data in this study.

3.2. Time-domain rhythm measures

The speech datasets from all the groups were segmented at the breath group level using the Praat software. A breath group is considered as a basic unit of prosodic analysis [35]. It refers to a stretch of utterance between two pauses of sufficient length for an intake of breath [36]. The speech files corresponding to the breath groups were subjected to a MATLAB® code for vowel onset point (VOP) and vowel endpoint (VEP) detection [37]. Annotation and analysis were limited to fluent utterances, i.e., breath groups including seven syllables or more, and without any pause or hesitation. The duration of vowels and consonants was determined from VOPs and VEPs. Thereafter, a perl script was used to find out the rhythm metrics. The time-domain rhythm measures calculated for a breath group used in the above study are %V (percentage of vocalic interval) [38], ΔV (standard deviation of the duration of vocalic intervals), ΔC (standard deviation of the duration of intervocalic intervals) [38], $\text{Varco}\Delta V$ (percentage of standard deviation of vocalic interval duration (ΔV) of the average duration of vocalic intervals (mean V)), $\text{Varco}\Delta C$ (percentage of standard deviation of consonantal interval duration (ΔC) of the average duration of consonantal intervals (mean C)) [39] nPVI-V (rate normalized measures of the durational variation of two consecutive vocalic intervals), nPVI-V (rate normalized measures of the durational variation of two consecutive vocalic intervals r-PVI (raw measure of the durational variation of two consecutive consonantal intervals) [40, 16].

3.3. Frequency-domain rhythm measures

The rhythm formant analysis (RFA) is a novel and quantitative modulation-theoretic process of analyzing the dynamic long-term rhythm variation in long narratives. This technique ad-

¹<https://github.com/dafydd/RFA>

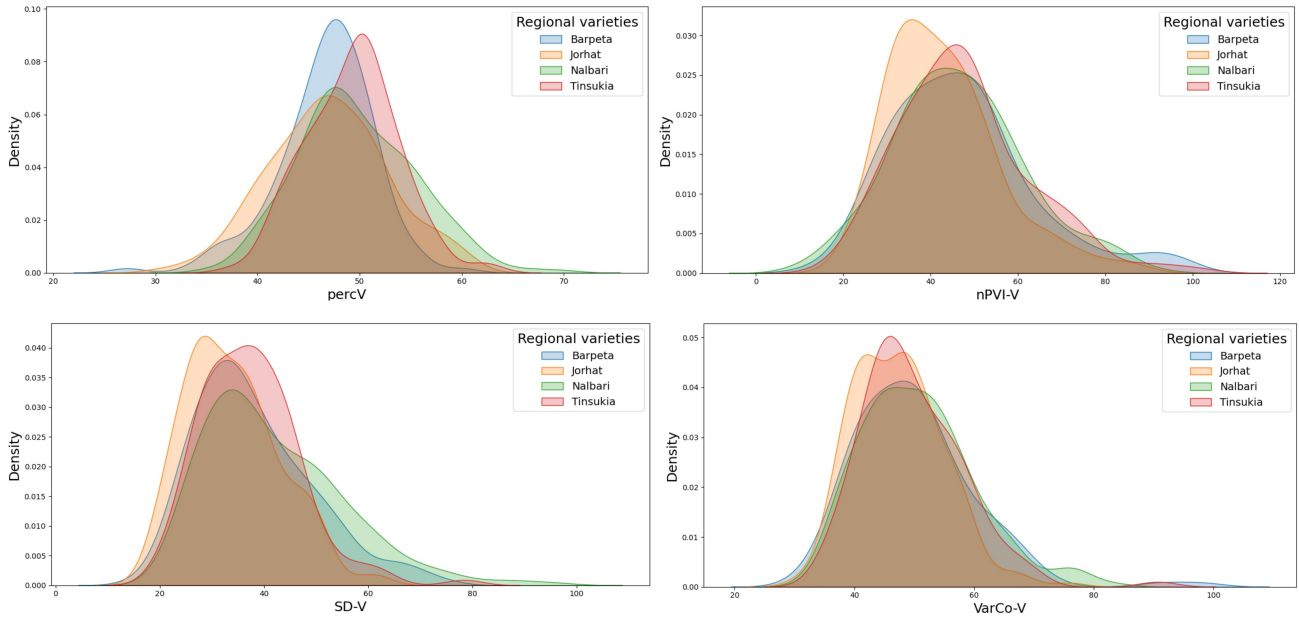


Figure 2: Kernel density estimation (KDE) plots for four different vocalic time-domain rhythm features.

dresses the temporal variation aspect of LF spectrogram analysis for both AM and FM. Speech rhythms can be identified by spectral analysis as magnitude peaks of the amplitude and frequency modulation envelopes of the speech [9]. To distinguish between tokens of various speech styles and language varieties, very long-term rhythms of rhythm are found [41] and subjected as input to the hierarchical clustering (a most fundamental unsupervised machine learning algorithm) [42]. Hierarchical Clustering is employed as it is unsupervised and does not require cluster mentions. It is versatile takes a variety of distances and produces robust results for small data. RFA has been employed in several disciplines, including linguistics [9], studies on discriminating and learning second languages [43], etc. In contrasting English with Mandarin, the interaction of amplitude modulation (AM) and frequency modulation (FM) rhythm factors are explored [44]. The maximum magnitude frequency trajectory of the LF spectrogram is extracted and applied to classify readings in different languages (English and Mandarin) and German dialects using a number of distance-based hierarchical clustering approaches [9].

The RFA procedure involves downsampling of the signal and AM and FM demodulation [45]. The fundamental idea is that in FM, the source signal's frequency, originating from the larynx, is modulated as a result of factors like pitch accent, tone, and intonation. In AM, on the other hand, the amplitude of the resulting FM signal is modulated due to the interaction of foot, syllable, and phrase sonority patterns [9] [46]. This results in the speech signal, which exhibits both FM and AM characteristics. The primary objective of RFA is to extract the FM and AM information from this modulated signal. In this study, we have considered the AM trajectory magnitude and frequency values as frequency-domain rhythm features.

The AM information is retrieved and assessed through a three-step process: demodulation of AM, LF spectrum, and spectrogram generation, and the comparison of AM spectral attributes among various data elements using distance-based hierarchical clustering [41]. A fifth-order Butterworth filter with

a 10 Hz cutoff is used. Each spectrum was created using a 3s moving FFT window (50 overlapping steps, 60ms each step) to generate the spectrogram to demonstrate minute, steady rhythmic shifts. The rhythm formants are identified as the spectral frequency zones around the magnitude peaks in the LF spectrum and are used as input to the hierarchical clustering (a most fundamental unsupervised machine learning algorithm) [42]. In this study, the rhythm features are extracted and a QDA is performed to classify the dialects.

4. Results and Discussion

This section presents the experimental results of Assamese regional dialect classification performed using time- and frequency-domain rhythm measures.

4.1. Time-domain rhythm measures

To see the discriminability of broad Assamese dialects (EA and WA) and the regional varieties (Barpeta, Nalbari, Jorhat, and Tinsukia) using temporal rhythm metrics, a QDA was conducted by combining the rhythm measures as features. The QDA classification is performed using 5-fold cross-validation with 80% of the files as training and 20% as testing. Figures 3 and 4 show the confusion matrices for regional and broad dialects of Assamese. It can be seen from Figure 3, that all regional varieties are correctly categorized with an average accuracy of 52%. Broad dialect groups EA and WA are correctly categorized with an average accuracy of 65%. In a previous study, the authors classified 5 Assamese regional varieties using the acoustic properties of vowel [6]. When it comes to temporal rhythm measures, the separation between rhythm types is best when plotted on a %V and nPVI-V plane [8]. To assess the same, the kernel density estimation plots for four vocalic time domain features are plotted and shown in Figure 2. From this figure, it can be observed that geographically distant varieties have shown better discrimination, unlike the proximally closed ones.

True label	Recognized as			
	Barpeta	Jorhat	Nalbari	Tinsukia
Barpeta	49.4	19.8	13.6	17.3
Jorhat	14.9	43.6	28.0	13.5
Nalbari	11.3	20.5	62.8	5.4
Tinsukia	26.5	8.0	12.4	53.1

Figure 3: Confusion matrix for the classification of regional variety of Assamese using time-domain features.

True label	Recognized as	
	EA	WA
EA	58.7	41.3
WA	28.5	71.5

Figure 4: Confusion matrix for broad Assamese dialect group classification using time-domain features

4.2. Frequency-domain rhythm measures

We analyze the speech files corresponding to the said four dialects, considering the AM rhythm formant frequency. The mean accuracy obtained in classifying the dialects is 27.5%. A previous study mentioned that combining both frequency and magnitude may lead to better discrimination among closely similar dialects of a language [9]. So, we changed the input file to include both the trajectory frequency and magnitude values. This new feature resulted in a mean classification accuracy of 35% as observed from the confusion matrix shown in Figure 5. Then, we try to assess the discriminability of the Assamese regional dialects pairwise. We find that the mean accuracies for Barpeta-Nalbari, Jorhat-Tinsukia, and Jorhat-Nalbari pair are 50%, 55%, and 75%, respectively. For the Barpeta-Nalbari pair, most of the files belonging to Barpeta are misclassified as Nalbari. Similarly, for the Jorhat-Tinsukia pair, most of the files belonging to Tinsukia are misclassified as Jorhat. Since Jorhat and Tinsukia varieties belong to the EA dialectal group, while Nalbari and Barpeta belong to WA dialectal groups, we can say that the misclassification is not random and depends on the dialectal regions. Consequently, the misclassification appears to correlate with the distinct dialectal regions, implying a non-random association between misclassification and dialectal affiliation.

We also analyzed the speech files corresponding to broad dialects (EA and WA). The confusion matrix of QDA classification is shown in Figure 6. It can be noted from the figure that the mean accuracy in classifying the dialects corresponding to the dialectal regions based on the rhythm features turns out to be 67%.

5. Conclusion

This study attempts to classify four regional varieties of Assamese based on rhythm measures. For the same, the quadratic discriminant analysis has been performed using both traditional time-domain rhythm features and recently proposed RFA-based frequency-domain rhythm features. The study reveals that

True label	Recognized as			
	Barpeta	Jorhat	Nalbari	Tinsukia
Barpeta	20.0	30.0	20.0	30.0
Jorhat	30.0	50.0	20.0	0.0
Nalbari	30.0	20.0	50.0	0.0
Tinsukia	0.0	40.0	40.0	20.0

Figure 5: Confusion matrix for the classification of regional variety of Assamese using frequency-domain features.

True label	Recognized as	
	EA	WA
EA	75.0	25.0
WA	41.7	58.3

Figure 6: Confusion matrix for broad Assamese dialect group classification using frequency-domain features

the frequency-domain rhythm measures yield an average dialect classification accuracy of 35% which is slightly lower but compares well with that of time-domain rhythm measures. It is worth mentioning that time- and frequency-domain rhythm measures capture different attributes of the speech signal. Therefore it will be interesting to see if the two features can be combined, which can be pursued as future work.

6. References

- [1] G. A. Grierson, *Linguistic survey of India*. New Delhi: Motilal Banarasisidass, 1927, vol. 5, no. 1,2.
- [2] B. Kakati, *Assamese its Formation and Development*. Government of Assam, 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] M. Sharma and K. K. Sarma, "Dialectal Assamese vowel speech detection using acoustic phonetic features, KNN and RNN," in *2015 2nd International Conference on Signal Processing and Integrated Networks (SPIN)*, 2015, pp. 674–678.
- [5] M. Sarma and K. K. Sarma, "Dialect identification from Assamese speech using prosodic features and a neuro fuzzy classifier," in *2016 3rd International Conference on Signal Processing and Integrated Networks (SPIN)*, 2016, pp. 127–132.
- [6] P. Sarmah and L. Dihingia, "Assamese dialect identification from vowel acoustics," in *Data Engineering for Smart Systems*, P. Nanda, V. K. Verma, S. Srivastava, R. K. Gupta, and A. P. Mazumdar, Eds. Singapore: Springer Singapore, 2022, pp. 313–322.
- [7] H. C. Das and U. Bhattacharjee, "Assamese dialect identification using static and dynamic features from vowel," *Journal of Advances in Information Technology*, vol. 15, no. 2, 2024.
- [8] L. Dihingia and P. Sarmah, "Rhythm and Speaking Rate in Assamese Varieties," in *Proc. Speech Prosody 2020*, 2020, pp. 561–565.
- [9] D. Gibbon, "The Rhythms of Rhythm," *Journal of the International Phonetic Association*, pp. 1–33, 2021.

- [10] P. A. Torres-Carrasquillo, T. P. Gleason, and D. A. Reynolds, "Dialect Identification using Gaussian Mixture Models," in *ODYSSEY04-The Speaker and Language Recognition Workshop*, 2004.
- [11] E. Singer, P. A. Torres-Carrasquillo, T. P. Gleason, W. M. Campbell, and D. A. Reynolds, "Acoustic, Phonetic, and Discriminative Approaches to Automatic Language Identification," in *Interspeech*, 2003.
- [12] M. A. Zissman and K. M. Berkling, "Automatic Language Identification," *Speech Communication*, vol. 35, no. 1-2, pp. 115–124, 2001.
- [13] F. Biadsy and J. Hirschberg, "Using prosody and phonotactics in Arabic dialect identification," in *Interspeech*, 2009. [Online]. Available: <https://api.semanticscholar.org/CorpusID:6290967>
- [14] H. C. Das and U. Bhattacharjee, "Identification of Four Major Dialects of Assamese Language using GMM with UBM," in *Pattern Recognition and Data Analysis with Applications*. Springer, 2022, pp. 311–319.
- [15] M. Sarma and K. K. Sarma, "Dialect Identification from Assamese Speech Using Prosodic Features and a Neuro Fuzzy Classifier," in *2016 3rd International Conference on Signal Processing and Integrated Networks (SPIN)*, 2016, pp. 127–132.
- [16] L. Dihingia, "Vowel acoustics of Assamese spoken in five geographical locations," Ph.D. dissertation, Indian Institute of Technology Guwahati, 2020.
- [17] K. J. Kohler, "The perception of prominence patterns," *Phonetica*, vol. 65, no. 4, pp. 257–269, 2009.
- [18] F. Cummins, "Rhythm as an affordance for the entrainment of movement," *Phonetica*, vol. 66, no. 1-2, pp. 15–28, 2009.
- [19] C. Alan, "Tonal effects on perceived vowel duration," *Lab. Phonol.*, vol. 10, no. 4, pp. 151–168, 2010.
- [20] S. Tilsen and A. Arvaniti, "Speech rhythm analysis with decomposition of the amplitude envelope: Characterizing rhythmic patterns within and across languages," *The Journal of the Acoustical Society of America*, vol. 134, no. 1, pp. 628–639, 2013.
- [21] H. Kim and J.-S. Park, "Automatic language identification using speech rhythm features for multi-lingual speech recognition," *Applied Sciences*, vol. 10, no. 7, 2020. [Online]. Available: <https://www.mdpi.com/2076-3417/10/7/2225>
- [22] F. Biadsy and J. Hirschberg, "Using prosody and phonotactics in Arabic dialect identification," in *Interspeech*, 2009.
- [23] G. A. Grierson, *Linguistic Survey of India: Vol. V. Indo-Aryan Family, Eastern Group; Pt. I. Specimens of the Bengali and Assamese Languages*. Office of the superintendent of government printing, India, Reprinted 1967 by Motilal Banarsidass, 1903.
- [24] G. C. Goswami and J. Tamuli, "Asamiya," in *The Indo-Aryan languages*, D. Jain and G. Cardona, Eds. Routledge, 2003, pp. 391–443.
- [25] G. C. Goswami, *Structure of Assamese*. Gauhati University, 1982.
- [26] S. Mahanta, "Some aspects of prominence in assamese and assamese english," MPhil Thesis, Central Institute of English and Foreign Languages, 2001.
- [27] A. I. Twaha, "Intonational phonology and focus in two varieties of Assamese," Ph.D. dissertation, Dissertation submitted at the Department of HSS, IIT Guwahati, 2017.
- [28] K. L. Pike, *The intonation of American English*. ERIC, 1945.
- [29] D. Abercrombie *et al.*, *Elements of general phonetics*. Edinburgh University Press Edinburgh, 1967, vol. 203.
- [30] M. Han, "The feature of duration in Japanese," *Onsei no kenkyu*, vol. 10, pp. 65–80, 1962.
- [31] R. F. Port, S. Al-Ani, and S. Maeda, "Temporal compensation and universal phonetics," *Phonetica*, vol. 37, no. 4, pp. 235–252, 1980.
- [32] 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.
- [33] P. S. Ray, "A reference grammar of Bengali." Center for Applied Linguistics, Washington DC, Tech. Rep., 1966.
- [34] 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.
- [35] Y.-T. Wang, J. R. Green, I. S. Nip, R. D. Kent, and J. F. Kent, "Breath group analysis for reading and spontaneous speech in healthy adults," *Folia Phoniatrica et Logopaedica*, vol. 62, no. 6, pp. 297–302, 2010.
- [36] Y.-C. Tsao and G. Weismer, "Interspeaker variation in habitual speaking rate: Evidence for a neuromuscular component," *Journal of Speech, Language, and Hearing Research*, vol. 40, no. 4, pp. 858–866, 1997.
- [37] B. D. Sarma, S. M. Prasanna, and P. Sarmah, "Consonant-vowel unit recognition using dominant aperiodic and transition region detection," *Speech Communication*, vol. 92, pp. 77–89, 2017.
- [38] F. Ramus, M. Nespor, and J. Mehler, "Correlates of Linguistic Rhythm in the Speech Signal," *Cognition*, vol. 73, no. 3, pp. 265–292, 1999.
- [39] V. Dellwo, *Language and Language-Processing*. Peter Lang, 2006, ch. Rhythm and Speech Rate: A Variation Coefficient for deltaC, pp. 231–241.
- [40] E. Grabe, E. L. Low *et al.*, "Durational Variability in Speech and the Rhythm Class Hypothesis," *Papers in Laboratory Phonology*, vol. 7, no. 515-546, pp. 1–16, 2002.
- [41] D. Gibbon, "Speech Rhythms: Learning to Discriminate Speech Styles," *Proc. Speech Prosody 2022*, pp. 302–306, 2022.
- [42] F. Nielsen and F. Nielsen, "Hierarchical Clustering," *Introduction to HPC with MPI for Data Science*, pp. 195–211, 2016.
- [43] A. Arvaniti, "The Usefulness of Metrics in the Quantification of Speech Rhythm," *Journal of Phonetics*, vol. 40, no. 3, pp. 351–373, 2012.
- [44] D. Gibbon, "Rhythm Formants of Story Reading in Standard Mandarin," *Chinese Journal of Phonetics*, vol. 14, 2020.
- [45] ———, "The Future of Prosody: It's about Time," in *Proc. Speech Prosody 2018*, 2018, pp. 1–9.
- [46] H. Traunmüller, "Conventional, Biological and Environmental Factors in Speech Communication: A Modulation Theory," *Phonetica*, vol. 51, no. 1-3, pp. 170–183, 1994.