Estudios de Fonética Experimental
Journal of Experimental Phonetics
ISSN: 1575-5533 · ISSN-e: 2385-3573

How language-specific and cross-linguistic factors affect speech rhythm: Evidence from Bengali and English

Nafiseh Taghva [1] 0000-0001-9715-4505
Shouvik Chaudhuri [2] 0000-0001-8957-5086

[1] Shiraz University (Iran)
[1] University of Southern Denmark (Denmark)

DOI: 10.1344/efe-2024-33-169-190

Corresponding address: taghvanafiseh@gmail.com

Received: 12/01/2024 Accepted: 02/09/2024 Published: 23/09/2024

Taghva, N., & Chaudhuri, S. (2024). How language-specific and cross-linguistic factors affect speech rhythm: Evidence from Bengali and English. *Estudios de Fonética Experimental*, *33*, 169–190. https://doi.org/10.1344/efe-2024-33-169-190

Abstract

This study examined the durational aspects of speech rhythm in Bengali (L1) and English spoken by educated Bengali advanced English learners (L2) to represent the effect of language-specific and cross-linguistic factors on speech rhythm. Employing metrics such as rateSyl, ΔC, nPVI-C, nPVI-V, VarcoV, %V, and ΔPeakLn, the study revealed that L1 exhibited a faster tempo, shorter consonants, and longer vowels relative to consonants, while L2 speech demonstrated greater variability in consonant and vowel durations, which are language-specific factors of L1 being a syllable-timed language and L2 being a stress-timed language. However, sonority patterns between syllables seemed consistent across L1 and L2 suggesting the cross-linguistic impact of L1 on L2. Moreover, the results identified the proportion of vocalic intervals (%V) as the most effective metric for differentiating between L1 and L2 rhythms.

Keywords

durational rhythmic metrics; Bengali (L1); English (L2); language-specific; cross-linguistic

Com afecten els factors lingüístics i específics de la llengua el ritme de la parla: proves del bengalí i l'anglès

RESUM

L'estudi examina el ritme de parla del bengalí (L1) i de l'anglès parlat per aprenents avançats bengalís (L2), i pretén representar l'efecte de factors específics de llengua i interlingüístics sobre el ritme de la parla. Utilitzant mètriques com rateSyl, ΔC, nPVI-C, nPVI-V, VarcoV, %V i ΔPeakLn, l'estudi revela que la L1 mostra un tempo més ràpid, consonants més curtes i vocals més llargues, mentre que la parla L2 presenta més variabilitat en la durada de consonants i vocals, que són factors específics d'idioma (L1 de ritme sil·làbic i L2 de ritme accentual). Els patrons de sonoritat entre síl·labes, al seu torn, semblen coherents entre L1 i L2, cosa que suggereix l'impacte interlingüístic de la L1 en la L2. Finalment, els resultats identifiquen la proporció d'intervals vocàlics (%V) com la mètrica més eficaç per diferenciar entre els ritmes L1 i L2.

MOTS CLAU

mètriques rítmiques de durada; bengalí (L1); anglès (L2); específic de l'idioma; interlingüístic

1. Introduction

Recently the number of studies dealing with the interaction of the two sound system of languages used by the same speakers has been increasing. These studies have examined the impact of language-specific features of the first language (L1) on the production and perception of the second language (L2) which is referred to as cross-linguistic influence or transfer (Fuchs, 2023). These studies have focused on variables such as age, order of acquisition, language dominance, proficiency, and the amount of L2 input received, as well as other non-linguistic factors (Amengual, 2021). Jarvis and Pavlenko (2008, p. 62) defined Cross-linguistic phonetic/phonological influence as "the way in which a person's knowledge of the sound system of one language can affect that person's perception and production of speech sounds in another language".

However, the terminology employed to refer to speakers who command two or more languages remains a contentious topic among researchers (Gass, 2020). The term "bilingual" frequently is used to encompass the broader concept of multilingualism (Bhatia & Ritchie, 2006). Edwards (2006) posited that bilingualism can be ascribed at any stage of second language acquisition (SLA), while Bhatia's (2006) viewpoint is centered on the ultimate attainment in L2 without specifying that this end must be native-level proficiency. Valdés (2001) suggested that it is more accurate to consider bilingualism as a continuum where varying degrees of proficiency in L1 and L2 are depicted. From the standpoint of L2 researchers, the term bilingual is designated for individuals who have achieved a stable proficiency in both languages. However, due to their interest in unraveling SLA process, these researchers may instead concentrate on individuals who are near-native speakers or advanced language learners (Gass, 2020). Kroll and Sunderman (2003) described these individuals as "skilled adult bilinguals," which is likely synonymous with the concept of advanced language learners.[1]

Generally, it is agreed that L2 learners face challenges in acquiring L2 rhythm due to the intricate interplay of language-specific and cross-linguistic influences. (Flege et al., 2003, Kinoshita & Sheppard, 2011; Rodríguez-Vázquez & Roseano, 2023). Rasier and Hiligsmann (2007) elucidate the concept of prosodic transfer, where the rhythmic patterns of L1 can permeate into their L2, thereby affecting the rhythm of the L2. Similarly, Rodríguez-Vázquez and Roseano's (2023) pilot study on Galician learners of English demonstrates that rhythmic transfer is indeed observable and that the degree of this transfer diminishes as proficiency in the L2. These findings show that rhythm quantification in L2 cannot be disentangled from the linguistic background of the speakers, as their L1 rhythm can have a lasting impact on their L2 rhythm, especially at lower levels of L2 proficiency. Consequently, rhythm quantification in a multilingual context must account for the interplay between the native and target language rhythms, which is a multifaceted process absent in monolingual rhythm analysis. Therefore, L2 rhythm acquisition and linguistic transfer have attracted the attention of some researchers (Fuchs, 2012, 2016, 2023; Ordin & Polyanskaya, 2015; Oñate, 2019; Sailaja, 2010, 2012; Van Maastricht et al., 2019; White & Mattys, 2007a).

Rhythm is a component of language prosody that, along with intonation and stress, can distinguish languages from both production and perception perspectives (Cummins & Port, 1996; Gussenhoven & Chen, 2020; Schön & Tillmann, 2015). White and Mattys (2007a) showed that English-French bilinguals exhibit stress-timed rhythms in both languages, attributing this to the dominance of English, a stress-timed language. In contrast, Sailaja (2010, 2012) explored Indian English, spoken by bilinguals of English and various Indian languages, revealing a hybrid rhythm influenced by regional

[1] Based on what is mentioned here regarding the terminology for individuals speak two or more languages, the term 'advanced English learner' is designated for the participants of this study based on the participants information (See 2.1). Conversely, the term 'bilingual' is predominantly utilized in the introduction section to align with the terminology preferred by the original authors.

languages. Fuchs (2012) compared German-French bilinguals to German monolinguals, noting that bilingual group displayed syllable-timed rhythms. Ordin and Polyanskaya (2015) studied Russian-English bilinguals, finding syllable-timed rhythms in English, possibly due to Russian influence. Oñate (2019) investigated Spanish-English bilinguals, who exhibited stress-timed rhythms in Spanish, attributed to English influence. Van Maastricht et al. (2019) and Fuchs (2023) explored rhythmic patterns in Dutch-English and German-French bilinguals respectively, highlighting the varying degrees of cross-linguistic influence based on proficiency and exposure and proposing a bilingual rhythm acquisition model incorporating linguistic and paralinguistic factors.

On the other hand, forensic phoneticians (Dellwo et al., 2007; Dellwo et al., 2015; McDougall, 2004, 2006; Nolan, 2002) reported the influence of speakers on the rhythmic metrics besides the effects of age, style and speech rate on rhythmic metrics (Asadi et al., 2018; Gibbon, 2022; Pellegrino, 2019). To the best knowledge of the authors, no research has been undertaken to examine the rhythmic distinctions between languages employed by advanced language learners, while accounting for the variations arising from different speakers. Therefore, this study aimed to explore the differences between the rhythmic properties of Bengali (L1) and English spoken by Bengali advanced learners of English (L2), using durational rhythmic metrics and controlling for influencing factors such as between-speaker differences, while also minimizing the impact of age, style, and speech rate.

Bengali is one of the 1,683 languages and more than 30,000 dialects spoken in India (Chaudhary, 2009). It belongs to the Eastern Indo-Aryan branch of languages, which evolved from eastern Middle Indo-Aryan dialects of Magadhi Prakrit and Pali. Bengali is the official language of West Bengal, a state in Eastern India, and Bangladesh, a neighboring country (Lewis et al., 2009). It is the main language of Bangladesh (spoken by 142 million people, 98.8% of the population, Bangladesh Bureau of Statistics, 2012) and the official language of three Indian states, West Bengal, Tripura, and Assam (spoken by 97 million people, 8.3% of the population, Office of the Registrar General & Census Commissioner, India, 2011). Bengali is also spoken by many Bengali immigrants (from India and Bangladesh) in the United States, the United Kingdom, the Middle East and other Western countries. Bengali is regarded as the 7th most spoken language in the world; more than 265 million people use Bengali as their first or second language in their daily life (Statista, 2020). On the other hand, India has the largest number of English speakers or learners in the world, both native and nonnative (Visceglia et al., 2009). The data for the current study was produced by speakers from the urban metropolis of Kolkata, India, in West Bengal (Figure 1).

1.1. Phonetic and phonological contrasts between English and Bengali: Implications for rhythmic variation and cross-linguistic influence

English and Bengali have different phonetic and phonological properties that influence their rhythm. According to the phonetic table of British English and Bengali languages,[2] British English has 12 vowels and 24 consonants (Roach, 2009), while Bengali has 7 vowels (not counting nasal vowels) and 32 consonants (Bhowmik & Mandal, 2018; Chatterji, 1926). So British English has a larger vowel system than Bengali, but Bengali has more consonant sounds than British English.

English has complex syllable structures (Roach, 2009); it can have up to three consonants in the onset and four in the coda, like (C)(C)(C)V(C)(C)(C)(C). Bengali can have maximum two vowels as the nucleus (except diphthong), three consonants in an onset and one consonant in a coda, like (C)(C)(C)V(V)(C). Most of the syllables in Bengali are open (Roy et al., 2008).

[2] Here British-English is compared to Bengali since the participants of the study are mostly exposed to British-English rather than the other varieties of English.

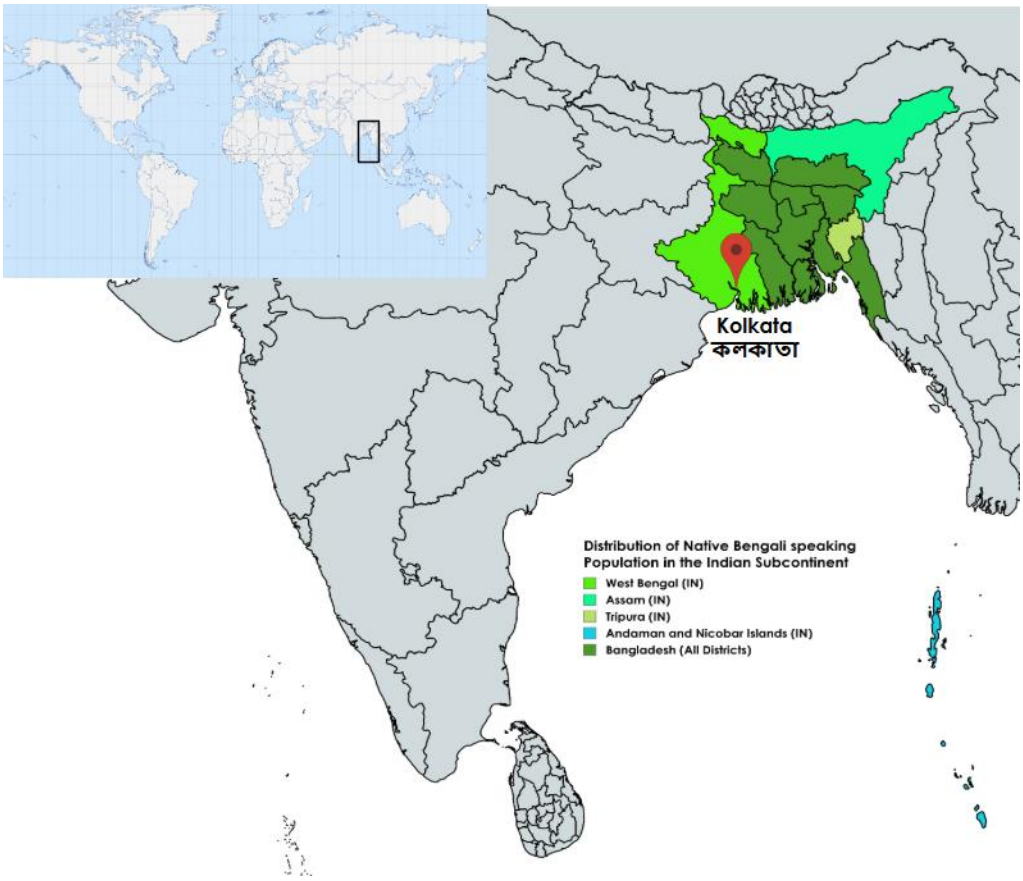

Figure 1. Distribution of native Bengali speaking population in the Indian subcontinent. The participants of the study originate from the locales in and around the city of Kolkata (West Bengal, India) (created with: mapchart.net).

English and Bengali have different rules for stress placement. In English, stress can fall on any syllable of the word (Beckman, 2012). In Bengali, stress is always on the first syllable of a word (Khan, 2008). Therefore, it implies that Bengali is more regular in stressed-interval duration and can have transfer on the English produced by Bengali speakers.

Vowel reduction is a key feature in English, where vowels in unstressed syllables are in a reduced form. The central vowel /ə/ usually occurs in unstressed syllables in English as a result of vowel reduction (Crystal, 2003). Thus, vowel reduction is a process that affects vowel quality and English rhythm by making vowels more central when they are unstressed (Lindblom, 1963). But vowel reduction in unstressed syllables does not happen in Bengali.

The mentioned differences between Bengali and English resulted in some transfer of L1 on L2 of Bengali-English speakers (Payne & Maxwell, 2018; Saha & Mandal, 2018). The example of these transfers includes the lack of differentiation between vowels that are long and short in British-English, in both unstressed and stressed syllables. It was concluded that the general lack of tense-lax vowel in English produced by Bengali advanced English learners is due to the influence of Bengali, which does not have vowel length as a phonological feature, unlike the other L1s (Payne & Maxwell, 2018).

Bengali has contrastive length in consonants but not in vowels (Payne & Maxwell, 2018). Therefore, although English does not have contrastive consonant length, the overall durational rhythmic features of Bengali may influence how Bengali-English speakers produce prosodic structure (Payne & Maxwell, 2018). Nonetheless, consonant lengthening hasn't been reported in Bengali-English (Payne & Maxwell, 2018). Moreover, Saha and Mandal (2015) showed that Bengali speakers may add a vowel to separate English consonant clusters or to prevent a consonant at the end of a syllable.

Saha and Mandal (2018) indicated that native Bengali speakers could adjust the duration, intensity, and F0 of stressed and unstressed vowels in English. However, the results of formant-based analysis indicated that native Bengali speakers had some difficulties in producing vowels with native-like quality in some cases, even though they were fluent in English. This implies that vowel quality is a challenging aspect of English pronunciation for native Bengali speakers which is the result of interference from their native phonology and native vowel system.

1.2. Durational approach to speech rhythm and its application to SLA

Speech rhythm has been explored through three linguistics approaches including durational variabilities in different phonetic intervals, modulations, and prominence (He, 2022). Although durational approaches calculate the variation of the duration of vowels and consonants in speech (Arvaniti, 2012; Dellwo, 2010; Dellwo et al., 2015; Grabe & Low, 2002; Ramus et al., 1999; White & Mattys, 2007a), modulations approaches analyze the changes of the loudness and pitch of speech and extracts the frequencies and phases of different rhythmic units such as syllables and stresses (Gibbon, 2023; Leong et al., 2014; Malisz et al., 2017; O'Dell & Neimenen, 1999; Barbosa, 2002; Tilsen & Johnson, 2008) and prominence approaches examine the intensity or spectral variability of speech and uses them to identify the rhythmic pattern of speech (Cummins & Port, 1998; Lee & Todd, 2004; Todd, 1985). In spite of recent criticism of durational rhythmic metrics (Arvaniti, 2012; Kohler, 2009; Rathcke & Smith, 2015), the cross-linguistic differences in rhythmic patterns between L1 and L2 speech have continued to be investigated through durational variabilities in different phonetic intervals (Fuchs, 2023; Gabriel & Kireva, 2014; Kawase et al., in press; Rodríguez-Vázquez & Roseano, 2023). And as this study aimed at exploring the durational properties of Bengali (L1) and English spoken by Bengali advanced English learner (L2), a brief overview of durational approach regarding language learners is provided.

The durational consideration of speech rhythm categorizes languages in two groups of stress-timed and syllable-timed languages (Dauer, 1983; Ramus et al., 1999, p. 270; Schiering, 2007). Stress-timed languages often have different consonant clusters up to three or four consonants, and vowels in stress-timed languages are often shortened in unstressed syllables. Syllable-timed languages, on the other hand, tend to prevent consonant clusters and vowel reduction. As a result, stress-timed languages usually display more syllable types, mostly with complex onsets and codas, but also open syllables with simple onsets (Dauer, 1983; Ramus et al., 1999; Schiering, 2007). The different syllable types of stress-timed languages cause large variation in the duration of consonant intervals (ΔC), being higher in stress-timed languages and lower in syllable-timed languages (Ramus et al., 1999). Ramus et al. (1999) also proposed a vocalic measure, the percentage of vowel durations in the whole utterance (%V) which was expected to be lower in stress-timed languages than in syllable-timed languages due to the more consonant clusters in stress-timed languages. Ramus et al. (1999) and Nespor et al. (2011) showed that high C-interval variability is reflected by a higher ΔC and high V-interval variability is reflected by a lower %V in stress-timed languages; however, a low ΔC and high %V are the features of syllable-timed languages (Figure 2).

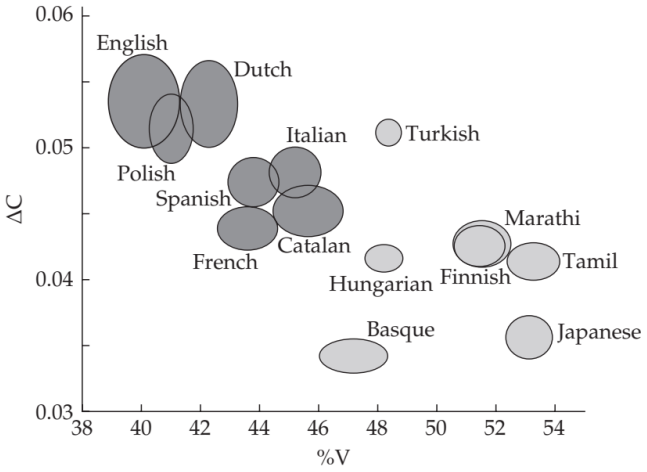

Figure 2. Distribution of languages over %V and ΔC, derived from Nespor et al. (2011), p.1153.

Grabe & Low (2002) introduced the pairwise variability index (PVI) to measure the pairwise durational variability of vocalic and consonantal intervals for 16 different languages. They concluded that stress-timed languages present higher nPVI-V than syllable-timed one. However, Ramus et al. (1999) proposed global rhythm metrics that compare the duration of each interval with the durations of all other intervals in the recording, whereas the PVI, introduced by Grabe and Low (2002), applies local rhythm metrics, comparing the duration of each interval with that of the subsequent interval (Rodríguez-Vázquez & Roseano, 2023).

Dellwo (2010) and Dellwo et al. (2015) proposed other normalization methods to investigate the influence of speech rate on speech rhythmic features, including the coefficient of variation (Varco) as well as the natural logarithmic transformation (ln) which is a method by which numeric data points are expressed as logarithms. The results (Dellwo, 2010) demonstrated that speech rate affects the segmental domain, such as segment duration and segment quality. Dellwo et al. (2012) established a method for measuring variability of intervals between peaks, where these peaks represent the highest points of amplitude in the speech's amplitude envelope. Following that Dellwo et al. (2015) measured ΔPeakLn (Standard deviation of peak intervals based on log transformations of the raw durations) which relates to the sonority dimension of speech rhythm, capturing differences in the degree of sonority contrasts between consonants and vowels in an utterance.

Since speech rhythm is a complex phenomenon that involves multiple acoustic aspects, various rhythm metrics have been suggested to compare the rhythm of L1 and L2. White and Mattys (2007a) recorded native and non-native speakers of French, Spanish, Dutch and English to find out which of seven rhythm metrics (VarcoV, %V, nPVI-V, ΔV, rPVI-C, ΔC, VarcoC) can distinguish best between native speaker groups and between each L1 and the corresponding L2 groups. Their data included three females and three males who read five sentences. Their results showed that VarcoV was the best metric to show significant differences both between L1 groups and between L1 and L2 groups; %V differences among L1 groups were all significant or close to significant, and for L1 vs. L2 comparisons two were significant. nPVI-V among L1 speakers showed very significant differences for all pairs, but for the L1 and L2 comparisons, only one was significant. The other metrics did not perform as well; although rPVI-C showed significant differences between all L1 pairs except those with Dutch, none of the L1 vs. L2 comparisons were significant. ΔV was significant only for Spanish and ΔC only between Spanish and English, and VarcoC showed no significant differences at all.

Jang (2008), Loukina et al. (2011) and White and Mattys (2007a, 2007b) showed that %V, VarcoV and nPVI-V are robust to variation in speech rate and relatively robust to between-speaker and between-sentence variations. Moreover, they indicated that ΔV, ΔC, rPVI-C and VarcoC, do not discriminate well between languages and show poor consistency between speakers and sentences.

1.3. Cross-linguistic influences on speech rhythm in English as a second language varieties

Fuchs (2016) stated that the ESL (English as a Second Language, where English is a secondary language in public sectors, like in India) are proved to be more syllable-timed than ENL (English as a Native Language, where English is the primary language, as in the UK). The reason for a tendency towards syllable-timing in ESL is often attributed to cross-linguistic influence from locally spoken syllable-timed languages (Fuchs, 2023).

Some studies have suggested that Indian-English has more syllable-timed, regardless of the L1s and locations of the speakers (Fuchs, 2016; Gargesh, 2004; Hickey, 2004; Lange, 2009; Masica, 1972) or a more syllable-timed rhythm than the stress-timed rhythm of British-English (Sailaja, 2010, 2012). Fuchs (2012) compared British-English with an educated variety of Indian-English that had Hindi, Bengali, Telugu or Malayalm as their L1. Fuchs (2012) found some evidence that Indian-English

had a syllable-timed speech rhythm, because the vocalic intervals in Indian-English had less variation in their durations than those in British-English.

The other cross-linguistic influence in speech rhythm is the lack of a consistent difference between lax and tense vowels in Indian-English. This is relevant for the measurement of rhythm because some of the metrics that have been proposed, such as the Syllable Ratio, depend on this difference, as do some methods of syllabification (Fuchs, 2016). The rate of speech in syllable-timed languages has also been claimed to be higher, which would imply that Indian-English is faster than British-English (Fuchs, 2016). However, ESL rate of speech is usually lower than ENL (Deschamps, 1980; White & Mattys, 2007a; Raupach, 1980).

Another cross-linguistic influence is sonority and the duration of sonorant sounds that are believed to vary less for syllable-timed languages, therefore the average sonority and the proportion of sonorant duration over the whole speech duration are expected to be higher in syllable-timed languages than in stress-timed languages which has effect on English spoken by Indians (Fuchs, 2016).

Rodríguez-Vázquez and Roseano (2023) conducted a study investigating the acquisition of English speech rhythm by Galician learners and the potential transfer effects from their native Galician rhythm. Their findings shed light on the influence of L1's rhythmic patterns on SLA rhythm and the challenges faced by Galician learners in mastering the stress-timed rhythm of English.

A related phonological feature of some English varieties in the ESL is how glottal stops before vowels at the start of syllables affect the absence of linking between words (Setter, 2000, 2006), which can affect the duration of vocalic and consonantal intervals. As an example, Deterding (2007) proposed that the presence of glottal stops at the beginning of words in Singapore-English might create the perception of a syllable-timed rhythm.

1.4. Research questions and objectives

The current study aims to respond to the following research questions by exploring the differences between the rhythmic properties of Bengali (L1) and English spoken by Bengali advanced English learners (L2), using durational rhythmic metrics produced by educated Bengali advanced English learners:

RQ1: How does language-specific and cross-linguistic factors impact the rhythmic metrics of Bengali (L1) and English spoken by Bengali advanced English learners (L2)?

RQ2: What is the most effective durational metric for distinguishing between L1 and L2 speech rhythms?

Therefore, the first question tackles with understanding the influence of language-specific and cross-linguistic factors on the variability of rhythmic metrics in the speech of Bengali advanced English learners, examining the variability in speech rhythms between the participants' native language and their second language. The second question focused on analyzing durational metrics to determine which one is most effective in differentiating between the rhythmic patterns of the two languages spoken by language learners.

2. Method

To reach the aims of the study we endeavored to minimize the potential influence of confounding factors such as age, style, speech rate and speaker (Asadi et al., 2018; Gibbon, 2022; Pellegrino, 2019). To control for age, we selected speakers from the same age group (see 2.1). Speaking style was standardized by having all participants read the same story (see 2.2). Speech rate was regulated by instructing participants to read at a normal speed (see 2.2). Finally, the impact of individual differences on the metrics was assessed using statistical tests (see 2.3). The information about the participants, experimental task, metrics and statistical tools are as follows.

Participant Initials	Sex	Age	Highest Educational Qualification	School Education & Board	Level of English Proficiency
SC	M	35	PhD (Engineering)	Science (CBSE)	C1
DD	F	35	PhD (Science)	Science (CBSE)	C1
SB	F	25	Bachelor (Engineering)	Science (ICSE)	C1
US	F	34	PhD (Science)	Science (CBSE)	C1
KS	M	37	MBA	Science (WBBSE)	C1
MC	F	26	Masters (Science)	Science (CBSE)	C1
PC	M	27	Masters (Science)	Science (CBSE)	C1
SP	F	35	PhD (Science)	Science (WBCHSE)	B2

Table 1: List of participants' details and educational attributes. Mean age: 31.75 (SD = 4.862, CV = 0.153).

2.1. Participants

The participants of this study were eight Bengali advanced English learners, five females and three males. They all originated from the urban metropolis of Kolkata in West Bengal, India who were pursuing their postgraduate degrees or research. They had all learned English as a second language as they were exposed to English since their childhood and had lived in an English-speaking environment for up to five years. They identified Bengali as their native language, which was the predominant language spoken at home. A list of these participants including their details and educational attributes are provided in Table 1. The Coefficient of Variation (CV) for participant ages is 15.3% (0.153 when expressed as a percentage), signifying that the standard deviation is a modest proportion of the mean age, representing a narrow spread of ages. The age range spans 12 years (from 25 to 37), with a standard deviation of 4.862 years, less than one-sixth of the mean age of 31.75 years, indicating that the ages are closely clustered around the mean (Table 1). Therefore, the impact of age factor was minimized in this study. Also, it should be mentioned that the level of English proficiency was self reported by the participants. To the best of authors' knowledge, these levels are based on proficiency certificates issued by competent authorities like the British Council (https://www.britishcouncil.in/).

2.2. Experimental task

Gibbon (2022) stated that the speech rhythm can change based on the style or context in which language is used. In more formal or structured settings, such as public speeches, poetry recitations, and reading aloud, speech tends to be more rhythmical. On the other hand, in less formal and more spontaneous contexts, such as planning discussions, speech may exhibit less rhythmic regularity. Hence, to minimize the impact of style in rhythmic metrics this study used the English and Bengali versions of the story "The North Wind and the Sun" as the source of speech data as it is well established in the speech-science, repeatedly used for both analyzing sound segments and prosodies, and besides representing a standard text for the phonetic documentation of languages by the International Phonetic Association (Baird et al., 2022). The story "The North Wind and the Sun" contains 6 English sentences and 9 Bengali sentences (Appendix 1). The data collection took place in a quiet room, where the speakers were recorded using a Zoom H4 audio recorder with an external lapel microphone. The recordings had a sampling rate of 44.1 kHz. To minimize the impact of speech rate the most natural reading of each speaker was chosen for further analysis. The selected recordings were divided into the separate sentences and then segmented and annotated in 5 tiers of Praat TextGrids (Boersma & Weenink, 1992–2022). The first tier determined the boundaries of every sound using the WebMAUS services, Munich AUtomatic Segmentation Service MAUS (Kisler et

al., 2017). The second, third, and fourth tiers identified the consonant (C) and vowel (V) intervals, while the fifth tier separated the syllables. The sixth tier indicated the peak intervals, based on the principle of sonority. The second to sixth tiers were automatically generated using Dellwo's script which can be found at https://www.cl.uzh.ch/de/people/team/phonetics/vdellw/software.html. All the segmentations were manually verified and adjusted by the first author of this study (NT).

The data set of this study consists of 72 sentences from L1 (9 sentences × 8 speakers) and 48 sentences from L2 (6 sentences × 8 speakers). Figure 3 presents the TextGrid for the second sentence of the L1 version of the selected story as an example.

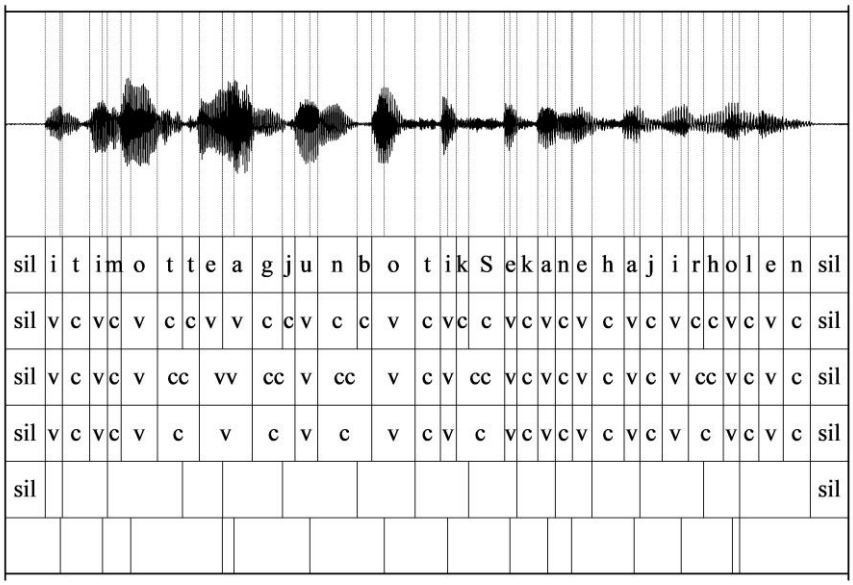

Figure 3. An example of the TextGrid (the second sentence of "The North Wind and the Sun", Bengali version).

2.3. The metrics

The metrics were selected to account for potential language-specific factors in both L1 and L2, as well as the possible cross-linguistic influences of L1 on L2, as indicated by the literature. Considering the distinct syllable structures of English and Bengali (Roach, 2009; Roy et al., 2008), and the fact that a language with a greater variety of syllable types tends to exhibit more variability in the number of consonants and their total duration within the syllable, thereby leading to a higher ΔC (Ramus et al., 1999), we hypothesize that there are differences in their ΔC and nPVI-C values. The former captures the variability of consonant intervals in a speech signal. It is calculated by taking the standard deviation of the durations of consonant intervals in an utterance. Hence, it reflects the degree of variation or irregularity in the distribution of consonant durations in speech. Also, nPVI-C stands for normalized pairwise variability index of consonants, and it is a measure of how much the duration of consonantal intervals varies locally in speech. It is calculated by taking the average of the absolute differences between adjacent consonant intervals, divided by their mean, and multiplied by 100. As the following equation:

$$(1)\quad nPVI\_C = \frac{100}{m-1} \times \sum_{k=1}^{m-1} \left| \frac{d_k - d_{k+1}}{\frac{d_k + d_{k+1}}{2}} \right|$$

Where the number of vowel intervals is presented as m and the duration of vowel intervals is presented as d_k (Grabe & Low, 2002).

Moreover, the presence of two vowels in the nucleus of some Bengali syllables (Roy et al., 2008) leads to the hypothesis that there is a difference in nPVI-V when compared with English, as this may affect the duration of vocalic intervals. nPVI-V is calculated by taking the average of the absolute differences between adjacent vowel intervals, divided

by their mean, and multiplied by 100. It reflects the normalized degree of variation or irregularity in the distribution of vowel durations in speech. As the following equation:

$$(2)\quad nPVI\_V = \frac{100}{m-1} \times \sum_{k=1}^{m-1} \left| \frac{d_k - d_{k+1}}{\frac{d_k + d_{k+1}}{2}} \right|$$

Where the number of consonant intervals is presented as m and the duration of vowel intervals is presented as d_k (Grabe & Low, 2002).

Also, Bengali's two vowels in the nucleus besides the distinct syllable structures of English and Bengali can affect %V which is the reflection of the relative duration of vocalic intervals in speech (Ramus et al., 1999). It is calculated by taking the sum of vowel durations, divided by the total duration of the text, and multiplied by 100. As in the following equation:

$$(3)\quad \%V = \frac{\sum_{i=1}^{n_v} v_i}{\sum_{i=1}^{n_v} v_i + \sum_{i=1}^{n_c} c_i} \times 100\%$$

Where the number of vowel intervals is n_v, is the number of consonant intervals is n_c, is the duration of the vowel is v_i, and c_i is the duration of the consonant (Ramus et al., 1999, Taghva et al., 2023).

As Bengali-English does not have a consistent difference between lax and tense vowels due to the interference of L1, it is plausible that this could result in a varying speech rate (Fuchs, 2016). This makes our hypothesis to calculate rateSyl of L1 and L2. RateSyl captures the average number of syllables per second in our corpus. In addition, inconsistence difference between lax and tense vowels can affect VarcoV which captures the coefficient of variation of vowel intervals. It is calculated by taking the standard deviation of vowel intervals, divided by their mean, and multiplied by 100. It reflects the degree of variation or irregularity in the distribution of vowel durations in an utterance. RateSyl and VarcoV are calculated as follow:

$$(4)\quad rateSyl = \frac{N_{Syl}}{d}$$

Where the number of syllable intervals in the sentence is presented as N_{Syl}, and the sentence duration without considering the pauses is presented as d (Dellwo, 2010; Taghva et al., 2023).

$$(5)\quad \text{VarcoV} = 100 \times \frac{\Delta \text{V}}{\bar{\text{v}}}$$

where ΔV is the standard deviation of vowel intervals and $\bar{V}$ is the mean duration of vowel intervals (Dellwo, 2010).

We investigated ΔPeakLn following the findings of Fuchs (2016), which suggested that sonority and the proportion of sonorant duration over the whole speech duration are expected to be higher in L1 than in L2 (Fuchs, 2016). ΔPeakLn quantifies the variability in peak amplitude/sonority between consonantal and vocalic intervals, which can differ across languages based on their rhythmic timing patterns. It is calculated as:

$$(6)\quad \Delta PeakLn = \sqrt{\frac{n_{Peak} \sum_{i=1}^{InPeak} (lnInPeak_i)^2 - [\sum_{i=1}^{InPeak} (lnInPeak_i)^2]}{N_{InPeak} \cdot (N_{InPeak} - 1)}}$$

Where *InPeak* is peak intervals and *N* is the number of Peak intervals (Dellwo et al., 2015).

Therefore, ΔC, nPVI-C, nPVI-V, VarcoV, %V, rateSyl and ΔPeakLn examined in both L1 and L2.

2.4. Statistical tools

To compare the rhythmic durational features of L1 and L2 of the speakers, we used the mean values of the metrics, as well as linear-mix model, a post hoc Tukey analysis and box plots. All rhythm metrics were calculated using a Praat script created by Dellwo. The script is available on his website [https://www.cl.uzh.ch/de/people/team/phonetics/vdellw.html].

Afterward data analysis was performed using JASP version 0.18.3, as developed by the JASP Team (2024). In this study, we investigated the significance of the language variable within L1 and L2

through the application of linear mixed-effects models. L1 and L2 was considered as a fixed factor, while speakers were treated as random factors to eliminate the effect of each speaker on the metrics. The assessment of the impact being studied involved two key aspects: determining whether it should be treated as a fixed or random effect within the comprehensive model, and examining its absence in the reduced model. Statistical significance was determined via likelihood ratio tests, which compared the full model, including the effect in question, with the reduced model, which excluded it. *P*-values were obtained from these statistical tests. Further, we utilized a post hoc Tukey analysis to conduct comparisons and determine the specific variations in the selected rhythmic metrics between L1 and L2. This statistical procedure allowed us to discern and quantify the significant differences in rhythmic metrics that exist between L1 and L2, providing a more detailed insight into language-specific and cross-linguistic variations.

3. Results

The descriptive analysis of rateSyl, ΔC, nPVI-C, nPVI-V, VarcoV, %V and ΔPeakLn of L1 and L2 are represented in Table 2.

Table 2 shows that L1 had a higher rate of syllables per second (rateSyl) compared to the L2, indicating faster speech tempo in L1. ΔC was lower for L1, suggesting shorter consonants in Bengali speech. Also, nPVI_C was higher for L2, indicating more variable consonant durations in L2. The nPVI and varco measures for vowels (nPVI_V and VarcoV) also showed higher variability in L2 vowel durations compared to L1. The variability of nPVI_V values for L1 is lower than that for L2, meaning that there is less variation in vowel variability among L1 speech than among L2 speech. %V was higher for L1, indicating longer vowel durations relative to consonants in L1 compared to L2. On the other hand, the variability of %V values for L1 are lower than that for L2, meaning that there is less variation or diversity in vowel proportion among Bengali than English spoken by advanced English learners. According to Table 2, the mean ΔPeakLn value for L1 is 0.805, while the mean ΔPeakLn value for L2 is 0.796. This means that both languages have similar levels of peak amplitude variability, with L1 having slightly more variation than L2. However, the difference is not very large. Table 3 shows the significance of rateSyl, ΔC, nPVI-C, nPVI-V, varcoV, %V and ΔPeakLn over L1 and L2 as fixed factors, while speakers were treated as random factors.

Table 3 shows that L1 and L2 has a significant effect on all rhythmic measures, except for ΔPeakLn, which is not significant. The clearest significant effects are observed for rateSyl, ΔC and %V which have *p*-values less than 0.001. Moreover, %V has the highest *F*-statistic value of 83.632, indicating a very strong effect of language on this particular rhythmic measure. The AIC values across the measures further support the model's adequacy in capturing the effects of language on rhythm, with negative values indicating particularly strong model fits for ΔC, VarcoV, and ΔPeakLn.

To further investigate the nuances of cross-linguistic influence in rhythm metrics, we conducted post hoc Tukey tests. These tests allow us to pinpoint specific variations between L1 and L2 each rhythmic measure. The results of the Tukey tests (Table 4) are presented for the following rhythmic metrics.

The Tukey test results from Table 4 highlight the significant impact of L1 and L2 on the rhythmic metrics. The analysis revealed distinct differences between the two languages across multiple dependent variables. Remarkably, the estimates for L1 and L2, along with their corresponding standard errors, *z*-scores, and *p*-values, demonstrated significant variations in rhythmic patterns.

In conclusion, the comparison of speakers and languages, along with their interactions, demonstrates that all the considered metrics show significant differences between L1 and L2, with the exception of ΔPeakLn. Furthermore, %V is the most effective metric for differentiating between L1 and L2. Figure 4 visualizes the distribution of the rhythmic metrics across L1 and L2.

	L1		L2	
Metric	***M***	***SD***	***M***	***SD***
rateSyl	5.083	.574	3.982	.416
ΔC	.049	.009	.075	.013
nPVI_C	59.773	11.043	65.293	10.986
nPVI_V	38.247	7.927	52.895	10.949
varcoV	.370	.067	.460	.085
%V	37.483	3.771	30.778	2.895
ΔPeakLn	.805	.182	.795	.237

Table 2. Descriptive statistics of durational rhythmic measures in Bengali (L1) and English spoken by Bengali advanced English learners (L2).

Metric	***df***	***f***	***AIC***	***p***
rateSyl	1, 6.00	53.215	150.887	< .001
ΔC	1, 6.00	49.538	-619.25	< .001
nPVI-C	1, 15.66	6.020	806.157	.026
VarcoV	1, 6.00	12.890	-237.08	.012
nPVI-V	1, 6.00	21.021	754.150	.004
%V	1, 7.64	83.632	528.489	< .001
ΔPeakLn	1, 17.85	0.063	-28.114	.805

Table 3. Summary of the linear mixed-effect model for rhythmic metrics in L1 and L2 as fixed variables and Speakers as a random-effects grouping factor.

	L1				L2			
Metric	***Estimate***	***SE***	***z***	***p***	***Estimate***	***SE***	***z***	***p***
rateSyl	5.08	015	31.99	< .001	3.98	0.09	44.21	< .001
ΔC	0.05	0.00	30.61	< .001	0.07	0.00	23.01	< .001
nPVI-C	59.77	1.52	39.14	< .001	65.29	2.18	29.85	< .001
VarcoV	0.37	0.01	28.87	< .001	0.46	0.02	18.70	< .001
nPVI-V	38.24	1.12	33.96	< .001	52.89	3.16	16.70	< .001
%V	37.48	1.12	33.44	< .001	30.77	0.69	44.57	< .001
ΔPeakLn	0.805	0.04	19.71	< .001	0.796	0.05	14.17	< .001

Table 4. Tukey Analysis showing statistical summary for the effect of languages (L1 and L2) as fixed model and speakers as a random effect grouping factor on dependent variable (metrics).

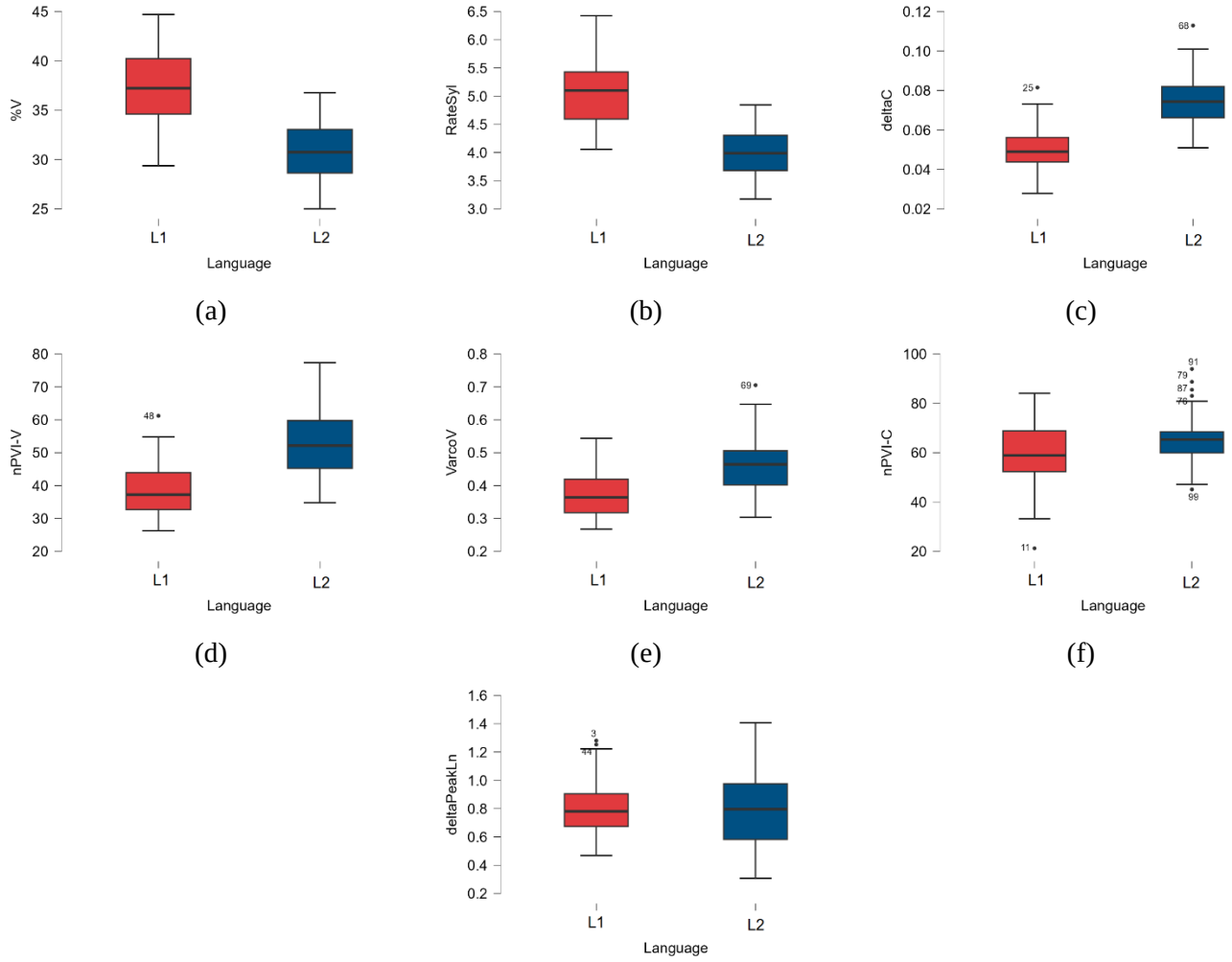

Figure 4. Box plots, visualizing the distribution of seven rhythmic metrics across L1 and L2: (a) %V, (b) rateSyl, (c) ΔC, (d) nPVI-V, (e) varcoV, (f) nPVI-C, (g) ΔPeakLn

4. Discussion

The main aim of this study was to explore the influence of language-specific and cross-linguistic factors on speech rhythm by investigating the differences between durational rhythmic measures of Bengali and English produced by educated Bengali advanced English learners (RQ1). Therefore 7 durational rhythmic metrics (rateSyl, ΔC, nPVI-C, nPVI-V, VarcoV, %V and ΔPeakLn) were analyzed by minimizing the possible influence of speakers, age, style and speech rate.

Regarding the vocalic intervals, our results indicate that L1 demonstrated significantly higher %V compared to L2, indicating that vowels constitute a larger proportion of speech in L1 (Arvaniti, 2012). This finding aligns with Ramus et al. (1999), who observed higher %V in syllable-timed languages compared to stress-timed languages. The analysis of nPVI-V revealed differences between L1 and L2. nPVI-V measures the variability in the duration of vowel sounds in speech, with higher values indicating more timing variation locally (Arvaniti, 2012). The findings suggested that L2 exhibits more variability in the timing of vowel sounds compared to L1, indicating greater flexibility and diversity in vocalic intervals within L2. This observation is consistent with Grabe & Low (2002), who reported higher nPVI-V in stress-timed languages relative to syllable-timed languages. VarcoV, which measures the relative variability in the duration of vocalic intervals when speech rate is controlled, showed that L2 has more diversity in the duration of vocalic sounds compared to L1. A higher VarcoV indicates greater variation in the duration of vocalic sounds relative

to their average duration (Arvaniti, 2012). This implies that English produced by Bengali advanced English learners exhibits more variability in the relative duration of vocalic intervals compared to Bengali. The presence of dual vowels in the nucleus of certain Bengali syllables (Roy et al., 2008), may contribute to the lower nPVI-V and VarcoV, alongside a higher %V in L1 compared to L2. This suggests that the duration of vocalic intervals in L1 is influenced by the structure of Bengali syllables.

Regarding consonantal intervals, our results revealed a higher amount of ΔC in L2 compared to L1. As ΔC quantifies the variability in the duration of consonantal intervals, with higher values indicating greater variation in the length of consonant sounds (Arvaniti, 2012), which means that L2 exhibits more diversity in the duration of consonants compared to L1. This finding aligns with Ramus et al. (1999), who reported higher ΔC values in stress-timed languages compared to syllable-timed languages. Significant differences were also found in nPVI-C between L1 and L2. nPVI-C measures the variability in the duration of consonantal intervals, with higher values indicating greater variation in the length of consonant sounds between neighboring intervals (Arvaniti, 2012). The higher nPVI-C values observed in L2 indicate more variability in the normalized duration of consonantal intervals compared to L1. The greater variety of syllable types in English compared to Bengali may contribute to the higher ΔC and nPVI-C values observed in L2. This suggests that the number of consonants and their duration within syllables varies more in L2 due to the structural differences between the languages. This finding is consistent with Ramus et al. (1999), who observed differences in ΔC and nPVI-C between stress-timed and syllable-timed languages.

Moreover, the rate of syllable production (rateSyl) was found to be higher in L1 compared to L2, indicating a faster speech rate in L1. This may be attributed to the absence of a distinction between lax and tense vowels in L1 (Fuchs, 2016). This finding supports the claim by Fuchs (2016) that syllable-timed languages have a higher speech rate compared to stress-timed ones.

Interestingly, no significant difference was observed in ΔPeakLn between L1 and L2. A higher ΔPeakLn value indicates greater variability in the peak amplitudes between consonants and vowels. Lower values mean the peak amplituds are more similar across consonantal and vocalic intervals (Dellwo et al., 2015). Therefore the similarity of this metric indicates that the sonority and the proportion of sonorant duration over the whole speech duration are similar in both L1 and L2. This suggests that Bengali speakers may not adjust sonorant duration in the same way as in L2, indicating a cross-linguistic influence of L1 on L2 which may be related to the insertion of glottal stops before vowels at the start of syllables, that affects the absence of linking between words (Setter, 2000, 2006), or the inconsistent difference between lax and tense vowels in L2 (Fuchs, 2016). The results also confirmed the observations of Payne & Maxwell (2018), who found that Bengali-English speakers did not differentiate between vowels that are phonetically long and short in British English, in both unstressed and stressed syllables. Though to comprehend the reason for the comparable ΔPeakLn between L1 and L2, further exploration is required.

Thus, the findings showed that Bengali speech has a faster tempo, shorter consonants, and longer vowels relative to consonants compared to English spoken by Bengali advanced English learner. Meanwhile, English spoken by Bengali advanced English learners' speech shows more variable timing of consonants and vowels compared to Bengali. The sonority changes between syllables seem comparable between the two languages. These findings support the hypothesis that speech rhythm is influenced by both language-specific and cross-linguistic factors. The faster tempo, shorter consonants, and longer vowels in Bengali are consistent with the syllable-timed nature of this language, as proposed by Khan (2010). The higher variability of consonant and vowel durations in L2 reflects the influence of English, which is considered a stress-timed language with more complex syllable structures and stress patterns Roach (2009). However, the results suggest that peak amplitude variability is not a relevant feature of speech rhythm for L1 and L2. It shows the

similar sonority changes between syllables in both languages suggest that there is a common underlying mechanism for syllable organization and prominence across L1 and L2 that possibly shows the cross-linguistic impact of L1 on L2.

Furthermore, it aligns with the findings of Saha and Mandal (2018), which indicate that despite their fluency in English, Bengali speakers occasionally struggle to produce vowels with the quality of a native English speaker, suggesting that achieving native-like vowel pronunciation in English poses a significant challenge for native Bengali speakers, likely due to interference from their native phonology and native vowel system.

The results of this study demonstrate that speech rhythm is a multidimensional phenomenon that involves various acoustic features, such as tempo, consonant and vowel duration variability, proportion of vocalic intervals, and sonority changes. The results show that these features vary across languages and across speakers. The results also show that these features are interrelated and influenced by both language-specific and cross-linguistic factors, such as syllable structure, vowel quality, vowel reduction, stress placement, and L1 interference. Therefore, speech rhythm cannot be reduced to a single metric or dimension.

In conclusion, to answer the second question of this study (RQ2) and to seek the most effective durational metric for distinguishing between Bengali and English produced by English advanced learner speech rhythms, the interaction of speakers and languages results suggested that significant differences were observed in all metrics due to the languages, except for ΔPeakLn (Table 4). Overall, the findings support the assertion that VarcoV, %V, nPVI-V, ΔV, rPVI-C, ΔC are effective measures for distinguishing between L1 and L2 (White & Mattys, 2007a, Kohler, 2009; Arvaniti, 2012; Rathcke & Smith, 2015; Gabriel & Kireva, 2014; Rodríguez-Vázquez & Roseano, 2023, Fuchs, 2023, Kawase et al., in press).

Additionally, relevant significant effects are observed for rateSyl, ΔC and %V with *p*-values < 0.001. %V has the highest *F*-statistic value (Table 4), indicating a very strong effect of language on this particular rhythmic measure. Consequently, this study identifies the average proportion of vocalic as the most effective metric for distinguishing between Bengali and English produced by Bengali advanced English learners. Ramus et al. (1999) and Nespor et al. (2011) demonstrated that a greater variety of syllable types in a language leads to increased variability in the number of consonants, resulting in a higher ΔC and lower %V. The comparable peak amplitude variability (ΔPeakLn) suggests some transfer of native sonority patterns into the learners' L2 English production. Hence, the syllable types of the languages play an important role in language-specific and cross-linguistic factors of L1 and L2.

These results illustrate how rhythmic properties are influenced by an intricate mix of language-specific factors like syllable structure and vowel quality, as well as cross-linguistic effects from the speakers' L1 phonological system and experience.

5. Conclusion

This study investigated the influence of language-specific and cross-linguistic factors on speech rhythm by comparing durational rhythmic measures in Bengali (L1) and English spoken by Bengali advanced English learners (L2). The results provide evidence that both language-specific characteristics and cross-linguistic transfer effects shape the rhythmic patterns observed in L1 and L2 speech.

The findings show that L1 speech exhibited characteristics typical of syllable-timed languages, with a faster tempo, shorter consonants, and longer vowels relative to consonants. In contrast, the L2 displayed more variable timing of consonants and vowels, reflecting the stress-timed nature and complex syllable structures of English.

Crucially, the proportion of vocalic intervals (%V) proved most effective at distinguishing the rhythms of L1 and L2, highlighting how syllable structure

and vowel duration features are pivotal in defining a language's rhythmic patterns. However, the sonority changes across syllables (ΔPeakLn) appeared comparable between the two languages. which suggests the presence of an underlying cross-linguistic mechanism governing sonority, where L1 phonological patterns influence the learners' L2 production.

Overall, these findings support the hypothesis that speech rhythm arises from an interplay of language-specific factors, such as the syllable-timed vs. stress-timed typology, as well as cross-linguistic transfer effects from the speaker's native language. While most rhythmic properties differed between L1 and L2, the similarity in sonority patterning points to some rhythmic features being shaped by universal phonological tendencies across languages.

The results of this study have implications for second language teaching and learning, as well as speech recognition and synthesis systems. For second language learners, it may be beneficial to raise their awareness of the rhythmic differences between Bengali and English, and provide them with feedback and practice on how to adjust their speech tempo, consonant and vowel durations, and stress patterns according to the target language. For speech technology developers, it may be useful to incorporate rhythmic features into their models and algorithms, as they may affect the accuracy and naturalness of speech recognition and synthesis across languages.

Ethics statement

This study involved human voices belong to the Bengali advanced English learners. All participants gave their written informed consent to participate in the study after receiving detailed information about the purpose, procedures, risks, benefits, and confidentiality of the study. The study had no potential risks or harms for the participants or others. The potential benefits of the study included advancing our understanding of speech rhythm in SLA and contributing to linguistic theory and practice. The privacy and confidentiality of the participants and their data were protected throughout the study. The participants' names and other identifying information were replaced by codes in the data analysis and reporting. The audio recordings of their speech samples were stored in a password-protected computer and deleted after transcription. The participants were informed that they could withdraw from the study at any time without any consequences.

Conflict of interest statement

There are no conflicts of interest that may have influenced the results obtained or the interpretations proposed.

References

Amengual, M. (2021). The acoustic realization of language-specific phonological categories despite dynamic cross-linguistic influence in bilingual and trilingual speech. *The Journal of the Acoustical Society of America*, *149*(2), 1271–1284. https://doi.org/10.1121/10.0003559

Arvaniti, A. (2012). The usefulness of metrics in the quantification of speech rhythm. *Journal of Phonetics*, *40*(3), 351–373. https://doi.org/10.1016/j.wocn.2012.02.003

Asadi, H., Nourbakhsh, M., He, L., Pellegrino, E., & Dellwo, V. (2018). Between-speaker rhythmic variability is not dependent on language rhythm, as evidence from Persian reveals. *International Journal of Speech, Language and the Law*, *25*(2), 151–174. https://doi.org/10.1558/ijsll.37110

Baird, L., Evans, N., & Greenhill, S. J. (2022). Blowing in the wind: Using 'North Wind and the Sun' texts to sample phoneme inventories. *Journal of the International Phonetic Association*, *52*(3), 453–494. https://doi.org/10.1017/S002510032000033X

Bangladesh Bureau of Statistics (2012). Population and Housing Census—2011 (Vol. 3: Socio-economic and demographic report). Ministry of Planning, Dhaka.

Barbosa, P. A. (2002). Explaining cross-linguistic rhythmic variability via a coupled-oscillator model of rhythm production. In D. Hirst & M. Rossi (Eds.), *Proceedings of Speech Prosody 2002* (pp. 163–166). International Speech Communication Association. https://doi.org/10.21437/SpeechProsody.2002-26

Beckman, M. E. (2012). *Stress and non-stress accent*. De Gruyter Mouton.

https://doi.org/10.1515/9783110874020 (Original work published 1996)

Bhatia, T. (2006). Introduction. In T. Bhatia & W. Ritchie (Eds.), *The Handbook of Bilingualism* (pp. 5–6). Blackwell. https://doi.org/10.1002/9780470756997.part1

Bhatia, T. & Ritchie, W. (Eds.) (2006). *The Handbook of Bilingualism.* Blackwell. https://doi.org/10.1002/9780470756997

Bhowmik, T., & Mandal, S. K. D. (2018). Manner of articulation based Bengali phoneme classification. *International Journal of Speech Technology*, *21*(2), 233–250. https://doi.org/10.1007/s10772-018-9498-5

Boersma, P., & Weenink, D. (1992–2022). Praat: doing phonetics by computer (Version 6.3.03) [Computer program]. http://www.praat.org/

Chatterji, S. K. (1926). *The origin and development of the Bengali language* (Part 2). Calcutta University Press.

Chaudhary, S. (2009). *Foreigners and foreign languages in India: A sociolinguistic history*. Cambridge India. https://doi.org/10.1017/UPO9788175968493

Crystal, D. (2003). *English as a global language*. Cambridge university press. https://doi.org/10.1017/CBO9780511486999

Cummins, F., & Port, R. (1996). Rhythmic commonalities between hand gestures and speech. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *18*, 415–419.

Cummins, F., & Port, R. (1998). Rhythmic constraints on stress timing in English. *Journal of Phonetics*, *26*(2), 145–171. https://doi.org/10.1006/jpho.1998.0070

Dauer, R. M. (1983). Stress-timing and syllable-timing reanalyzed. *Journal of phonetics*, *11*(1), 51–62. https://doi.org/10.1016/S0095-4470(19)30776-4

Dellwo, V. (2010). *Influences of speech rate on the acoustic correlates of speech rhythm: An experimental phonetic study based on acoustic and perceptual evidence* [Doctoral dissertation, Universität Bonn]. bonndoc - The Repository of the University of Bonn. https://bonndoc.ulb.uni-bonn.de/xmlui/handle/20.500.11811/4239

Dellwo, V., Huckvale, M., & Ashby, M. (2007). How is individuality expressed in voice? An introduction to speech production and description for speaker classification. In C. Müller (Ed.), *Speaker classification I: Fundamentals, features, and methods* (pp. 1–20). https://doi.org/10.1007/978-3-540-74200-5_1

Dellwo, V., Leemann, A., & Kolly, M.-J. (2012). Speaker idiosyncratic rhythmic features in the speech signal. In R. Sproat (Ed.), *Proceedings of Interspeech 2012*. International Speech Communication Association. https://doi.org/10.21437/Interspeech.2012-342

Dellwo, V., Leemann, A., & Kolly, M. J. (2015). Rhythmic variability between speakers: Articulatory, prosodic, and linguistic factors. *The Journal of the Acoustical Society of America*, *137*(3), 1513–1528. https://doi.org/10.1121/1.4906837

Deschamps, A. (1980). The syntactical distribution of pauses in English spoken as a second language by French students. In H. Dechert & M. Raupach (Eds.), *Temporal variables in speech* (pp. 255–262). De Gruyter Mouton. https://doi.org/10.1515/9783110816570.255

Deterding, D. (2007). *Singapore English*. Edinburgh University Press.

Edwards, J. (2006). Foundations of bilingualism. In T. Bhatia & W. Ritchie (Eds.), *The Handbook of Bilingualism* (pp. 7–31). Blackwell. https://doi.org/10.1002/9780470756997.ch1

Flege, J., Schirru, C., & MacKay, I. (2003). Interaction between the native and second language phonetic subsystems. *Speech Communication*, *40*(4), 467–491. https://doi.org/10.1016/S0167-6393(02)00128-0

Fuchs, R. (2012, July 28). *A duration-based account of speech rhythm in Indian English* [Poster presentation]. LabPhon 13: The 13th Conference on Laboratory Phonology, Stuttgart, Germany.

Fuchs, R. (2016). *Speech rhythm in varieties of English: Evidence from educated Indian English and British English*. Springer. https://doi.org/10.1007/978-3-662-47818-9

Fuchs, R. (2023). A synthesis of research on speech rhythm in native, learner and second language varieties of English: Introduction to the Volume. In R. Fuchs (Ed.), *Speech Rhythm in Learner and Second Language Varieties of English* (pp. 1–14). Springer. https://doi.org/10.1007/978-981-19-8940-7_1

Gabriel, C., & Kireva, E. (2014). Speech rhythm and vowel raising in Bulgarian Judeo-Spanish. In N. Campbell, D. Gibbon, & D. Hirst (Eds.), *Proceedings of Speech Prosody 2014* (pp. 728–732). https://doi.org/10.21437/speechprosody.2014-133

Gargesh, R. (2004). Indian English: Phonology. *A handbook of varieties of English* (Vol. 1, pp. 992–1002). De Gruyter Mouton. https://doi.org/10.1515/9783110197181-064

Gass, S., Behney, J., & Plonsky, L. (2020). *Second language acquisition: An introductory course* (5th ed.). Routledge. https://doi.org/10.4324/9781315181752

Gibbon, D. (2022). New perspectives on Ibibio speech rhythm. In M. E. Ekpenyong & I. I. Udoh (Eds.), *Current issues in descriptive linguistics and digital humanities* (pp. 457–486). Springer. https://doi.org/10.1007/978-981-19-2932-8_34

Gibbon, D. (2023). The rhythms of rhythm. *Journal of the International Phonetic Association*, *53*(1), 233–265. https://doi.org/10.1017/S0025100321000086

Grabe, E., & Low, E. L. (2002). Durational variability in speech and the rhythm class hypothesis. In C. Gussenhoven & N. Warner (Eds.), *Laboratory Phonology 7* (pp. 515–546). De Gruyter Mouton. https://doi.org/10.1515/9783110197105

Gussenhoven, C., & Chen, A. (Eds.) (2020). *The Oxford handbook of language prosody*. Oxford University Press.

He, L. (2022). Characterizing first and second language rhythm in English using spectral coherence between temporal envelope and mouth opening-closing movements. *The Journal of the Acoustical Society of America*, *152*(1), 567–579. https://doi.org/10.1121/10.0012694

Hickey, R. (Ed.). (2005). *Legacies of colonial English: Studies in transported dialects*. Cambridge University Press. https://doi.org/10.1017/CBO9780511486920

Jang, T. Y. (2008). Speech rhythm metrics for automatic scoring of English speech by Korean EFL learners. *Malsori*, (66), 41–59.

Jarvis, S. & Pavlenko, A. (2008). *Crosslinguistic influence in language and cognition*. Routledge.

JASP Team (2024). JASP (Version 0.18.3) [computer software]. https://jasp-stats.org/

Kawase, S., Davis, C., & Kim, J. (in press). Impact of Japanese L1 rhythm on English L2 speech. *Language and Speech*. https://doi.org/10.1177/00238309241247210

Khan, S. U. D. (2008). *Intonational phonology and focus prosody of Bengali* [Unpublished doctoral dissertation]. University of California.

Khan, S. U. D. (2010). Bengali (Bangladeshi Standard). *Journal of the International Phonetic Association*, *40*(2), 221–225. https://doi.org/10.1017/S0025100310000071

Kinoshita, N., & Sheppard, C. (2011). Validating acoustic measures of speech rhythm for second language acquisition. In W.-S. Lee & E. Zee (Eds.), *17th International Congress of Phonetic Sciences* (Vol. 17, pp. 1086–1089). International Phonetic Association.

Kisler, T., Reichel, U., & Schiel, F. (2017). Multilingual processing of speech via web services. *Computer Speech & Language*, *45*, 326–347. https://doi.org/10.1016/j.csl.2017.01.005

Kohler, K. (2009). Rhythm in speech and language. *Phonetica*, *66*(1–2), 29–45. https://doi.org/10.1159/000208929

Kroll, J., & Sunderman, G. (2003). Cognitive processes in second language learners and bilinguals: The development of lexical and conceptual representations. In C. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 104–129). Blackwell. https://doi.org/10.1002/9780470756492.ch5

Lange, C. (2009). Review of Indian English, by Pingali Sailaja. In R. Singh (Ed.), *Annual Review of South Asian Languages and Linguistics* (pp. 213–216). De Gruyter Mouton.

Lee, C., & Todd, N. M. (2004). Towards an auditory account of speech rhythm: application of a model of the auditory 'primal sketch' to two multi-language corpora. *Cognition*, *93*(3), 225–254. https://doi.org/10.1016/j.cognition.2003.10.012

Leong, V., Stone, M. A., Turner, R., & Goswami, U. (2014). A role for amplitude modulation phase relationships in speech rhythm perception. *The Journal of the Acoustical Society of America*, *136*(1), 366–381. https://doi.org/10.1121/1.4883366

Lewis, P. (Ed.) (2009). *Ethnologue: Languages of the World* (16th ed.). SIL International. https://www.ethnologue.com

Lindblom, B. (1963). Spectrographic study of vowel reduction. *The Journal of the Acoustical Society of America*, *35*(11), 1773–1781. https://doi.org/10.1121/1.1918816

Loukina, A., Kochanski, G., Rosner, B., Keane, E., & Shih, C. (2011). Rhythm measures and dimensions of durational variation in speech. *The Journal of the Acoustical Society of America*, 129(5), 3258–3270. https://doi.org/10.1121/1.3559709

Malisz, Z., O'Dell, M., Nieminen, T., & Wagner, P. (2017). Perspectives on speech timing: Coupled oscillator modeling of Polish and Finnish. *Phonetica*, *73*(3–4), 229–255. https://doi.org/10.1159/000450829

Masica, C. (1972). The sound system of Indian English (Monograph no. 7). Central Institute of English and Foreign Languages.

McDougall, K. (2004). Speaker-specific formant dynamics: An experiment on Australian English/aI. *International Journal of Speech, Language and the Law*, *11*(1), 103–130. https://doi.org/10.1558/sll.2004.11.1.103

McDougall, K. (2006). Dynamic features of speech and the characterization of speakers: Toward a new approach using formant frequencies. *International Journal of Speech, Language and the Law*, *13*(1), 89–126. https://doi.org/10.1558/ijsll.v13i1.89

Nespor, M., Shukla, M., & Mehler, J. (2011). Stress-timed vs. syllable-timed languages. *The Blackwell companion to phonology: Volume II. Suprasegmental and Prosodic Phonology* (pp. 1147–1159). Wiley. https://doi.org/10.1002/9781444335262.wbctp0048

Nolan, F. (2002). Intonation in speaker identification: An experiment on pitch alignment features. *International Journal of Speech, Language and the Law*, *9*(1), 1–21. https://doi.org/10.1558/sll.2002.9.1.1

O'Dell, M., & Nieminen, T. (1999). Coupled oscillator model of speech rhythm. In J. Ohala, Y. Hasegawa, M. Ohala, D. Granville, & A. Bailey (Eds.), *Proceedings of the 14th International Congress of Phonetic Sciences* (pp. 1075–1078). International Phonetic Association.

Office of the Registrar General & Census Commissioner, India. (2011). Census of India 2011: Provisional population totals.

Oñate, K. C. (2019). *Testing the effect of synchronous speech tasks in the production of L2 speech rhythm in learners of Spanish as a second language* [Doctoral dissertation, Pontificia Universidad Catolica de Chile]. Repositorio UC. https://doi.org/10.7764/tesisUC/LET/27208

Ordin, M., Polyanskaya, L., & Ulbrich, C. (2011). Acquisition of timing patterns in second language. In P. Cosi & R. De Mori (Eds.), *Proceedings of Interspeech 2011* (pp. 1129–1132). International Speech Communication Association. https://doi.org/10.21437/Interspeech.2011-336

Payne, E., & Maxwell, O. (2018). Durational variability as a marker of prosodic structure in Indian English(es). In K. Klessa, J. Bachan, A. Wagner, M. Karpiński, & D. Śledziński (Eds.), *Proceedings of Speech Prosody 2018* (pp. 749–753). International Speech Communication Association. https://doi.org/10.21437/SpeechProsody.2018-152

Pellegrino, E. (2019). The effect of healthy aging on within-speaker rhythmic variability: A case study on Noam Chomsky. *Loquens*, *6*(1), e060. https://doi.org/10.3989/loquens.2019.060

Ramus, F., Nespor, M., & Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition*, *73*(3), 265–292. https://doi.org/10.1016/S0010-0277(00)00101-3

Rasier, L., & Hiligsmann, P. (2007). Prosodic transfer from L1 to L2: Theoretical and methodological issues. *Nouveaux Cahiers de Linguistique Française*, *28*, 41–66.

Rathcke, T. & Smith, R. (2015). Speech timing and linguistic rhythm: On the acoustic bases of rhythm typologies. *The Journal of the Acoustical Society of America*, *137*(5), 2834–2845. https://doi.org/10.1121/1.4919322

Raupach, M. (1980). Temporal variables in first and second speech production. In H. Dechert & M. Raupach (Eds.), *Temporal variables in speech* (pp. 255–262). Mouton. https://doi.org/10.1515/9783110816570.263

Roach, P. (2009). English phonetics and phonology: A practical course (4th ed.). Cambridge University Press, 10–52.

Rodríguez-Vázquez, R., & Roseano, P. (2023). L2 rhythm acquisition and rhythmic transfer in Galician learners of English: A pilot study. *Loquens*, *9*(1–2), e087. https://doi.org/10.3989/loquens.2022.087

Roy, R., Basu, T., Saha, A., Basu, J., & Mandal, S. K. D. (2008). Duration modeling for Bangla text to speech synthesis system. In M. Rayanakorn & K. T. Lua (Eds.), *Proceedings of IALP 2008: International Conference on Asian Language Processing*. Chinese and Oriental Languages Information Processing Society.

Saha, S. N., & Mandal, S. K. D. (2015). Study of acoustic correlates of English lexical stress produced by native (L1) Bengali speakers compared to native (L1) English speakers. In S. Möller & H. Ney (Eds.), *Proceedings of Interspeech 2015* (pp. 815–819). International Speech Communication Association. https://doi.org/10.21437/Interspeech.2015-257

Saha, S. N., & Mandal, S. K. D. (2018). Phonetic realization of English lexical stress by native (L1) Bengali speakers compared to native (L1) English speakers. *Computer Speech & Language*, *47*, 1–15. https://doi.org/10.1016/j.csl.2017.06.006

Sailaja, P. (2010). The standard, (non-)rhoticity and rhythm in Indian English: A response to Lange. In Rajendra Singh (Ed.), *Annual Review of South Asian Languages and Linguistics: 2010* (pp. 183–186). https://doi.org/10.1515/9783110254976.183

Sailaja, P. (2012). Indian English: Features and sociolinguistic aspects. *Language and Linguistics Compass*, *6*(6), 359–370. https://doi.org/10.1002/lnc3.342

Schiering, R. (2007). The phonological basis of linguistic rhythm: cross-linguistic data and diachronic interpretation. *STUF-Language Typology and Universals*,

60(4), 337–359. https://doi.org/10.1524/stuf.2007.60.4.337

Schön, D., & Tillmann, B. (2015). Short-and long-term rhythmic interventions: Perspectives for language rehabilitation. *Annals of the New York Academy of Sciences*, *1337*(1), 32–39. https://doi.org/10.1111/nyas.12635

Setter, J. E. (2000). *Rhythm and timing in Hong Kong English* [Unpublished doctoral dissertation]. University of Reading.

Setter, J. (2006). Speech rhythm in world Englishes: The case of Hong Kong. *Tesol Quarterly*, *40*(4), 763–782. https://doi.org/10.2307/40264307

Statista (2020). The most spoken languages worldwide 2023. https://www.statista.com/statistics/266808/the-most-spoken-languages-worldwide/

Taghva, N., Moloodi, A., Abolhasanizadeh, V., & Tabei, R. (2023). A corpus study of durational rhythmic measures in the Kalhori variety of Kurdish. *Loquens*, *10*(1–2), e098. https://doi.org/10.3989/loquens.2023.e098

Tilsen, S., & Johnson, K. (2008). Low-frequency Fourier analysis of speech rhythm. *The Journal of the Acoustical Society of America*, *124*(2), EL34-EL39. https://doi.org/10.1121/1.2947626

Todd, N. P. M. (1994). The auditory "primal sketch": A multiscale model of rhythmic grouping. *Journal of New Music Research*, *23*(1), 25–70. https://doi.org/10.1080/09298219408570647

Valdés, G. (2001). Heritage language students: Profiles and possibilities. In J. Peyton, D. Ranard, & S. McGinnis (Eds.), *Heritage languages in America: Preserving a national resource*. Center for Applied Linguistics.

Van Maastricht, L., Krahmer, E., Swerts, M., & Prieto, P. (2019). Learning direction matters: A study on L2 rhythm acquisition by Dutch learners of Spanish and Spanish learners of Dutch. *Studies in Second Language Acquisition*, *41*(1), 87–121. https://doi.org/10.1017/S0272263118000062

Visceglia, T., Tseng, C. Y., Kondo, M., Meng, H., & Sagisaka, Y. (2009). Phonetic aspects of content design in AESOP (Asian English Speech cOrpus Project). In *Oriental COCOSDA 2009: International Conference on Speech Database and Assessments* (pp. 60–65). IEEE. https://doi.org/10.1109/ICSDA.2009.5278376

White, L., & Mattys, S. L. (2007a). Calibrating rhythm: First language and second language studies. *Journal of Phonetics*, *35*(4), 501–522. https://doi.org/10.1016/j.wocn.2007.02.003

White, L., & Mattys, S. L. (2007b). Rhythmic typology and variation in first and second languages. In P. Prieto, J. Mascaró, & M.-J. Solé (Eds.), *Segmental and prosodic issues in Romance phonology* (pp. 237–257). John Benjamins. https://doi.org/10.1075/cilt.282.16whi

Appendix

A. The English version of "The North Wind and the Sun"

The North Wind and the Sun were disputing which was the stronger, when a traveler came along wrapped in a warm cloak.

They agreed that the one who first succeeded in making the traveler take his cloak off should be considered stronger than the other.

Then the North Wind blew as hard as he could, but the more he blew the more closely did the traveler fold his cloak around him;

and at last the North Wind gave up the attempt.

Then the Sun shined out warmly, and immediately the traveler took off his cloak.

And so the North Wind was obliged to confess that the Sun was the stronger of the two.

B. "The North Wind and the Sun" in Bengali transcription.

উত্তরে হাওয়া এবং সূর্য একবার তর্ক করছিল যে তাদের মধ্যে কে বেশি শক্তিশালি। ইতিমধ্যে একজন পথিক সেখানে হাজির হলেন। তিনি একটি গরম চাদর জড়িয়ে ছিলেন।

তাকে দেখে তারা পরামর্শ করলো, তাদের মধ্যে যে সর্বপ্রথম পথিক কে গরম চাদরটি খুলে রাখতে বাধ্য করবে, তাকেই দুজনের মধ্যে বেশি শক্তিশালি বলে মানা হবে।

এরপর উত্তরে হাওয়া যতটা সম্ভব জোরে বইতে আরম্ভ করে । কিন্তু সে যত জোরে বয়, পথিক তার চাদরটিকে আরো আষ্টে-পৃষ্টে জড়িয়ে ধরে।

শেষমেশ উত্তরে হাওয়া থেমে যেতে বাধ্য হয়।

তারপর, সূর্য তার উজ্জ্বল কিরনের উত্তাপ ছড়াল এবং অবিলম্বে পথিক তার চাদরটি খুলে রাখল।

তখন উত্তরে হাওয়া বিনীত ভাবে স্বীকার করল যে তাদের মধ্যে সূর্যই বেশি শক্তিশালি ।

C. The phonemic transcription of "The North Wind and the Sun" in Bengali

uttore hawɑ ebong ʃurjo ekbɑr torko kortʃilo dʒe tɑder moddhej ki beʃi ʃoktiʃali. itimodhe ekjon pothik ʃekhane hajir holen. tini ekti groom ʃaɑdor dʒoriye ʧilen.

tɑkej dekhe tara poramorʃo korlo, tɑder modhej dʒe sorboprothom pothik ke groom ʃadorti khule rakhtej bɑddho korbe, tɑkei dudʒoner modhej beʃi ʃoktiʃali bole manɑ hobe.

erpor uttore hawɑ dʒototɑ ʃombhob dʒore boitej aarambho korej. kintu ʃey dʒoto ʃorej boje, pothik tar ʧadortike ɑro ɑʃthej-priʃthej dʒorije dhorej.

ʃeʃmeʃ uttorey hawɑ themej dʒetej baddho hoje.

tarpor, ʃurdʒo tar udʒdʒol kironer uttap ʧoralo ebong obilombe pothik tar ʧadorti khule rakhlo.

toxon uttore hawɑ binito bhabe ʃikar korlo dʒe tɑder modhej ʃurdʒoj beʃi ʃoktiʃali.