

Envelope-Based Metrics Reveal Speech Rhythm Development Tendencies in Brazilian L2 English Learners

Leonardo Antonio Silva Teixeira¹, Universidade Federal do Ceará

Sam Tilsen², Cornell University

Ronaldo Manguiera Lima Jr.³, Universidade de Brasília

Pablo Arantes⁴, Universidade Federal de São Carlos

Abstract

Speech rhythm in second language (L2) development remains underexplored, despite previous reports indicating its impact on intelligibility, comprehensibility, and perceived degree of foreign accent (Munro & Derwing, 2001; Silva Jr. & Barbosa, 2019). While most studies rely on interval-based rhythm metrics, envelope-based approaches such as spectral analysis of the envelope (Tilsen & Johnson, 2008) and *Empirical Mode Decomposition* (EMD; Tilsen & Arvaniti, 2013) provide alternative tools to capture rhythmic periodicities in the amplitude envelope of speech. This study investigates speech rhythm development in Brazilian learners of English (EN-L2), comparing them with monolingual native speakers of English (EN-L1) and of Brazilian Portuguese (BP-L1). We analyzed 1,060 speech chunks (1–2.5s) extracted from controlled readings. Filtered vocalic energy envelopes were extracted via Voice Activity Detection and decomposed into Intrinsic Mode Functions (IMFs). Three envelope-based rhythm metrics were analyzed: *Spectral Band Power Ratio* ($sbpr_1$), *Spectral Centroid* ($scntr_1$), and the IMF_2/IMF_1 power ratio (imf_ratio_{21}). Mixed-effects models assessed the influence of language and chunk duration while accounting for speaker variability. Contrary to expectations, results indicated no significant cross-linguistic differences between BP-L1 and EN-L1. The most robust effect was observed in L2 learners, who gradually amplified the stress-level timescale over time. By the latest stage, learners exhibited rhythm patterns even more stress-dominant than either L1 baseline group, consistent with prosodic hyper-articulation (Cho, Lee & Kim, 2011; Smiljanić & Bradlow, 2009). A key contribution of this study lies in the analysis of between-group differences in correlations between envelope-based rhythm metrics. This relational analysis reveals that while L1 speakers exhibit relatively stable patterns of spectral energy coordination, L2 learners show evolving patterns of co-variation that complement the analysis of groupwise differences observed in individual metric values. These results point to the suitability of envelope-based approaches to the study of L2 rhythm development, suggesting that learners undergo a non-linear process of adjustment and redistribution of acoustic energy along different timescales.

Keywords: speech rhythm; L2 acquisition; envelope-based metrics; spectral analysis; intermetric correlation

¹ E-mail: leonardo.teixeira@ifce.edu.br

² E-mail: tilsen@cornell.edu

³ E-mail: ronaldo.junior@unb.br

⁴ E-mail: pabloarantes@ufscar.br

1. Introduction and background

Rhythm appears to be one of the most elusive constructs in linguistic theory. As Albert & Grice (2024) point out, rhythm resists objective formalization and remains difficult to “accommodate as a clear construct in language systems.” Regarding L2 development, despite evidence pointing to the influence of rhythm on the aspects of intelligibility, comprehensibility, and perceived degree of foreign accent (Munro & Derwing, 2001; Silva Jr. & Barbosa, 2019), rhythm emerges as a rather underexplored construct.

Early accounts (Abercrombie, 1967; Jassem, 1950) of speech rhythm proposed isochrony-based frameworks and were rooted in intuitions about surface regularity and the existence of rhythm classes such as syllable-timed and stress-timed languages. However, several attempts have failed to demonstrate the existence of such neatly evenly recurring intervals, like syllable or foot. Later empirical approaches (Ramus, Nespor & Mehler, 1999; Dellwo, 2006; Grabe & Low, 2002) aimed to operationalize cross-linguistic rhythm distinctions through interval-based metrics derived from consonant-vowel intervals, such as %V, ΔC , Varcos, and PVIs. Although these metrics offered useful initial insights into cross-linguistic differences, they generally ignore the dynamic properties of rhythmic modulation and coordination, treating rhythm as a static temporal artifact that mainly reflects language-specific phonotactic constraints.

Recent theoretical advances have reframed rhythm as a multidimensional, dynamic, and perceptually driven phenomenon involving coordination across articulatory gestures, prosodic hierarchies, and oscillatory timing systems (Cummins, 2009; Kohler, 2008; Tilsen, 2019). Within this paradigm shift, spectral approaches have been proposed as tools to capture rhythm modulations in the amplitude envelope across frequency bands. For example, Tilsen & Johnson (2008) present an approach to studying speech rhythm by employing low-frequency Fourier analysis of the amplitude envelope. Building on this, Tilsen & Arvaniti (2013) proposed the Envelope Spectral Empirical Mode Decomposition (EMD) of vocalic energy, which decomposes the amplitude envelope into Intrinsic Mode Functions (IMFs) that reflect rhythmic energy patterns at different timescales. Thus, envelope-based metrics derived from these methods offer a dynamic spectral characterization of rhythm that does not depend on *a-priori* linguistic units.

However, envelope-based metrics remain underexplored in the context of second language (L2) development. We aim to bridge this gap by analyzing three envelope-based metrics: *sbpr*₁, *scntr*₁, and *imf_ratio*₂, across native speakers of Brazilian Portuguese (BP-L1), Canadian English (EN-L1), and Brazilian learners of English tracked longitudinally over four semesters (EN-L2-1 to EN-L2-4).

Crucially, we introduce a novel methodological contribution: a correlation analysis that examines how these envelope-based metrics co-vary within and across speaker groups. This approach allows for the assessment of both metric-specific trends and intermetric co-variation patterns as indices of rhythm coordination and development. By analyzing patterns of correlation between envelope-based metrics, we provide a new perspective for understanding prosodic development and cross-linguistic rhythmic organization.

Unexpectedly, significant differences were absent between the two native baselines, suggesting that generally similar rhythmic patterns are present in BP-L1 and EN-L1, at least under the elicitation method of reading. As for the learners, however, a robust tendency emerged across the four semesters: rhythmic periodicities became increasingly more stress-timed, culminating in prosodic hyper-articulation at the most advanced stage, that is, an exaggeration of acoustic cues in the speech signal, probably reflecting learners' attempt to make speech more intelligible and improve listeners' ability to access and comprehend utterances (Smiljanić & Bradlow, 2009, p. 236).

Finally, the intermetric correlation analysis revealed that the dynamics among rhythm metrics, particularly between *imf_ratio*₂₁ and both *sbpr*₁ and *scntr*₁, gradually became more native-like over time. This suggests that L2 rhythm development does not involve a simple accommodation to native norms, but rather an ongoing process of fine-tuning certain acoustic cues with progressive reorganization of how these cues co-vary within the prosodic system.

2. Methodology

We analyzed 1,060 speech chunks (1–2.5 seconds in duration) derived from controlled readings of a diagnostic passage from Celce-Murcia *et al.* (2010) by five native speakers of Brazilian Portuguese (BP), five native speakers of Canadian English (EN-L1), and five Brazilian learners of English (EN-L2). Over four academic semesters, learners were recorded longitudinally (EN-L2-1 through EN-L2-4). Recordings were conducted in a quiet environment using a Shure Beta 150B supercardioid microphone and a Zoom Hn4 recorder (44.1 kHz, 16-bit resolution).

Preprocessing involved four steps as follows: (i) applying a 4th-order Butterworth bandpass filter (80–10,000 Hz) to the acoustic waveform; (ii) extracting the vocalic amplitude envelope; (iii) low-pass filtering at 10 Hz to track only rhythmic modulation relevant to characterizing vocalic energy fluctuations; (iv) chunking the envelopes via Voice Activity Detection (VAD) (only pause-free stretches of the envelope were included in analyses).

Each envelope was spectrally analyzed and decomposed into Intrinsic Mode Functions (IMFs) resulting from Empirical Mode Decomposition (EMD), a method that isolates oscillatory modes at different timescales (Huang *et al.*, 1998).

Three envelope-based metrics were analyzed:

- *sbpr*₁ (Spectral Band Power Ratio 1): This represents the ratio of spectral power in the 1 - 3.5 Hz range to power in the 3.5 - 10Hz range, that is the ratio of power on the stress timescale to power on the syllable timescale. Higher *sbpr*₁ values indicate a greater degree of alternation between stressed and unstressed syllables.
- *scntr*₁ (Spectral Centroid 1): This metric refers to the weighted mean frequency of the envelope's spectrum between 0 and 10 Hz. In other words, it indicates the "center of gravity" of spectral energy. A lower centroid reflects energy concentrated in slower

frequencies (e.g., stress/phrase timing), while a higher centroid suggests a greater contribution from faster, syllable-based modulations.

- imf_ratio_{21} (IMF2/IMF1 Energy Ratio): Derived from EMD, it captures the relative strength of slower temporal modulations (IMF2) compared to faster ones (IMF1). A higher ratio may be indicative of greater dominance of stress-timed rhythms, while a lower ratio suggests that faster, more rapid oscillations play a stronger role.

To assess the effect of group on each envelope-based metric analyzed, we adjusted linear mixed-effects models using the *lmer* function from the *lme4* R package (Bates et al., 2015). The dependent variable was each rhythm metric, and the main predictor was *group*, a categorical variable with six levels representing each speaker set: EN-L1, BP-L1, and four stages of EN-L2 development (EN-L2-1 to EN-L2-4, which correspond to the four semesters of instruction). We included *speaker* as a random intercept and *chunk duration* as a covariate. For each of the three metrics (sbpr_1 , scnr_1 and imf_ratio_{21}), we compared the following three models:

- Maximal model: $\text{metric} \sim \text{group} + \text{dur} + (1 \mid \text{speaker})$
- No duration: $\text{metric} \sim \text{group} + (1 \mid \text{speaker})$
- No language: $\text{metric} \sim \text{dur} + (1 \mid \text{speaker})$

We conducted model comparison through the *compare_performance* function from the *performance* package (Lüdtke et al., 2021). For each metric, AIC, BIC, and R^2 values were used to determine the best-fitting model. For all metrics, the maximal model showed superior fit. This suggests that both group and chunk duration contribute to rhythmic variation in the amplitude envelope, beyond what can be attributable to speaker variability.

Additionally, we conducted a correlation analysis, a novel contribution of this study. First, we extracted residuals from mixed-effects models of the form $\text{metric} \sim 1 + (1 \mid \text{speaker})$, fitted separately for each metric. This is in line with our interest in covariation that is not explained by speaker variability. Following this, we computed Pearson correlations between each pair of metrics and within each language group to explore whether envelope-based rhythm metrics have different patterns of covariation across groups.

3. Results and Discussion

3.1. Groupwise analysis: no significant differences between L1s and hyper-articulation effects among L2 learners

Figure 1 shows groupwise distributions for the three envelope-based rhythm metrics. Visual inspection shows consistent differences between L1 and L2 speakers, with L2 learners exhibiting stronger stress-timed modulation than either native baseline. The metrics indicate an unexpected exaggeration of stress rhythm, or prosodic hyper-articulation, in the most advanced learner group (EN-L2-4), which produced utterances with greater stress-time scale rhythmicity than native speakers.

This phenomenon may suggest an over-application of L2 rhythmic features, particularly those associated with stress periodicities. In other words, learners attempt to conform to native-like timing by exaggerating stress cues beyond native norms, potentially as a strategy to enhance intelligibility or due to instructional emphasis on stress. Such behavior aligns with findings on prosodic hypercorrection (Cho, Lee & Kim, 2011; Smiljanić & Bradlow, 2009), that is, when learners overcompensate for perceived L2 deficits by amplifying salient prosodic features. This finding supports the view that rhythm development may involve an active process of reweighting of rhythmic cues, rather than simple accommodation to native norms.

Figure 1. Violin plots showing the distribution of spectral rhythm metrics across L1 and L2 groups. Descriptive statistics are provided in Table 1.

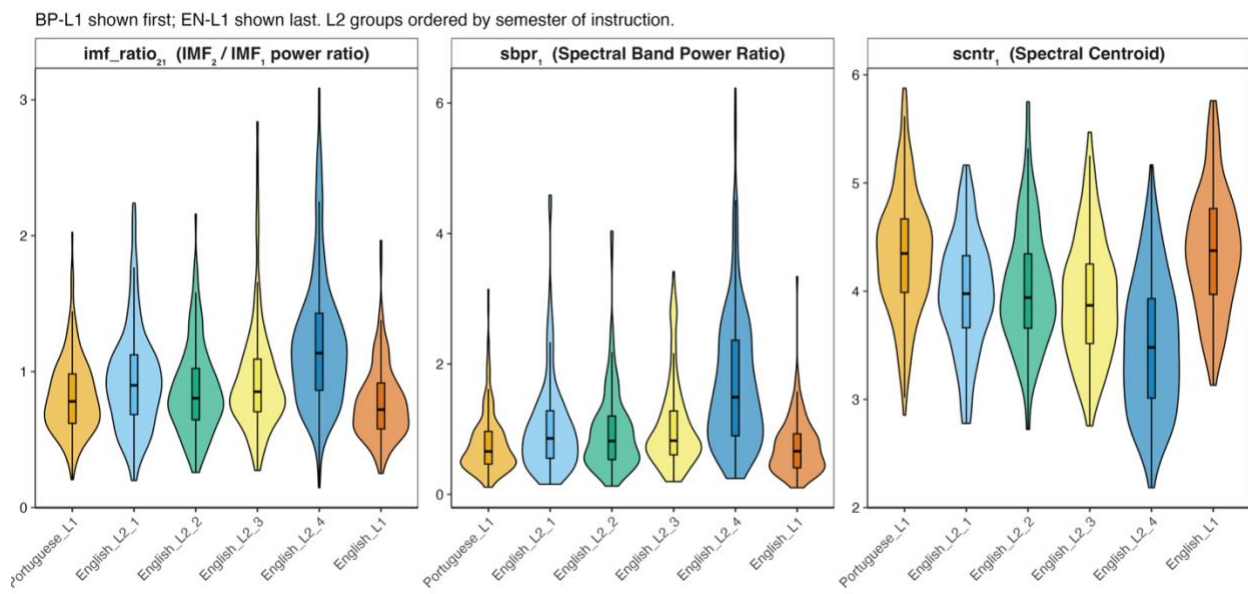


Table 1. Mean and standard deviation (SD) of envelope-based metrics by group.

Metric	English_L1	Portuguese_L1	English_L2_1	English_L2_2	English_L2_3	English_L2_4
sbpr_1	0.73 ± 0.43	0.77 ± 0.46	1.03 ± 0.75	0.97 ± 0.65	1.06 ± 0.68	1.71 ± 1.03
scnr_1	4.37 ± 0.56	4.34 ± 0.56	3.98 ± 0.52	4.03 ± 0.56	3.93 ± 0.56	3.5 ± 0.62
imf_ratio21	0.77 ± 0.27	0.82 ± 0.28	0.94 ± 0.38	0.86 ± 0.33	0.94 ± 0.41	1.2 ± 0.48

Regarding $sbpr_1$, which refers to the ratio of slower stress-level rhythmic modulation over faster syllabic pulses, mean values were somewhat higher for BP-L1 than for EN-L1, though the model indicates that cross-linguistic differences were not significant. Importantly, subsequent correlation analyses (see Section 3.2) reveal some differences in how rhythmic energy is distributed across spectral bands in the two native systems. This indicates that even when global values appear similar, stress-level and syllable-level activity may co-vary differently within each system. As for the L2 stages, $sbpr_1$ values increased steadily over time, reaching the highest mean along with large dispersion in EN-L2-4 (1.71 ± 1.03).

Similarly, $scntr_1$ exhibited the prosodic hyper-articulation effect. This metric captures energy dispersion across frequency bands, and lower values indicate a greater contribution of stress/phrase-timescale periodicity in the envelope. Again, EN-L1 and BP-L1 did not differ significantly, and the L2 groups displayed the lowest (most stress-timed in this case) values. Crucially, EN-L2-4 exhibited the lowest centroid (3.5 ± 1.03), which might reflect learners' attempts to encode stress prominence more frequently at later stages of L2 prosodic development.

The imf_ratio_{21} metric further supports this interpretation: means gradually increased throughout the learning stages, reaching the highest value in EN-L2-4 (1.2 ± 0.48). Considering that this metric reflects the energy ratio of faster (IMF₂) to slower (IMF₁) oscillations, higher values indicate a relative dominance of slower, stress-level pulses over faster, syllable-level modulation. The extreme values in EN-L2-4 suggest that learners may overproduce local stress cues without sufficient integration across larger prosodic domains. Again, differences between the L1s for imf_ratio_{21} were not significant.

Table 2. Coefficients from the maximal mixed-effects model (Estimate 95% CI), by metric. Models: $metric \sim group + dur + (1 | speaker)$

Predictors	$sbpr_1$	$scntr_1$	imf_ratio_{21}
(Intercept) = English L1	0.55 [0.28, 0.82], $p < .001$	4.66 [4.44, 4.87], $p < .001$	0.82 [0.70, 0.93], $p < .001$
Portuguese L1	0.03 [-0.30, 0.35], $p = 0.873$	-0.01 [-0.27, 0.25], $p = 0.942$	0.05 [-0.07, 0.18], $p = 0.396$
L2_1	0.33 [-0.02, 0.67], $p = 0.065$	-0.40 [-0.69, -0.12], $p = 0.006$	0.17 [0.03, 0.30], $p = 0.017$
L2_2	0.24 [-0.09, 0.57], $p = 0.143$	-0.33 [-0.60, -0.07], $p = 0.015$	0.09 [-0.03, 0.22], $p = 0.146$
L2_3	0.37 [0.04, 0.70], $p = 0.029$	-0.48 [-0.75, -0.22], $p < .001$	0.17 [0.04, 0.30], $p = 0.011$
L2_4	0.86 [0.54, 1.19], $p < .001$	-0.76 [-1.02, -0.50], $p < .001$	0.42 [0.29, 0.54], $p < .001$

Maximal models showed significant group effects ($sbpr_1$: $F = 6.66$, $p < .001$; $scntr_1$: $F = 8.59$, $p < .001$; imf_ratio_{21} : $F = 9.86$, $p < .001$), with L2 groups in the last two stages being significantly different from either native baselines across all metrics

For $sbpr_1$, positive coefficients for L2_3 ($\beta = 0.37$ [0.04, 0.70], $p = .029$) and L2_4 ($\beta = 0.86$ [0.54, 1.19], $p < .001$) corroborate an increase of this parameter over time, resulting in exaggerated stress-level periodicity. This adds to the descriptive trend that EN-L2-4 has the highest contrast values (Figure 1), consistent with hyper-articulation. For $scntr_1$, negative coefficients for L2 groups (e.g., L2_2 $\beta = -0.33$ [-0.60, -0.07], $p = .015$; L2_4 $\beta = -0.76$ [-1.02, -0.50], $p < .001$) reflect lower spectral centroids, that is, a redistribution of rhythmic energy toward slower frequency bands. This shift aligns with the interpretation of increasingly stress-oriented timing at advanced L2 stages. Finally, imf_ratio_{21} also increases with proficiency (L2_3 $\beta = 0.17$ [0.04, 0.30], $p = .011$; L2_4 $\beta = 0.42$ [0.29, 0.54], $p < .001$), suggesting greater relative energy in slower oscillatory modes (IMF₂) compared to faster ones (IMF₁). The L1 comparison revealed no statistically significant difference between English and Brazilian Portuguese speakers, indicating similar rhythmic organization in the two native languages. In contrast, L2 speakers exhibited a clear trajectory: all three spectral metrics revealed an increased weight of stress-time periodicities in the amplitude

envelope, with the most advanced learners emphasizing stress-timed periodicities beyond the native baselines.

To sum up, L2 groups showed systematic differences from both native baselines, with a developmental trajectory that suggests an increasing exaggeration of stress periodicities over the four semesters of instruction. The last stage (EN-L2-4) did not show full convergence with EN-L1 but rather exhibited the highest values across metrics⁵, suggesting a prosodic system that exaggerates native stress-timed patterns, possibly relying on the amplification of stress cues.

3.2 Intermetric Correlations

3.2.1 L1 Systems: slightly different patterns of co-variation

The correlations of rhythm metrics for the two native speaker groups (see Table 3) did not show significant differences, although the *imf_ratio*₂₁ correlation analysis indicates different spectral co-variation patterns. In both groups, a strong negative correlation emerged between *sbpr*₁ and *scntr*₁: $r = -0.86$ for BP-L1 and $r = -0.87$ for EN-L1. This pattern suggests that as stress-level oscillations become more prominent (higher *sbpr*₁), spectral energy shifts toward slower frequency bands (lower *scntr*₁).

Table 3. Intermetric Pearson correlations for spectral rhythm metrics pairs in BP-L1 and English-L1.

L1 group	Metric Pair	r
BP-L1	<i>sbpr</i> ₁ ~ <i>scntr</i> ₁	-0.86
	<i>sbpr</i> ₁ ~ <i>imf_ratio</i> ₂₁	+0.51
	<i>scntr</i> ₁ ~ <i>imf_ratio</i> ₂₁	-0.39
English-L1	<i>sbpr</i> ₁ ~ <i>scntr</i> ₁	-0.87
	<i>sbpr</i> ₁ ~ <i>imf_ratio</i> ₂₁	+0.38
	<i>scntr</i> ₁ ~ <i>imf_ratio</i> ₂₁	-0.48

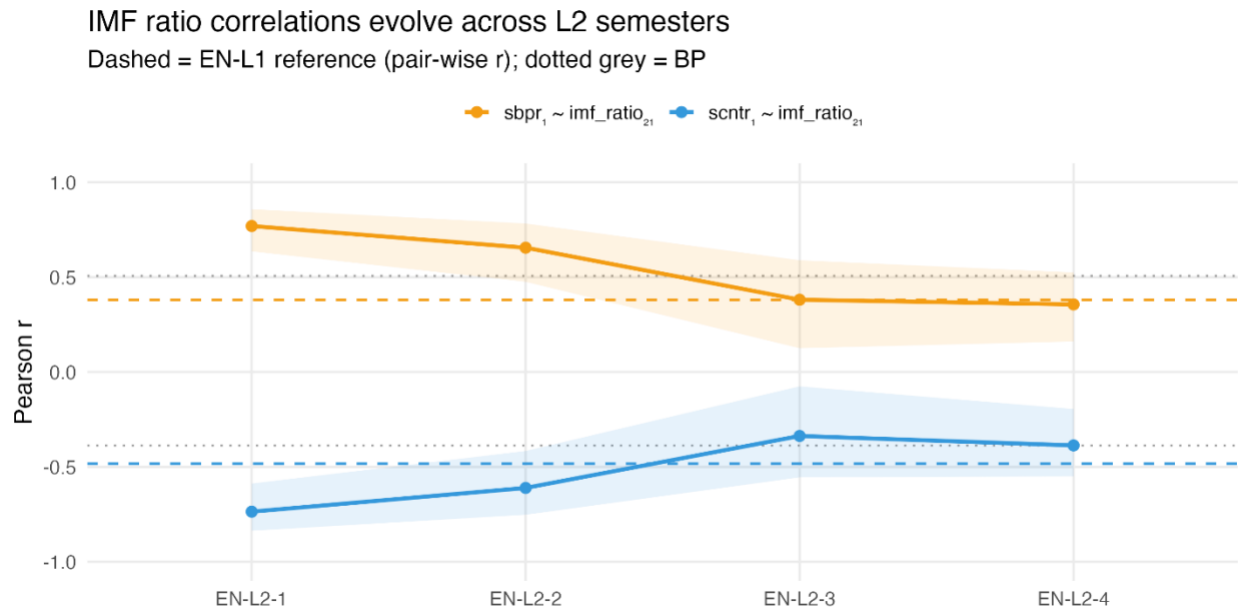
However, cross-linguistic differences appeared when examining *imf_ratio*₂₁, which indicates the dominance of slower, stress-level oscillations over faster syllabic ones. In BP-L1, the positive correlation between *sbpr*₁ and *imf_ratio*₂₁ is somewhat stronger ($r = +0.51$) than in English ($r = +0.38$), implying a tendency for stress-like oscillations to occur alongside increased stress-level energy. This pattern should be viewed with caution due to the modest difference. Conversely, English shows a slightly stronger negative correlation between *scntr*₁ and *imf_ratio*₂₁ ($r = -0.48$ versus -0.39 in BP), suggesting that greater dominance of IMF₂ is more consistently linked to reduced spectral dispersion in English.

⁵ Except for *scntr*₁, for which low values point to a center of gravity concentrated in slower, stress-timed periodicities. EN-L4

3.2.2 L2 Systems: the emergence of native-like intermetric coordination

As shown in Figure 2, during the earliest stage of learning (EN-L2-1), correlations among the metrics differ from those in both L1 baselines. The correlations among all three metrics are strong: $sbpr_1 \sim scnr_1 = -0.88$, $sbpr_1 \sim imf_ratio_{21} = +0.77$, and $scnr_1 \sim imf_ratio_{21} = -0.74$. This suggests that learners initially depend on a rigid prosodic strategy, where increases in stress-band energy ($sbpr_1$) are closely linked with broader spectral dispersion ($scnr_1$) and relative IMF activity (imf_ratio_{21}).

Figure 2. Intermetric Pearson correlations for spectral rhythm metrics in English-L2-1 to English-L2-4.



Over time, these correlations begin to weaken as learners advance, indicating a loosening of internal constraints. In EN-L2-2, the correlation between $sbpr_1$ and imf_ratio_{21} drops (+0.64) and continues declining in EN-L2-3 (+0.38), while the $scnr_1 \sim imf_ratio_{21}$ link also becomes weaker (-0.34). These changes may indicate that learners are beginning to differentiate the functional roles of prosodic features, moving toward a more complex organization. By EN-L2-4, the correlation pattern becomes even more similar to that of EN-L1. While the $sbpr_1 \sim scnr_1$ correlation remains strong (-0.88), the association between $sbpr_1$ and imf_ratio_{21} decreases even more (+0.36), almost equivalent to the English-L1 value (+0.38). The $scnr_1 \sim imf_ratio_{21}$ correlation remains moderate (-0.37), again approximating native levels (-0.48). This convergence in relational patterns, rather than absolute metric values alone, highlights an important aspect of rhythm development: the emergence of native-like intermetric coordination.

These findings suggest that L2 rhythm development can be observed through both changes in individual metric behavior as well as in intermetric correlations among rhythmic cues. Early L2

speech may rely on overcoupled rhythmic strategies, which gradually weaken and restructure as learners gain proficiency.

However, it is worth pointing out that these group-level differences in the correlation analysis could in part be explained by chunk duration, since longer or shorter utterances may affect the relative distribution of spectral components. This potential confound suggests future analyses should control for duration effects.

4. Conclusion and Future Directions

This study provides new evidence for the usefulness of envelope-based metrics for studying L2 rhythm development. Correlation analyses also emerged as a complementary tool to reveal the internal coordination among multiple envelope parameters.

The unexpected lack of a difference between the L1 baselines could be due to several potential factors: (i) the data were elicited via read speech rather than spontaneous speech; (ii) the content of the utterances in the English corpus may have produced some bias toward syllable periodicity; (iii) there might be a subtle difference between the L1s, but, at this stage, selected data lacks statistical power to detect it; (iv) Canadian English may be less stress-timed than other varieties of English; and (v) some of our hyperparameters may have introduced variability or bias that influences the result, particularly the choice of chunk-duration could be contributing a factor.

As for L2 learners, groupwise results revealed a complex developmental pattern: L2 learners exhibited signs of prosodic hyper-articulation by exaggerating stress-timed periodicities. At the same time, the correlation analysis revealed an important parallel trajectory: learners progressed from an overly coupled rhythmic system in early stages to a more varied and native-like relational structure in later stages.

Future work should expand the metric repertoire to include other envelope-based metrics, such as rate metrics and rhythmic stability metrics, to provide a more comprehensive profile of rhythmic development. Additionally, increasing the sample size across both L1 and L2 groups would allow for more robust generalizations and validation of the observed trends, which should be interpreted with caution at this point. Finally, future work involving highly proficient bilinguals could clarify whether prosodic rhythm in advanced L2 users ultimately converges to native-like dynamics.

5. References

Abercrombie, D. (1967). *Elements of general phonetics*. Edinburgh: Edinburgh University Press.

Albert, A., Grice, M. (2024). Rhythm is a timescale. [10.31234/osf.io/h3g7t](https://doi.org/10.31234/osf.io/h3g7t).

Celce-Murcia, M.; Brinton, D. M.; Goodwin, J. M., & Griner, B. (2010). *Teaching Pronunciation: a course book and reference guide*. Cambridge: Cambridge University Press.

Cho, T., Lee, Y. & Sahyang, K. (2011). Communicatively driven versus prosodically driven hyper-articulation in Korean. *Journal of Phonetics*, 39(3), 344-361, <https://doi.org/10.1016/j.wocn.2011.02.005>.

Cummins, F. (2009). Rhythm as affordance for the entrainment of movement, *Phonetica* 66, 15–28.

Dellwo, V. (2006). Rhythm and speech rate: A variation coefficient for deltaC. In P. Karnowski & I. Szigeti (Eds.), *Language and language-processing* (pp. 231–241). Peter Lang.

Fuchs, R. *Speech rhythm in varieties of English*. Singapore: Springer, 2016.

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). Berlin: Mouton de Gruyter.

Huang, N. E., Shen, Z., Long, S. R., Wu, M. C., Shih, H. H., Zheng, Q., Yen, N.-C., Tung, C. C., and Liu, H. H. (1998). The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis, *Proc. R. Soc. London, A* (454), 903–995.

Jassem W. 1950. Indication of speech rhythm in the transcription of educated Southern English. *le Maitre Phonétique*. III/92, 22-24.

Kohler, K. J. (2008). The perception of prominence patterns, *Phonetica* 65, 257–269.

Lüdecke et al., (2021). performance: An R Package for Assessment, Comparison and #> Testing of Statistical Models. *Journal of Open Source Software*, 6(60), 3139. #> <https://doi.org/10.21105/joss.03139>

Munro, M. J., & Derwing, T. M. (2001). Modeling perceptions of the accentedness and comprehensibility of L2 speech: The role of speaking rate. *Studies in Second Language Acquisition*, 23(4), 451–468.

Ramus, F., Nespor, M., & Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition*, 73(3), 265–292.

Silva Jr., B. L., & Barbosa, P. A. (2019b). A new method for extracting prosodic metrics and its application to L2 speech analysis. *Proceedings of Interspeech 2019*, 1353–1357.

Smiljanić, R., Bradlow, A. R. (2009). Speaking and Hearing Clearly: Talker and Listener Factors in Speaking Style Changes. *Language and Linguistics Compass*, 3(1), 236–264, <https://doi.org/10.1111/j.1749-818X.2008.00112.x>

Teixeira, L. A.S. (Teixeira). A Multidimensional Analysis of English-L2 Rhythm Development. *Ilha do Desterro*, 76 (3), 223-250, <https://doi.org/10.5007/2175-8026.2023.e94742>

Tilsen, Sam & Johnson, K. (2008). Low-frequency Fourier analysis of speech rhythm. *The Journal of the Acoustical Society of America*. 124. EL34-9. 10.1121/1.2947626.

Tilsen, S., & Arvaniti, A. (2013). Speech rhythm analysis with empirical mode decomposition. *Journal of the Acoustical Society of America*, 134(2), 1520–1533.

Tilsen, S. (2019). Space and time in models of speech rhythm. *Annals of the New York Academy of Sciences*. <https://doi.org/10.1111/nyas.14102>.