Korean Journal of English Language and Linguistics, **Vol 25, June 2025, pp. 894-911**
DOI: 10.15738/kjell.25..202506.894

KOREAN JOURNAL OF ENGLISH LANGUAGE AND LINGUISTICS

ISSN: 1598-1398 / e-ISSN 2586-7474

http://journal.kasell.or.kr

An Envelope-Based Analysis of Utterance Rhythmicity in Korean-English Bilinguals

Seung-Eun Kim (Northwestern University)

Received: May 13, 2025
Revised: June 10, 2025
Accepted: June 17, 2025

Kim, Seung-Eun
Postdoctoral Fellow,
Department of Linguistics,
Northwestern University
2016 Sheridan Road, Evanston,
IL 60208, USA
Email:
seungeun.kim@northwestern.edu

ABSTRACT

Kim, Seung-Eun. 2025. An envelope-based analysis of utterance rhythmicity in Korean-English bilinguals. *Korean Journal of English Language and Linguistics* 25, 894-911.

This study investigated rhythmicity in the speech of Korean-English bilinguals and English monolinguals, utilizing envelope-based rhythm metrics. Unlike most previous studies that analyzed consonantal and vocalic intervals to study rhythm, this study examined the stability of syllabic- and stress-related oscillations derived from the amplitude envelope of filtered speech. The rhythm metrics were obtained from short and simple Korean and English sentences, and the average metrics calculated for each speaker and language were analyzed. The results found a significant difference in foot-level rhythmicity between L1 Korean and L1 English: specifically, L1 English speakers were in general more rhythmic at the foot-level than L1 Korean speakers. In addition, L2 English exhibited an intermediate rhythmic pattern, which was not significantly different from either L1 Korean or L1 English. Analysis of L1 and L2 rhythmicity within bilinguals found a correlation between the two measures, suggesting that the bilingual's L1 rhythmicity predicts their L2 rhythmicity at the foot-level. Analyses of the syllable-level rhythmic metric did not exhibit systematic patterns in all comparisons. Overall, this study adopted relatively underexplored metrics to characterize rhythm in L1 and L2 speech, highlighting the need for their broader application across diverse speaker groups and speech materials.

KEYWORDS

speech production, second-language speech, rhythmicity, amplitude envelope, syllable-level oscillations, stress-related oscillations, Korean-English bilinguals

1. Introduction

Traditional descriptions of speech rhythm have categorized languages into stress-timed, syllable-timed, and mora-timed, based on the notion of isochrony. The main hypothesis underlying these categories is that rhythm arises from a regular recurrence of a speech unit that has equal duration (Abercrombie 1967, Pike 1945). In stress-timed languages such as English or German, that speech unit is the stress-delimited foot, while in syllable-timed languages like Spanish or French, the syllable is the unit; in case of mora-timed languages like Japanese, the speech unit is the mora. The hypothesis of isochrony (and also the validity of different rhythmic categories) has been empirically tested both in speech production and perception, and yet, previous studies have failed to find concrete evidence for isochrony; see Arvaniti (2012) for a review.

Despite limited empirical support, the rhythmic categories have continued to be adopted by researchers studying speech rhythm. In particular, researchers have proposed various *rhythm metrics* (e.g., %V, ΔV, ΔC, rPVI, nPVI, VarcoV, VarcoC) and compared them across different languages to examine whether they reflect a distinction among stress-timed, syllable-timed, and mora-timed languages. Ramus et al. (1999), for example, proposed three rhythm metrics that are based on consonantal and vocalic intervals in an utterance: %V (the proportion of vocalic intervals within an utterance; i.e., the sum of vocalic intervals divided by the total utterance duration), ΔV (the standard deviation of vocalic intervals within an utterance), ΔC (the standard deviation of consonantal intervals within an utterance). Grabe and Low (2002) (see the related metrics proposed in Low et al. 2000), on the other hand, proposed a metric called Pairwise Variability Index (PVI). The PVI calculates differences between pairs of successive consonantal or vocalic intervals, which could be analyzed in raw values (rPVI; mostly for consonantal intervals) or in normalized values controlling for speech rate variation (nPVI; for vocalic intervals). Another set of metrics is VarcoV and VarcoC, which were proposed by Dellwo (2006) as a revision to the metrics of Ramus et al. (1999); these measures are the normalized standard deviations of consonantal and vocalic intervals (i.e., the standard deviation divided by the mean).

These *interval*-based rhythm metrics have been tested on a variety of languages; however, the resulting classifications did not align with the traditional rhythmic categories in some cases, even for languages that are considered as canonical examples of stress-timed or syllable-timed rhythm (e.g., Grabe and Low 2002, White and Mattys 2007). In addition, the classification results often varied depending on which metrics were used. Arvaniti (2012), for example, conducted a comprehensive study of rhythm, analyzing six different interval-based metrics – %V, ΔC, rPVI, nPVI, VarcoV, VarcoC – on speakers of six languages (eight speakers per language): English and German (prototypical stress-timed languages), Italian and Spanish (syllable-timed), and Korean and Greek (non-prototypical). Data were also collected via three different elicitation tasks: sentence reading, passage reading, and spontaneous speech. Arvaniti (2012) observed that different rhythm metrics did not classify the languages in the same way. For example, for the consonantal metrics, ΔC significantly distinguished German, Korean, and Italian, but no differences were found in these languages when VarcoC was analyzed. A similar pattern was observed for the vocalic measures. Arvaniti (2012) also demonstrated a substantial inter-speaker variation as well as robust effects of the elicitation task, further complicating the interpretation of the studies involving the interval-based rhythm metrics.

In this context, this study adopts an alternative way of characterizing speech rhythm and examines cross-linguistic differences. Namely, instead of utilizing rhythm metrics that are based on consonantal or vocalic intervals (durations of sequences of sounds), this study utilizes metrics that are based on *speech amplitude envelope*, that are developed by Tilsen and Arvaniti (2013). In this method, speech waveforms are filtered, and amplitude envelopes are obtained from the filtered waveforms. The envelopes then undergo the process of Empirical Mode

Decomposition (Huang et al. 1998). The output of this process is a set of intrinsic mode functions (IMFs); the first two IMFs are assumed to reflect syllable- (first IMF) and stress-driven (second IMF) fluctuations in the envelope. The Hilbert transform is then applied to each IMF, from which the instantaneous frequencies are obtained. The present study particularly focuses on the variance of the instantaneous frequencies of the first (var. ω_1) and second IMFs (var. ω_2), which is considered to represent the stability/periodicity of syllabic- and stress-related oscillations (i.e., rhythmicity at each time-scale), respectively. Specific technical details of Tilsen and Arvaniti (2013) and some practical advantages (especially compared to the interval-based measures) are presented in Section 2 below.

Using the *envelope*-based rhythmic metrics, this study examines rhythmic patterns of the two languages: Korean and English. While English is a prototypical stress-timed language, the rhythmic status of Korean is not straightforward. Previous studies that examined interval-based metrics in production data or listeners' segmentation of speech suggest that Korean aligns most closely with syllable-timing (e.g., Kim et al. 2008, Mok and Lee 2008); yet, the studies also noted that Korean cannot be definitively categorized as a syllable-timed language. Moreover, some researchers found evidence for stress-timing (e.g., Lee and Seong 1996, Lee et al. 1994) or mora-timing (Cho 2004) in Korean. The current study thus examines var. ω_1 and var. ω_2, which represent rhythmicity at the level of syllable and foot, respectively, in Korean and compare them with English. Given that Korean is (relatively) syllable-timed, the variance of the frequencies of the first IMF (var. ω_1) would be smaller in Korean (i.e., more stable syllabic oscillations) than in English; yet, this may not hold true due to the complex nature of Korean rhythm. On the contrary, English would have a smaller variance in the frequencies of the second IMF (var. ω_2) (i.e., more periodic supra-syllabic oscillations) than Korean.

The second question that is examined in the current study is the rhythmicity of second-language (L2) English produced by L1 Korean speakers. L2 speakers can exhibit a range of rhythmic patterns. For instance, when the rhythmic properties of L1 and L2 differ (e.g., English vs. Spanish), L2 speakers may adapt to the rhythm of L2 (e.g., L1 English speakers exhibiting syllable-timing rather than stress-timing when producing L2 Spanish), but they may also retain L1-like rhythmic features (e.g., maintaining stress-timing in L2 Spanish). Previous studies on L2 rhythm have observed both patterns, which were often related to speakers' level of L2 proficiency or types of speech materials (e.g., spontaneous vs. read speech) (e.g., Lee and Song 2019, Lin and Wang 2005, Low et al. 2000, Oh and Park 2024, Wenk 1985). Some researchers found that L2 speakers exhibit neither L1-like nor L2-like rhythm but instead produce intermediate patterns (e.g., Wenk 1985, White and Mattys 2007).

In terms of L2 English produced by L1 Korean speakers, previous studies found that L2 English exhibits rhythmic patterns that are distinct from Korean. For example, Oh and Park (2024) elicited Korean speakers reading the story *The North Wind and the Sun* (read speech) and retelling the story (spontaneous speech) in Korean and English and calculated the normalized Pairwise Variability Index for vocalic intervals (nPVI-V). Note that nPVI-V is generally higher in stress-timed languages than in syllable-timed languages, due to frequent alternation between stressed and unstressed vowels in the former. In their analysis, Oh and Park (2024) found significantly lower durational variability in L1 Korean compared to L2 English (across speech styles), suggesting that their speakers produced more English-like rhythm rather than Korean-like rhythm in the L2 mode. The durational variability of L2 English also exhibited a significant difference from L1 English – interestingly, the former exhibited greater variability than the latter; this suggests that Korean speakers produced more English-like rhythm in L2 English compared to L1 English speakers.

On the other hand, Kim et al. (2007) (see the related study of Lee and Kim 2005) investigated pairwise variability of vocalic intervals and syllables in L2 English produced by L1 Korean speakers and compared it against L1 English. They observed a significantly greater durational variability in L1 English than in L2 English; notably, however, when speakers received a 5-week English pronunciation training (including an explicit instruction about

rhythmic differences between Korean and English) and were tested again, their durational variability significantly increased, becoming more English-like (though it did not reach the level of L1 English speakers). Although both of these studies showed that L2 English exhibits rhythmic patterns that are distinct from L1 English, this was not always the case. Jang (2008), for instance, examined a wider range of rhythm metrics (not just the pairwise variability as in the studies above) and found that only some of them reliably distinguished L2 English from L1 English, highlighting the need for further investigation.

Alongside *group*-level differences between L1 and L2 rhythmicity, substantial variation may also exist at the *individual*-level within each L1 and L2 group. Within L1 English group, some speakers may be more rhythmic than others due to speaker-specific characteristics (e.g., speech habits, musical training). Likewise, within L2 group, speakers can exhibit different levels of rhythmicity, which may stem from individual characteristics as well as differences in L2 proficiency. The final question of the current study addresses the effect of *speaker-specific* rhythmic characteristics on L2 rhythmicity: namely, within bilingual speakers, does their L1 rhythmicity predict their L2 rhythmicity? In other words, if a speaker is more rhythmic in their L1, are they also more rhythmic in their L2? A positive relationship between L1 and L2 rhythmicity would suggest the presence of speaker-internal rhythmic characteristics that persist across languages.

Some acoustic parameters were found to be correlated between bilinguals' L1 and L2 production, raising the possibility that rhythmicity may also constitute a speaker-specific feature. For example, Bradlow et al. (2017) examined whether there is a significant relation between L1 and L2 speaking rate. They measured speaking rate (i.e., average number of syllables per second) in read and spontaneous speech in over 80 bilingual speakers from a variety of L1 backgrounds and found that L1 rate significantly predicts L2 rate within individuals. This result provided evidence for a *speaker-specific trait* of speaking rate that persists across different languages. Expanding this result, Bradlow et al. (2018) found a significant relation between bilinguals' L1 and L2 speech intelligibility (i.e., the extent to which listeners can correctly identify the intended words of a speaker). That is, a speaker who is more intelligible in L1 was also found to be more intelligible in their L2, again showing *speaker-specific trait* characteristics playing a role. Similar predictive relationships within speakers' L1 and L2 have also been observed in oral fluency (see references in Bradlow et al. 2017) as well as acoustic measures that are more directly related to vocal anatomy and physiology, such as fundamental frequency (F0) mean or range (Bradlow et al. 2018).

Based on the previous research, this study examines a relation between rhythmicity of L1 Korean and L2 English within Korean-English bilinguals, both at the time-scale of syllable and foot. In the literature of L2 rhythm, researchers have primarily examined the effect of L2 proficiency level, testing the hypothesis that more proficient L2 speakers will exhibit more target language-like rhythm. Analysis of the correlation between L1 and L2 rhythmicity can thus shed light on a different source of variation in rhythmicity among L2 speakers – namely, speaker-specific trait characteristics. Note that Oh and Park (2024) claimed that there are speaker-internal rhythmic characteristics that are persistent in their participants' L1 (Korean) and L2 (English). However, what they found is the *tendency* of bilinguals being more durationally variable in spontaneous speech (retelling the story *The North Wind and the Sun*) than in read speech (reading the story); L1 Korean speakers produced more variable durations in spontaneous speech regardless of whether they are speaking in their L1 or L2. This is clearly different from what the current study investigates; the focus of this study is whether L1 rhythmicity is a *predictor* of L2 rhythmicity, not whether a speaker exhibits a similar rhythmic *tendency* across languages.

Overall, the current study examines rhythmicity of Korean-English bilinguals and English monolinguals, using the envelope-based metrics developed by Tilsen and Arvaniti (2013). First, I examine the stability of syllabic- and supra-syllabic oscillations (i.e., rhythmicity at the syllable- and foot-level) in L1 Korean vs. L1 English. As mentioned above, considering the rhythmic characteristics of the two languages, I predict var. ω_1 (syllable-level)

to be lower in L1 Korean than L1 English, and var. ω_2 (syllable-level) to be lower in L1 English than L1 Korean. However, since Korean is rhythmically ambiguous, different patterns may emerge.

Second, I examine rhythmicity of L2 English both at the level of syllable and foot. Specifically, I compare var. ω_1 and var. ω_2 measures taken from L2 English to (i) L1 English (across-speaker comparison: different speakers, same language) and also to (ii) L1 Korean (within-speaker comparison: same speaker, different languages). These analyses will allow for a comprehensive examination of L1 vs. L2 rhythmicity. The var. ω_1 and var. ω_2 measures in L2 English may pattern with L1 Korean (indicating a failure to accommodate to L2 rhythm) or with L1 English (indicating full accommodation). Alternatively, the rhythm metrics of L2 English may pattern with neither, suggesting a blend of L1 and L2 characteristics or an intermediate pattern.

Lastly, I examine whether L1 rhythmicity is a predictor of L2 rhythmicity within bilinguals at both time-scales. A positive relationship between the two would provide evidence for the speaker-specific rhythmic trait that influences L1 and L2 production. On the other hand, the absence of a relationship would suggest that rhythm is more dependent on the specific language that is spoken (i.e., the specific rhythmic patterns of a language overwhelm the speaker-specific trait).

Section 2 presents a brief summary of technical details of the envelope-based analysis method of speech rhythm (Tilsen and Arvaniti 2013). The following section (Section 3) introduces speech materials assessed in the current study, along with analysis methods. Section 4 presents the analysis results, which are further discussed in Section 5.

2. Envelope-Based Rhythm Metrics (Tilsen and Arvaniti 2013)

Tilsen and Arvaniti (2013) proposed seven rhythm metrics – broadly categorized into *power distribution metrics* (3 metrics), *rate metrics* (2), and *rhythmic stability metrics* (2) – that are derived from the amplitude envelope of filtered speech. In this method, rhythm is conceptualized as "periodicity in the envelope, and greater stability of that periodicity corresponds to greater rhythmicity" (Tilsen and Arvaniti 2013 p.629). This assumes that all utterances exhibit a certain degree of rhythmicity. Among these metrics, the current study focuses on the *rhythm stability metrics* and compares them across languages and L1/L2 mode. Below, I briefly illustrate how the stability metrics are derived, summarizing the procedures of Tilsen and Arvaniti (2013); for the other metrics, see Tilsen and Arvaniti (2013) for details.

The first step is to bandpass-filter a speech signal. Specifically, they used a fourth-order Butterworth filter with cut-off frequencies of 400 and 4,000 Hz (though they note that the exact cut-off points in all of their filtering processes are somewhat arbitrary). This is done (i) to reduce the contribution of F0, thereby decreasing the extent of voicing to be directly represented in the signal (this makes, for example, voiced consonants to be more similar as voiceless consonants and differentiates them from vocalic nuclei, whose resonances are preserved in filtered speech) and (ii) to reduce the representation of sibilants or stop bursts. The output is then low-pass filtered using a fourth-order Butterworth filter with a 10 Hz cut-off in order to extract an envelope that varies on the time-scale of alternation between vocalic nuclei and consonantal margins – i.e., the *syllable* time-scale. (Note that the 10 Hz cut-off assumes that the duration of a syllable is no less than 100 ms.) The filtered output is referred to as *vocalic energy amplitude envelope* (or simply, *envelope*), which is then further processed (i.e., normalized, downsampled, Tukey-windowed).

The amplitude envelopes then undergo Empirical Mode Decomposition (EMD; Huang et al. 1998). EMD decomposes a speech signal into a number of basis functions using a sifting procedure. (Note that EMD is similar

to Fourier analysis in that both methods decompose a speech signal; yet, Fourier analysis breaks the signal into predefined sinusoidal components, whereas EMD decomposes the signal according to its own characteristics.) The resulting components that satisfy specific conditions are referred to as Intrinsic Mode Functions (IMFs). Thus, when a speech signal undergoes EMD, it is transformed into a set of IMFs, each representing oscillations at different time-scales; the first IMF represents the fastest time-scale of oscillation, the second IMF corresponds to the next fastest oscillation, and so on.

Crucially, Tilsen and Arvaniti (2013) assumed that the first IMF (IMF_1) is associated with syllable-level oscillations, and the second IMF (IMF_2) with supra-syllabic (specifically, foot-level) oscillations. Although these associations are not based on a priori evidence, they argued that it is a reasonable assumption. First, because the envelope was low-pass filtered at 10 Hz, this makes the fastest time-scale (IMF_1) to align at the level of the syllable. Second, because stressed syllables introduce additional amplitude modulations in the envelope, and because there is no linguistic unit intervening between syllable and foot, these stress-related modulations are likely to be captured by the second IMF. Tilsen and Arvaniti (2013) confirmed that these associations – i.e., IMF_1 with syllable-level and IMF_2 with foot-level – held true in the majority of cases that they examined. They further hypothesized that higher IMFs would correspond to oscillations at larger linguistic domains, such as the level of phrase, but these were not tested in their study.

Each IMF is further analyzed using Hilbert transform, that yields instantaneous phase and instantaneous frequency (ω) of that component. The *rhythmic stability metrics* are derived from the instantaneous frequencies of IMF_1 and IMF_2. Specifically, the variance of the instantaneous frequencies within an utterance captures *rhythmic stability*: the variance of within-utterance instantaneous frequency of IMF_1 (var. ω_1) represents the stability of syllabic oscillations, while the variance of within-utterance instantaneous frequency of IMF_2 (var. ω_2) represents the stability of stress-related (foot-level) oscillations. In both measures, a lower variance (i.e., smaller var. ω_1 or var. ω_2) indicates greater periodicity and thus a higher degree of rhythmicity at the syllabic or supra-syllabic level. The current study examines these two stability measures – i.e., var. ω_1 and var. ω_2 – and compares them across speakers and languages.

There are several advantages of the envelope-based analysis method, especially compared to the interval-based methods mentioned in the prior section. First, segmentation of consonants and vowels is not required. Because this method analyzes fluctuations in the amplitude envelope rather than durations of phonological units, a speech signal can be analyzed without segmentation (which often requires manual inspection and revision). See Campbell et al. (2025) for a similar discussion. Second, syllable-level and stress-level periodicity can be examined independently. The interval-based metrics are used to infer syllable-timing or stress-timing, by comparing how different languages that are known to be rhythmically distinct pattern on the same metric. Yet, the envelope-based metrics are based on the assumption that all utterances exhibit a certain degree of rhythmicity across multiple time-scales. This enables researchers to isolate and analyze rhythmicity at specific time-scales by selecting the appropriate measures. However, a downside of the envelope-based metrics is that they could be sensitive to specific cut-off points chosen for filtering. Finally, the envelope-based method yields three different *types* of rhythm metrics – i.e., power, rate, and rhythmic stability. Tilsen and Arvaniti (2013) pointed out that these metrics capture different aspects of speech rhythm. Although the present study only examines the rhythmic stability metrics, future research could explore how rate and power metrics contribute to an understanding of speech rhythm.

3. Methods

3.1 Speakers

The envelope-based rhythm metrics were tested on L1 Korean speakers producing L1 Korean and L2 English and L1 English speakers producing in their L1. Two sets of previously collected recordings were assessed. The first set was drawn from the Archive of L1 and L2 Scripted and Spontaneous Transcripts and Recordings (ALLSSTAR) corpus (Bradlow, n.d.). It was composed of 11 Korean-English bilinguals (4 men, 7 women) and 25 English monolinguals (11 men, 14 women). The second set was the Korean-English Intelligibility (KEI) corpus, collected under the protocol of Bradlow et al. (2018). This set consisted of 10 Korean-English bilinguals (5 men, 5 women) and 10 English monolinguals (5 men, 5 women). Both of these corpora are available in *Speechbox* (https://speechbox.linguistics.northwestern.edu); the first set is in the ALLSSTAR Corpus section, and the second set is in the Scripted Speech Corpora section. In total, recordings from 21 Korean-English bilinguals and 35 English monolinguals were tested. See Table 1 for the summary.

Table 1. Organization of the Data Used in the Current Study

corpus	L1	number of speakers	task language
ALLSSTAR	Korean	11	Korean
			English
	English	25	English
KEI	Korean	10	Korean
			English
	English	10	English
total (ALLSSTAR + KEI)	Korean	21	Korean
			English
	English	35	English

3.2 Materials

In the ALLSSTAR dataset, each bilingual produced 120 English and 120 Korean sentences taken from the Hearing in Noise Test (HINT) set of each language (Soli and Wong 2008); the English monolinguals produced the same 120 English sentences. These sentences are short, simple sentences that are widely used in audiology or clinical settings. Examples for English sentences are “The towel fell on the floor”, “The car is going too fast”; examples of Korean sentences are “날씨가 굉장히 흐렸어요 (The weather was very cloudy)”, “내일은 내 생일입니다 (Tomorrow is my birthday)”. Note that most Korean sentences in this dataset are not direct translations of English sentences, although some are. A total of 5,640 sentences were analyzed: 120 sentences × (25 L1 English speakers + 11 L1 Korean speakers producing L1 + 11 L1 Korean speakers producing L2).

In the KEI dataset, each bilingual produced 112 English and 112 Korean sentences taken from the revised Bamford-Kowal-Bench (BKB-R) Standard sentence set (Bamford and Wilson 1979); the monolinguals produced the same 112 English sentences. Similar to HINT, the BKB-R sentences are also short, simple sentences that are often used in clinical settings; some English examples are “The green tomatoes are small”, “The car engine is running”. The Korean sentences in this dataset are direct translations of English sentences, created by one of the authors in Bradlow et al. (2018) and cross-checked by two additional L1 Korean speakers. In the KEI dataset, 26 sentences were missing (0.77%), which resulted in 3,334 sentences in total: 112 sentences × (10 L1 English

speakers + 10 L1 Korean speakers producing L1 + 10 L1 Korean speakers producing L2) – 26 missing sentences.

3.3 Measurements

All sentences were normalized in loudness within each speaker and language. The two rhythm metrics – var. ω_1 and var. ω_2 – were then obtained from each individual sentence, following the procedures outlined in Section 2 above. After calculating var. ω_1, var. ω_2 of each sentence, following Tilsen and Arvaniti (2013), outlier values were identified and removed. Specifically, within each corpus, the ω_1 values (from which the var. ω_1 is calculated) of all sentences and speakers were pooled, and the values that are outside the three standard deviations from the mean were removed; the same process was done for ω_2. This step allowed for more reliable estimation of var. ω_1 and var. ω_2, which is presumably preferable to calculating var. ω_1 and var. ω_2 first for all sentences and then removing outliers based on those values. All of these processes were done in Matlab, using the scripts available in https://github.com/tilsen/EnvelopeMetrics.

Because the rhythm metrics are dependent on sentence durations, as pointed out by Tilsen and Arvaniti (2013), the duration of each utterance was also measured. Tilsen and Arvaniti (2013) found a significant effect of utterance duration on the two metrics tested in the current study: in particular, the instantaneous frequencies become more variable – i.e., greater var. ω_1 and var. ω_2 – in longer utterances. Because of this reason, duration of each utterance was measured in seconds and added as a covariate in the statistical models (see Section 3.4 for details). Within each corpus, sentences with durations exceeding three standard deviations from the mean were excluded: a total of 62 sentences (0.69%) – 17 from the KEI corpus and 45 from the ALLSSTAR corpus – were excluded, leaving 8,912 sentences in total for analysis.

The rhythm metrics and durations were then averaged across individual sentences of each speaker in each language mode, thus resulting in one syllable-level rhythm score (average of var. ω_1), one foot-level rhythm score (average of var. ω_2), and one sentence duration value for each speaker in each language: i.e., two sets of these values were derived for Korean-English bilinguals (one for L1 Korean and the other for L2 English) and a single set for English monolinguals. These measures reflect how rhythmic a speaker is *on average* at the syllable and foot levels as well as how long their utterances are. This study thus examines how rhythmicity varies at the *speaker*-level in L1 Korean vs. L1 English and in L1 vs. L2 settings. Deriving speaker means is also essential for addressing whether speaker-internal trait characteristics persist across their L1 and L2.

3.4 Statistical Analysis Methods

Before introducing specific statistical models, it is important to first mention that all of the analyses were conducted on a combined dataset including speakers from both ALLSSTAR and KEI corpora, without accounting for corpus differences (i.e., did not include corpus membership – whether a speaker came from the ALLSSTAR or KEI corpus – as a variable in the statistical models). As mentioned in Section 3.2 above, the two datasets differ in sentence content and in the way the Korean sentences are constructed compared to the English sentences (KEI: direct translations vs. ALLSSTAR: mostly distinct sentences). Nevertheless, the current study viewed the two datasets as quite similar and assessed rhythmicity in the combined dataset; the reasons are outlined below.

First, the *types* of sentences used in the two datasets are largely similar: both consist of short, syntactically simple sentences. The HINT and BKB-R sentences contain roughly 3 to 5 content words along with several function words. Their vocabulary is also simple and accessible, as they are commonly used to assess hearing, even for children. Given this structural similarity, the specific words that are used in each dataset may not substantially

influence the assessment of speaker rhythmicity.

Second, the focus of the current study is to examine rhythmic characteristics of *speakers* across different languages (L1 Korean vs. L1 English) and different language settings (L1 vs. L2 mode). That is, the target of the analysis is the *average* rhythm metric of each speaker in each language. Although differences in sentence content between the corpora could influence speaker-level rhythmicity (e.g., certain words could make some speakers more rhythmic than others), given that sentence structure and vocabulary are largely controlled, variation in rhythmicity is more likely to arise from speaker-related factors (e.g., overall rhythmic tendencies, L2 proficiency level) rather than from sentences themselves.

Third, the ALLSSTAR and KEI datasets were collected in the same location by the same research team. They were both recorded at the sound-attenuated booth at Northwestern University in the United States by the same research group. The speakers were also from the student population of Northwestern. (See also Bradlow et al. 2018 and Chernyak et al. 2024 who considered the two corpora as comparable and conducted analyses on the two datasets.)

Lastly, the distribution of the rhythm metrics and sentence durations was similar across the two datasets. Figure 1 shows the histograms of (a) var. ω_1, (b) var. ω_2, and (c) duration of individual sentences in the KEI (orange) and ALLSSTAR (blue) corpus. The histograms of the two corpora are quite similar in all three measurements. More specifically, the mean and standard deviation of var. ω_1 in the KEI were 7.88 and 2.47, and they were 7.43 and 2.41 in the ALLSSTAR. In terms of var. ω_2, the mean and standard deviation in the KEI were 1.06 and 0.56 and were 0.95 and 0.51 in the ALLSSTAR. Tilsen and Arvaniti (2013) noted that the instantaneous frequency of the first IMF is in general more variable than the frequency of the second IMF which changes more slowly; consistent with this, the var. ω_1 values in the datasets were larger than the var. ω_2 values. The mean and standard deviation of sentence durations were 1.56 and 0.29 in the KEI and 1.58 and 0.29 in the ALLSSTAR.

Overall, considering the similarities in sentence structure, recording conditions, and distributions of the measures between the two corpora, and also considering that the focus of the current study is on speaker-level rhythmic characteristics, the analyses were conducted on a combined dataset without accounting for differences in the corpus materials. Importantly, conducting separate analyses for each dataset would substantially reduce statistical power, potentially leading to misleading results.

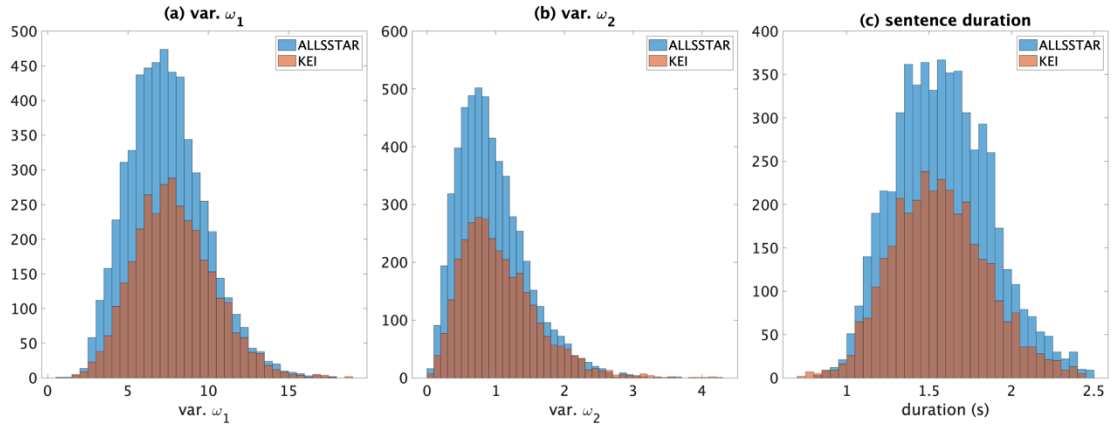

Figure 1. Distributions of the Rhythm Metrics (var. ω_1, var. ω_2) and Sentence Durations in Each Corpus

The statistical analyses were conducted using R (R Core Team, 2023). To investigate the cross-linguistic differences of rhythmicity – i.e., L1 Korean vs. L1 English, a linear regression model was fit to the average rhythm

scores of Korean and English speakers with L1 contrast (sum-coded with L1 Korean as -0.5 and L1 English as +0.5) as the predictor. The average sentence duration of each speaker (centered) was included as the covariate. The model was fit to var. ω_1 and to var. ω_2, separately. For the comparison of rhythmicity in L1 vs. L2 mode, two models were tested for each rhythm metric: (i) L1 English vs. L2 English (i.e., same language; different speakers), and (ii) L1 Korean vs. L2 English (i.e., different languages; same speaker). In both models, L1-L2 status was the main predictor (sum-coded with L1 English or L1 Korean as -0.5 and L2 English as +0.5), and average durations of the target speakers (centered) were used as the covariate. For the second model, a random intercept for each speaker was included as well (note: the model with random slopes did not converge). Lastly, to examine whether Korean-English bilinguals' L1 rhythm score predicts their L2 score, a linear regression model was fit to the bilinguals' L2 rhythm score with their L1 score as the predictor (centered); the model also had average durations of speakers in their L1 and L2 (centered) as covariates. In all analyses, a log-likelihood ratio test was conducted to determine whether the factor of main interest – i.e., L1 contrast, L1-L2 status, L1 rhythmicity – was significant.

4. Results

In the cross-linguistic investigation, L1 English speakers were found to be more rhythmic than L1 Korean speakers at the level of the stress-delimited foot; but, speakers of the two languages did not show significant differences in rhythmicity at the level of the syllable. Figure 2 shows distributions of average rhythm scores of speakers in each language, panel (a) showing the distribution of var. ω_1 (syllable-level) and panel (b) showing that of var. ω_2 (stress-level); each dot indicates the average rhythm metric of each speaker. The panel (a) shows that L1 Korean speakers exhibited varying rhythmicity at the level of the syllable, making the two language groups not clearly distinguishable. On the other hand, the panel (b) shows substantial differences between the two languages: L1 English speakers showed overall lower variance than L1 Korean speakers. This suggests that the foot-level oscillations were more regular for L1 English speakers. These observations were confirmed in the statistical models. For var. ω_1, there was no significant main effect of L1 contrast (Korean vs. English) ($\beta = 0.14$, s.e. $\beta = 0.17$, $\chi^2(1) = 0.68$, $p = 0.411$). For var. ω_2, a significant main effect of L1 contrast was observed ($\beta = -0.07$, s.e. $\beta = 0.03$, $\chi^2(1) = 5.11$, $p < 0.05$), with L1 English speakers exhibiting lower var. ω_2 than L1 Korean speakers.

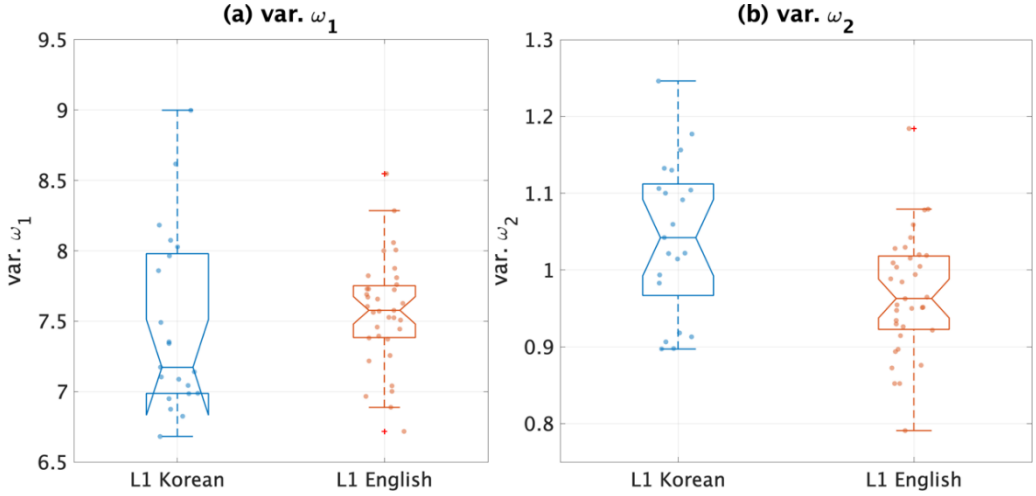

Figure 2. Cross-linguistic Differences (L1 Korean vs. L1 English) in Rhythmicity at the (a) Syllabic and (b) Supra-syllabic Time-scales

The comparison of L1 vs. L2 status was done in two ways – (i) L1 English vs. L2 English (across speaker groups; same language), (ii) L1 Korean vs. L2 English (within speaker; different languages); both (i) and (ii) did not find substantial differences between L1 and L2 mode, at the syllable and foot time-scales. Figure 3 shows distributions of speakers' var. ω_1 (left columns) and var. ω_2 (right column); the top row represents (i) across-group comparison, and the bottom row represents (ii) within-speaker comparison, where each dot indicates each individual's average rhythm score in a given language. (L2 English boxplots in the top and bottom panels are identical within each column.) The var. ω_1 (syllable-level) was not significantly different between L1 English vs. L2 English; see panel (a)-1 ($\beta = 0.20$, s.e. $\beta = 0.14$, $\chi^2(1) = 1.94$, $p = 0.164$). It was also not significantly different within Korean-English bilinguals (L1 Korean vs. L2 English); see panel (b)-1 ($\beta = 0.35$, s.e. $\beta = 0.14$, $\chi^2(1) = 3.52$, $p = 0.061$). Similarly, the var. ω_2 (foot-level) was not significantly different between L1 English vs. L2 English, as shown in panel (a)-2 ($\beta = 0.02$, s.e. $\beta = 0.03$, $\chi^2(1) = 0.48$, $p = 0.489$). It also did not exhibit significant differences between L1 Korean vs. L2 English within bilinguals, shown in panel (b)-2 ($\beta = -0.05$, s.e. $\beta = 0.02$, $\chi^2(1) = 1.48$, $p = 0.224$). Altogether, rhythmicity at both time-scales did not differ significantly by whether a speaker is in L1 or in L2 mode, regardless of the comparisons made across or within speakers.

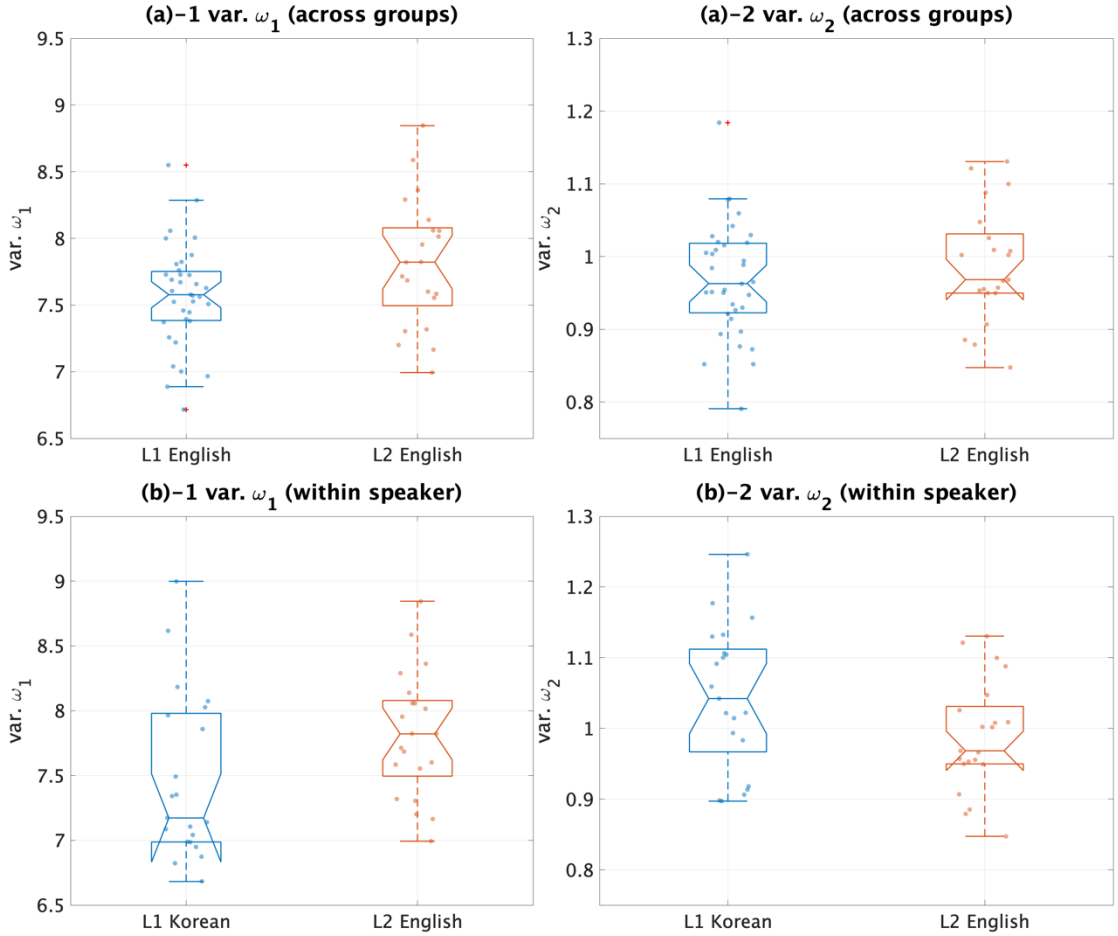

Figure 3. Rhythmicity Differences in L1 vs. L2 Mode (a) across Speaker Groups and (b) within Speaker at the Syllabic (left column) and Supra-syllabic (right column) Time-scales

Although rhythmicity scores did not differ by L1 vs. L2 settings, there was a significant relation between L1 and L2 rhythmicity within bilinguals at the foot-level. Figure 4 shows each bilingual's average rhythm metric calculated from their L1 (x-axis) and L2 speech (y-axis), with panel (a) and panel (b) showing var. ω_1 and var. ω_2, respectively. The red dashed line shows y = x line; if the dots (represent individual bilinguals) lie on this line, it indicates that L1 and L2 rhythmicity are perfectly correlated for those bilinguals. For var. ω_1, there was no significant main effect of L1 rhythm score on L2 score; see panel (a) ($\beta = 0.30$, s.e. $\beta = 0.17$, $\chi^2(1) = 3.57$, $p = 0.059$). However, for var. ω_2, a significant main effect of L1 rhythm score was observed ($\beta = 0.49$, s.e. $\beta = 0.15$, $\chi^2(1) = 10.01$, $p < 0.01$). As shown in the panel (b) of Figure 4, L1 and L2 rhythm scores showed a positive relationship: when a bilingual is more rhythmic in L1 (i.e., low variance in L1), that person is likely to be more rhythmic in L2 (i.e., low variance in L2).

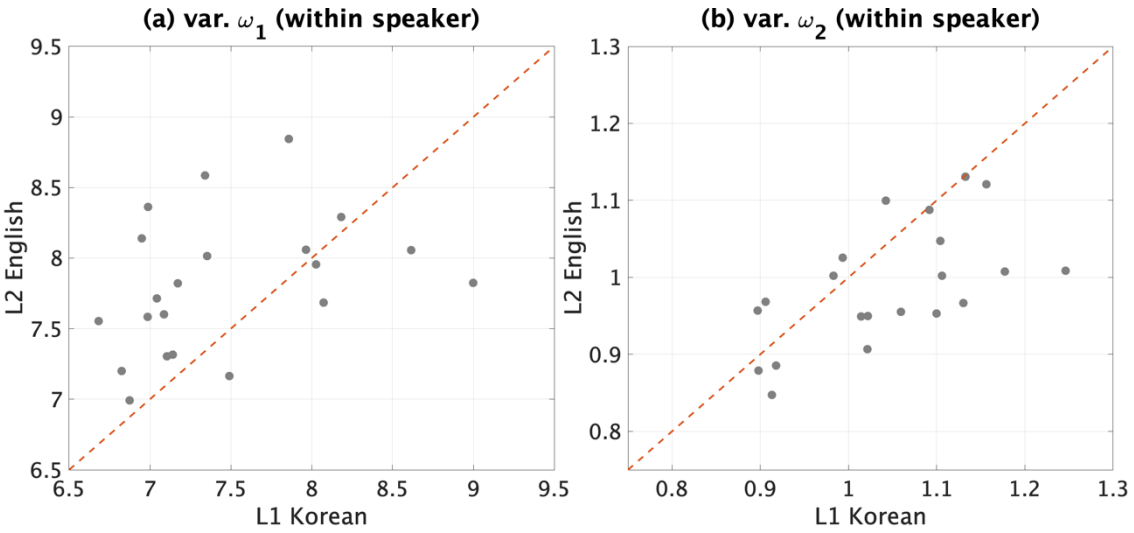

Figure 4. Relationships between L1 Korean and L2 English Rhythmicity within Speaker at the (a) Syllabic and (b) Supra-syllabic Time-scales

5. Discussion

Using the envelope-based rhythm metrics developed by Tilsen and Arvaniti (2013), this study examined rhythmicity of speakers of two languages, Korean and English, which are known to be rhythmically distinct. Specifically, the present study analyzed the variance of the instantaneous frequencies of the first and second IMFs – var. ω_1, var. ω_2 – which is considered to reflect stability of syllable-level and foot-level oscillations, respectively. Three specific questions were examined: (i) whether the rhythm metrics differ by L1 Korean vs. L1 English; (ii) whether they differ by L1 vs. L2 settings; (iii) whether Korean-English bilinguals' L1 rhythm score predicts their L2 score. These questions were examined utilizing recordings of short and simple Korean and English sentences.

The first analysis found that L1 English speakers show greater stability at the foot-level oscillations than L1 Korean speakers, but the two speaker groups did not differ in terms of the regularity at the syllable-level. Given that English is a prototypical stress-timed language and Korean is relatively closer to syllable-timing, the prediction was that var. ω_1 will be lower (i.e., more regular, rhythmic at the syllable-level) in L1 Korean than L1 English, while L1 English will exhibit lower var. ω_2 (i.e., more regular, rhythmic at the foot-level) than L1 Korean. The

prediction about stress-related-regularity was borne out: English speakers had overall lower var. ω_2, consistent with stress-timing of English.

However, the prediction about the syllable-regularity was not confirmed. This is indeed not too surprising, when the rhythmic characteristics of Korean are considered. In Figure 2-(a), there was wide variability of var. ω_1 values among L1 Korean speakers. The interquartile range of the L1 Korean data (i.e., the distance between the edges of the box in the boxplot) was 0.993, while it was 0.368 in L1 English data. This suggests that, as claimed by previous studies, Korean is close to syllable-timing, but not as robustly as prototypical syllable-timed languages such as Spanish or French. Some studies even found evidence that Korean is stress-timed or mora-timed (see Section 1). The variability among Korean speakers thus supports the view that Korean has an ambiguous rhythmic status.

The second analysis showed that rhythmicity is not significantly different between L1 and L2 settings at either time-scale. Notably, L1 and L2 did not differ when the metrics were compared either across speaker groups (L1 English vs. L2 English) or within bilinguals (L1 Korean vs. L2 English), highlighting the absence of systematic differences in rhythmicity between L1 and L2 mode. Focusing on the var. ω_2 measure, which showed a significant difference between L1 Korean and L1 English in the previous analysis, L2 English did not differ significantly from either L1 Korean or L1 English. When the distribution of this metric is compared across L1 English, L2 English, and L1 Korean (as illustrated in Figure 5, which is same as Figure 3 but presented in a different layout), L2 English fell between L1 Korean and L1 English. This suggests that the stress-related rhythmicity of L2 English is intermediate between that of L1 English and L1 Korean, though the differences are not statistically significant.

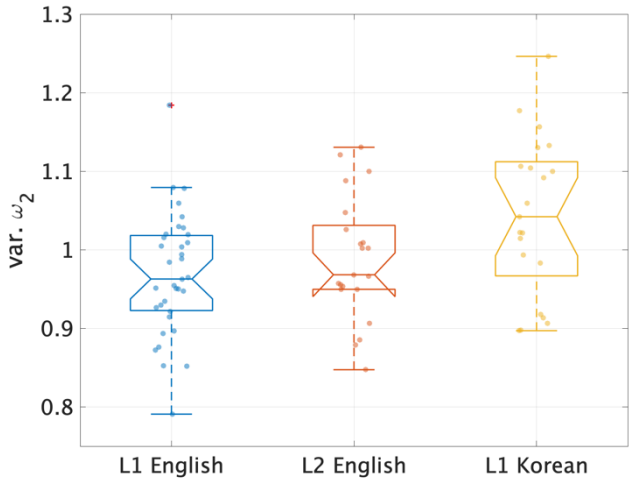

Figure 5. Comparison of Stress-related Rhythmicity (var. ω_2) across Different Speaker/Language Groups (revised boxplot layout from Figure 3)

This result is in contrast with some prior studies that examined Korean-English bilinguals. For example, Oh and Park (2024) found that L2 English has a significantly higher durational variability (i.e., more stress-timed) than L1 Korean. They additionally found that durational variability of L2 English is different from L1 English, though it was not in the expected direction, as L2 English showed greater durational variability than L1 English – i.e., L2 English being more stress-timed than L1 English. Similarly, Kim et al. (2007) found significant differences in pairwise variability indices between L2 English and L1 English both before and after instructions on English pronunciation. Some possible sources of differences between these studies and the current one include the type of

measures tested (envelope-based metrics vs. interval-based metrics), speech materials (e.g., read vs. spontaneous speech; short/simple sentences vs. morphologically-related word pairs), or speakers involved.

It is, however, similar to the findings of White and Mattys (2007), who demonstrated that L2 Spanish speakers (L1: English) as well as L2 English speakers (L1: Spanish) exhibit intermediate rhythmic patterns between the first language and the target language (see also Wenk, 1985). The difference is that L2 rhythm scores were significantly different from both L1 scores in White and Mattys (2007), but they were not in the current study. This could be due to different sample sizes or materials (there were more speakers and sentences per speaker in the current study), but also different languages involved. Spanish and English, the target languages of White and Mattys (2007), are examples of prototypical syllable-timed and stress-timed languages and may have resulted in a much clearer distinction between L2 vs. L1, while Korean and English, the target of the current study, yielded a less clear pattern. Another possibility is that the Korean-English bilinguals of the present study had varying L2 proficiency levels (in White and Mattys (2007), speakers' proficiency was relatively homogeneous; all of them were competent in L2 but had clear non-native accent). That is, the extent of differences between L2 English and L1s varies by the bilingual's English proficiency (as shown in, for example, Lee and Song 2019, Ordin and Polyanskaya 2015). It is therefore possible that the lower ends of the L2 English boxplot in Figure 5 (i.e., more similar to L1 English) represent speakers with higher L2 proficiency, compared to those ones at the top ends. This could be further tested in future studies.

The last analysis found that the Korean-English bilinguals' rhythmicity in L1 predicts their rhythmicity in L2 at the foot level. That is, in stress-related oscillations, a bilingual who was more rhythmic in their L1 was likely to be more rhythmic in their L2. Combined with the finding about within-speaker L1 vs. L2 comparison above (Figure 3 bottom row), this suggests that the foot-level rhythmicity does not significantly differ when bilinguals switch language modes, but their L1 rhythmicity is somewhat transferred to their L2. This finding is important, as it identified a factor that accounts for variation in stress-related rhythmicity in the L2 group – i.e., the speaker-inherent rhythmicity in their L1. Given that variation in L2 rhythmicity has been analyzed mostly in relation to L2 proficiency level (e.g., Jang 2008, Lee and Song 2019, Ordin and Polyanskaya 2015, Wenk 1985), this shows that speaker-internal rhythmic characteristics can matter in L2 speech. As mentioned above, the current study differs from Oh and Park (2024) in that it found a significant correlation between L1 and L2 rhythm scores per se, rather than observing a rhythmic tendency associated with speech style that persists across L1 and L2 (Oh and Park 2024).

In contrast, the var. ω_1 measure in bilinguals' L1 did not significantly predict the var. ω_1 measure in L2 (Figure 4-(a)); the relation between the two measures was only marginally significant ($p = 0.059$). Together with the high degree of variability in var. ω_1 observed among Korean speakers (Figure 2), this lack of a strong relation may suggest that individuals' syllable-level rhythmic strategies in their L1 do not persist in their L2. One possible explanation is that the linguistic properties of L2 inhibit the transfer of individuals' L1 rhythmic characteristics at the syllable-level. That is, speakers must adapt to English syllable structures as well as its phonological and phonotactic constraints, and that may override any syllable-level rhythmic tendencies carried over from their L1. In contrast, the var. ω_2 measure may reflect a speaker's general tendency to group syllables into larger prosodic units. As such, this (relatively) higher-level prosodic pattern may be less constrained by language-specific structure and more reflective of speaker-driven timing habits that can persist across languages. These interpretations remain largely speculative and warrant further investigation.

The current findings about foot-level rhythmicity within-bilinguals – i.e., the L1 and L2 var. ω_2 metrics do not differ in mean values but are correlated – are similar to what Bradlow et al. (2018) observed for the fundamental frequency (F0) measures and the slope of the long-term average speech spectrum (LTASS) in the mid-frequency range. In their study, the distributions of F0 mean, F0 range, and LTASS slope largely overlapped between L1 and

L2 (i.e., no dissociation by L1 vs. L2 in absolute terms), but L1 and L2 values within speakers were highly correlated (i.e., association in relative terms). This pattern clearly differed from the speech intelligibility measure in Bradlow et al. (2018) and the speaking rate of Bradlow et al. (2017), both of which showed significant dissociation between L1 and L2 in absolute terms (i.e., overall mean values differed by L1 vs. L2) as well as their association in relative terms (i.e., an individual's relative position for a given acoustic parameter remained consistent for L1 and L2).

Bradlow et al. (2018) viewed the association between the L1 and L2 measures in relative terms as arising from *speaker-specific trait* characteristics that persist across L1 and L2. For example, a speaker with overall low pitch or slow tempo would exhibit low F0 and slow speaking rate both in L1 and L2 production. The presence/absence of dissociation in absolute terms is then relevant to whether a given acoustic parameter is more closely related to the *source* vs. *articulatory patterns*. The *source*-related parameters (i.e., acoustic variables that are direct consequences of vocal tract anatomy and physiology or are related to overall vocal effort) such as F0 or LTASS slope are not so much affected by whether speakers are in L1 or L2 mode, resulting in no dissociation in absolute sense. On the other hand, the parameters that are related to the *patterns of speech articulation* such as speaking rate or intelligibility could differ by L1 vs. L2 settings, leading to substantial dissociation in absolute terms (e.g., L2 speakers' articulatory patterns may be less precise than those of L1 speakers and result in lower intelligibility). The distinction between the *articulation patterns* vs. *source* can also be conceptualized as the characteristics that are (relatively) more under *speaker-control* vs. those that are *automatic* consequences of vocal anatomy and physiology.

In this context, the present finding raises an interesting point about the framework of L1/L2 (relative) association and (absolute) dissociation. I believe that rhythmicity is a parameter that is more closely related to the articulation of individual sounds and thus is under speaker-control, rather than being a more automatic, source-related parameter. Speakers, for instance, dynamically alter articulations of speech sounds (temporal intervals in particular) which makes speech more or less rhythmic – e.g., English speakers reduce durations of vowels when they are unstressed but expand them when they are stressed (although see Cummins (2009) who viewed rhythm as entrainment of movement). If this assumption is correct, it suggests that the pattern of "L1/L2 (relative) association without (absolute) dissociation" could be observed in non-source-related factors as well. Further research needs to be done to find out why rhythmicity does not exhibit dissociation between L1 and L2, testing various languages and its relation to L2 proficiency level, and to examine the nature of speech rhythm (e.g., whether it is more related to the source vs. articulatory patterns).

The important contribution of the current study is that it utilized envelope-based metrics to assess speech rhythm, rather than interval-based measures (which were the focus of previous studies), both in L1 and L2 production. Yet, at the same time, it suggests several directions for future studies. One possible direction is to incorporate bilinguals' L2 proficiency level or speech intelligibility in the analysis. As mentioned above, similarities or differences between L2 English vs. L1 Korean/English could vary as a function of speakers' proficiency/intelligibility. One could also examine which factor better accounts for variation in L2 rhythm – i.e., L1 rhythmicity (demonstrated in the third analysis) vs. L2 proficiency/intelligibility.

Another potential direction is to analyze speech elicited with different materials and tasks; this point has been emphasized by many previous studies of speech rhythm (e.g., Arvaniti 2012, Oh and Park 2024, Tilsen and Arvaniti 2013). It is remarkable that this study found rhythmic differences across languages even with short and simple sentences; however, at the same time, it is possible that the use of short sentences actually facilitated detection of meaningful results. That is, when longer and more complex sentences are tested, speakers (or L2 speakers in particular) may exhibit less rhythmicity overall. In addition, the current study tested read speech, but

spontaneous speech may show different patterns. Along with testing different types of speech materials, one could test other envelope-based metrics – power distribution or rate metrics – and also compare them with the interval-based metrics. To gain a more comprehensive understanding of the envelope-based metrics and L2 rhythmic patterns, further analyses should be conducted on speakers from a variety of L1 backgrounds and across different L1-L2 language pairs.

On a final note, it would be interesting to examine whether the rhythmic differences demonstrated in this study are associated with any perceptual differences. Tilsen and Arvaniti (2013) mentioned that the relation between envelope-based metrics and rhythm perception have not been examined; however, later work by Robinson (2022) found that the envelope-based rhythm metrics accounted for variation in speech-in-noise recognition. Specifically, when the signal-to-noise ratio (SNR) was low (i.e., speech was accompanied by loud noise), sentences with higher rhythmicity were more accurately recognized; yet, utterance rhythmicity did not affect speech recognition at a higher SNR (i.e., speech was accompanied by less noise). On the other hand, in a more recent study by Kim et al. (2025), rhythmicity did not account for variation in speech-in-noise recognition, either when listeners were presented with L1 or L2 English speech. In this context, it would be interesting to test whether listeners are sensitive to differences among speakers of the same language who vary in their rhythm scores (e.g., whether they perceive speakers with comparable rhythm scores as more similar) or to compare the relationship between envelope-based rhythmicity and perception with that of other rhythm metrics or other acoustic parameters.

6. Conclusion

Overall, this study examined rhythmic patterns of Korean-English bilinguals and English monolinguals using metrics that are derived from the amplitude envelope of filtered speech. The stability of the syllable-level and foot-level oscillations was examined, although significant findings were observed only at the foot-level. Specifically, L1 English speakers showed overall lower variance in the foot-level oscillations than L1 Korean speakers, reflecting the stress-timing characteristics of English. The rhythmicity of L2 English was not significantly different from L1 Korean or L1 English. The present study has also identified one source of variation in L2 rhythmicity, which is the bilingual's L1 rhythmicity at the same time-scale. Future research examining the envelope-based metrics on a wider range of speech materials, L1s, and L1-L2 pairings as well as their relation to perception will facilitate our understanding of how speakers produce and listeners perceive speech rhythm. The current findings also offer new insights and suggest directions for future research about the relationship between L1 and L2 speech production.

References

Abercrombie, D. 1967. *Elements of General Phonetics.* Edinburgh University Press.

Arvaniti, A. 2012. The usefulness of metrics in the quantification of speech rhythm. *Journal of Phonetics* 40, 351-373.

Bamford, J. and I. Wilson. 1979. Methodological considerations and practical aspects of the BKB sentence lists. In J. Bench and J. Bamford, eds., *Speech-hearing Tests and the Spoken Language of Hearing-impaired Children,* 148-187. Academic Press.

Bradlow, A. R. n.d. *ALLSSTAR: Archive of L1 and L2 Scripted and Spontaneous Transcripts and Recordings.*

Retrieved from https://speechbox.linguistics.northwestern.edu/allsstar

Bradlow, A. R., M. Blasingame and K. Lee. 2018. Language-independent talker-specificity in bilingual speech intelligibility: Individual traits persist across first-language and second-language speech. *Laboratory Phonology* 9(1), 1-20.

Bradlow, A. R., M. Kim and M. Blasingame. 2017. Language-independent talker-specificity in first-language and second-language speech production by bilingual talkers: L1 speaking rate predicts L2 speaking rate. *The Journal of the Acoustical Society of America* 141(2), 886-899.

Campbell, J., D. Byrd and L. Goldstein. 2025. The stability of articulatory and acoustic oscillatory signals derived from speech. *JASA Express Letters* 5(4), 045203.

Chernyak, B. R., A. R. Bradlow, J. Keshet and M. Goldrick. 2024. A perceptual similarity space for speech based on self-supervised speech representations. *The Journal of the Acoustical Society of America* 155, 3915-3929.

Cho, M.-H. 2004. Rhythm typology of Korean speech. *Cognitive Processing* 5, 249-253.

Cummins, F. 2009. Rhythm as an affordance for the entrainment of movement. *Phonetica* 66, 15-28.

Dellwo, V. 2006. Rhythm and speech rate: A variation coefficient for delta C. In P. Karnowski and I. Szigeti, eds., *Language and Language-processing*, 231-241. Peter Lang.

Grabe, E. and E. L. Low. 2002. Durational variability in speech and the rhythm class hypothesis. In C. Gussenhoven and N. Warner, eds., *Laboratory Phonology 7*, 515-546. De Gruyter Mouton.

Huang, N.E., Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N.-C. Yen, C. C. Tung and H. H. Liu. 1998. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proceedings of the Royal Society London. Series A: Mathematical, Physical and Engineering Sciences,* 903-995.

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

Kim, J., C. Davis and A. Cutler. 2008. Perceptual tests of rhythmic similarity: II. Syllable rhythm. *Language and Speech* 51(4), 343-359.

Kim, J.-m., S. Flynn and M. Oh. 2007. Non-native speech rhythm: A large-scale study of English pronunciation by Korean learners. *Studies in Phonetics, Phonology, and Morphology* 13(2), 219-250.

Kim, S.-E., B. R. Chernyak, J. Keshet, M. Goldrick and A. R. Bradlow. 2025. Predicting relative intelligibility from inter-talker distances in a perceptual similarity space for speech. *Psychonomic Bulletin & Review.*

Lee, H.-Y. and J. Song. 2019. Evaluating Korean learners' English rhythm proficiency with measures of sentence stress. *Applied Psycholinguistics* 40, 1363-1376.

Lee, H., N.-t. Jin, C.-j. Seong, I.-j. Jung and S.-m. Lee. 1994. An experimental phonetic study of speech rhythm in Standard Korean. *Proceedings of the 3rd International Conference on Spoken Language Processing (ICSLP 94)*, 1091-1094.

Lee, H. and C.-j. Seong. 1996. Experimental phonetic study of the syllable reduction of Korean with respect to the positional effect. *Proceedings of the 4th International Conference on Spoken Language Processing (ICSLP 96)*, 1193-1196.

Lee, O.-h. and J.-m. Kim. 2005. Syllable-timing interferes with Korean learners' speech of stress-timed English. *Speech Sciences* 12(4), 95-112.

Lin, H. and Q. Wang. 2005. Vowel quantity and consonant variance: A comparison between Chinese and English. *Proceedings of the Between Stress and Tone.*

Low, E. L., E. Grabe and F. Nolan. 2000. Quantitative characterizations of speech rhythm: Syllable-timing in

Singapore English. *Language and Speech* 43(4), 377-401.

Mok, P. and S. I. Lee. 2008. Korean speech rhythm using rhythmic measures. *Proceedings of the 18th International Congress of Linguistics (CIL18)*.

Oh, S. and H. Park. 2024. The impact of native language on second language rhythm acquisition: Insights from a cross-linguistic and intra-language corpus study. *Linguistics Research* 41(3), 391-429.

Ordin, M. and L. Polyanskaya. 2015. Acquisition of speech rhythm in a second language by learners with rhythmically different native languages. *The Journal of the Acoustical Society of America* 138, 533-544.

Pike, K. L. 1945. *The Intonation of American English.* University of Michigan Press.

R Core Team. 2023. *R: A language and environment for statistical computing* [Computer software]. R Foundation for Statistical Computing. https://www.R-project.org/

Ramus, F., M. Nespor and J. Mehler. 1999. Correlates of linguistic rhythm in the speech signal. *Cognition* 73, 265-292.

Robinson, A. 2022. *A Role for Rhythmicity in Speech Intelligibility*. Unpublished master's thesis, Northwestern University.

Soli, S. D., and L. L. N. Wong. 2008. Assessment of speech intelligibility in noise with the Hearing in Noise Test. *International Journal of Audiology* 47(6), 356-361.

Tilsen, S. and A. Arvaniti. 2013. Speech rhythm analysis with decomposition of the amplitude envelope: Characterizing rhythmic patterns within and across languages. *The Journal of the Acoustical Society of America* 134, 628-639.

Wenk, B. J. 1985. Speech rhythms in second language acquisition. *Language and Speech* 28(2), 157-175.

White, L. and S. L. Mattys. 2007. Calibrating rhythm: First language and second language studies. *Journal of Phonetics* 35, 501-522.

Examples in: English, Korean
Applicable Languages: English, Korean
Applicable Level: Tertiary

