A MULTIDIMENSIONAL ANALYSIS OF ENGLISH-L2 RHYTHM DEVELOPMENT

Leonardo Antonio Silva Teixeira[1*]
[1]Universidade Federal do Ceará, Fortaleza, CE, Brasil

Abstract

This work aims to discuss the development of L2 English rhythm by Brazilian learners (L2ers) through rhythmic metrics and acoustic parameters. Five L2ers were recorded reading a text in English at the beginning of their college studies in English Language and Literature, and again, four semesters later, after having taken two English phonology courses. They were also recorded reading a Portuguese version of the same text. As for the control group, five native speakers of a North American variety of English were recorded reading the text in English. Data were manually segmented in PRAAT, and the indexes were automatically extracted. As for English-L2, different developmental paths were captured as function of the rhythm dimension, with L1 influencing more L2ers' rhythm in the f_0 and intensity dimensions. Crucially, L2 rhythm patterns converged to the target language in all dimensions, suggesting positive effects of explicit pronunciation teaching.

Keywords: phonological acquisition; prosody; rhythm metrics; acoustic parameters; explicit instruction

* E-mail: leoteixeira.professor@gmail.com. Orcid: https://orcid.org/0000-0002-4905-4259.

1. Introduction

Most linguistic environments in the world are bilingual (Ferreira et al., 2018). Therefore, bilingualism seems to be the rule, not the exception, and there has been a great deal of research investigating the oral production of bilinguals, and speech differences arising from the realization of native and non-native systems (henceforth L1 and L2 respectively), at multiple levels of analysis. Within this body of research, segmental aspects rather than prosodic ones have been the most investigated (Li & Post, 2014; Thomson & Derwing, 2015; Teixeira, 2021). This tendency is also confirmed by L2 acquisition models available in the literature, such as the Speech Learning Model (Flege, 1995) and its revised version - Revised Speech Learning Model (Flege & Bohn, 2021), as well as the L2 - Perceptual Assimilation Model of Second Language Speech Learning (Best & Tyler, 2007). By emphasizing segmental aspects, those models offer little support to the understanding of L2 prosody development. Crucially, some studies have demonstrated the persistence of atypical prosodic patterns in L2 production even in advanced-level users (Li & Post, 2014), such as variations of fundamental frequency, inadequate stress marking, and speech timing differences, which attests the need for more research on L2 prosody development.

The production of those atypical prosodic patterns by the L2 speaker (henceforth L2er) affects comprehensibility (i.e., the level of effort required to understand spoken utterances) to a greater extent (Moreno, 2000) and impair communication more seriously (Celce-Murcia et al., 2010) than those arising from the segmental level. Furthermore, prosodic competence can facilitate understanding even when L2ers make mistakes at the lexical, stylistic and grammatical levels (Moreno, 2000).

Our working definition of prosody is based on Barbosa (2012), who defines it as an umbrella term that entails linguistic features, such as stress, intonation and rhythm, paralinguistic features, such as discourse and social markers, and extralinguistic factors such as emotions instantiated in enunciative acts. Those factors, combined with social and biological variables "shape our enunciation by imprinting on "what is spoken" a "way of speaking" that is intentionally or unintentionally directed at the listener" (Barbosa, 2012, p.14).

Among those prosodic aspects, rhythm is the least explored (Whitworth, 2002; Cumming, 2010; Gut, 2012), despite evidence that it can influence the communication process globally, affecting degrees of foreign accent, intelligibility (Silva Jr. & Barbosa, 2019b) and comprehensibility (Munro & Derwing, 2001; Ordin & Polyanskaya, 2015). Furthermore, rhythm can offer learners acoustic cues that guide them in the process of speech segmentation. The prosodic constituents resulting from that operation perform multiple functions in the linguistic, paralinguistic and expressive fields (Barbosa, 2012).

As for the psycholinguistics of rhythm, empirical data suggest that babies can perceive interlinguistic differences as a function of rhythm (Mehler et al., 1988). The work of psycholinguists also suggests that the development of

phonological analysis by infants is based on rhythmic differences, as there seems to be a correlation between rhythmic typology and speech signal segmentation (Ramus, Nespor & Mehler, 1999). In the early stages of L1 development, rhythm also attends to the detection of word boundaries (Cutler, 1996), and later in the reading process (Holliman & Wood, 2010), with indications that it might remain operating as an acoustic cue in word perception even in adulthood.

There have been at least three waves of research on linguistic rhythm. Underlying the first one, the Hypothesis of Isochrony thought of rhythm as the effect of prominence arising from the isochronous (of near-equal duration) recurrence of some kind of speech segment (syllables, morae or interstress intervals). Abercrombie (1967) suggested that all languages in the world would fall into one exclusive rhythm category, either syllable-timed or stress-timed. In stress-timed languages (e.g., English and German) interstress intervals should be near-equal, while in syllable-timed languages (e.g., Spanish and Italian), syllables would be such isochronous speech units. However, many studies have demonstrated the implausibility of the isochrony paradigm (Dauer, 1983; Borzone de Manrique & Signorini, 1983; Dellwo, 2006).

In the second wave of studies, rhythm is approached as a gradient rather than a categorical phenomenon. Within this paradigm, rhythm is investigated through statistical indexes called rhythmic metrics, which calculate the degree of durational variability of a given speech unit and place languages on a gradient plane. It is worth mentioning that such approach was backed up by a series of psycholinguistic studies that showed the role of prominence in early language acquisition and processing: (a) Mehler et al. (1996), who point out that babies perceive vowels as they have more spectral energy, last longer than consonants, carry accent and signal stress; b) Bertoncini et al. (1989), who suggest that infants pay more attention to vowels; c) Van Ooijen (1994), who claims that infants are capable of identifying the number of syllables, regardless of syllabic structure or weight, perceiving speech as a sequence of vowels that vary in terms of duration and intensity and alternate with periods of unanalyzed noise (consonants). Based on those findings, Ramus, Nespor & Mehler (1999) proposed that, by statistically analyzing durational patterns of consonants and vowels in the speech signal, one could account for rhythmic typology, offer support to the understanding of language perception observed in infants and provide acoustic correlates for rhythm. Ramus, Nespor & Mehler (1999) proposed a bidimensional approach in which ΔC, the standard deviation of consonant intervals, and %V, the percentage of vowel segments in the utterance, were able to spatially discriminate languages considered syllable-timed (French, Spanish, Italian, and Catalan), stress-timed (English, Polish, and Dutch) and mora-timed (Japanese) (Ladefoged, 1975) on a plane with ΔC and %V on each axis, as can be seen in Figure 1, reproduced from the original paper:

Figure 1: Distribution of languages over the (%V, ΔC) plane. Error bars represent 1 standard error.
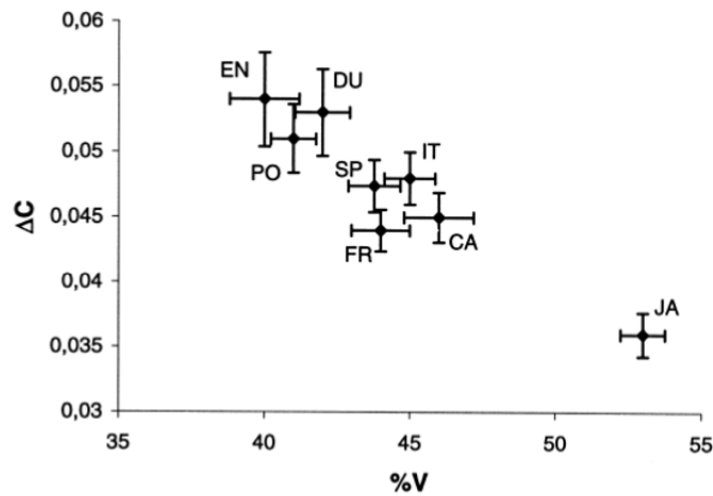
Note: Elaborated by Ramus, Nespor & Mehler (1999).

All metrics that were later proposed by other researchers based on that line of research (cf. Fuchs, 2016; Teixeira; 2021, for a comprehensive account of those metrics) can predict gradience in rhythm because languages can behave more syllable-timed in relation to a metric, and more stress-timed in relation to another.

Finally, a third wave of studies, with which the present study is aligned, seeks to investigate rhythm through correlates of prominence other than duration, such as intensity, fundamental frequency (henceforth f_0) and speech rate. Hence, our work defines rhythm as a function of the distribution of prominent elements in the acoustic signal, that is, the systematic patterning of these features throughout the speech signal, which involves several acoustic dimensions – duration, f_0, intensity and speech rate, and may be influenced by the native language of the speaker (Cumming, 2010; Fuchs, 2016; Silva Jr., & Barbosa, 2019b). For the purpose of this work, the indexes computing prominence in the dimensions of intensity and f_0, as well as speech rate, will be referred to as *acoustic parameters*, whereas durational-based indexes will be called *rhythm metrics*, or simply *metrics*. The separation of speech rate as a distinct acoustic dimension from duration is primarily due to the difference in focus and interpretation of these two measures in linguistic analysis. Speech rate is indeed calculated based on duration, specifically as the number of syllables produced per second of speech (Laver, 2005). However, it is treated as a separate parameter because it reflects the pace or speed of speech production, rather than the absolute duration of individual linguistic units (e.g., phonemes, syllables, words). In summary, while speech rate is calculated using durational information, it serves a different analytical purpose.

This work aims to apply durational-based metrics and acoustic parameters (cf. Tables 1 and 2) to the acoustic analysis of four corpora of oral production: Brazilian Portuguese (henceforth BP) - L1, English-L1 (henceforth Eng-L1),

and English-L2 (henceforth Eng-L2) in two distinct stages of development. Given that few works on rhythm investigate L2ers, and an even smaller number investigate English-L2ers who are native speakers of BP, this work is also intended to bridge a gap of studies that consider rhythm development in L2 in correlation to explicit pronunciation teaching, in a longitudinal perspective. Recently, Silva Jr. (2023) demonstrated the potential of *BeatMaker*, a technology designed to aid L2 pronunciation teaching by focusing on prosodic elements such as f_0 and duration. The research showed that L2ers, particularly those whose L1 differs significantly from the target L2 in terms of prosody, can benefit from training that emphasizes prosodic aspects, improving their intonation and f_0 contours. These findings reveal the importance of incorporating the explicit pronunciation teaching of prosody into language instruction.

The main purpose of this work is to describe the rhythm development of Eng-L2 by Brazilian L2ers through rhythmic metrics and acoustic parameters that characterize their oral production in different stages, in a multidimensional perspective. We follow Lima Jr. (2016) and claim that the process of phonological development in L2 must be studied through the analysis of individual and longitudinal data, because of its intrinsically dynamic and nonlinear character. Thus, this study was guided by the following questions: (i) how do the 30 metrics and 14 acoustic parameters (see Tables 1 and 2) place North American Eng-L1, Eng-L2, and BP-L1 in the rhythmic space? (ii) What is the influence of BP-L1 on the development of Eng-L2 rhythm patterns? (iii) What is the effect of explicit pronunciation teaching on learners' Eng-L2 rhythm? The following hypotheses were raised: (i) BP-L1, Eng-L1 and Eng-L2 are rhythmically different systems; (ii) there will be rhythmic differences between the Eng-L2 of the speakers in the two different stages of development analyzed; (iii) the Eng-L2 of the first recording should be more dissimilar to Eng-L1 due to a greater L1 influence and lack of explicit instruction.

2. Methods

It should be noted that this work analyzed speech data from a previously constituted database. Therefore, the procedures related to participants, speech data collection, and stimuli choice did not aim to directly analyze speech rhythm, but rather constitute a corpus to serve a variety of phonetic-phonological studies, both at segmental and prosodic levels.

2.1 Participants

The experimental group makes up a total of five L2ers, undergraduate students of English Language and Literature, four males and one female, aged between 18 and 24 years, all sharing the same L1 variety, that is, the Brazilian Portuguese spoken in the city of Fortaleza, located in the northeastern part of the country. The criteria for participation in the research included: a) not

having stopped or failed the undergraduate course; b) not having traveled to an L1-English speaking country; and c) not having had continuous contact with a native English speaker (self-reported). Students were recorded four times, in four consecutive semesters. In the present study, we analyzed the first and fourth recordings [henceforth Eng-L2(1) and Eng-L2(4) respectively], which were obtained respectively before and after they took two undergraduate courses in English Phonetics and Phonology. The control group consisted of five Eng-L1 speakers, all Canadians, aged between 23 and 34, one man and four women. Data collection was authorized by the Research Ethics Committee (CAAE: 40985414.1.0000.5054) and the participants signed a free consent form.

2.2 Materials

Four corpora of oral production were then analyzed: Eng-L1, BP-L1, Eng-L2 (1), and Eng-L2 (4). Eng-L2 data were obtained by recording the L2ers reading the diagnostic text of Celce-Murcia et al. (2010). The present work analyzed only the first paragraph read (see Appendix A). This text was read by the control group to collect Eng-L1 data. The L2ers also read the Portuguese version of the same text so that BP-L1 data could be obtained (see Appendix B). The text was translated by a linguist who is a native BP speaker with high proficiency in English. Both text versions contain seven sentences, including simple and complex ones, declarative affirmatives, and closed and open-ended questions.

2.3 Procedures

The recordings were conducted in a silent room using a cardioid Shure MX150B lapel microphone connected to a Zoom 4HnSP recorder. The audio was recorded in mono at a sampling rate of 44.1 kHz and a quantization rate of 16 bits, and subsequently saved in .wav format. For the acoustic analysis, the data were manually segmented in PRAAT (Boersma & Weenink, 2021), in vowels (V), consonants (C), phonetic syllables (VV), sentences (s), syntactic-prosodic units (chunks) and pauses (#). As can be seen in Figure 2, the VV unit corresponds to the interval between the acoustic onset of a vowel and the onset of the adjacent vowel, integrating phones of two distinct syllables (Barbosa, 2007).

The sentences were segmented according to syntactic-prosodic criteria, with the goal of differentiating strong and weak tonal boundaries: verb phrases formed by copula; elements in complementizer position; declarative sentences; open and closed interrogative sentences; complex sentences; and subordinate clauses. For the chunks, longer pauses (above 600ms) marking the end of a group of sentences, were used as a segmentation reference. The application of this segmentation protocol resulted in 10 sentences and 5 chunks per speaker in each corpus analyzed.

Figure 2: Partial waveform, broadband spectrogram, and four tiers, respectively, segmented and labeled as: 1) phonetic syllables (VV); 2) vowels and/or consonants (V/C); 3) sentences (S); 4) chunks (CH) produced by an L2er.
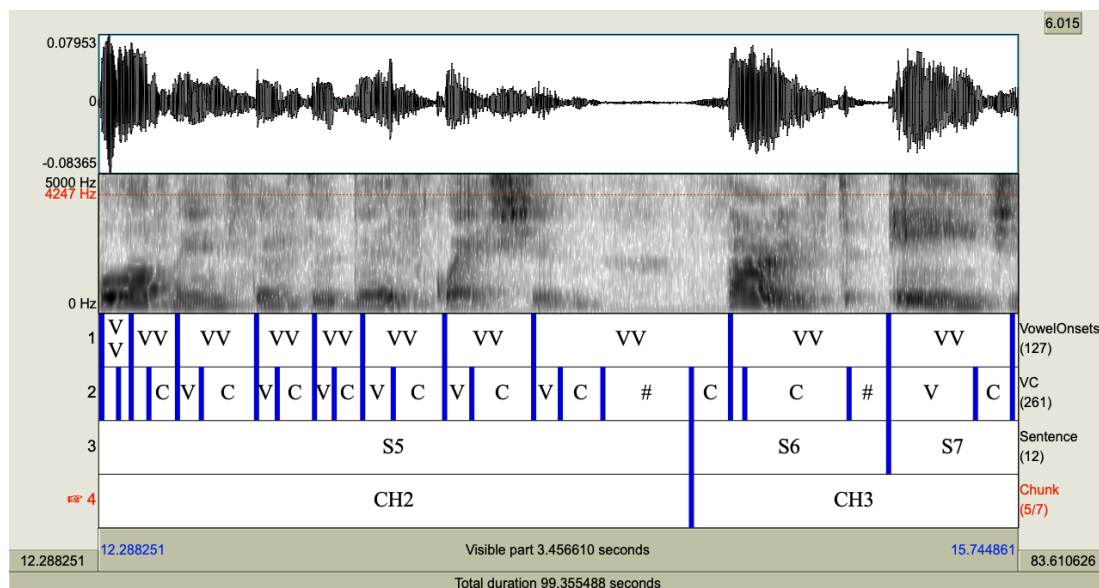
Note: Elaborated by the author.

For the extraction of the 30 metrics and 14 acoustic parameters, a PRAAT script, the Metrics&AcousticsExtractor (Silva Jr., & Barbosa, 2019a), was used (except for *Scope of* f_0, which was manually calculated). Tables 1 and 2 specify the metrics and acoustic parameters that were used in this study, the segments to which they were applied, and what they mean in terms of rhythm.

Table 1: Rhythm metrics

Metrics	*Application*	*Definition*
Percentual (%)	V, C	Proportion of the total duration of the reference segment in the utterance.
Standard-deviation (Δ)	V,C, VC, VV	Standard deviation of the segment duration divided by the mean, multiplied by 100.
Variation coefficient (Varco)	V,C, VC, VV	Mean of the differences between successive segments.
Raw pairwise variability index (r- PVI)	V,C, VC, VV	Mean of the differences between successive segments divided by their sum, multiplied by 100.
Normalized pairwise variability index (n-PVI)	V,C, VC, VV	Mean of pairwise quotients of adjacent segment durations, where the duration of the shorter is divided by the duration of the longer one and multiplied by 100.
Rhythm ratio (RR)	V,C, VC, VV	Mean of pairwise quotients of adjacent segment durations, where the duration of the shorter is divided by the duration of the longer one and multiplied by 100.
Variability index (VI)	V,C, VC, VV	Mean of the differences between successive segments where the duration of each segment is normalized through division by the mean of all segments' durations.

Yet another rhythm determination (z-score duration) (YARD)	V,C, VC, VV	Mean of the differences between successive segments where the durations are normalized by z-transformation.

Note: Elaborated by the author.

Table 2: Acoustic Parameters

Acoustic Parameters	*Application*	*Definition*
f_0 median	S, CH	It provides insights into the regularity and timing patterns of vocalizations. A high f_0 median is typically an indicator of stress-timed rhythm.
f_0 peak	S, CH	The highest f_0 value in the f_0 contour. The presence and location of f_0 peaks can also indicate variations in rhythm, including pauses, phrase boundaries, and other prosodic features.
f_0 minimum	S, CH	The lowest f_0 value in the f_0 contour. It contributes to the division of speech into distinct syllabic units, influencing syllable timing and pacing.
Scope of f_0	S, CH	A larger scope of f_0 (f_0 peak - f_0 minimum) indicates a greater difference between the highest and lowest pitches, while a smaller pitch range suggests a more limited range of pitch variation.
f_0 standard deviation	S, CH	A higher f_0 standard deviation suggests greater pitch variability within the speech segment.
f_0 skewness	S, CH	The measure of asymmetry in the distribution of f_0 values within a segment of speech. Left skewness suggests a concentration of lower-pitched f_0 values in the segment, while right skewness suggests the prevalence of higher-pitched ones.
Mean of f_0 first derivative ($\mu\Delta1$-f_0)	S, CH	The mean rate of change in pitch for that segment, $\mu\Delta1$-f_0 quantifies how quickly the pitch is rising or falling over time. A higher $\mu\Delta1$-F0 indicates a faster rate of pitch change and a more dynamic rhythm.
Standard deviation of f_0 first derivative ($\sigma\Delta1$-f_0)	S, CH	It quantifies the variability or dispersion of the rate of change of f_0. A higher $\sigma\Delta1$-f_0 indicates that the pitch changes within the segment are more diverse and less uniform.
Skewness of f_0 first derivative ($sk\Delta1$-f_0)	S, CH	It quantifies whether the rate of change of pitch is skewed to one side (left or right) relative to the mean rate of change.
Speech rate (SR)	VV, S, CH	It indicates the average number of syllables a speaker produces in one second of speech.
f_0 rate (f_0-R)	S, CH	It quantifies how quickly or slowly the pitch of a speaker's voice is changing.
Spectral emphasis	S, CH	It measures the energy or amplitude of specific frequency ranges within the speech signal. It is associated with phrasal stress marking.
Mean of normalized syllable- peak duration (μdur- Sil)	VV, S, CH	It provides information about the timing patterns within speech, specifically related to the duration of syllable peaks.
Mean duration of pauses (μdur-#)	S, CH	It reflects how speakers pace their speech and the organization of temporal intervals between speech segments.

Note: Elaborated by the author.

Data obtained from the extraction of those metrics and acoustic parameters were statistically analyzed via R (R Core Team, 2021) through the implementation of mixed-effect regression models, which treated the variable 'Lang' [Eng-L1, BP-L1, Eng-L2(1) and Eng-L2(4)] as fixed effect, with BP as the intercept, serving as the focal point for investigating the influence of the L1s and L2s on the rhythm metrics and acoustic parameters. By including 'Lang' as a fixed effect, we aimed to discern whether those distinct systems exhibited significant variations in rhythm. The magnitude of the effect of 'Lang' on the rhythm indexes was captured by the estimated coefficients in our models. Positive or negative coefficients indicate the direction and strength of the relationship, with larger coefficients signaling a more substantial effect of language type. In our analysis, random effects were specified for both the 'Chunk' and 'Speaker' (participant) variables to account for potential correlations and variability within these grouping factors.

3. Results

The results presented in this section are organized into two parts: *metrics* and *acoustic parameters*.

3.1 Metrics

Twenty out of the thirty employed metrics reached statistical significance for at least two of language systems, as can be seen in Table 3:

Table 3: Absolute means for the statistically significant metrics and standard deviation (between parentheses) for BP, Eng-L1, Eng-L2 (1), Eng-L2(1) and Eng-L2(4)

Metric	*BP*	*Eng-L1*	*Eng-L2(1)*	*Eng-L2(4)*
%V	48.56 (3.16)	38.88 (4.96)	36.24 (5.36)	46.48 (8.02)
%C	51.44 (3,16)	61.12 (4.96)	63.76 (5.36)	68.416 (14.55)
ΔV	40.08 (10.81)	41.16 (11.82)	51.81 (12.79)	105.192 (32.6)
ΔC	46.48 (8.02)	68.41 (14.55)	105.192 (32.6)	84.088 (36.51)
ΔS	133.4 (45.27)	198.53 (77.44)	217.46 (97.65)	184.75 (56.72)
VarcoV	49.04 (8.49)	54.80 (12.52)	58.80 (11.30)	56.32 (14.81)
VarcoC	49.84 (7,00)	59 (11.89)	66.80 (13.69)	59.72 (22.02)
rPVI-V	65.1 (11.71)	70.74 (14.84)	96.98 (19.27)	81.21 (15.75)
rPVI-C	48.22 (8.91)	86.16 (20.23)	116.84 (46.64)	88.7 (21.69)
rPVI-VC	64.73 (18.72)	83.58 (11.12)	114.4 (38.18)	89.1 (14.53)
rPVI-S	102.85 (40.24)	130.66 (33.59)	176.27 (90.03)	137.96 (44.62)

nPVI-C	53.96 (7.21)	68.56 (11.72)	72.36 (12.14)	64.84 (11.46)
nPVI-VC	59.76 (7.69)	68.96 (11.09)	72.6 (9.40)	65.08 (8.12)
RR-C	61.17 (4.36)	53.13 (6.29)	50.97 (6.59)	54.59 (6.42)
RR-VC	58.07 (4.35)	52.8 (5.65)	50.91 (4.97)	54.55 (4.59)
VI-V	0.818 (0.166)	0.981 (0.322)	1.128 (0.302)	0.894 (0.188)
VI-C	0.830 (0.037)	0.924 (0.049)	0.929 (0.052)	0.934 (0.059)
VI-VC	0.684 (0.101)	0.834 (0.157)	0.859 (0.120)	0.746 (0.116)
VI-S	0.516 (0.120)	0.606 (0.136)	0.615 (0.160)	0.538 (0.126)
YARD-VC	0.717 (0.150)	0.695 (0.133)	0.869 (0.113)	0.848 (0.123)

Note: Elaborated by the author.

One example of the mixed-effect regression models that were implemented via R can be seen in Table 4, which was adjusted, in this example, for the standard deviation of the duration of consonantal intervals (ΔC) and the percentage of vocalic intervals (%V).

Table 4: Coefficients, confidence intervals (95%) and p-Values for the two linear mixed-effect regression models adjusted for ΔC and %V. models: deltaC ~ Lang + (1|Chunk) + (1|Speaker) and percV ~ Lang + (1|Chunk) + (1 / Speaker)

Predictors	ΔC			%V		
	Estimates	CI	p	Estimates	CI	p
(Intercept)	46.48	36.17 – 56.79	*<0.001*	48.78	46.02 – 51.55	*<0.001*
Lang [Eng-L1]	21.94	7.35 – 36.52	*0.004*	-9.66	-12.97 – -6.35	*<0.001*
Lang [Eng-L2 (1)]	58.71	44.13 – 73.30	*<0.001*	-11.92	-14.23 – -9.61	*<0.001*
Lang [Eng-L2 (4)]	37.61	23.02 – 52.19	*<0.001*	-9.28	-11.47 – -7.09	*<0.001*

Note: Elaborated by the author.

Figure 3 shows the distribution of the 4 corpora over the classic bidimensional planes formed by the pairs ΔC-%V, VarcoC-VarcoV and rPVI-C and nPVI-V in comparison to the data reviewed and obtained by Arvaniti (2012).

Figure 3: Present study data (dark blue) amid all the data reviewed and obtained by Arvaniti (2012) (light blue) for ΔC - %V (3a), VarcoC-VarcoV (Figure 3b) and rPVI-C-n-PVI-V (3c), in which Eng = English, Ger = German, Gre = Greek, Spa = Spanish, UI = Italian, Kor = Korean.
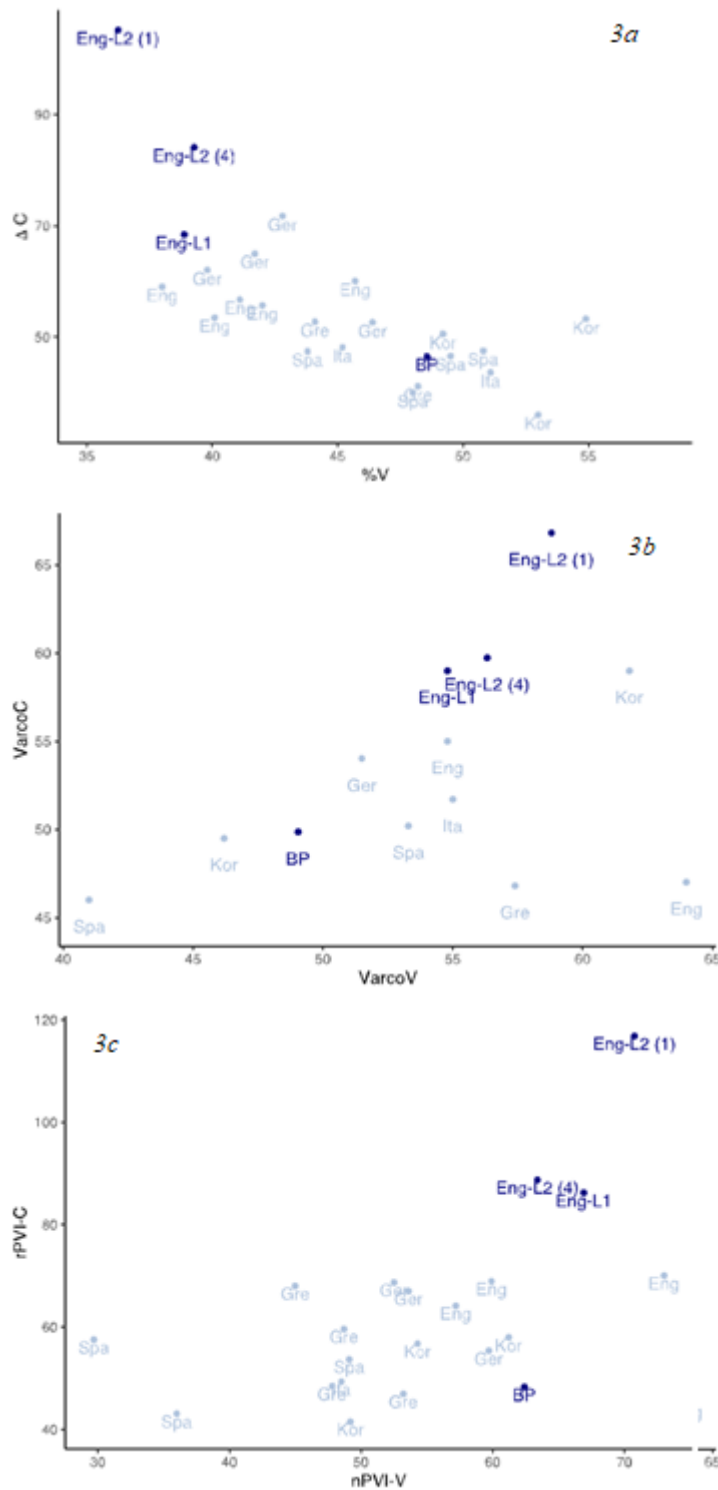
Note: Elaborated by the author.

According to Fuchs (2016), percentage metrics, such as %V and %C (cf. Table 1) represents the substantial contribution of the segment of reference in the composition of the oral production corpus. It is expected that the greater the number of syllabic arrangements a language allows (e.g., allowing for complex consonant clusters), the greater the stress-timing rhythm tendency will be. As for standard deviation metrics, they globally compute the degree of variation in the duration of the reference intervals, and greater variability signals a tendency towards a stress-timing organization. As can be seen in Figure 3a, Eng-L1 presented a greater standard deviation of consonantal duration ($\Delta C_{\text{Eng-L1}}$ = 68.41) compared to BP (ΔC_{BP} = 46.48); and BP presented a greater proportion of vocalic constitution (%V_{BP} = 48.56) compared to Eng-L1 (%$V_{\text{Eng-L1}}$ = 38,88). Eng-L2 (1) data were positioned far from the two L1s, scoring quite high ΔC values ($\Delta C_{\text{Eng-L2 (1)}}$ = 105.19) and the lowest proportion of vowel segments (%$V_{\text{Eng-L2 (1)}}$ = 36.24). On the other hand, Eng-L2 (4) values were much closer to Eng-L1 in relation to both axes ($\Delta C_{\text{Engl-L2(4)}}$ = 84.08; %$V_{\text{Eng-L2(4)}}$ = 39.28).

As for Varcos, they globally calculate the standard deviation of the reference interval duration and normalize them through the division by the mean, multiplied by a 100 afterwards. Such methodology is meant to mitigate the effect of speech rate. Languages that lean towards stress-timing rhythm are expected to display high Varco means. As for VarcoC-VarcoV (Figure 3b), Eng-L1, Eng-L2 (1), Eng-L2 (4) and BP were distributed analogously to the plane ΔC-%V, with the BP data recording the lowest values for both VarcoC ($\text{VarcoC}_{\text{BP}}$ = 49.84) and VarcoV ($\text{VarcoV}_{\text{BP}}$ = 49.04) axis, and Eng-L2 (1) presenting the highest scores both in the VarcoC axis ($\text{VarcoC}_{\text{Eng-L2(1)}}$ = 66.8) and in relation to the VarcoV axis ($\text{VarcoV}_{\text{Eng-L2(1)}}$ = 58.8). The fact that Eng-L2(1) assumed values far from the L1 (BP) indicates no objective influence of durational prosodic patterns to the L2 systems. On the other hand, the approximation between Eng-L2(4) and Eng-L1 indicates a possible effect of explicit instruction, among other factors, that may have influenced the temporal (re)organization of the L2ers' speech towards the prosodic patterns of the target language.

Paired variability indexes (PVI), in their turn, calculate the mean of the differences between adjacent reference intervals. These indexes are considered local and can be raw (rPVIs) or normalized (nPVIs) for speech rate. A lower PVI indicates less durational variability between adjacent reference intervals, which favors the implementation of syllable-timed rhythm. In the plane formed by the axes rPVI-C and nPVI-V (Figure 3c), Eng-L1, Eng-L2 (1), Eng-L2 (4) and BP also assumed distinct positions, although the greatest distances were recorded on the rPVI-C axis (rPVI-C_{BP} = 48.22; rPVI-$C_{\text{Eng-L1}}$ = 86.16; rPVI-$C_{\text{Eng-L2(1)}}$ = 116.84; rPVI-$C_{\text{Eng-L2(4)}}$ = 88.70). BP and Eng-L2(1) once again occupied opposite positions, and Eng-L1 and Eng-L2 (4) occupying intermediate and very close positions, indicating a developmental path towards the prosodic patterns of the non-native language, at least in relation to the variability of adjacent consonant intervals. In relation to the nPVI-V axis, the results are less elucidative, although the data indicate a greater tendency to syllabic rhythm for BP (nPVI-V_{BP} =

62.4) compared to Eng-L1 data (nPVI-V$_{Eng\text{-}L1}$ = 66.92) and the L2 system data (nPVI-V$_{Eng\text{-}L2(1)}$ = 70.76; nPVI-V$_{Eng\text{-}L2(4)}$ = 63.4).

In general, regarding the data from Arvaniti (2012), these bidimensional analyses grouped BP with languages considered more syllable-timed, that is, with more durational regularity among the segments of reference, such as Spanish and Italian. Eng-L1 results were also consistent with the literature, gathering with the results for English and German from other studies, which are considered languages with more stress-timing tendency.

The hierarchy of values for ΔC-%V, VarcoV-VarcoC and rPVI-C and nPVI-V illustrates the dominant positioning pattern captured by the metrics that reached statistical significance, that is [+stress-timed] Eng-L2 (1) > Eng-L2 (4) > Eng-L1 > BP [+ syllable-timed]), as can be seen in Table 3. The boxplots for the remaining statistically significant metrics are shown in Figure 4 and visually demonstrate this dominant tendency of distribution.

Figure 4: Boxplots of the means of %C (4a), ΔS (4b), rPVI-V (4c), rPVI-VC (4d), rPVI-S (4e), nPVI-C (4f), nPVI-VC (4g), RR-C (4h), RR-VC (4i), VI-V (j), VI-C (4k), VI-VC (4l), VI-S (4m) and YARD-VC (4n) for Eng-L1, English- L2(1), Eng-L2(4) and BP. The blue dots and lines represent the means and standard errors respectively.
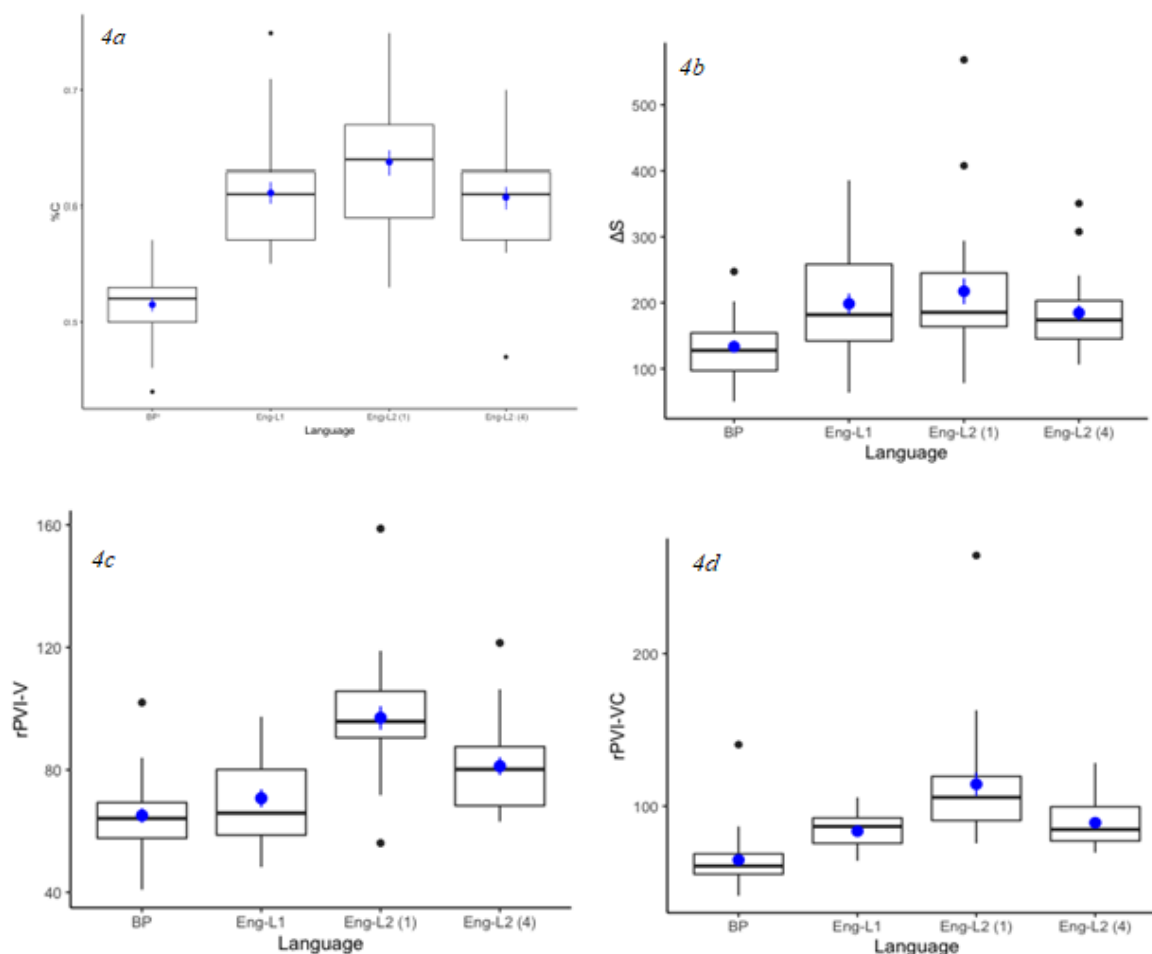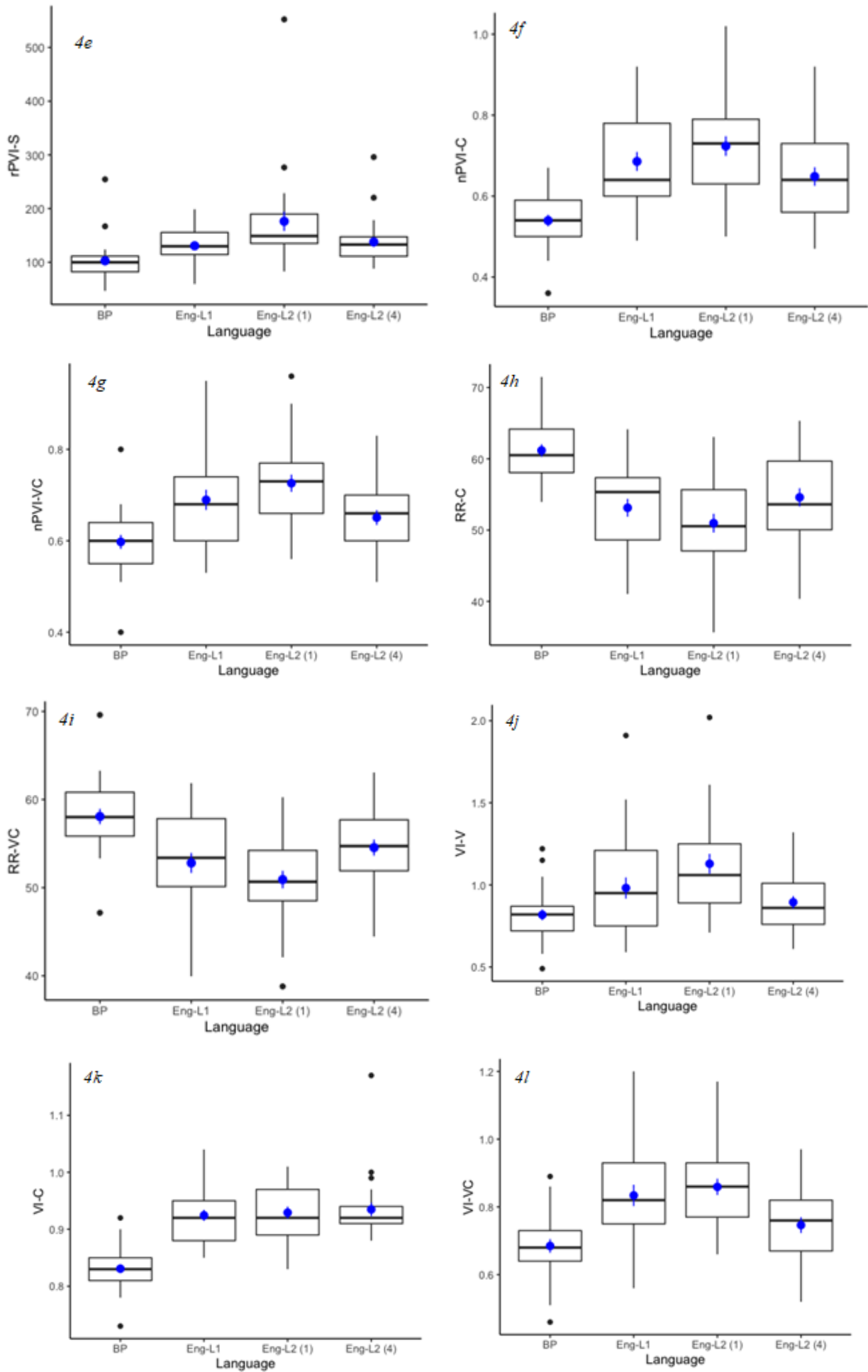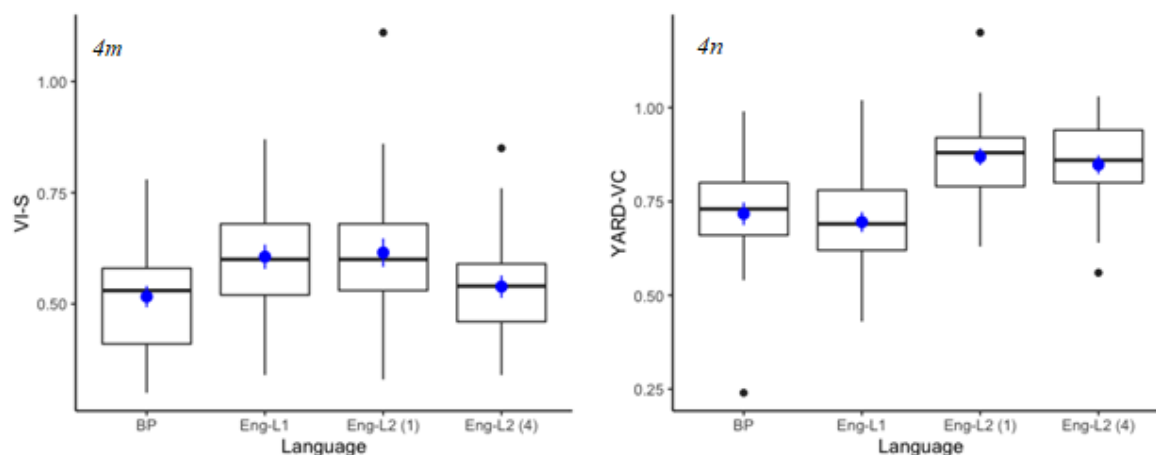
Note: Elaborated by the author.

Figures 4h and 4i show the boxplots with the means and standard errors for rhythm ratio (RR), which is a variation of the paired variability index (PVI), differing only in terms of the normalization technique. Lower RR values are expected for languages with a more stress-timed rhythm. In the comparison between the L1s, the results for RR applied to all segments suggest that BP has a greater tendency to syllabic rhythm (RR-C$_{BP}$ = 61.17; RR-VC$_{BP}$ = 58.07) compared to Eng-L1 (RR-C$_{Eng-L1}$ = 53.13; RR-VC$_{Eng-L1}$ = 52.8). The L2 system data, on the other hand, are less elucidating, with Eng-L2(1) emerging as the corpus with the greatest stress-timing tendency (RR-C$_{Eng-L2(1)}$ = 50.97; RR-VC$_{Eng-L2(1)}$ = 50.91), and Eng-L2(4) occupying an intermediate position between BP and Eng-L1 (RR-C$_{Eng-L2(4)}$ = 54.59; RR-VC$_{Eng-L2(4)}$ = 54.55). Although RR means did not differ radically among the analyzed corpora, the pattern observed for Eng-L2(1) as the system with more extreme values was also replicated for this metric.

The results for the variability index (VI) are shown in Figures 4j, 4k, 4l and 4m. This metric is similar to PVI, but with a different normalization technique: the duration of each reference interval is divided by the mean duration of all intervals of the same type in the utterance. In theory, higher values for VI represent greater variability in the duration of consecutive reference intervals and, therefore, signal a tendency to stress-timing rhythm. The analysis of the results obtained for VI positioned BP as the system with the most syllabic tendency among all the corpora analyzed (VI-V$_{BP}$ = 0.81; VI-C$_{BP}$ = 0.83; VI-VC$_{BP}$ = 0.68; VI-S$_{BP}$ = 0.51). When comparing those means with those of Eng-L1 (VI-V$_{Eng-L1}$ = 0.98; VI-C$_{Eng-L1}$ = 0.92; VI-V$_{Eng-L1}$ = 0.83; VI-S$_{Eng-L1}$ = 0.6), it can be noticed that the VI metrics fulfilled their role in separating the L1s, especially VI-C and VI-VC.

Finally, as for YARD, a local metric that can also be considered a variation of PVI (Figure 4n) shows the only segment with which significant results were found. These results do not follow the dominant pattern that emerged from the distribution of the other durational indexes and say little about L2ers' developmental path. PB and Eng-L1 recorded the lowest means for these metrics (YARD-VC$_{BP}$ = 0.717; YARD-VC$_{Eng-L1}$ = 0.695, with Eng-L1 showing a

slightly more syllabic tendency than PB. Eng-L2 (1) presented the highest values (YARD-VC$_{BP}$ = 0.869), followed by Eng-L2 (4) (YARD-VC$_{Eng-L2(4)}$ = 0.848).

Finally, as for YARD, a local metric that can also be considered a variation of PVI (Figure 4n) shows the only segment with which significant results were found. These results do not follow the dominant pattern that emerged from the distribution of the other durational indexes and say little about L2ers' developmental path. PB and Eng-L1 recorded the lowest means for these metrics (YARD-VC$_{BP}$ = 0.717; YARD-VC$_{Eng-L1}$ = 0.695, with Eng-L1 showing a slightly more syllabic tendency than PB. Eng-L2 (1) presented the highest values (YARD-VC$_{BP}$ = 0.869), followed by Eng-L2 (4) (YARD-VC$_{Eng-L2(4)}$ = 0.848).

3.2 Acoustic Parameters

Five out of the fourteen employed acoustic parameters reached statistical significance for at least two of the (inter) languages: σf_0, $\sigma\Delta 1$-f_0, f_0 peak, spectral emphasis (emph) and speech rate (SR).

As for the standard deviation of f_0 (Figure 5a), Eng-L1 presented the highest standard deviation among the corpora analyzed ($\sigma f_{0Engl-L1}$ = 3.79), followed by Eng-L2(4) ($\sigma f_{0Engl-L2(4)}$ = 3.34), Eng-L2(1) ($\sigma f_{0Engl-L2(4)}$ = 2.71) and BP (σf_{0BP} = 2.62). The results for this parameter suggest a gradual prosodic development of L2ers towards the f_0 variation patterns of the target language. The standard deviation of f_0 first derivative ($\sigma\Delta 1$-f_0) (Figure 5b) was also successful in the separation of the L1s and captured a similar developmental path to σf_0. The highest mean was scored by Eng-L1($\sigma\Delta 1$-$f_{0Eng-L1}$ = 5.51), the lowest mean was scored by BP ($\sigma\Delta 1$-f_{0BP} = 3.61). The L2 systems registered intermediate values, but the mean Eng-L2(4) was much closer to Eng-L1 ($\sigma\Delta 1$-$f_{0Eng-L2(1)}$ = 3.73 < $\sigma\Delta 1$-$f_{0Eng-L2(4)}$ = 4.61).

Figure 5: Boxplots of the means σf0 (Figure 5a) and Δ1- f0 (Figure 5b) for Eng-L1, English- L2(1), Eng-L2(4) and BP. The blue dots and lines represent the means and standard errors respectively.
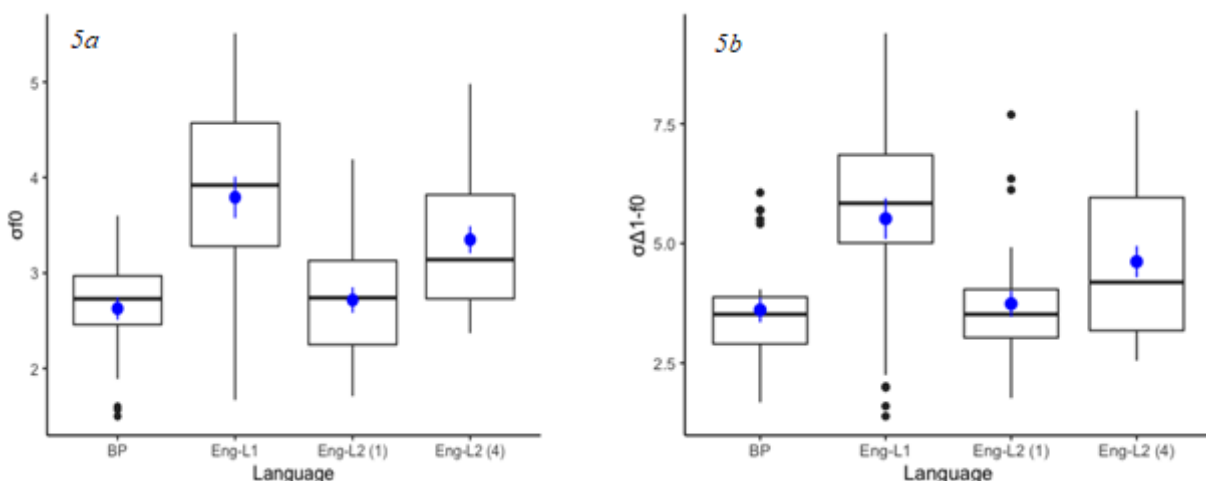
Note: Elaborated by the author.

The results for the f_0 dimension must be interpreted with caution, since there was an unbalance between male and female participants in both groups (control group: 1 male, 4 female; experimental group: 4 males, 1 female). As there is evidence that the scope of f_0 variation is wider for women than for men (Cumbers, 2013), presenting the results only by language system would not be realistic. Thus, we decided to calculate the scope of f_0 variation individually to verify the individual behavior of the participants and the possible influence of the variable sex on the results. To this end, we subtracted f_0 minimum from f_0 peak, as can be seen in Table 4. Participants in the experimental group were identified by a capital letter, and a combination of three letters was assigned to those who make up the control group. Participant N is the only female in the experimental group and participant Ros is the only male in the control group.

Table 5: Means of f_0 peak, f_0 min, and scope of f_0 in semitons for BP, Eng-L1, Eng-L2 (1) and Eng-L2 (4) per speaker.

Language	*participant*	f_0*peak*	f_0*min*	f_0*range*
	A	92.49	79.33	13.16
	F	90.63	75.52	15.11
PB	K	92.67	78.64	14.02
	L	86.22	77.69	8.52
	N	97.16	79.97	17.18
	Car	98.12	78.56	19.55
	Rho	98	77.19	20.8
Eng-L1	Ros	87.34	79.06	8.27
	Roz	98.54	78.77	19.76
	Van	98.33	80.82	17.5
	A	91.12	78.84	12.28
	F	92.88	77.19	15.68
Eng-L2 (1)	K	94.8	79.29	15.5
	L	91.72	78.34	13.37
	N	98.06	82.47	15.59
	A	91.76	78.95	12.81
	F	90.95	75.86	15.09
Eng-L2 (4)	K	92.39	77.96	14.43

|			
L	92.1	77	15.1
N	98.3	79.24	19.06

Note: Elaborated by the author.

In fact, the correlation between f_0 and sex is evident when individual results are taken into consideration. For instance, in the experimental group, it was observed that participant N, the only female, is the one that recorded the highest f_0 peaks (97.16), as well as the widest scopes of f_0 for BP (17.18) and Eng-L2(4) (19.06). There was also a smaller variation between the male learners f_0 scope of Eng-L2(1) (A = 12.28; F = 15.68; K = 15.5; L 13.37) and Eng-L2(4) (A = 12.81; F = 15.09; K = 14.43; L = 15.1), in comparison to the variation of the female participant, who went from 15.59 to 19.06 in the last recording.

In the dimension of intensity, as visually demonstrated in Figure 6a, spectral emphasis separated the L1s, with the Eng-L1 recording the highest mean among the analyzed corpora ($\text{emph}_{\text{Eng-L1}}$ = 4.34), higher than BP (emph_{BP} = 2.73). If we consider works that show the correlation between spectral emphasis and phrasal stress (Heldner, 2001), this result suggests that native English speakers make more effort as an acoustic cue in stress marking than Portuguese speakers. Regarding the L2 systems, Eng-L2 (1) obtained the lowest mean of spectral emphasis, very close to BP values, ($\text{emph}_{\text{Eng-L2(1)}}$ = 2.56), and Eng-L2 (4) was much closer to Eng-L1 ($\text{emph}_{\text{Engl-L2 (4)}}$ = 3.23). This indicates L1 influence at the intensity dimension, and a tendency towards the prosodic patterns of Eng-L1 in the last recording.

Figure 6: Boxplots of the means spectral emphasis (Figure 6a) and speech rate (Figure 6b) for Eng-L1, English- L2(1), Eng-L2(4) and BP. The blue dots and lines represent the means and standard errors respectively.
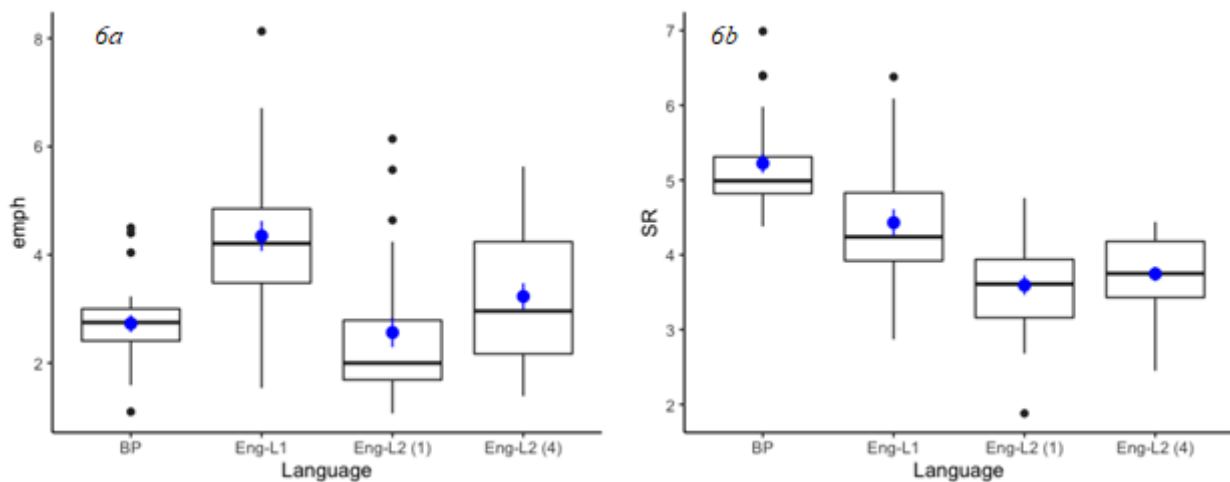
Note: Elaborated by the author.

As expected, the L1s presented higher speech rates (Figure 6b), with the BP registering a higher mean compared to Eng-L1 (SR_{BP} = 5.22 > $SR_{Engl-L1}$ = 4.43). In addition, Eng-L2 (1) presented the lowest speech rate among the corpora analyzed ($SR_{Engl-L2(1)}$ = 3.59) and Eng-L2(4) registered a slightly higher mean, closer to Eng-L1 ($SR_{Eng-L2(4)}$ = 3.74). The increase in the speech rate of the L2 systems between the first and last recording may be related to the effects of explicit instruction.

3. Discussion

Three research questions guided this work. The first one concerned how the acoustic parameters and metrics would position Eng-L1, Eng-L2(1), Eng-L2(4) and BP-L1 in the rhythmic space.

At the durational dimension, the metrics positioned BP, Eng-L1, Eng-L2 (1) and Eng-L2 (4) as rhythmically different, confirming our hypothesis. This became evident in the two-dimensional planes formed by the classical pairs %V-ΔC, VarcoV-VarcoC, and nPVI-V-rPVI-C. In addition, the boxplots and the distribution of the results for the other metrics that reached statistical significance support the hypothesis that the language systems analyzed are rhythmically different. Moreover, rhythmic differences were also detected in the two developmental stages of Eng-L2, with much closer values between Eng-L2 (4) and Eng-L1, which could be partly attributed to the effect of explicit instruction. Surprisingly, the durational patterns of Eng-L2 (1) were even more dissimilar to BP values: 14 out of the 20 durational metrics that reached statistical significance, grouped the language systems with this pattern of positioning.

While it may be tempting to interpret this hierarchy as a straightforward reflection of the stress-timed versus syllable-timed nature of speech, we must exercise caution. Ranking L2 English speakers in such a hierarchy does not necessarily imply that they perfectly represent these timing patterns. Several factors influence the distribution of vowels and consonants, as well as syllable structures in L2 speech, with vocal load being a prominent example (Albert & Obler, 1978). For instance, when vocal load increases, speech rate tends to decrease, potentially distorting other timing measures (Gut, 2012). Therefore, while our analysis identifies a hierarchy of values for those metrics, it is essential to recognize these indexes are not isolated from other factors, including vocal load, that may have influenced the observed patterns.

Regarding the acoustic parameters, five reached statistical significance, namely, f_0 peak (f_0 peak), f_0 standard deviation (σf_0), standard deviation of f_0 first derivative ($\sigma\Delta1$-f_0), spectral emphasis (emph) and speech rate (SR). In these dimensions, different positioning patterns emerged. The L1s were well separated by the acoustic parameters σf_0, $\sigma\Delta1$-f_0, emph and SR, with BP always recording the lowest values compared to Eng-L1. In relation to the L2s, except for f_0 peak, whose values for both L2 systems remained virtually unchanged, the acoustic parameters captured a developmental path towards Eng-L1, with

Eng-L2 (1) closer to BP (considering σf_0, $\sigma\Delta 1\text{-} f_0$ and spectral emphasis), and Eng-L2 (4) closer to Eng-L1. This corroborates the idea that orthogonal patterns of rhythmic development seem to coexist as function of the different dimensions of prominence, a hypothesis that has already been raised in a preliminary report of this study (Teixeira & Lima Jr., 2021). In other words, although L2ers displayed more extreme values in comparison to both L1 and L2 in the durational dimension, this behavior was not replicated for the other dimensions of prominence. Such behavior aligns with the multidimensional and gradient perspective of rhythm adopted in this work. In this sense, both the rhythmic metrics and the acoustic parameters positioned BP, Eng-L1, Eng-L2(1) and Eng-L2 (4) as rhythmically different systems, but the positioning of the L2 systems varied according to the dimension of prominence.

This brings us to the second guiding question of this research, which concerns the influence of BP on the rhythm development of L2ers. On the durational dimension, an unexpected path emerged, Eng-L2 (1) presenting more extreme values towards stress-timed rhythm, unlike studies that argue in favor of the existence of a universal path of rhythm development, which should initially display syllable-timed values, regardless of the L1 (Bunta & Ingram, 2007; Kehoe, Lléo & Rakow, 2011; Ordin & Polyanskaya, 2015). Thus, at least in the durational dimension, BP-L1 patterns did not seem to have influenced Eng-L2 (1). This developmental path is consistent with the definition of L2 systems as relatively independent from both L1 and L2 (Li & Post, 2014). This result supports a complex dynamic perspective on rhythm development, given its nonlinear character (De Bot, 2008; Larsen-Freeman & Cameron, 2008).

However, that result does not mean that L2 durational patterns have not resembled those of L1 at some point. It is worth mentioning that this study adopted semester as one of its predicting variables, which does not necessarily indicate an elementary level of proficiency. In addition, the fact that they were pursuing a degree in English Teaching at the time data was collected could have fostered an effort to phonetically realize Eng-L2 in a markedly different way from BP. Learners may have mobilized a process of dissimilation, imprinting exaggerated durational values to maintain distinction between L1 and L2, in a similar way to what is predicted by the Speech Learning Model for the segmental level (Flege, 1995; Flege & Bohn, 2021). As L2ers advanced in L2 development, and after having taken two undergraduate courses in Phonology, the temporal organization of the L2 system may have been rearranged towards L2 patterns. The influence of the prosodic patterns of BP was more evident in the dimension of f_0, given that almost all acoustic parameters that reached statistical significance at this level revealed a greater proximity between BP and Eng-L2(1), which registered the lowest means among the analyzed corpora. The analysis of individual means of σf_0 and $\sigma\Delta 1\text{-} f_0$ revealed variability and complexity levels of f_0 contours very similar between BP and English- L2(1).

These results align with studies that suggest the influence of L1 on L2 f_0 patterns. For example, Silva Jr. and Barbosa (2019b) detected less variability in the

melodic trajectory of Brazilian English-L2ers, with lower complexity in f_0 contours in comparison to the speech data of the control group (native speakers of American English). According to the authors, this is due to a greater tendency on the part of non-native speakers to pay more attention to segmental information rather than to prosodic features. Urbani (2012), in her research with Italian English-L2ers, was not able to reach a conclusion on the possible L1 influence of f_0 patterns to L2, due to the difference between men and women in the experimental group: the data from the male subjects showed L1 effect on the range of f_0 contours in L2, while women's data did not. As the author conducted a cross-sectional study, different developmental stages could not be compared as well.

Our data support the hypothesis of L1 influence in the dimension of f_0, considering that the scope of f_0 variation for Eng-L2(1) was very close to BP values. In addition, the scope of f_0 variation for the male subjects discreetly changed between the first and fourth recordings, while the only female participant recorded the greatest increase for this parameter, reaching, in the last recording, an analogous mean to those recorded by the female participants of the control group. Thus, we raise the hypothesis that the influence of L1's f_0 patterns on L2 exists and tends to be more persistent and long-lasting in male subjects. This would explain the gender-based differences found by Urbani (2012).

Before proceeding to the discussion of the third guiding question of this study, some considerations shall be made regarding the possibility of L1 attrition, that is, the L2 system affecting the L1 system in its turn. In other words, it is possible that the data of BP is also being influenced by the L2 under development. However, as BP data were collected just once, and no data from BP monolingual speakers were collected, attrition could not be assessed in this study. Future studies should address this issue, considering that the influence among languages does not occur from one's L1 to L2 only, but rather in a multidirectional manner (Larsen-Freeman & Cameron, 2008; De Bot, 2008; Kupske, 2016).

The third and final guiding question of this study concerned the effects of explicit pronunciation teaching on L2ers' rhythm development. The developmental path captured by both metrics and acoustic parameters in this analysis suggest that the explicit instruction may have positively impacted the prosodic development of the L2ers, given that Eng-L2 (4) data were collected after the speakers had taken two undergraduate courses in English phonology. On the durational dimension, a greater temporal organization was detected in Eng-L2(4), whose values got much closer to Eng-L1. In the dimension of f_0, the acoustic parameters captured more complexity in L2ers' melodic contours in the last recording. Moreover, the individual analysis of f_0 scope showed that three of the five L2ers increased their f_0 interspace in comparison to the first recording, although this effect was more discrete for the male participants.

In the dimension of intensity, the increase in L2ers' spectral emphasis means demonstrates a higher degree of vocal/respiratory effort, which may be related to an improvement in the ability to contrast stressed and non-stressed syllables and

/ or in the marking of sentence stress, since, according to Heldner (2001), spectral emphasis operates in English as an acoustic cue for those suprasegmental features.

Finally, the highest speech rate detected in Eng-L2 (4) data suggests an improvement in L2ers' fluency, which may have been possible because of an attention displacement from segmental aspects (supposedly more automated at this stage of development) to prosodic features. This increase in speech rate may be related to the effects of explicit instruction. Moreover, the fact that speech rate displayed the most significant effects agrees with the results reported by Silva Jr. and Barbosa (2019b), for which speech rate emerged as the most consistent acoustic parameter. According to Pellegrino (2012), at the suprasegmental level, fluency, along with a wide scope of f_0 and a low number of silent pauses contribute to a lower degree of perceived foreign accent. Therefore, it is possible that explicit pronunciation teaching in L2 classes may exert a positive effect in the development of prosodic aspects, similarly to what occurs at the segmental level (Lima Jr., 2010; Lima Jr., & Alves, 2019).

It is also crucial to assess the performance of the rhythmic metrics and acoustic parameters employed in this study. Of the 30 rhythmic metrics employed in this study, 13 revealed effects of all systems: %V, %C, ΔC, Δs, VarcoC, rPVI-C, rPVI-VC, nPVI-C, nPVI-VC, RR-C, RR-VC, VI-C and VI-VC (Cf. Table 1). Based on the assumption that the four corpora analyzed here are rhythmically different, most rhythmic metrics were able to place each of the systems in distinct areas of the rhythmic space. Among those metrics, the dominant pattern of distribution was ([+ stress-timed] Eng-L2 (1) > Eng-L2 (4) > Eng-L1 > BP [+ syllable-timed]), with Eng-L2(1) exhibiting the most extreme means towards stress-timing, and BP, conversely, with the most extreme values towards syllable-timing. Eng-L2 (4) displayed closer means to Eng-L1, but still leaning a bit more towards the stress-timing direction of the continuum.

Despite the greater variability, the consonants turned out to be the reference interval with the highest number of metrics that revealed effects for the language systems. In this sense, this result corroborates Arvaniti (2012), in which more stable results were obtained from the consonant intervals and diverges from White and Mattys (2007) and Wiget et al. (2010), for which interlinguistic differences were more efficiently detected by vowel metrics. We argue according to Arvaniti (2012) that a possible explanation for the efficiency of consonant metrics lies in the fact that syllabic structure variability reflects more directly on duration, which can be relatively efficiently captured by consonant metrics. The acoustic parameters were also useful in distinguishing the language systems in the gradient rhythmic space and proved to be useful in rhythm studies.

Despite the different developmental paths of L2ers in each dimension, the results for the acoustic parameters and rhythm metrics converge in the sense they all lean towards L2 prosodic patterns over the time. However, the influence of the L1 on the L2 varied as function of the dimension. This reinforces the thesis raised in this study that the development of a prosody dimension in L2 is orthogonal with respect to the other dimensions. Thus, our data support a multidimensional

perspective on rhythm, as function of the distribution of elements of prominence in multiple dimensions of the acoustic signal. Future works are necessary to detect correlations among these indexes and refine the analysis.

Finally, some brief considerations can be made regarding the teaching of the phonetic-phonological component with regard to rhythm. Considering the phonological differences between BP and English, teaching rhythm in the classroom could: (i) raise L2ers' awareness about the durational differences between stressed and unstressed syllables in the phonetic realization of English and Portuguese; (ii) emphasize the process of syllable stress marking in English, conducting phonetic training for the perception and production of these features; (iii) draw attention to the inadequacy of inserting empethetic vowels in some phonological contexts, as the beginning of words starting with the letter ' s ' or to the end of words ending in stop consonants or silent 'e'; (iv) emphasize the distinctive function of stress in language, both at semantic and pragmatic levels; (v) promote the perception of pitch in its correlation with the syntactic-discursive structure, and the production of more varied intonational contours. These teaching aims can be achieved within a communicative approach framework, as proposed by Celce-Murcia et al. (2010), which can be developed in five stages: (i) description and analysis; (ii) auditory discrimination; (iii) controlled practice and feedback; (iv) guided practice and feedback; and (v) communicative practice. Crucially, as pointed out by Lima Jr. and Alves (2019), the teaching of the phonetic-phonological component must be implemented in an integrated manner with other linguistic components, such as syntax, morphology, semantics and pragmatics, converging to previously established communicative aims, as it is part of the language in use.

4. Final considerations

All metrics and acoustic parameters that reached statistical significance positioned Eng-L1, BP-L1, Eng-L2(1) and Eng-L2(4) in distinct positions of the rhythmic gradient space, confirming the first hypothesis that these are rhythmically different systems. On the other hand, the course of L2ers' development assumed different trajectories depending on the dimension of prominence considered. In the durational dimension, this route started from an opposite direction to the expected one, since the values of Eng-L2 (1) were in general far from those displayed by BP. This means that in terms of duration, our data did not point to a noticeable influence of prosodic patterns from BP to Eng-L2(1). This atypical development path corroborates the idea of the L2 system (and language) as a dynamic and nonlinear system. Additionally, despite the temporal disorganization of Eng-L2 (1), with high levels of variability and dispersion, Eng-L2(4) values showed a tendency towards the rhythmic patterns of Eng-L1.

Conversely, especially in the f_0 dimension, the values of Eng-L2(1) were much closer to BP. This leads us to the second guiding question of this study

– "what is the influence of PB-L1 rhythm on learners' Eng-L2 development?". Our data did not reveal a direct influence of rhythmic patterns of BP in the durational dimension, as initially hypothesized. However, further investigations are necessary, since the participants of the experimental group did not necessarily present a profile of beginner proficiency in the first semester and may have mobilized a process of phonetic dissimilation with the aim of performing L1 and L2 in a markedly different way.

As for the third and final question, there are indications that the explicit teaching of pronunciation may have had a positive effect on the development of L2ers' rhythm, as both rhythmic metrics and acoustic parameters confirmed the hypothesis that Eng-L2 (1) data was more dissimilar to Eng-L1 in comparison to Eng-L2 (4). The smaller difference between Eng-L1 and Eng-L2 (4), after L2ers' have taken two undergraduate courses in English Phonology, suggests that the development of prosody in L2ers can benefit from the explicit teaching of pronunciation. It is important to highlight that, due to the absence of a control group for this variable (i.e., a group that did not receive explicit pronunciation instruction), explicit pronunciation teaching emerges in this study only as one of the possible factors that may have had an influence on the development path of the learners.

One limitation of this study concerns the number of sentences analyzed, only seven. It is possible that the analysis of a larger number of sentences in the future will result in more stability in the metrics and acoustic parameters values (Arvaniti, 2012). Thus, the future analyses should also include the corpora of recordings 2 and 3, for a more longitudinal approach, in addition to including the other two paragraphs of the text read. In addition, the perceptual validity of rhythm should be explored, employing judges to evaluate the recordings modified through low-pass filters, as a means of neutralizing segmental information. Thus, it will be possible to investigate L2ers' rhythm development in correlation to aspects such as intelligibility, comprehensibility and degree of foreign accent. Finally, L1 attrition must be assessed in future longitudinal studies by collecting L1 data at different stages of L2 development, as well as L1 data from monolinguals, in order to investigate the mutual influence of L1 and L2 systems in rhythm patterns.

The results obtained in this study demonstrate the usefulness of rhythmic metrics and acoustic parameters for the description of rhythm development in L2, and the need to consider multiple dimensions in investigations of this nature. We believe that in the future stages of this research, the expansion of the analyzed material and the inclusion of the perceptual dimension will allow an even more precise characterization of the rhythm development in L2ers. Consequently, it will be possible to support the preliminary findings to which this research has arrived, that is, the coexistence of orthogonal patterns of rhythmic development as function of the different dimensions of prominence considered, and the validity of the metrics and acoustic parameters employed.

References

Abercrombie, D. (1967). *Elements of general phonetics*. Aldine.

Albert, M.L.; Obler, L.K. (1978*). The bilingual brain: Neuropsychological and neurolinguistic aspects of bilingualism.* New York: Academic Press.

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

Barbosa, P. A. (2007). Análise e modelamento dinâmicos da prosódia do português brasileiro. *Revista de Estudos da Linguagem*, *15* (2), 127-137.

Barbosa, P.A. (2012). Conhecendo melhor a prosódia: aspectos teóricos e metodológicos daquilo que molda nossa enunciação. *Revista de Estudos da Linguagem*, *20*, (1), 11-27. http://www.periodicos.letras.ufmg.br/index.php/relin/article/view/2571/2523

Bertoncini, J., Morais, J., Bijeljac-Babic, R., McAdams, S., Peretz, I., & Mehler, J. (1989). Dichotic perception and laterality in neonates. *Brain and Language*, *37* (4), 591-605.

Best, C., & Tyler, M. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In O. Bohn & M. Munro (Eds.), *Language experience in second language speech learning: In honor of James Emil Flege* (pp. 13-34). John Benjamins.

Boersma, P., & Weenink, D. (2021). Praat: doing phonetics by computer. (Version 6.0). http://www.praat.org

Borzone De Manrique, A. M., & Signorini. (1983). A. Segmental duration and rhythm in Spanish. *Journal of Phonetics*, *11* (2), 117–128.

Bunta, F., & Ingram, D. (2007). The acquisition of speech rhythm by bilingual Spanish- and English speaking 4- and 5-year-old children. *Journal of Speech, Language, and Hearing Research*, *50* (4), 999–1014.

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

Cumbers, B. A. (2013). *Perceptual correlates of acoustic measures of vocal variability.* Master's thesis. University of Wisconsin, The United States of America.

Cumming, R. E. (2010). *Speech rhythm: the language-specific integration of pitch and duration.* Doctoral dissertation, Universidade de Cambridge, The United Kingdom.

Cutler, A. (1996). Prosody and the word boundary problem. In J. L. Morgan, & K. Demuth (Eds.), *Signal to syntax: Bootstrapping from speech to grammar in early acquisition* (pp. 87-99). Mahwah, NJ: Erlbaum.c

De Bot, K. (2008). Introduction: Second language development as a dynamic process. *The Modern Language Journal, 92* (2), 166-178.

Dellwo, V. (2006). Rhythm and Speech Rate: A Variation Coefficient for deltaC. In Karnowski, P.; Szigeti, I. (Eds.), *Language and language-processing.* (pp. 231–241) Peter Lang.

Ferreira, G. C., Torres, E. M. O., Garcia, M. V., Vasconcellos, S. J. L., Frizzo, N. S., & Costa, M. J. (2018). The effect of bilingualism on cognitive and auditory abilities in normally hearing adults. *Revista CEFAC.* 20(1), 21-28. https://www.scielo.br/j/rcefac/a/wTNb4GD9D5dSbX44mPtC75c/abstract/?lang=en

Flege, J. (1995). Second Language Speech Learning: Theory, Findings and Problems. In: Strange, W. (Ed.). *Speech perception and linguistic experience: issues in cross-language research.* (p. 233-277) York Press.

Flege, J.; Bohn, O. (2021) The Revised Speech Learning Model (SLM-r). In Wayland, R. (Ed.). *Second Language Speech Learning: Theoretical and Empirical Progress.* (pp. 03-83) Cambridge University Press.

Fuchs, R. (2016). *Speech rhythm in varieties of English.* Springer.

Gut, U. Rhythm in L2 speech. (2012). *Speech and Language Technology, 14* (15), 83-94.

Heldner, M. Spectral emphasis as an additional source of information in accent detection. (2001). *Prosody 2001: ISCA Tutorial and Research Workshop on Prosody in Speech Recognition and Understanding, July.*

Holliman, A.J., & Wood, C. (2010). Does Speech Rhythm Sensitivity Predict Children's Reading Ability 1 Year Later?. *Journal of Educational Psychology*, *102* (2), 356 –366.

Kehoe, M., Lleo, C., & Rakow, M. (2011). Speech Rhythm in the Pronunciation of German and Spanish Monolingual and German-Spanish Bilingual 3-Year-Olds. *Linguistische Berichte.* 323-352.

Kupske, F. (2016). *Imigração, Atrito e Complexidade: a produção das oclusivas surdas iniciais do Inglês e do Português por Sul-Brasileiros residentes em Londres.* Doctoral dissertation, Federal University of Rio Grande do Sul, Brazil.

Ladefoged, P. (1975). *A Course in Phonetics.* Harcourt Brace Jovanovich Inc.

Larsen-Freeman, D. & Cameron, L. (2008). *Complex Systems and Applied Linguistics.* Oxford University Press.

Laver, J. (2005). *Principles of Phonetics.* Cambridge University Press.

Li, A.; Post, B. (2014). L2 acquisition of prosodic properties of speech rhythm: evidence from L1 Mandarin and German Learners of English. *Studies in Second Language Acquisition*, *36*, (2), 223-255. https://doi.org/10.1017/S0272263113000752.

Lima Jr., R. M. (2010). Uma investigação dos efeitos do ensino explícito da pronúncia na aula de inglês como língua estrangeira. *RBLA - Revista Brasileira de Linguística Aplicada*, *10* (3), 747-771.

Lima Jr.., R.M. (2016). A necessidade de dados individuais e longitudinais para análise do desenvolvimento fonológico de L2 como sistema complexo. **ReVEL**, *14* (27), 203-225.

Lima Jr., R. M; Alves, U. K. (2019). A dynamic perspective on L2 pronunciation development: bridging research and communicative teaching practice. *Revista do GEL, 16* (2), 27-56.

Mehler, J., Jusczyk P., Lambertz, G., Halsted, N., Bertoncini, J., & Amiel-Tison, C. (1988). A precursor of language acquisition in young infants. *Cognition 29* (2), 143-178.

Mehler, J., Dupoux, E., Nazzi, T., & Dehaene-Lambertz, G. (1996). Coping with linguistic diversity: The infant's viewpoint. In J. L. Morgan & K. Demuth (Eds.), *Signal to syntax: Bootstrapping from speech to grammar in early acquisition* (pp. 101-116). Mahwah, NJ: Lawrence Erlbaum Associates.

Moreno, M. C. (2000). Sobre la adquisición de la prosodia en lengua extranjera: estado de la cuestión. *Didactica (Lengua y Literatura). 12*, 91–119. https://revistas.ucm.es/index.php/DIDA/article/view/DIDA0000110091A

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.

Ordin, M. Polyanskaya, L. (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* (2), 533–544.

Pellegrino, E. (2012). The perception of foreign accented speech. Segmental and suprasegmental features affecting the degree of foreign accent in L2 Italian. *Proceedings of the VIIth GSCP International Conference: Speech and Corpora*, *1* (1), 261-267.

R Core Team. *A Language and Environment for Statistical Computing.* (2021). R Foundation for Statistical Computing. https://www.R-project.org/

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

Silva Jr., L. (2023). BeatMaker: a computational system for foreign language pronunciation teaching based on speech prosody. *Revista Novas Tecnologias na Educação*, 21(1), 341–352.

Silva Jr., L., & Barbosa, P. (2019a.). *Metrics & Acoustics Extractor version 1.0.* Script for Praat.

Silva Jr., L., & Barbosa, P. (2019b). Speech rhythm of English as L2: as investigation of prosodic variables on the production of Brazilian Portuguese speakers. *Journal of Speech Sciences*, *8* (2), 37–57.

Urbani, M. (2012). Pitch Range in L1/L2 English: An analysis of F0 using LTD and linguistic measures. *Methodological Perspectives on Second Language Prosody*, 79–83.

Teixeira, L.A.S.; Lima Jr., R.M. (2021). An Analysis of the Development of the Rhythm of English-L2 by Brazilian Learners through Three Rhythmic Metrics. *Revista X*, *16* (5), 1258-1292.

Teixeira, L.A.S. (2021). *Análise do Desenvolvimento do Ritmo de Inglês-L2 de Aprendizes Brasileiros.* Master's thesis, Federal University of Ceará, Brazil.

Thomson, R. I.; Derwing, T. M. (2015). The Effectiveness of L2 Pronunciation Instruction: A narrative Review. *Applied Linguistics*, *36* (3), 326–344.

Van Ooijen, B. (1994). The processing of vowels and consonants. Doctoral dissertation, University of Leiden, The Netherlands.

White, L.; Mattys, S. L. (2007). Calibrating rhythm: First language and second language studies. *Journal of Phonetics*, *35* (4), 501–522.

Whitworth, N. (2002) Speech rhythm production in three German-English bilingual families. *Leeds Working Papers in Linguistics and Phonetics*, *9*, (3), 175–205.

Wiget, L., White, L., Schuppler, B., & Grenon, Izabelle. (2010). How stable are acoustic metrics of contrastive speech rhythm? *The Journal of the Acoustical Society of America, 127 (*3), 1559–1569.

Appendix A

First Paragraph of Celce-Murcia et al. (2010) Diagnosis Text (original text in English)

Is English your native language? If not, your foreign accent may show people that you come from another country. Why is it difficult to speak a foreign language without an accent? There are a couple of answers to this question. First, age is an important factor in learning to pronounce. We know that young children can learn a second language with perfect pronunciation. We also know that older learners usually have an accent, though some older individuals also have learned to speak without an accent.

Appendix B

Portuguese version of the First Paragraph of Celce-Murcia et al. (2010) Diagnosis Text

O inglês é a sua língua nativa? Caso não seja, o seu sotaque estrangeiro pode mostrar para as pessoas que você vem de outro país. Por que é difícil falar uma língua estrangeira sem sotaque? Existem algumas respostas para essa pergunta. Primeiro, idade é um fator importante na aprendizagem da pronúncia. Nós sabemos que crianças pequenas conseguem aprender uma segunda língua com pronúncia perfeita. Também sabemos que aprendizes mais velhos normalmente têm sotaque, apesar de alguns aprendizes mais velhos também conseguirem aprender a falar sem sotaque algum.