

SPEECH RHYTHM OF ENGLISH AS L2: AN INVESTIGATION OF PROSODIC VARIABLES ON THE PRODUCTION OF BRAZILIAN PORTUGUESE SPEAKERS

SILVA JR, Leônidas J. ^{1*}
BARBOSA, Plínio A. ²

¹State University of Paraíba – University of Campinas

²University of Campinas

Abstract: *When speaking a foreign language (L2), non-native speakers (NNS) produce different phonetic features perceivable by the native speakers (NS) of that language. Such features are based on the production of phonetic gestures characteristic of their mother tongue (L1), and differ from those of the L2 in terms of the segmental (vowels and consonants) and prosodic (stress, rhythm and intonation) features. Causes such as neuro-plasticity and length of residence (Flege, 1995), for example, have been claimed to interfere in L2 production. This work aims to analyze how L2 speech rhythm of English is produced by Brazilian Portuguese (BP) speakers and how prosodic variables such as, metric and acoustic correlates, interfere in the production. This research is based on Barbosa (2006) for the dynamic determination of speech rhythm in addition to Ramus et al. (1999) and so, on the choice of metrics and segmentation procedures. As for the Methods, we collected phonetic data from twenty BP and four American speakers. Next, the data were segmented and labeled into six different units; vowels, consonants, pauses, (phonetic) syllables, sentences and higher units for the acoustic analysis. From these units, we extracted values from metric and acoustic parameters for the statistical analysis, in which we ran one-way ANOVA statistics to check the variability between both groups. Results pointed out to a significant difference between L1 and L2 rhythm of English produced by the groups. These results have confirmed our hypothesis and sparked some implications for understanding L2 English rhythm produced by Brazilian speakers, such as lower speech rate, syllabic and F0 variability. This study has contributed to fill a gap on studies of L2 rhythm of English spoken by L1 BP speakers as well as to L2 (experimental) prosody in Brazil.*

Keywords: L2 speech rhythm; Prosodic variables; L2 prosody.

*Corresponding author: leonidas.silvajr@gmail.com

1 Introduction

When speaking a foreign language (L2), non-native speakers' (NNS) produce different phonetic features that are perceivable by the native speakers (NS). Such features are based on the production of phonetic gestures characteristic of their mother tongue (L1), and differ from those of the L2 in terms of the segmental (vowels and consonants) and prosodic (stress, rhythm and intonation) features (see Flege, 1995). The interference that the L1 exerts on L2 phonetic production and perception is undoubtedly a factor that determines a great deal of difficulties faced by NNS.

Either segmental or prosodic features' acquisition are debatable points from the literature of Second Language Acquisition (SLA). Roach (1982) states that the success in producing oral communication is based on rhythmic aspects of L2. On the other hand, Barry (2007) argues that in SLA, the concept of speech rhythm should not be taken into account, and the practitioner should concentrate on underlying phonological processes, such as the reduction of vowels in weak forms, contrasts of duration and quality of vowels as well as reduction of consonant clusters. Let us notice, thereupon, that the mentioned aspects by Barry are part of a prosodic set (see Barbosa, 2000) and are present in the temporal and dynamic speech domain.

Since the early 1970s, L1 interference on phonological and phonetic aspects in SLA was pointed out in studies, (Cf. Selinker, 1972; Stampe, 1973; and in Brazil, Mascherpe, 1970). In fact, SLA literature has sustained several probable causes that explain the interference of L1 over L2 target productions, such as neurological maturation of NNS followed by the decrease of neuro-plasticity and neuromotricity during L2 sound production, and awkward phonetic input from an early age in addition to the amount of input, length of residence (LOR), aptitude, proficiency level, motivation, etc. (see Piske, MacKay & Flege, 2001). These causes are also influenced by inaccurate perception of L2 sounds resulting in the production of foreign accent.

Neurological factors, other than motricity and plasticity, may affect SLA. Moyer (2004) points out for asymmetrical relation between perceptual and production abilities for late learners, for example. Flege's (1991, 1995) Speech Learning Model (SLM) asserts foreign learners may rely on preexistent phonetic categories from their L1. These categories prevent the development of phonetically accurate targets for L2 sounds due to the decline in neuroplasticity to perform specific phones out from their L1 inventory. On the other hand, Dupoux (2003) argues that perception is not plastic, but that production might be. That is, adult learners, for example, can sound native-like even without forming accurate perceptual distinctions between L1 and L2 phonemic categories.

The aforementioned diversity of causes gives an outline of the complexity of speech production (and perception), especially in the prosodic domain (see Flege *op. cit.*), that is, they occur at the suprasegmental level and can substantially hinder NNS speech productions' intelligibility. Such prosodic phenomena are laid from the syllable structure to higher levels (lexical/phrasal stress position) and set the intonational and rhythmic aspects in a speech turn, helping listeners to structure the speech signal and process segmental, prosodic, syntactic, and semantic information. Shedding light on L2 speech rhythm, this research aims to analyze how English as L2 is produced by Brazilian Portuguese (BP) speakers from (semi) spontaneous speech and how acoustic correlates such as duration, fundamental frequency (F0) and intensity may influence these productions.

Our main hypothesis is that L2 English speech rhythm is produced by BP speakers with a lower speech rate (due to lack of fluency), lower syllable variability (due to regularity patterns on both syllabic and stress group levels), less variability of F0 (due to attention to segmental aspects rather than prosodic).

This paper is divided into the following sections: *Introduction*; *Theoretical Framework*, in which we review the researches about production of L1 and L2 rhythm and intonation. *Methods*, in which it is presented the subjects' characteristics and how they were chosen to be part of the experimental and control group as well as the research corpus. The acoustic analysis, where it is explained the procedures and techniques for extracting rhythmic parameters. The statistical analysis, where it is demonstrated a stepwise procedure for the application of ANOVA statistics that accounted for the data. *Results*, in which acoustic and statistical analyses for the duration-based and acoustic-based distribution of the data are presented. *Discussion*, in which it is explained the implications of the results to a better understanding of the English L2 rhythm spoken by Brazilians. *Conclusions*, where we have summarized and drawn inferences based on the results of this research and pointed out for future studies; and the References used for this research.

2 Theoretical framework

According to Major (2001), there are three categorical levels that allow us to manage the phonological and phonetic productions of a L2; segments (individual sounds), syllables (onset and coda complexity and architecture) and prosody (rhythm and intonation). The combination of these three mentioned levels results in a fourth level called *global accent*. In fact, the so-called *global accent* concerns to whether or not one can sound native/non-native in a specific L2 target.

We shall agree, therefore, that although Major's combination consists of well-defined categorical elements for determining whether phonetic production comes from a NS or NNS, the phonetic elements are governed by a dynamic system, which yields segmental and prosodic aspects to be transferred from L1 to L2 and these aspects determine the performance of a speaker. Based on that dynamicity, we aim to examine how speech rhythm configuration is set in English as L1 and L2 from the next section on.

2.1 Speech rhythm of L1 and L2

Since ancient times in Greece, man was fascinated by the very idea of rhythm and has sought to define its nature. Adams (1979, p. 9) described rhythm as a rule, instead of the phenomenon itself. Her belief comes, despite its all-pervading presence in the systems' dynamic universe, because the numerous manifestations of the phenomenon, which cause it to be apprehended in different ways, happen according to one's sensory perception and experience.

The literature, over time, has conceptualized speech rhythm as a given movement marked by successions of strong and weak beats. According to Lloyd James, (1940 apud Couper-Kuhlen, 1993), Pike (1945), Abercrombie (1967), Hayes (1984, 1995), Roach (2005, 2009) and so, speech rhythm would be characterized by regular time intervals of linguistic units like syllables and stressed syllables. Studies interchange rhythm concept by means of production, perception or both.

Pike's (1945) traditional account for speech rhythm, proposes a categorical distinction into two rhythmic types to classify the languages of the world: stress-timed rhythm (alternation between stressed and unstressed syllables), such as, English; and syllable-timed rhythm (succession of syllables equally spaced in time), such as Spanish. Barbosa and Bailly (1994) propose that rhythm is the sensation caused by the succession of different degrees of syllabic prominence alternated with non-prominent syllables throughout the utterance.

According to the Barbosa (2000, 2006), in his study from the dynamicity of (BP) speech rhythm, a coupled-oscillator model (COM) in which the strength of the coupling between the syllable oscillator and the stress group oscillator is used to infer if languages tend towards syllable-timing or stress-timing presenting rhythm as a dynamic organism. From his COM, Barbosa proposes BP rhythm as having an alternation of prominent/non-prominent syllables whose contrast is determined by linguistic information (semantics, syntax) and regularity constraints on both the syllabic and stress group levels.

Accounting for SLA, studies in Prosody involving acoustic and experimental phonetics have emerged since the late 1970s with the verification of the realization of English stress by NS and NNS from F0, intensity and duration of connected speech (Adams and Munro, 1978). In L2 speech rhythm, Adams' work was grounded on acoustic correlates in order to verify if NS's and NNS's (English foreign learners) productions in spoken English are held by the serial recurrence of more or less isochronous intervals marked off by stressed syllables. The author also investigated the periodic movement associated with the rhythmic pulse in interface with the respiratory muscles (Adams, 1979).

Yet with respect to L2 speech rhythm, phonetic literature has laid on different mathematical parameters for its characterization since the middle of the 1990s. Deterding (1994, 2001), Nootboom (1997), Ramus, *et al.* (1999), Grabe & Low (2002), Barbosa (2006, 2012), Ordin and Polyanskaya (2015) and other researchers proposed and have made use of the so-called *rhythm metrics* (in a vast majority, based on the duration of segments) to measure L2 speech rhythm. Until the late 2000s, studies have used bi-parametric models in order to run bi-dimensional statistical analyses.

Loukina *et al.* (2009) presented a multidimensional analysis in a large corpus for five different languages (English, French, Greek, Russian, Mandarin) and proposed the application of seven metrics instead of the traditional two-metric studies (in fact, there was a total of fifteen metrics if it is considered vowel and consonant regions). The authors applied these metrics into five two-dimensional-based *classifiers* that showed the most efficient combinations (Cf. pp. 1532-1533). The most effective metric combinations were (%V-nPVI-V); (V/Cdur-nPVI-V); (rPVI-C-nPVI-V) and the best one (%V- Δ C) *classifier* which proved to be the most robust intersection between the five languages. Their findings showed that *within* language variation is large and one can compare to the *between* language variation. The authors concluded that speech rhythm appears to be best described under a bi-dimensional perspective and different published rhythm metrics capture different aspects of the rhythm. They also evaluated the success of different measures in separating languages and showed that the efficiency of these parameters depends on each of the language included in the corpus. When testing the effect of speech rate on the other metrics, the authors assert that it failed to account for perceived differences between languages and results were therefore non-significant since the most efficient in combination with speech rate were also the ones, which were most efficient on their own.

As far as segmentation concerns, it was taken into account automatic speech segmentation based on acoustic (loudness and irregularity) parameters claiming for identical segmentation criteria to all languages involved and to calculate which metrics reveal more details about each of the languages. Their automatic segmentation algorithm yields three types of segments: *a*) silences/pauses, *b*) vowel-like segments (with a nearly periodic waveform), and, *c*) segments where the waveform is not periodic (frication and/or regions with rapid changes in the waveform) controlled by smoothed and normalized loudness and irregularity. Loukina *et al.* (op. cit) claim this is broadly consistent with most metrics on the basis of vocalic and intervocalic intervals.

One of the research's limits lays on the extraction of uniquely vowel(-like) and consonant(-like) intervals with the idea that intensity parameters by themselves will account for the whole representation of vowels and consonants (their metrics tested for both of these segments). Another problem is that the smoothing process tends to suppress very short segments (unstressed vowels, semivowels, liquid consonants) that account for the phonetic utterance and it is clearly perceivable. As well as the mentioned segments, the authors do not hold a segmentation of higher acoustic units such as; syllables and/or sentences or even use melodic parameters based on the F0 for the measurements.

More recently, Ding and Xu (2016) conducted a contrastive investigation-oriented research comparing target language (English/L1), source language (Mandarin/L1) and L2 interlanguage with ten Chinese students and ten native English speakers. The authors measured the rhythmic correlate in passage readings of Mandarin and L2 English produced by the native Chinese subjects, and those of English by the native British speakers. Comparison of the widely used rhythm metrics %V, ΔC , ΔV , nPVI, rPVI, VarcoV and VarcoC pointed out that Mandarin Chinese is a highly syllable-timed language. Results suggested that vowel-related metrics were better indexes to classify L2 English rhythm produced by Chinese speakers as being more syllable-timed than stress-timed (p. 2696).

White and Mok (2019) conducted a longitudinal study for investigating L2 speech rhythm of English in Cantonese/L1 immigrants. Seven Hong Kong students were recorded five times throughout a two-year period while they were living abroad in English-speaking countries. The research also accounted for the amount of L1/L2 speech spoken during their time abroad. Speech rhythm of the read utterances was then, measured using some of the durational variability metrics (speech rate, vorco-C, and PVI-V). Results suggest that L2 English speech rhythm may change into more stress-timing condition after immigration to an English-speaking country, but only under certain sociophonetic conditions such as; accommodation of L2 and dialectal configuration of the English-speaking population and demography of the city to which one immigrates.

With reference to the reliability of these metrics, Gut (2012) mentions that each metric generates different results in different studies. In the classical two-dimensional studies, one metric prevails over the other (see Ramus *et al.*, 1999; Wagner and Dellwo, 2003; Dellwo, 2006). Loukina *et al.* (2009) concluded that the use of metrics vary independently within the same or between different languages. This inconsistency observed for the metrics may be largely due to the different segmentation procedures used in the studies, which are very debatable over the literature as well as the influence of speech style and material selection.

One way to minimize the observed inconsistency from the resulted metric values would be to assemble and test the effect for all of those rhythmic metric measures proposed so far into a complex combinatory matrix of metrics. Silva Jr. and Barbosa (2019) coded these classical metrics of L2 speech rhythm (since Deterding, 1994) in addition to applying these duration-based parameters into other segments (phonetic syllables and pauses). Out of the metrics, the authors proposed acoustic parameters, such as speech rate and melodic parameters based on F0 centrality, dispersion and derivative as well as the intensive parameter of spectral emphasis (that represents the vocal effort produced by the speaker) into a script for Praat. The authors' algorithm automatically extracts each parameter (see Table 1 for the description of the parameters captured from the script) from annotated tiers manually segmented and labeled described in the Methods section.

Based on a dynamic, rather than a categorical analysis for modeling speech rhythm, F0 has played a significantly important role not only by means of differences between intonation from native or foreign language but, on the determination of speech recognition from NS of the

target L2. According to Moreno (2000), intonation is considered one of the most difficult prosodic aspects to be assimilated. Magen's (1998) study on perception by English natives reveals that F0 was considered the most relevant acoustic parameter on the foreign accent discrimination from different subjects.

When taken into account in pronunciation classes, intonation does not come into a spotlight and it is considered to be irrelevant at best. Commonly, what we observe in these classes is that when students consciously or unconsciously, hear recordings in L2, they concentrate on other aspects (semantic, grammatical, segmental, etc.) and mostly do not perceive intonation patterns. This remains the learners increasingly distant from the phonetic production (and perception) of intonational gestures such as; pitch accent position in phrasal stress, pitch accent for determining broad or narrow focus in intonational phrase and pitch accent to signal *yes-no* questions and/or *wh*-questions.

In studies across L2 English rhythm, Urbani (2012) found that Italian NNS of English speak with a narrower F0 range and less variation than British English NS speakers. Mennen *et al.* (2008, 2012), by comparing German NNS of English with British NS, proposed a new methodological approach to quantify the gap between the two languages. The study showed that British speakers have a more extended and varied F0 range than German NNS's. Mennen *et al.*'s study also found that German listeners tend to interpret under the perceptual domain, high frequency English voices (especially for females) as "overly excited" or even "aggressive" because of the amount of variation.

As well as Urbani's and Mennen *et al.*'s studies based on the F0 range and the L2 speech intonation studies above cited, let us compare, from our corpus, a BP NNS of English production to an AmE speaker's as showed in Figure. 1. We can see a much higher melodic variation (rising-falling trajectory contour) for the NS speech from the beginning to the end of the utterance and a shift of the movement when it reaches phrasal stress position between the speakers' productions (high-low (HL) for the NS and low-high (LH) tone for the BP speaker):

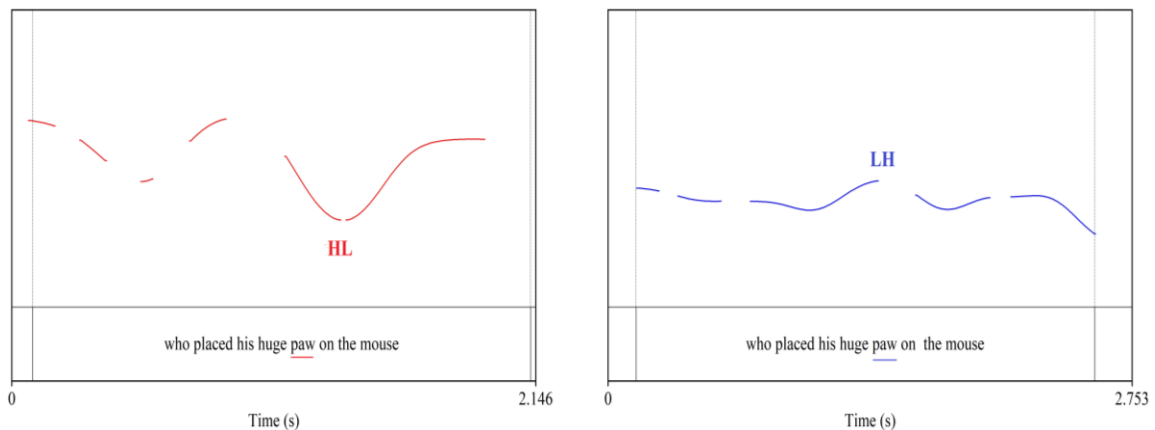


Figure 1: F0 contour and phrase stress over time for the utterance: [*who placed his huge PAW on the mouse*]U spoken by a NS of AmE (red contour on the left portion) and by a BP speaker (blue contour on the right portion).

We have seen so far how prosodic aspects such as rhythm (and intonation) are relevant for the production of L2. Besides the rhythmic metrics, an investigation over the acoustic parameters may also reveal promising results in the verification of speech rhythm in terms of how F0 trajectories can discriminate production. Once such parameters come in hand, we might check the methodological choices and protocols herein adopted.

3 Methods

In this section, we present: i) the subjects' selection; ii) the research corpus and how data were collected and their recording procedures; iii) the acoustic analysis and the criteria for the units' segmentation and labeling as well as each of the metric and acoustic parameter presented in Table 1; and, iv) the statistical analysis.

3.1 Subjects

For the subjects' selection, we collected data from twenty BP speakers of English as L2 (experimental group) and four American native speakers of English (AmE - control group). The experimental group was divided into two categorical proficiency levels of English (ten subjects per level): *high intermediate* and *advanced*, as determined through the Oxford Online Placement Test (OOPT) (<<https://www.oxfordenglishtesting.com/>>).

The experimental group is composed by Brazilian undergraduate students of *Letras-ingles* (Modern Languages with qualification in English as a Foreign Language - EFL) randomly chosen from the first to the fourth year of graduation. Students were communicated about the research by e-mail and volunteered to participate. The grand total of fifty-two students applied for the OOPT and we have selected the ones that scored B1-B2 (Threshold-Vantage; *high intermediate*) and B2-C1 (Vantage-Effective Operational Proficiency; *advanced*) levels of EFL (Cf. <<https://support.italki.com/hc/en-us/articles/206352458-Oxford-Online-Placement-Test-OOPT>> for level rating details and criteria).

The control group is composed by American monolingual speakers of English from the United States that were visiting Brazil by the time we have collected the data. The participants are graduate and have distinct professions: a missionary, a farmer, a businessperson and a dentist. Once they were communicated about the research, they promptly volunteered to participate.

3.2 Corpus

Speakers were told that their task would be to read aloud a text to be recorded. We chose for text a simplified version of one of the Aesop's fable, "The Lion and the Mouse". They were showed the text in advance to be familiarized with and to avoid anxiety while reading. The text was segmented into thirty-three syntactic sentences and four paragraph-delimited chunks that contained at least one pause in the chunk-to-chunk transition besides vowel, consonant, phonetic syllable and pause units (Cf. section 3.3 and Figure 2 for details and purposes of segmentation and labeling into distinct intervals). A grand total of 96 chunks, 660 sentences and more 17000 vocalic, consonantal, syllabic and pause units were computed for the research herein.

When participants were ready to read the text, they were taken to a soundproof room. Speakers were recorded from a Boss BR 1600CD Recorder and a unidirectional electromagnetic-isolated Shure SM7B microphone, at sampling rate of 48 kHz and 16-bit quantization to ensure high quality and noise interference reduction of the recordings that would be later used for acoustic analysis.

3.3 Acoustic analysis

To perform the acoustic analysis, data were first manually segmented and labeled in Praat software (Boersma and Weenink, 2019) into six distinct units located in four text tiers (see

Figure 2 for a detailed tier determination and unit segmentation). *Tier 1*, phonetic syllable (VV)¹ units; *Tier 2*, vowel (V), consonant (C) and pause (#) units; *Tier 3*, syntactic sentence units (S) and *Tier 4*, syntactic-prosodic larger units (chunks - CH).

For the segmentation and annotation protocols, we have adopted Barbosa (2006) for the phonetic syllable units, where the author accounts for the successions of CV transition (Cf. footnote 1); Ramus *et al.* (1999), where there was a phonetic segmentation of the sentences into phonemic-sized units, and labeled these units into vowels, consonants and pauses (V, C and # respectively. Carroll (1994) for sentences and chunks, where he established syntactic ordering for sentence (breath groups) segmentation and prosodic ordering for higher units than a sentence (more than one breath group) containing at least a pause (#) before the next unit (CH1#, CH2# ... CHn#).

The units of each tier have specific functions for the extraction of the metric-based parameters (that calculate duration of the segments such as V, C and VV) and for the acoustic-based parameters (that calculate melodic and intensive values of the segments such as; #, VV, S and CH) from a sound (.WAV) file and a text (.TextGrid) file pair.

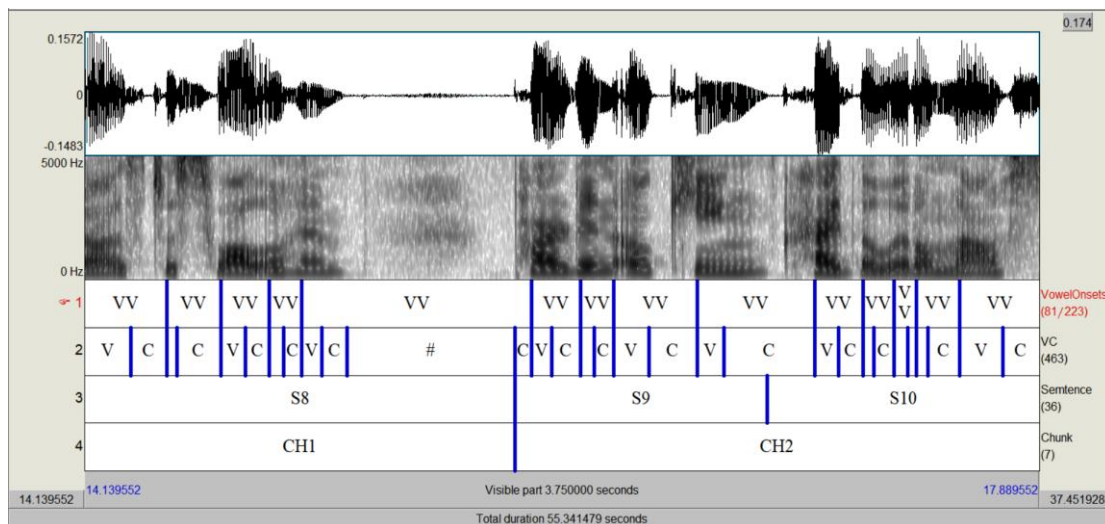


Figure 2: Partial waveform, broadband spectrogram and four tiers respectively segmented and labeled as: 1) phonetic syllable units (VV); 2) vowel and/or consonant phonemic-sized units (V/C); 3) syntactic-based sentence units (S); 4) utterance (*chunk*-based) units (CH) produced by a NS speaker.

¹ According to Stetson ([1928], 1951, p. 188), **phonetic syllable units** comprise intervals in between two vowel onsets (*onsetV-onsetV*) which represent the variation of intensity presented between the groups of sounds which one calls syllables. The author points out that a phonetic architecture of a syllable comes from the perceptual domain and denotes as being a group of sounds separated from the others by a sudden diminution of intensity - a diminution caused either by a decrease of the breath pressure (in the case of a one-vowel syllable), or by the presence of a sound of less sonority between two of greater sonority (in the case of a consonant, pause or other noise between two vowels) (Cf. Stetson, *op. cit.*, p. 190, Fig. 1 for details). The author still severs that if a sonority syllable (the one consisted by discontinuity or diminution of intensity) is assumed it must be admitted that stressed or unstressed, it is a rhythmic element, either an accented or unaccented beat (p. 207).

In addition to Stetson, Morton, Marcus & Frankish (1976) point out that words with regular acoustic onsets are not perceptually regular and establish the *perceptual centers* (P-centers), that is, how hearers of a certain language perceive regularity. They state that "The P-center of a word corresponds to its psychological moment of occurrence" (p. 405). In their experiments for the definition of P-center location, the authors found great evidence that the perceptual regularity occurs in vowel onset-to-onset position than of the ones in which the word onsets are aligned. The asynchrony of the latter would add a large extra variance when comparing to consonant-vowel syllables (p. 406-407).

For the paper herein, we have used Barbosa's (2006) protocol for the phonetic syllable (V-V) determination accounting for the successions of CV transition (Cf. p. 62-98 for detailed discussion).

For the extraction of the metric, melodic and intensive values, we ran a script for Praat (*Metrics & Acoustics Extractor* - Silva Jr. and Barbosa, 2019). As mentioned in previous section, the script extracts *metric* parameters such as percentuals (%) for V/C, standard deviation (σ) for V/C, variation coefficient (*varco*) for V/C, simple, raw and normalized variability indexes, and so; *melodic* parameters based on the F0 centrality/dispersion/derivative such as, regular and derivative peak, minimum, median, standard deviation, and so. It also extracts *intensive*-based parameters such as, spectral emphasis (Cf. Table 1 in this section). The script runs in Praat (Boersma and Weenink, *op. cit.*) and generates an output (.txt) file containing all of the parameters.

It is worth highlighting that for the sake of this research, we refer to the classical nomenclature ‘**metric(s)**’ as the duration-based parameters, and ‘**acoustic parameter(s)**’ as the F0, speech rate and intensive-related ones, which are more directly related with perception.

3.4 Statistical analysis

Once metric and acoustic measures were returned from the script, we ran the statistical analyses into three distinct parts.

- Firstly, we established the 42 prosodic variables that would assemble the statistical models and assured that the dataset could be analyzed performing the one-way ANOVA procedure by submitting the data under the one-way ANOVA assumptions. Our purpose then was to reduce the models discarding the ones that failed to meet at least one of the three statistical required assumptions (normality of the residuals, homoscedasticity of variances, and independence of the samples). At this point, we did not intend to perform any non-parametric equivalent statistics because the original values of the prosodic variables needed to be maintained in a first investigation of the differences between L1 and L2. A significance level (alpha) of 5% was established to test whether the metric and/or acoustic parameters varied significantly or not;
- Secondly, we trained the one-way ANOVA models that did not violate the assumptions in order to assess the effect for each of the language groups (AmE/BP) on each of the metric/acoustic parameters (see Table 1);
- Thirdly, we separated the significant from the non-significant models and took the significant ones into analysis for previous results and designed a scenario taking into account, which metric/acoustic parameter could be more reliable for distinguishing the both L1 and L2 English rhythm, and in which terms this fact might happen.

All steps of the statistical analyses were performed in R language (R CORE TEAM, 2019).

Table 1: Metrics and acoustic parameters extracted from the Silva Jr. and Barbosa’s (2019) algorithm.

METRICS		ACOUSTIC PARAMETERS	
Parameter	Segment of application	Parameter	Segment of application
Percentual (%)	V, C	F0 median	S, CH
Standard deviation (σ)	V, C, (V or C), VV	F0 peak	S, CH
Variation coefficient (Varco)	V, C, (V or C), VV	F0 minimum	S, CH
Raw pairwise variability index (r-PVI)	V, C, (V or C), VV	F0 standard deviation ($\sigma F0$)	S, CH
Normalized pairwise variability index (n-PVI)	V, C, (V or C), VV	F0 skewness	S, CH
Rhythm ratio (RR)	V, C, (V or C), VV	Mean of F0 first derivative ($\mu\Delta 1 - F0$)	S, CH
Variability index (VI)	V, C, (V or C), VV	Standard deviation of F0 first derivative ($\sigma\Delta 1 - F0$)	S, CH
Yet another rhythm determination (<i>z-score duration</i>) (YARD)	V, C, (V or C), VV	Skewness of F0 first derivative ($sk\Delta 1 - F0$)	S, CH
		Speech rate (SR)	VV, S, CH
		F0 rate (F0-R)	S, CH
		Spectral emphasis	S, CH
		Mean of normalized syllable-peak duration ($\mu dur - Sil$)	VV, S, CH
		Mean duration of pauses ($\mu dur - \#$)	S, CH

4 Results

In this section, we can check some of the significant results (*F-ratio* and *p-values*) withdrawn from the ANOVA test statistics and boxplots (Figures 3, 4, 5 for the metrics’ distribution and Figures 7 and 8 for the acoustic parameters’ distribution). Besides the boxplots, we can see plots from the effect of the language chunks (the independent variables) on the significant metrics and acoustic measures (the dependent variables) in Figure 6 (for the metrics) and Figure 9 (for the acoustic parameters).

4.1 Metrics distribution

For the metrics, 6 out of 30 parameters extracted from the data proved to be significantly reliable accounting variability between the factor’s groups. From these six parameters, five were extracted from the phonetic syllable (V-V) units: In Figure 3.1, Standard deviation (σ) and 3.2, Variation coefficient (Varco). Raw and variability index (r-PVI, VI respectively in Figures 4.1 and 4.2). Z-scored duration (YARD) in Figure 5.1, and the one extracted from consonants, the Rhythm ratio (RR) in Figure 5.2).

Speech rhythm of English as L2: an investigation of prosodic variables on the production of Brazilian Portuguese speakers

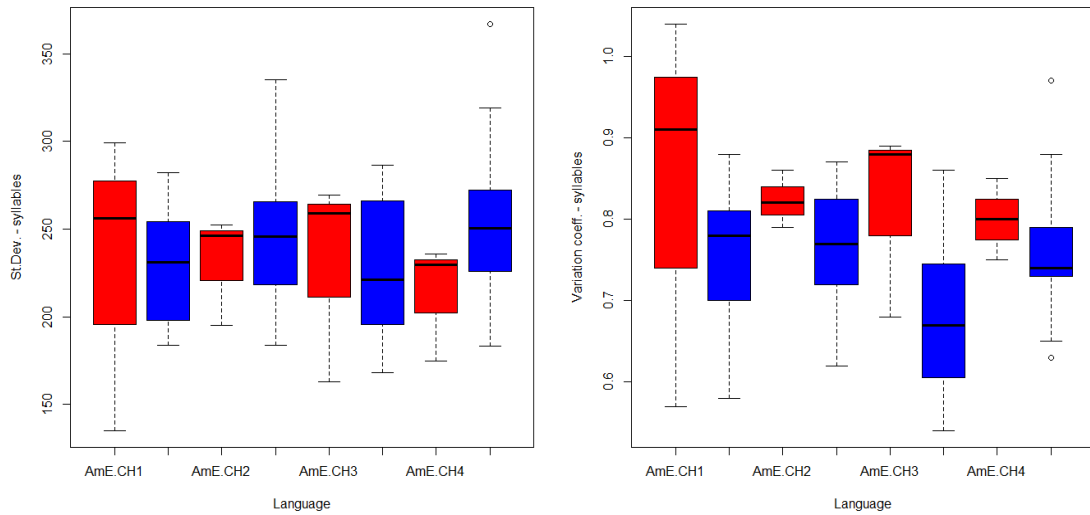


Figure 3: Boxplots of the syllabic standard deviation - σS (left portion, 3.1) and variation coefficient - VarcoS (right portion, 3.2) for the NS (red boxes) and NNS (blue boxes) speakers' productions.

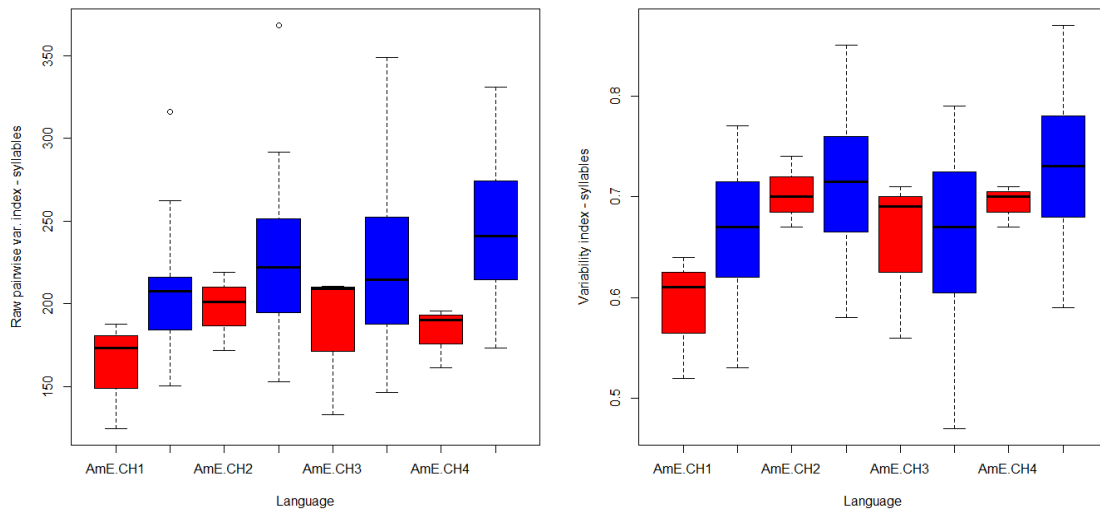


Figure 4: Boxplots of the syllabic raw variability index - r-PVIS (left portion, 4.1) and variability index - VIS (right portion, 4.2) for the NS (red boxes) and NNS (blue boxes) speakers' productions.

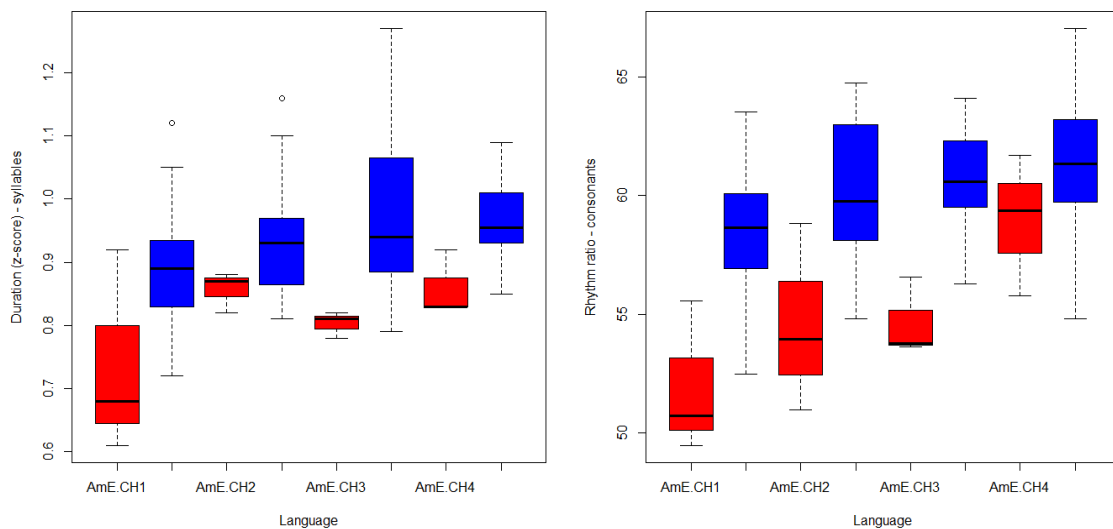


Figure 5: Boxplots of the syllabic z-scored duration - YARDS (left portion, 5.1) and for the consonantal rhythm ratio - RRC (right portion, 5.2) for the NS (red boxes) and NNS (blue boxes) speakers' productions.

From the metrics presented in Figures 3-5, it can be seen that both groups seem to be well discriminated. With the exception to the VarcoS (Figure 3.2), the other metrics show higher values towards NNS speakers either for local or for global parameters. It is worth noting that, considering we are laying discussion on durational values, it would be wise and reasonable to assign to BP speakers higher values since “timing” by means of duration plays the most important role on the metric values hither used.

Having an overview of the data in Figures 3, 4 and 5, let us see in Figure 6, the effect of the factor GROUP on these metrics.

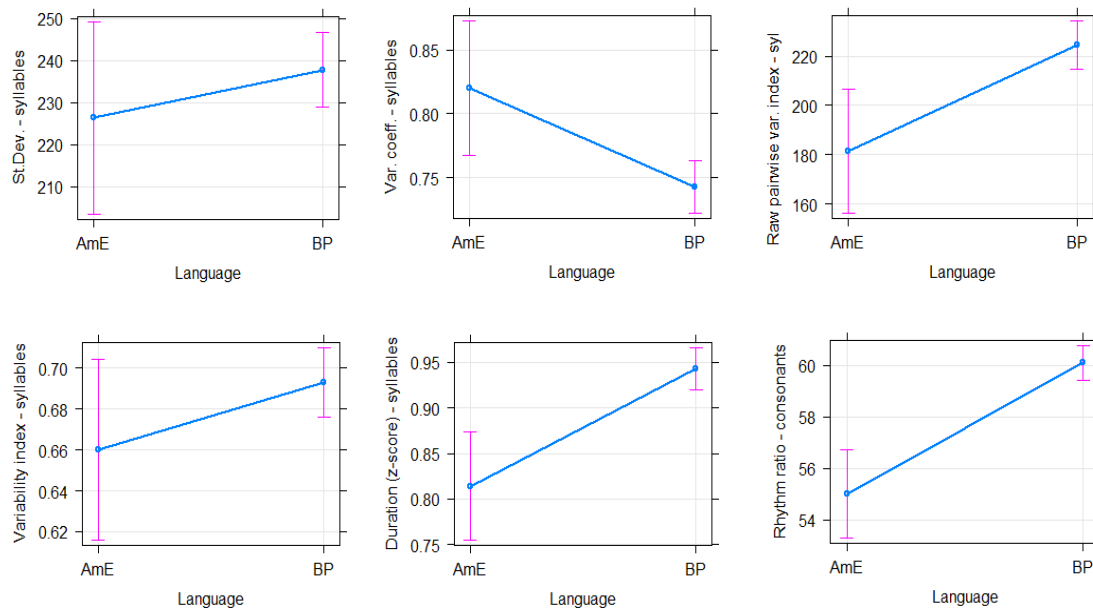


Figure 6: Linear effect plots of the AmE (L1) and BP (L2) speakers’ productions of English on the σS (6.1), VarcoS (6.2), r-PVIS (6.3), VIS (6.4), YARDS (6.5), RRC (6.6).

The reason for the effects presented from 6.1 to 6.6 are predictable to some extent. Most of these metrics are straightly related to variability and they are locally and/or globally calculated. As aforementioned, except the VarcoS (Figure 6.2), all of the other metric parameters have showed higher values for BP speakers’ productions of English. Mathematically, the Varco is a ratio between the standard deviation and the duration mean of the applied segment (V-V units, since the metric is the VarcoS). If on the one hand, standard deviation presented higher values for NNS, although the group has significantly lower variability (see whiskers for the standard deviation in Figure 6.1), on the other hand, the duration mean values for the syllables are yet very high even after z-scoring normalization as one can see in Figure 6.5. The duration mean is what keeps the VarcoS lowering for BP speakers of English.

Yet accounting for the V-V units, NS productions’ duration vary significantly more than NNS’s especially when *mean* is spotlighted. Returning to syllabic standard deviation, let us see how the values are described. In Figure 6.1 for σS ($F(1,94) = 3.60, p < 0.05$), we observe higher values for the NNS groups ($NS = 228; NNS = 238$). Even showing lower σS global value, NS group sets a significantly higher variation span. This fact happens because; once each individual value differs consistently, mean value tends to be higher. One of the reasons for higher absolute

value of σS for the NNS could be attributed to speech rate influence, higher for the NS group, and consequently, lower syllable duration. As previously mentioned, σS operates as one of the reasons for the VarcoS ($F(1,94) = 7.35, p < 0.005$), in Figure 6.2, to be inversely related to the other metrics (NS = 0.82; NNS = 0.74).

In Figure 6.3, the r-PVIS ($F(1,94) = 9.98, p < 0.002$) which is a local metric based on the duration differences between consecutive intervals consistently show that duration for the NS is significantly shorter (NS = 180; NNS = 226) and there is higher variability (see whiskers in Figure 6.3). The PVIs were proposed by Grabe and Low (2002) for locally measure vocalic and consonantal intervals. It was inspired in Deterding's (1994, 2001) variability index or simply, VI, in which its first application was in onset-to-coda syllable duration with the purpose of normalizing differences in speech rate. To calculate VIS, one needs to assure that before differences between successive syllables are computed, the duration of each syllable is divided by the mean duration of all syllables and then, these differences are divided by the total mean of syllabic intervals - 1.

On Figure 6.4, we can see results from the VIS ($F(1,94) = 3.07, p < 0.05$) (NS = 0.66; NNS = 0.69). As attested by YARD z-scored results on Figure 6.5 ($F(1,94) = 16.04, p < 0.002$), NS syllables consistently vary (see whiskers in Figure 6.5) and hold lower durational values in between stress groups. There is an opposed relation when YARD accounts for NS and NNS groups. While for phrasal stress and stress groups syllables highly vary in duration for the NS, the NNS shows regularity patterns on both syllabic and stress group levels, besides higher duration values (NS = 0.82; NNS = 0.94). It would not be nonsense to state that NNS holds more balanced and higher V-V units in durational terms to some extent.

For the consonants, a much higher rate is produced between the NS. As well as the metrics for the syllables, RRC ($F(1,94) = 29.87, p < 0.001$) on Figure 6.6 showed great consistency for the consonants. This metric is based on the relation between consecutive consonantal interval duration and the variation between those intervals, i.e., the more variation between intervals, the lower the RR. As expected, RR was lower for the NS (= 55) than for the NNS (= 60). Results set as attested Gut (2005) for the RR in onset-to-coda syllable. In other words, NS showed higher variation in successive consonant duration than NNS.

4.2 Acoustic parameters distribution

For the acoustic parameters, 4 out of 12 parameters extracted from the data proved to be significantly reliable accounting variability between the factor's groups. From these four parameters, three are F0-based: In Figure 7.1, F0 minimum (F0min.) and 7.2, Speech rate (SR). Mean and Standard deviation of the F0 1st derivative ($\mu\Delta 1-F0$ and $\sigma\Delta 1-F0$ respectively) in Figures 8.1 and 8.2.

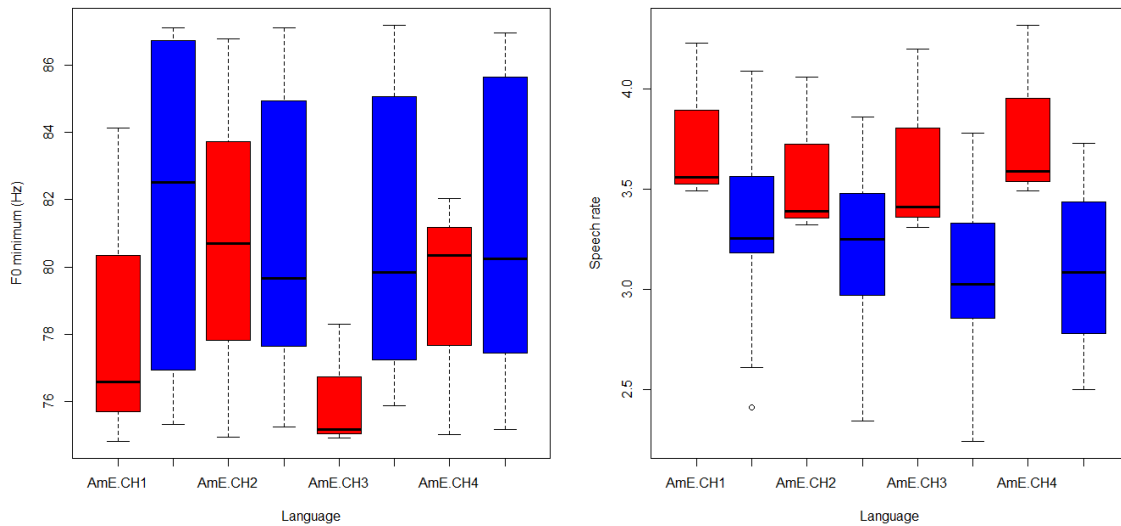


Figure 7: Boxplots of the F0 minimum (left portion, 7.1) and Speech rate (right portion, 7.2) for the NS (red boxes) and NNS (blue boxes) speakers’ productions.

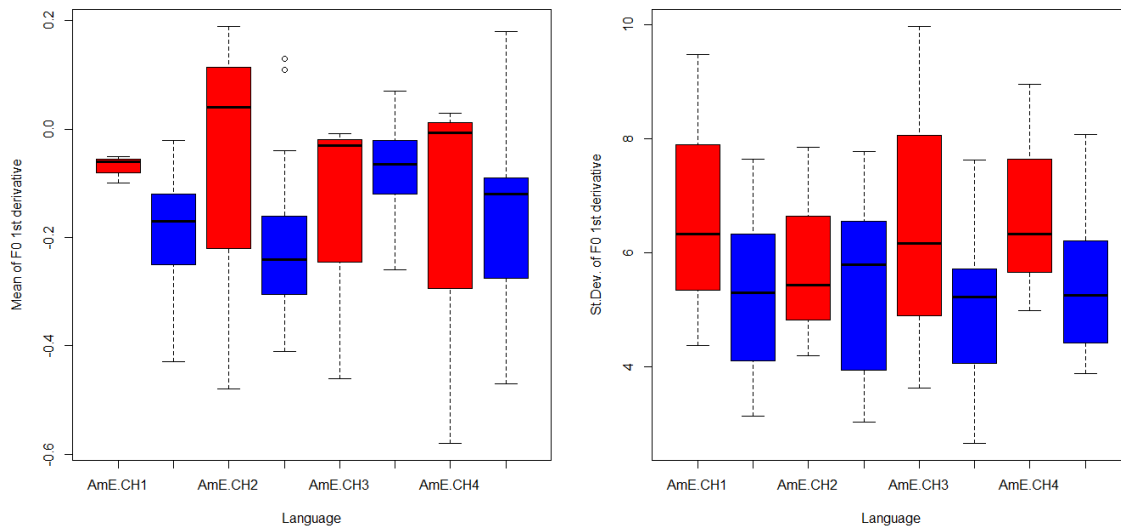


Figure 8: Boxplots of the Mean of F0 1st derivative (left portion, 8.1) and Standard deviation of F0 1st derivative (right portion, 8.2) for the NS (red boxes) and NNS (blue boxes) speakers’ productions.

From the acoustic measures presented in Figures 7 and 8, it can be seen that, such as the metrics presented in previous section, both language groups are well discriminated. As an inverse direction of the metrics, the acoustic measures presented higher absolute values for NS speakers with the exception for the F0 minimum (Figure 7.1) and, higher variability is maintained for the NS as well. Again, it is worth noting that; since this discussion is now upon the prosodic-acoustic parameters, it would be, once again, reasonable to assign to AmE speakers produce higher and more complex F0 contours since “trajectory” by means of variation in melody plays the most important role on the F0-related acoustic values hither used.

Having an overview of the data in Figures 7 and 8, let us see in Figure 9, the effect of the factor GROUP on these prosodic-acoustic parameters.

Speech rhythm of English as L2: an investigation of prosodic variables on the production of Brazilian Portuguese speakers

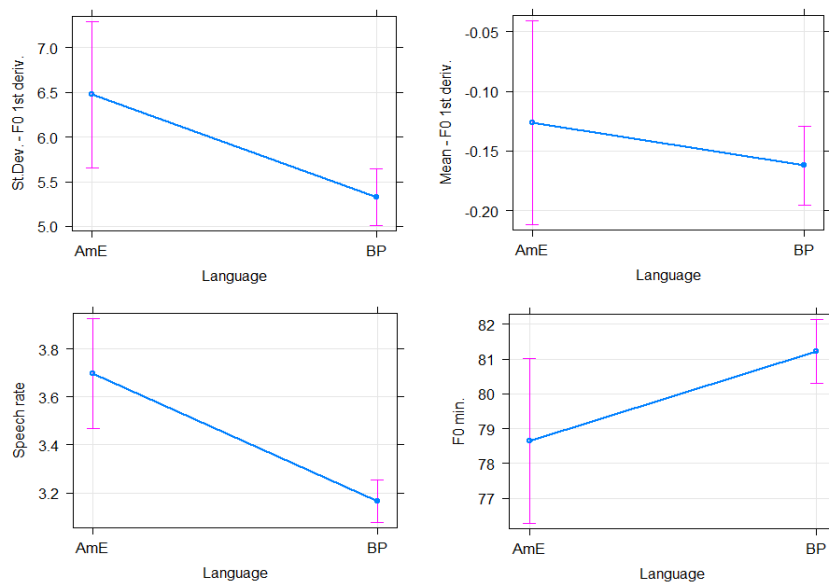


Figure 9: Linear effect plots of the AmE (L1) and BP (L2) speakers’ productions of English on the $\sigma\Delta 1$ -F0 (9.1), $\mu\Delta 1$ -F0 (9.2), Speech rate (9.3), F0 min. (9.4).

Accounting for the acoustic measures, NS’s productions set a significantly higher variation span (see whiskers in Figure 9) in comparison to NNS’s to whom the F0 contour remains in majority, *quasi*-monotonic (Cf. Figure 1). Higher values for the acoustic parameters described in this section can be attested for the NS (see Figure 9). In Figure 9.1, for $\sigma\Delta 1$ -F0 ($F(1,94) = 6.76, p < 0.02$) (NS = 6.49; NNS = 5.30); in Figure 9.1 for $\mu\Delta 1$ -F0 ($F(1,94) = 3.70, p < 0.05$) (NS = -0.13; NNS = -0.16) and Speech rate ($F(1,94) = 18.64, p < 0.001$) (NS = 3.7; NNS = 3.1) in Figure 9.3. As it was mentioned previously, F0 min. ($F(1,94) = 4.05, p < 0.05$) was the only parameter that had held higher values for the NNS group (NS = 78.8; NNS = 81.2) as we can see in Figure 9.4.

On the F0-based domain, Cumbers (2013, p. 4) points out that higher F0 min. is straightly related to lower F0 variability. This fact is also attested in the results hitherto presented. Yet as far as F0 min. in L2 English production is concerned, our findings are lined up with Urbani’s (2012) study, where she conducted a research on F0 range in L2 English speech; produced by Italian NNS, and L1 English speech; produced by American NS. The Italian NNS group produced significantly higher F0 min. values when comparing to American NS group. It is worth pointing out that the author also compared the influence of Italian’s pitch range with their L2 productions of English and found significant correlation in the F0 min. (p. 83). The paper herein did not compare as to whether there exists influence of BP F0 range in BP speakers L2 productions of English.

Along this section, we have presented the statistical values for the metrics and acoustic parameters and their relation with the distinction between the NS and NNS productions of English. In the next section, let us discuss about our hypotheses previously presented and explain the implications of our findings to a better understanding of the English L2 rhythm spoken by Brazilians.

5 Discussion

As previously mentioned, this study aims to find out whether L2 English speech rhythm is produced by BP speakers with a lower speech rate (due to lack of fluency), lower syllable

variability (due to regularity patterns on both syllabic and stress group levels) and less F0 variability (due to attention to segmental aspects rather than prosodic ones).

- Speech rate proved to be the most consistent acoustic parameter and it influences significantly BP English productions resulting in lack of fluency by the NNS although phonemic-sized vowel and consonant sounds were pronounced correctly most of the time. This may happen from a lack of attention to prosodic aspects such as speech rhythm;
- Syllabic variability also proved to be significantly lower in NNS productions according to our prediction. All of the metrics in Figure 6 point that out with the exception of the VarcoS. This metric is a ratio from standard deviation and the mean. Even though NNS present higher absolute standard deviation values, their mean values are also very high and consequently VarcoS was lower for BP speakers. This brings up the discussion that the regularity pattern of syllable production prevails for BP speakers on both syllabic and stress group levels over the distinction between these levels for the AmE speakers;
- Vocalic and consonantal metrics did not seem to be consistent measures, nevertheless the Rhythm Ratio for consonants showed significant variability;
- F0 variability was significantly lower for NNS as pointed out previously in Figure 1. There is a tendency to pay attention on segmental information rather than in the prosodic. By doing so, speakers can correctly produce categories of sounds in isolation. When the focus come to a vowel or consonant sound individually, prosodic sense by means of rhythmic and/or intonational patterns is misled (ROACH, 1982, 2009).

About the accuracy of the rhythm metrics, the study of Ding and Xu (op. cit.) claim that almost all classical rhythmic metrics can distinguish English from Mandarin Chinese. On the contrary, our investigation shows that only six, out of thirty classical rhythm metrics based on duration, can clearly distinguish the production of English as native language from BP speakers of EFL while four, out of twelve acoustic-based parameters can do this distinction as well. This fact leads us to assert, even if still early, that both metrics and acoustic parameters act simultaneously in speech rhythm without the need to forgo one or the other. We also assert that both categories of measurements are necessary for research refinements.

It is worth mentioning that from the six consistent metrics in our research, five are based on V-V unit intervals, which is a novelty in segmentation of L2 speech rhythm. Yet, these six metrics and four acoustic parameters are the ones that met the assumptions for the realization of the ANOVA statistics as aforementioned. We reiterate that we did not intend to perform any non-parametric equivalent statistics because the original values of the prosodic variables needed to be maintained since it is a first investigation of the differences between L1 and L2 productions.

About comparing L1-L2 rhythm from the use of metrics, Gut (2012, p. 91) yet asserts that, when describing L2 speakers' rhythm in comparison to native speech rhythm with the help of these metrics, it has to be borne in mind that the resulting values reflect not only methodological choices or protocols, but also the intrinsic differences in speech rate between both L1 and L2 groups. Besides, literature has posed the NNS groups at different proficiency levels may be characterized by differences in speech rate (Dellwo, 2006; Wong *et al.* 2019), our results did not cope with these differences between the BP group levels.

Differences in rhythm production as we have seen so far may have implications for pronunciation teaching where an accurate phonetic input plays a role on L2 pronunciation and speech acquisition. The rhythmic patterning of native versus non-native of English varies significantly in a great deal of the parameters herein presented. Brazilian patterns for L2 English speech rhythm showed significantly lower values in rhythmic metrics related to syllables and

consonantal units. If one compares these BP patterns to other EFL production such as Mandarin, Italian, French, Spanish speakers or so, one may encounter similar characteristics due to the language rhythmic pattern found therein. The metrics and acoustic parameters investigated in this research seems therefore, to be a more syllable-based tendency according to Low (2015, p. 132).

Taking Low's point of view of about how to teach pronunciation for English as an International Language (EIL), when one takes EFL pronunciation classes into account and considering instruction on speech rhythm in particular, teachers should be aware that if learners aspire towards a "globalist" orientation then "stress-based rhythm" should be taught. However, if learners aspire towards a "localist" orientation, then "syllable-based rhythm" should be the focus of the pronunciation classroom. For the author, the key is to introduce the element of choice to the learners, allowing them to decide their identity and orientation in the EFL pronunciation classroom. Low (op. cit.) makes clear that the argument in favor for the importance of teaching stress-based rhythm is to achieve fluency since in native varieties of English (British, American, and so), the presence or absence of reduced vowels forms the lowest level of the prosodic hierarchy (p. 133-134).

Yet in field of L2 pronunciation teaching, Celce-Murcia, *et al.* (1996, 2010, p. 2), come up with two approaches to be applied in pronunciation classes;

- An *intuitive-imitative approach*, where it depends on the learner's ability to listen to and imitate the rhythm, intonation and sounds of the target L2 without the intervention of any explicit instruction;
- An *analytic-linguistic approach*, where L2 speakers utilize information and tools such as a phonetic alphabet, articulatory descriptions, charts of the vocal apparatus, contrastive information, and other aids to supplement listening imitation and production. It explicitly informs the learner of and focuses attention on the segmental and prosodic information of the target language.

One of the authors' techniques to teach L2 rhythm is based on Morton, Marcus & Frankish's (1976) finger tapping experiment for the location of P-center. Celce-Murcia, *et al.* (1996) point out that learners from a syllable-based language background will have great difficulty in assigning greater length to the stressed syllables of content words within a sentence. The authors also emphasize that stress-based rhythm helps to improve the speech fluency of the of learners of English whose mother tongue is syllable-based considering rhythm to be one of the major structures that native speakers rely on to process speech.

By means of our research, prosodic aspects are aligned to Celce-Murcia *et al.* (1996, p. 324-325), if one compares BP speakers' productions of English to AmE speakers' productions characteristics in terms of rhythm. It can be concluded that BP speakers' productions present greater length of function words (articles, prepositions) and shorter length of content words (nouns, verbs) within the sentence and chunk levels, and that is one of the reasons for lower syllabic variability. They will also be challenged to appropriately reduce the length of function words an all unstressed syllables to maintain the characteristic stress-based rhythm of English.

The literature discussion between being syllable or stress-based is controversial and far to be finished. Intrinsic acoustic features of speech production (and perception) should be taken into account to determine what may or may not be taught in pronunciation classes other than external variables such as motivation, for example. In fact as aforementioned, the language oscillates (like a pendulum) dynamically to both sides in the continuum of speech in which the strength of the syllable oscillator and the stress group oscillator is used to infer if this language tends towards syllable-timing or stress-timing presenting rhythm as a dynamic organism as attested by Barbosa (2006). In addition, one must consider the melodic and intensive variables

applied to the dynamic nature of rhythm. These features coexist with the durational features and proved to be as robust as duration for the determination of differences between the speech rhythm of English as L1 and L2 productions. It makes us come up the inference that it is not a question of being stress or syllable-based but what other variables have implications in the rhythmic continuum when one speaks a foreign language.

We also corroborate Fuchs' (2016) statements that speech rhythm can be argued to be a multidimensional phenomenon with a range of acoustic correlates, and consequently a variety of rhythm metrics have been proposed. In order to fully capture the alternation of prominent/non-prominent units (such as vocalic/consonantal intervals, syllables or higher units) in a language, different correlates and dimensions of prominence should be taken into account for the sake of the coexistence from these parameters.

6 Conclusion

In this study, we compared L1 (AmE speakers) and L2 (BP speakers) English speech rhythm productions. By the results presented in this research, we conclude that duration and acoustic parameters may be taken into account in a great chance by far, since these measures significantly influence L2 rhythm production. On the basis of this investigation, it is possible to statistically train and establish reliable metric and acoustic models to enhance L2 prosody studies under distinct domains. More than training these models and keep the traditionally two-dimensional metric-based studies, the research results herein presented point out to a new "gateway" towards the direct acoustic parameters, which are indeed more sensitive and transparent to perception, since we perceive pitch and speech rate other than unique-duration units in L2 speech rhythm.

The novelty brought from this research for L2 speech rhythm studies is the inclusion of phonetic syllable intervals for the verification for the metrics and prosodic-acoustic parameters, as well as F0-based measures of centrality, dispersion, derivative and intensive parameters, such as spectral emphasis. Both durational metrics and F0 acoustic parameters showed robustness and reliability in the measurement of L2 rhythm. Moreover, this study could shed light towards (English as a) Foreign Language teaching. The implications could be high and effective on the implementation of acoustic-articulatory phonetics in the teaching context as one of its outcomes. Exploring prosodic aspects such as, rhythm and intonation, in (E)FL classes by pointing out and working speaking activities, can make a difference in order to aware learners and gradually improve pronunciation. Keeping in mind prosodic awareness and teaching phonetic abilities would be, by far, significant resources for pronunciation instruction enhancement.

Some non-target-like pronunciation features can make foreign language more challenging and adjusting prosodic production to the target L2 is undoubtedly desirable at ensuring pronunciation instruction. Furthermore, this research has great implications to be applied on field of forensic phonetics when verifying and assessing acoustic-prosodic and metric parameters that significantly interfere on the identification of foreign accent and/or speech disguise on speaker recognition investigations in context-based voice lineups, such as theft, rape, kidnapping and other crimes.

For future studies, we intend to check the perceptual phase of L2 speech rhythm and to test the alignment of its measures to the production ones performing a multiple-forced-choice experiment where the listener's perception will be based on AmE speakers rating different degrees of foreign accent.

Acknowledgements

We gratefully acknowledge the grant from CNPq (150143/2018-4) for the first author and (302657/2015-0) for the second author.

REFERENCES

1. Abercrombie, D. *Elements of General Phonetics*. Edinburgh, Edinburgh: University Press, 1967.
2. Adams, C. *English Speech Rhythm and the Foreign Learner*. The Hague: Mouton Publishers, 1979.
3. Adams, C; Munro, R. *In search of the acoustic correlates of stress. Fundamental frequency, amplitude and duration in the connected utterance of some native and non-native speakers of English*. *Phonetica* 35(3), 125-156, 1978.
4. Barbosa, P. *Panorama of Experimental Prosody Research*. In: Proc. of the GSCP Workshop, Belo Horizonte Firenze University Press, 33-42, 2012.
5. Barbosa, P. *Incursões em torno do ritmo da Fala*. Campinas: Pontes Editores, 2006.
6. Barbosa, P. *Syllable-Timing in Brazilian Portuguese: Uma Crítica a Roy Major (Tempo-silábico em Português do Brasil: a critic to Roy Major)*. *DELTA* 16 (2), 369-402, 2000.
7. Barbosa, P.; Bailly, G. *Characterisation of rhythmic patterns for text-to-speech synthesis*, *Speech Communication* 15, 127-137, 1994.
8. Boersma, P.; Weenink, D. *Praat: doing phonetics by computer*. (Version 6.0). <http://www.praat.org>, 2019.
9. Carroll, D. *Psychology of language*. Belmont: Wadsworth, 1994.
10. Celce-Murcia, M.; Brinton, D.; Goodwin, J. *Teaching Pronunciation: A course book and reference guide* 2ed. Cambridge: Cambridge University Press, 2010.
11. Celce-Murcia, M.; Brinton, D.; Goodwin, J. *Teaching Pronunciation: A Reference for Teachers of English to Speakers of Other Languages*. Cambridge: Cambridge University Press, 1996.
12. Couper-Kuhlen, E. *English Speech Rhythm: form and function in everyday verbal interaction*. Philadelphia: John Benjamins Publishing Co., 1993.
13. Cumbers, B. *Perceptual Correlates of Acoustic Measures of Vocal Variability*. Thesis (Master of Science in Communication Sciences and Disorders), The University of Wisconsin-Milwaukee, 2013.
14. Dellwo, V. *Rhythm and speech rate: A variation coefficient for deltaC*. In: *Language and language-processing*. In: Proceedings of the 38th linguistics colloquium, Frankfurt, Peter Lang, 231–241, 2006.
15. Deterding, D. *The rhythm of Singapore English*. In Proceedings of the fifth Australian international conference on speech science and technology, Perth, 316–321, 1994.
16. Deterding, D. *The measurement of rhythm: A comparison of Singapore and British English*. *Journal of Phonetics* 29, 217–230, 2001.
17. Ding, H.; Xu, X. *L2 English Rhythm in Read Speech by Chinese Students*. In. Proceedings of INTERSPEECH, San Francisco, 2696-2700, 2016
18. Dupoux, E. *Plasticity and non-plasticity in speech processing: Late learners and early forgetting*. Maryland, Paper presented to Linguistics Department, University of Maryland, 2003.
19. Flege, J. *Perception and production: The relevance of phonetic input to L2 phonological learning*. In T. Huebner and C. Ferguson (eds). *Crosscurrents: Second Language Acquisition and Linguistic Theory*. Amsterdam: John Benjamins, 249–289, 1991.
20. Flege, J. *Second Language Speech Learning: Theory, Findings and Problems*. In: W. Strange (eds), *Speech Perception and linguistic Experience: Issues in Cross-Language Research*. Timonium: York Press, 233-277, 1995.
21. Fuchs, R. *Speech Rhythm in Varieties of English: Evidence from Educated Indian English and British English*. Dordrecht: Springer Science Business Media, 2016.
22. Grabe, E; Low, E. *Durational variability in speech and the rhythm class hypothesis*. In: C. Gussenhoven and N. Warner, (eds), *Papers in Laboratory Phonology*, 515–546, 2002.

23. Gut, U. *Rhythm in L2 speech*. *Speech and Language Technology*. 14(15), 83-94, 2012.
24. Gut, U. *Nigerian English prosody*. *English World-Wide* 26(2), 153–177, 2005.
25. Hayes, B. *Metrical Stress Theory: Principles and Case Studies*. Chicago: University of Chicago Press, 1995.
26. Hayes, B. *The phonology of rhythm in English*. *Linguistic Inquiry*, 15(1), 33-74, 1984.
27. Loukina, A.; Kochanski, G.; Shih, C.; Keane, E.; Watson, I.; *Rhythm measures with language-independent segmentation*. In: *Proceedings of INTERSPEECH*, Brighton, 2009.
28. Low, E. *The Rhythmic Patterning of English(es): Implications for Pronunciation Teaching*, In. M. Reed and J. Lewis (eds). *The Handbook of English Pronunciation*. West Sussex: John Wiley & Sons, Inc., 125-138, 2015.
29. Magen, H. *The perception of foreign-accented speech*. *Journal of Phonetics*, (26), 381-400, 1998.
30. Major, R. *Foreign Accent: The Ontogeny and Phylogeny of Second Language Phonology*. Mahwah: Lawrence Erlbaum Associates, Inc., Publishers, 2001.
31. Mascherpe, M. *Análise comparativa dos sistemas fonológicos do inglês e do português*. São Paulo: Empresa Gráfica da Revista dos Tribunais S.A, 1970.
32. Mennen, I., Schaeffer, F., Docherty, G. *A methodological study into the linguistic dimensions of pitch range differences between German and English*. *Proc. 4th Speech Prosody*, University of Campinas, 527-530, 2008.
33. Mennen, I., Schaeffer, F., Docherty, G. *Cross-language difference in f0 range: a comparative study of English and German*. *JASA*, 131(3), 2249-2260, 2012.
34. Moreno, M. *Sobre la adquisición de la prosodia en lengua extranjera: Estado de la cuestión. Didáctica*. *Lengua y Literatura*, 91-119, 2000.
35. Morton, J., Marcus, S., Frankish, C. *Perceptual Centers (P-centers)*. *Psychological Review*, 83(5), 405-408, 1976.
36. Moyer A. *Age, accent, and experience in second language acquisition: an integrated approach to critical period inquiry*. Clevedon: Multilingual Matters; 2004.
37. Nooteboom, S. *The Prosody of Speech: Melody and Rhythm*. In. *The handbook of phonetic sciences*, Utrecht: UtdallasEdu, 640–673, 1997.
38. Ordin, M.; Polyanskay, L. *Acquisition of speech rhythm in a second language by learners with rhythmically different native languages*. *Journal of the Acoustical Society of America* 138, 223-255, 2015.
39. Pike, K. *The Intonation of American English*. Ann Arbor: University of Michigan Publications, 1945.
40. Piske, T.; Mackay, I.; Flege, J. *Factors affecting degree of foreign accent in an L2: a review*. *Journal of Phonetics* 29, 191-215, 2001.
41. Ramus, F.; Nespor, M; Mehler, J. *Correlates of linguistic rhythm in the speech signal*. *Cognition*. 73, 265–292, 1999.
42. Roach, P. *Rhythm and Production: A lecture by Peter Roach*. <https://www.youtube.com/watch?v=Xh_fiP6AT_I>, 2009.
43. Roach, P. *English Phonetics and Phonology*. Cambridge, Cambridge University Press, 2005.
44. Selinker, L. *Interlanguage*. *International Review of Applied Linguistics*, v.10, 209-231, 1972.
45. Silva Jr., L; Barbosa, P. *Metrics & Acoustics Extractor version 1.0*. Script for Praat. 2019.
46. Stampe, D. *A dissertation on natural phonology*. Thesis (Philosophy Doctor in Linguistics), Chicago: University of Chicago, 1973.
47. Stetson, R. *Motor Phonetics: A Study of Speech Movements in Action*, 2nd ed., Amsterdam: North Holland Publishing Co., [1928] 1951.
48. Urbani, M. *Pitch Range in L1/L2 English. An Analysis of F0 using LTD and Linguistic Measures*. Padova: Coop. Libreria Editrice Università di Padova, 2012.

*Speech rhythm of English as L2: an investigation of prosodic variables on the production of
Brazilian Portuguese speakers*

49. Wagner, P.; Dellwo, V. *Introducing YARD (yet another rhythm determination) and re-introducing isochrony to rhythm research*. In Proceedings of Speech Prosody, ISCA, 227–230, 2004.
50. White, D; Mok, P. *L2 Speech Rhythm and Language Experience in New Immigrants*. In: Proceedings of 19th ICPHS. Melbourne, 334-338, 2019.
51. Wong, S.; Dealey, J.; Leung V.; Mok, P. *Production of English connected speech processes: an assessment of Cantonese ESL learners' difficulties obtaining native-like speech*. In: The Language Learning Journal, 1-16, 2019.