L2 prosody effects on pronunciation teaching and oral communication: updated

Efeitos de prosódia L2 no ensino de pronúncia e comunicação oral: atualizado

Efectos de la prosodia L2 en la enseñanza de la pronunciación y la comunicación oral: actualizado

DOI: 10.54033/cadpedv21n5-158

Originals received: 04/19/2024
Acceptance for publication: 05/10/2024

Leônidas José da Silva Junior
Doctor of Linguistics
Institution: Universidade Estadual da Paraíba (UEPB)
Address: Guarabira, Paraíba, Brazil
E-mail: leonidas.silvajr@gmail.com

Plínio Almeida Barbosa
Doctor of Speech Science
Institution: Universidade Estadual de Campinas (UNICAMP)
Address: Campinas, são Paulo, Brazil
E-mail: pabarbosa.unicampbr@gmail.com

ABSTRACT
Pronunciation accuracy in a foreign language requires mastering production of both segmental (consonants and vowels) and prosody (features that operate from the syllable domain, such as lexical/phrasal stress, pitch accent, rhythm, intonation, and voice quality) of speech. This study aims to elucidate the significance of instruction in pronunciation, with a focus on L2 prosody as a pivotal
component in the enhancement of oral communication skills. The objective is to reduce pronunciation discrepancies and augment conversational interactions in spoken communication. The theoretical review of our research is based on studies such as, Munro and Derwing (2020 [1999]) Derwing and Munro (2015), Levis (2018), which underscore the role of prosody in the context of pronunciation instruction and in L2 oral communication. We also have supported in studies such as, Ortega-Lebaria et al. (2023), Silva Jr. (2023), for the selection of techniques and guidelines for extracting both durational and melodic parameters from L2 prosody-focused research. Regarding the methodology, we gathered data from Brazilian English speakers and native speakers of English from the USA. Each participant read a passage, from which we selected ten utterances for subsequent analysis. The acoustic analysis was conducted based on the duration of phonetic syllables (V-V units) and the fundamental frequency (F0), both normalized using Lobanov’s z-score. We also employed ANOVA statistical tests to examine the group effect for both duration and F0 parameters. Our findings indicate notable differences between native and Brazilian English speakers’ productions for both duration and F0. We conclude, even on a preliminary basis, that the application of L2 prosody is relevant in improving pronunciation instruction and enhancing oral communication.

**Keywords:** L2 Prosody. Pronunciation Teaching. Oral Communication. Acoustic Parameters.

**RESUMO**
A pronúncia acurada em uma língua estrangeira requer o domínio da produção de de segmentos (consoantes e vogais) e da prosódia (características que operam a partir do domínio silábico, como acento lexical/frasal, acento tonal, ritmo, entonação e qualidade da voz) da fala. Este estudo visa elucidar a importância da prosódia de L2 no ensino de pronúncia, sendo esta, um componente crucial no aprimoramento das habilidades de comunicação oral. O objetivo é reduzir discrepâncias na pronúncia e aumentar as interações conversacionais na comunicação oral. A revisão teórica de nossa pesquisa é baseada em trabalhos como os de Munro e Derwing (2020 [1999]) Derwing e Munro (2015), Levis (2018), que destacam o papel da prosódia no contexto do ensino de pronúncia e na comunicação oral L2. Também recorremos a estudos como os de Ortega-Lebaria et al. (2023) e Silva Jr. (2023), dedicados a seleção de técnicas e diretrizes para a extração de parâmetros duracionais e melódicos da pesquisa em prosódia de L2. Para a metodologia, coletamos dados de falantes brasileiros de inglês e falantes nativos de inglês dos EUA. Cada participante leu um texto, do qual selecionamos dez enunciados para análise subsequente. A análise acústica foi conduzida com base na duração das sílabas fonéticas (unidades V-V) e na frequência fundamental (F0), ambas normalizadas usando o z-score de Lobanov. Também empregamos testes estatísticos ANOVA para examinar o efeito do grupo para ambos os parâmetros de duração e F0. Os resultados indicam diferenças significativas entre falantes nativos e brasileiros de inglês para os parâmetros apresentados aqui. Concluímos, de forma preliminar, que a aplicação da prosódia L2 é relevante para o ensino da pronúncia e aprimoramento da comunicação oral do falante.

RESUMEN
La pronunciación precisa en un idioma extranjero requiere dominio de la producción segmentaria (consonantes y vocales) y la prosodia (características que operan desde el dominio silábico, como el acento léxico/frasal, el acento tonal, el ritmo, la entonación y la calidad de la voz) del habla. Este estudio pretende dilucidar la importancia de la prosodia L2 en la enseñanza de la pronunciación, componente crucial para mejorar las habilidades de comunicación oral. El objetivo es reducir las discrepancias en la pronunciación y aumentar las interacciones conversacionales en la comunicación oral. La revisión teórica de nuestra investigación se basa en trabajos como los de Munro y Derwing (2020 [1999]), Derwing y Munro (2015), Levis (2018), que destacan el papel de la prosodia en el contexto de la enseñanza de la pronunciación y en la comunicación oral L2. También recurrimos a estudios como los de Ortega-Lebaría et al. (2023) y Silva Jr. (2023), dedicados a la selección de técnicas y guías para la extracción de parámetros duracionales y melódicos a partir de investigación en prosodia L2. Para la metodología, recopilamos datos de hablantes de inglés brasileño y hablantes nativos de inglés de los Estados Unidos. Cada participante leyó un texto, del cual seleccionamos diez enunciados para su posterior análisis. El análisis acústico se realizó con base en la duración de las sílabas fonéticas (unidades V-V) y la frecuencia fundamental (F0), ambas normalizadas con la puntuación z de Lobanov. También empleamos pruebas estadísticas ANOVA para examinar el efecto de grupo para los parámetros de duración y F0. Los resultados indican diferencias significativas entre los hablantes nativos de inglés y brasileños para los parámetros aquí presentados. Se concluye, de manera preliminar, que la aplicación de la prosodia L2 es relevante para la enseñanza de la pronunciación y la mejora de la comunicación oral del hablante.


1 INTRODUCTION

According to Lengeris (2012), pronunciation accuracy in a foreign language (L2) requires mastering production of both segmental (consonants and vowels) and prosodic features of speech (features that extend over more than one segment such as lexical/phrasal stress, pitch accent, rhythm, intonation, and voice quality) but teaching pronunciation of the latter is traditionally neglected in language classrooms.
After the advent of the communicative approach to language teaching that prioritized language function over language form, the study of L2 prosody has admittedly experienced an increasing interest among language teachers. For the author (p. 25-26), (L2) prosody, and intonation in particular, plays a crucial role in communication by conveying not only linguistic information such as chunking the speech into phrases, signalling new and contrastive information, and disambiguating sentences that otherwise could sound ambiguous to the listener, but also paralinguistic information, i.e., information related to the identity, age, gender, and emotional state of the speaker.

From the abovementioned, the current research aims at presenting the importance of pronunciation of a L2 taking prosody as a central element to minimize pronunciation deviations and enhance conversational exchanges in oral communication. Our study corroborates most of the ideas of Levis (2018) on pronunciation teaching and prioritization of prosodic aspects when teaching the target-L2.

Roach, (1982); Celce-Murcia et al. (2010); Avery and Ehrlich (2012) assert that in L2 classes, phonological contents associated with phonetic practices have been neglected and, even when contemplated, the proposals are dedicated to phonetic-phonological contents. The authors highlight that it is common for most teachers to prioritize teaching pronunciation, essentially emphasizing the segmental features of the L2, that is, there is considerable classwork dedicated to the teaching of phonemic-sized units rather than higher prosodic units.

Oral production and comprehension of a language seem to be one of the main obstacles encountered by foreign speakers due to the differences between the mother tongue (L1), and the L2 phonetic-phonological systems. The teaching of L2 pronunciation has presented controversies about what, and how to teach. Although there is a considerable amount of research related to the importance and mediation of pronunciation teaching (Jenkins, 2000; Celce-Murcia Et Al., 2010; Avery & Ehrlich, 2012; Silva Jr., 2018; Silva Jr. & Barbosa, 2019, 2020; among others), it seems that this issue is still far from being somewhat consensus.
Cagliari (1978) mentions that it is essential that before learning grammatical skills such as syntax, students must train phonetic exercises of production and perception of sounds. As well as Cagliari (1978), Silva Jr. (2018) and Silva Jr. (2020a) highlight, from an experimental-based point of view, the importance of leading the learner to training L2 sound production and perception before other skills.

Roach (1982) also states that processes and methods used for the L2(-English) acquisition, that account for only pronunciation that prioritizes the production of phones in isolation, should be revisited. Derwing and Munro (2015) propose that, in fact, both aspects – segmental and prosodic – should be integrated during L2 pronunciation teaching practices to enhance oral communication.

This paper is divided into the following sections: Introduction, in which we present general questions related to L2 pronunciation; 2. Literature review, in which we approach works on the importance of L2 prosody for teaching pronunciation and its performance in oral communication; 3. Methods, in which we detail how data collection, acoustic and statistical analyses were carried out throughout this study; 4. Results and Discussion, in which we present the performance, description, and analysis of our data; Final remarks, in which we point out ways to use L2 prosody in teaching pronunciation, in addition to the references herein used.

This study aims to elucidate the significance of instruction in pronunciation, with a focus on L2 prosody as a pivotal component in the improvement of oral communication skills. The objective is to reduce pronunciation discrepancies and augment conversational interactions in spoken communication.

2 LITERATURE REVIEW

According to Levis (2018), Pronunciation, with all its intricacies and interesting features, does not exist per si. It is a servant skill, i.e., it provides subsidies so that oral communication may occur successfully. In other words, pronunciation teaching is a product from oral communication (listening and
speaking), and the conversational context in which one is inserted in. Fluent, or partially fluent L2 speakers produce syllables, words and utterances based on their lexical knowledge, and their particular grammars.

Celce-Murcia et al. (2010), point out that the speaker is able to process features and elements of the pronunciation of one’s interlocutor, whether native or L2 speaker. The authors mention that pronunciation may be followed by higher or lower speech rate, in addition to being more formal or casual. Such information reflects segmental and prosodic aspects of the target-L2 and determines the degree of foreign-accented speech.

In oral communication, the speaker’s higher or lower degree of foreign accent is the (first) anchoring point perceived by the native listener. It is the instance where the perception of deviations in the segmental and prosodic features occurs. Oral communication also functions as a trigger for social judgments due to perceived phonetic differences. As an example, we can name the deletion of the voiced glottal fricative consonant in English [h] (General American or British Received Pronunciation).

This phenomenon is likely to happen in L2 English production of Brazilian speakers in the initial phase of its acquisition by the interference from the L1 into the L2 phonetic-phonological system. The deletion (dropping) of the first sound in words like <hospital, hand>, produced as: [∅]ospital or [∅]and, for instance, are socially stigmatized (Mugglestone, 1996). Prosodic aspects such as little variability of the fundamental frequency (F0) in the L2 English productions is also socially regarded as a lack of mastering and knowledge of the language, as well as a non-confident and untrustful discourse (Lev-Ari & Keysar, 2010).

According to Levis (2018), among all approaches in L2 teaching, pronunciation figures as the central element. Although pronunciation is a skill that is a function of other variable aspects of L2 speech, it is far from being worked only as a linguistic variable. Instead, L2 pronunciation is essential as it is the skill that has the greatest impact on speech intelligibility, i.e., speakers are understood and can understand each other by oral conversational instances.

Levis (2018) claims that speech intelligibility is of great importance given that oral communication represents “the heart” of human interaction from a
conversational point of view. In addition, the author addresses that pronunciation is an important element of comprehensibility, i.e., the amount of work that listeners must do to understand one’s speech. Such comprehensibility reflects a fluid L2 speaker-listener accommodation in oral communication.

Regarding some of the aspects to be highlighted in L2 oral communication, the phonetic literature diverges to some extent. A part of the studies proposes that L2 prosody plays a more important role in oral communication and that, if neglected, may cause a significant impairment in the intelligibility of speech acts (Munro 1995; Derwing et al. 1998; Munro & Derwing 2020 [1999]; Derwing and Rossiner, 2003, among others).

On the other hand, a great deal of studies in the literature state that errors in the production of vowels and consonants are dominant and prevail as a cause of unintelligibility (Munro and Derwing, 2008, among others). Also, as reported in the Introduction of this study, studies such as those by Derwing and Munro (2015) point to results in which both prosodic and segmental errors in the L2 cause intelligibility problems during oral communication.

In more recent research, Trouvain et al., 2021, and Rosenberg and Hirschberg, 2021 point out that oral communication in L2 is strongly correlated with prosodic elements, not only from the point of view of intelligibility, but also, through an affective and a pragmatic point of view. For these authors, prosody plays an important role, and it is reflected in global acoustic parameters related to F0, intensity, pauses (silent or filled), duration, and voice quality in the suprasegmental domain. Their studies yet address that prosodic-oriented acoustic parameters are responsible for the perceptual effects of feelings, such as trust, empathy, leadership, enthusiasm, and charisma.

According to Brem and Niebuhr (2021), charisma is not substantially determined by voice quality, but rather, vocal quality in covariance with acoustic elements that reflect the rhythm of speech. Furthermore, it is claimed that, in the case of a foreign accent, the L2 factor would not have a negative effect on the speaker’s charisma and/or other qualities abovementioned, however the foreign accent would depend much more on the prestige relationship of the L1 and the target L2.
On the other hand, Lev-ari and Keysar (2010) state that when feelings are transmitted in L2 speech, they are judged as less reliable by native speakers regardless of the L1-L2 prestige relationship. The authors still attribute to pauses – silent or filled – and the use of hesitant elements of L2, as being the acoustic parameters with the greatest negative effect from the perceptual point of view.

2.1 L2 PROSODY AND PRONUNCIATION TEACHING

Prosody is dedicated to studying paralinguistic aspects such as: stress and pitch accent, rhythm, intonation, and voice quality, as well as other suprasegmentally determined elements of the speech acts (Barbosa, 2019).

In relation to L2 Prosody studies, acoustic and experimental phonetics have emerged since the late 1970s with the verification of English stress production by native and nonnative speakers based on F0, intensity and duration of connected speech (Adams & Munro, 1978). In their seminal study, the authors claim that inappropriate productions of the L2 prosodic features may lead to misunderstandings, both from a semantic and pragmatic point of view. Moving a bit further, Adams (1979) investigated the influence of L2 prosody in English as foreign language (EFL) productions and its correlations to respiratory parameters based on the amplitude returned by the chest and diaphragm muscular tension during L2-L2 speech.

According to Silva Jr. (2020a), it is inside the L2 prosody scope that we may find the most distinctive phonetic aspects in languages. It is not just about what is said linguistically, but about “how” the utterance is said (Silva Jr. & Barbosa, 2019). The allocation of prosodic stress, rhythm, and intonation of L2 speech play a significant role in the construction of L2 pronunciation. Munro (1995); Celce-Murcia, et al. (2010); Lengeris (2012); Derwing and Munro (2015), and so state that, in L1, such prosodic features are acquired in early childhood, making it more difficult to be overridden when producing L2 speech in a lifetime.

When one comes across L2 prosody (of English) in Brazilian speakers’ productions for lexical stress, Modesto (2019) points out that Brazilians present
difficulties in the perception and production of such stress patterns. Such difficulties can be effectively attributed to two factors:

a) floating perception of lexical and phrasal stress in English due to the interdependence of two acoustic correlates: $F_0$ (left-headed pitch accent) and duration (right-headed stress) as stated by Hewings (2007), and Hancock (2012) and;

b) in Brazilian Portuguese (BP), duration is the predominant acoustic correlate of both lexical and phrase stress (Barbosa, 2006).

Figure 1 presents duration (panel A) and $F_0$ (panel B), for both L1 and L2 English stress and pitch accent pattern (panels A and B respectively) extracted from phonetic syllables, that is, the intervals the comprehend the onset of a vowel until the onset of the next vowel. For the illustrated example in Figure 1, we used eight L1-English and eight L2-English productions of the sentence <Mobile, Alabama>, which were segmented into the following phonetic syllable units between square brackets: “m[OB] [ILE] [AL] [AB] [AM]a”. The L1 speakers were from the United States and the L2 speakers were from Brazil. The target sentences were extracted from larger chunks of spontaneous speech. Both duration and $F_0$ were normalized into Lobanov’s z-score method (Lobanov, 1971):

Figure 1: Duration (panel A) and $F_0$ (panel B) mean of the normalized V-V units for the English sentence <Mobile, Alabama> produced by eight native and BP speakers of English. Whiskers indicate 95% confidence interval.

Figure 2 shows both L1-L2 speech performance of intonational patterns, which presented much higher melodic variability (rising-falling F0 contour) for the L1 speech along the utterance, and a shift on the contour of the F0 curve when it reaches phrasal stress position for the L2 speech. Figure 2 yet presents an approximate time difference of 610 milliseconds (ms) for the utterance production (longer for L2 speech), as well as, different pause duration between the groups (L1 = 14 ms, and L2 = 75 ms) for the prosodic boundary between both phrases, i.e., \([\text{who placed his huge paw}]\phi [\text{on the mouse}]\phi\) (pause is represented by the blank part right after the arrows along the contour in panels A and B):

Figure 2: F0 contour of the utterance: \([\text{who placed his huge paw on the mouse}]\Upsilon\) spoken by both a native (red F0 contour on panel A), and a BP speaker of English (blue F0 contour on panel B). Arrows indicate contour trajectory towards phrase stress, where: ‘HL’ = high-low tone and ‘LH’ = low-high tone.

As far as melody of L2 pronunciation is concerned, Moreno (2000) addresses that, intonation is 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 by L1-English speakers.

In pronunciation classes, Moreno (2000); Reed and Michaud (2015) state that intonation does not come into a spotlight, and it is irrelevant at best due to its degree of difficulty. Commonly, what we observe in these classes is that when students consciously or unconsciously, hear recordings in L2, they concentrate on aspects such as semantic, grammatical, segmental, etc., and mostly do not perceive intonation patterns. This remains the learners 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, as well as pitch accent to signal sentence modality (yes-no questions and/or wh-questions).

To assess how L2 prosody acts in oral communication of foreign speakers, one will see in the next sections how data collection, and acoustic and statistical analyses were carried out.

3 METHODS

In this section, we describe the corpus of the study, that is, the screening of participants, data collection, as well as the acoustic and statistical data analyses carried out in this research.

3.1 PARTICIPANTS AND DATA COLLECTION

We collected audio data from two distinct groups:

a) an experimental group composed by ten Brazilian speakers of L2-English (henceforth, BRA), and;

b) a control group, composed by two native English speakers from the United States who have lived in Brazil for more than four years (henceforth, NAT).

The NAT group was also BP-L2 speakers.

The BRA group was submitted to a placement test to determine the English level of proficiency of participants. For this purpose, we ran the Oxford Online Placement Test (OOPT). BRA group was qualified as B2→C1 (mean score = 73), according to the Common European Framework of Reference (CEFR - see Pollitt (2019, p. 9 for details of mean scores and proficiency level transition).

Both groups read a text from which ten utterances (U) were extracted, as described from utterance 01 'U01' to 'U10' in Table 1:
Table 1: Utterances produced by both BRA and NAT group extracted from the text: “How good are my life, family and friends!”.

<table>
<thead>
<tr>
<th>Utterance</th>
</tr>
</thead>
<tbody>
<tr>
<td>U01 I want to stay at home, but I need to go to a library</td>
</tr>
<tr>
<td>U02 He was celebrating because he was approved</td>
</tr>
<tr>
<td>U03 I wanted to text you, but I don't have your cell phone number</td>
</tr>
<tr>
<td>U04 I go to the mall every week because I love window shopping</td>
</tr>
<tr>
<td>U05 The virus cannot live in immunized individuals, nor in nature</td>
</tr>
<tr>
<td>U06 Playing soccer in the park is entertaining in the winter, but it's better in the heat of summer.</td>
</tr>
<tr>
<td>U07 Nobody talks to Paul because he looks mad</td>
</tr>
<tr>
<td>U08 I always take a book to read, yet I never seem to turn a single page.</td>
</tr>
<tr>
<td>U09 She is very old but still attractive</td>
</tr>
<tr>
<td>U10 Paul thought he had a great chance to be accepted at YALE, because his grandfather taught him many things</td>
</tr>
</tbody>
</table>


Each participant was previously shown the text, so that they were able to be more familiar with the words, syntax and, consequently, input their prosody during recordings. In the recording moment, participants read the text once, that is, one text sample per participant. The statements were extracted from the complete audio file with Praat software (Boersma & Weenink, 2021). A total of 120 utterance tokens were selected for our study as detailed in equation (1):

\[
\text{Tokens} = (10_{\text{utterances}} \times 10_{\text{participants}} = 100_{\text{BRA group}}) + (10_{\text{utterances}} \times 2_{\text{participants}} = 20_{\text{NAT group}}), \quad 100_{\text{BRA group}} + 20_{\text{NAT group}} = 120_{\text{tokens}}
\]

Data collection was performed in a quiet room from a Zoom H1 Handy PCM Recorder, unidirectional on-board Zoom H1 microphone, at a response frequency from 30 to 16 kHz, a sampling rate of 44.1 kHz, and a 16-bit quantization rate. Signal-to-noise ratio was higher than 30 dB to ensure greater data quality and fidelity. The described settings could guarantee a better capture of F0 parameters, since we analyzed intonation aspects in our study.

3.2 ACOUSTIC ANALYSIS

Acoustic analysis was based on prosodic correlates, namely z-score-normalized duration and F0, (Lobanov, 1971) automatically extracted from a
Praat script: *(NormLobanov_DurF0F1F2 - Silva Jr., 2020b)*. This method takes the z-score statistical procedure over the units.

In the current study, in each utterance (Table 1), *Z-score* represents the difference between the *duration* (in milliseconds - ms), or the *F0* curve (in *Hertz*, Hz) of each V-V unit, and their mean values (µ*Duration* or µ*F0*), divided by their standard deviation values (SD*Duration* or SD*F0*). According to Barbosa and Madureira (2015, p. 211), z-score normalization is used to minimize micro-prosodic effects in duration (nature, number of phones and perceptually related peaks in a phonetic syllable), and in F0 (production of vowels and consonants) that do not have prosodic-linguistic function in addition to highlighting the real prosodic information. Z-score also smooths F0 differences between female (high-pitched) and male (low-pitched) voices.

For durational measurements, we adopted the duration of phonetic syllables, i.e., *onset-to-onset* of vowels units is based on Barbosa (2006), as well as Caroll’s (1994), and krivokapic (2012), for F0 measurements. The use of V-V units was successfully applied in L2 prosodic analysis by Silva Jr. and Barbosa (2019, p. 43-44). The protocol showed to be consistent and brought forward promising results concerning the use of V-V units as a syllable-level prosodic unit for measuring L2 speech rhythm by means of duration (by the use of the classical rhythm metrics, which showed to be more robust than the traditionally consonant- or vowel-like metrics), and intonation (by the used of F0-based prosodic parameters, which revealed plenty of the variability from native speakers of English when comparing to L2 speakers) (see p. 48 for a summary of the results using V-V units). Figure 3 shows how data were segmented into V-V units.

It is noteworthy mentioning that the symbol “V-V” used along the present study, does not stand for two vowel (V) segments, but an interval from the onset of a vowel to the next vocalic onset, i.e., “V-V” stands for V-to-V.
3.3 STATISTICAL ANALYSIS

As for the statistical analysis, we performed a one-way Analysis of Variance (ANOVA) between the groups BRA and NAT and checked for the effect size provided by the coefficient of determination (the adjusted-R²). An R² value of 0.5 (50%) of variance between groups is already considered a strong effect size (Brown, 1988).

We use R language (R Core Team, 2021) to run the tests and plots. A significance value (alpha) of 5% was used to check for variance between groups. The ANOVA statistics was performed after checking for the three statistical assumptions necessary for its application: normality of residuals, homogeneity of variances, and sample independence.

4 RESULTS AND DISCUSSION

In this section we present and discuss our results to check the performance of each group when comparing their duration and F0 means (BRA vs. NAT). We will also discuss which impact these prosodic differences may
cause in pronunciation and to oral communication, as well as how teachers can implement L2 prosody during pronunciation classes.

Table 2: Utterance mean values (m) and global mean from all of the utterances (mU) for normalized duration (Dur.) and F0 along the V-V units for both BRA and NAT groups; F statistic, and P-values for each utterance mean, and for the global mean all of the utterances.

<table>
<thead>
<tr>
<th>Utterance</th>
<th>Dur. z-score</th>
<th>F-value</th>
<th>P-value</th>
<th>F0 z-score</th>
<th>F-value</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>mU01</td>
<td>0.78</td>
<td>93.16</td>
<td>&lt;0.001</td>
<td>0.85</td>
<td>19.13</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>mU02</td>
<td>0.89</td>
<td>81.23</td>
<td>&lt;0.001</td>
<td>0.85</td>
<td>30.4</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>mU03</td>
<td>1.07</td>
<td>97.22</td>
<td>&lt;0.001</td>
<td>0.90</td>
<td>32.51</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>mU04</td>
<td>0.77</td>
<td>72.48</td>
<td>&lt;0.001</td>
<td>0.85</td>
<td>0.86</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>mU05</td>
<td>0.89</td>
<td>51.4</td>
<td>&lt;0.001</td>
<td>0.86</td>
<td>51.22</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>mU06</td>
<td>1.14</td>
<td>88.0</td>
<td>&lt;0.001</td>
<td>0.83</td>
<td>55.6</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>mU07</td>
<td>1.05</td>
<td>71.34</td>
<td>&lt;0.001</td>
<td>0.86</td>
<td>40.34</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>mU08</td>
<td>0.98</td>
<td>66.45</td>
<td>&lt;0.001</td>
<td>0.86</td>
<td>16.42</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>mU09</td>
<td>0.94</td>
<td>59.77</td>
<td>&lt;0.001</td>
<td>0.85</td>
<td>29.81</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>mU10</td>
<td>0.98</td>
<td>67.32</td>
<td>&lt;0.001</td>
<td>0.78</td>
<td>1.37</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>mU</td>
<td>0.95</td>
<td>64.5</td>
<td>&lt;0.001</td>
<td>0.85</td>
<td>21.01</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

Source: Personal collection.

Figure 4: Boxplot, mean values (larger black points), standard deviation (whiskers below the first quartile and above the third quartile of the boxes), individual values (shorter black points), and regression lines (in gray) for z-scored duration (panel A), and for z-scored F0 (panel B), for the L1 (native speech) and L2 (foreign speech) English speakers.

In Table 2, we may observe from the mean values for each utterance, as well as for the global mean of all of the utterances, that with the exception for the F0 mean of utterance 4, mU04= [F (1,18) = 0.86, p=.72], and utterance 10, mU10 = [F (1,18) = 1.37, p=.51], all of the values performed significantly different between both BRA and NAT groups. Figure 4(A) presents a significant variation...
between the factor GROUP regarding the duration of the V-V units, as well as for the F0 values as shown in Figure 4(B). Yet the global mean of the utterances for both duration – $mU = [F (1,118) = 64.5, p< .0001]$, and F0 – $mU = [F (1,118) = 21.01, p< .0001]$, results presented significant variation between groups.

As shown in Figure 4(A) and 4(B), and Table 2, the mean z-scored duration between utterances of BRA group = 0.95 - [F (1,18) = 0.86, p< .00, while for NAT= 0.61. As well as for duration, mean F0 z-score has significantly lower values for NAT group ($mF0= 0.59$), when compared to BRA group ($mF0= 0.85$). Another aspect to be highlighted is the variability (standard deviation represented by the whiskers in the plot) in duration, higher for BRA group due to L2 cognitive load in reading tasks) as showed in Figure 4A, and in F0, higher for NAT group – Figure 4B. Silva Jr. and Barbosa (2019, 2020) point to this F0 variability based on its slope. The F0 slope is perceptually related to liveliness and enthusiasm in the speech. In the case of a fairy tale or fable stories (our experiment), the F0 slope may reflect the role playing of good/bad characters.

From the results brought so far, it is possible, at least to some extent, to infer about the importance of L2 prosody for teaching pronunciation and, consequently, enhance oral communication:

1. aspects of pronunciation related to rhythm (long and irregular duration of V-V units due to L2 cognitive load), and intonation (little variation in the F0 contour) indicate difficulties for Brazilian speakers (of L2-English) regarding the use of L2 prosody. This fact can compromise the speaker's intelligibility in semantic aspects (the shift of lexical/phrasal stress and pitch accent may affect prosodic boundaries towards a morphosyntactic boundary, for example), leading to a pragmatic dimension of the discourse (misunderstandings during conversational turns);

2. as well as the study conducted by Silva Jr. and Barbosa (2019), the variability of F0 was significantly lower in the production of nonnative speakers. The literature reports for some decades (Roach, 1982), that there is a tendency to provide more attention to segmental information rather than the prosodic agenda. Thus, speakers can correctly produce phonetic categories in isolation, however, there may be a loss of comprehensibility in
higher domains of the speech due to, at best, putting aside L2 prosodic elements in teaching pronunciation context, as reported by Michaud and Reed (2015);

3. the z-scored duration of the V-V units reflects previous results in the literature (see LI, et al., 2018) from the point of view of variability between groups. As we can see in Figure 4(A), both duration and standard deviation values are significantly lower for L1-English speakers. The standard deviation (sd) for Brazilian productions is twice that of the native ones (sd BRA = 0.12; sd NAT = 0.06). Once we have a high-intermediate-English-speaking group (B2→C1), and, according to CEFR standards (see Methods), the speakers are quite close to an advanced level, they already know plenty of English lexicon.

   a) unknown lexicon can be initially produced, focusing on the segmental form (see Reed and Michaud, 2015, p. 456) and, thus, change the duration of acoustic parameters, such as a consistent increase of speech rate, and higher mean and standard deviation values for the pauses and phones that compound the V-V units.

   b) such strategies can affect prosodic domain and be one of the causes of longer syllables with high standard deviation as verified by Low (2015, p. 132) in productions of L2-English by speakers of Mandarin, Italian, French, Spanish as L1. Such as Silva Jr. and Barbosa (2019, 2020), Low’s findings may be linked to the rhythmic characteristics of the mentioned L1s. The author, however, used a phonological-based syllabic segmentation rather than phonetic syllables.

   Briefly recalling the results presented in Figure 4, the normalized mean duration for L2 English is higher and with greater variability compared to the L1 production. Likewise, the normalized F0 mean for L2 English is significantly higher although it performs little variability when compared to L1 English production. These results suggest that, although L2 segmental level may be produced without many difficulties if one considers the foreign group level of proficiency (B2→C1), prosody level showed to be not accurate as it is expected for this proficiency level. The relevance of the results herein presented is to point
out a path where the teaching of L2 prosody - by means of rhythm, intonation and other aspects - plays an important role on oral communication.

As suggested by Reed and Michaud (2015), we believe that the metacognitive aspects of L2 prosody ought to be considered in pronunciation classes no matter the proficiency level. The authors address that although high-qualified suprasegmental textbooks exist, with specific sections on intonation, these often focus on getting learners to produce the target intonation itself. Nevertheless, teaching intonation must include metacognitive awareness as well as productive and receptive skills if it is to be successful (p. 474).

For a definition of metacognitive process, Goh (2008, p. 23) states that: “Metacognition refers to listener awareness of the cognitive processes involved in comprehension, and the capacity to oversee, regulate, and direct these processes”. Reed and Michaud (2015) yet suggest that teachers must concentrate in giving instructions in the prosodic approach, in contrast to the grammatical approach where the teacher seeks to engage learners in mimicking the exaggerated prosody of the target L2, such as large pitch variations, high speech rate, among other aspects. There are great advantages on adopting L2 prosody in pronunciation classes. As noted by Gumperz (1982), intonation determines how a speaker’s message is understood, as well as L2 prosody may result in negative social evaluation if it is not produced properly.

The linear models defined by the present research are represented in the (linear) equations (2), for duration, and (3), for F0:

\[
\text{Duration} = 0.95 - 0.35 \times (\text{NAT GROUP}), \quad R^2 = 0.77 \tag{2}
\]

\[
F0 = 0.85 - 0.25 \times (\text{NAT GROUP}), \quad R^2 = 0.51 \tag{3}
\]

Since Eq. 2 and Eq. 3 show negative values for both Duration \((b = -0.35)\), and F0 \((b = -0.25)\) respectively, both models will perform descending slopes (gray regression lines in Figure 4). In other words, both z-scored duration and F0 go lower in values along the V-V units when we change from BRA to NAT
group production. ‘R²’ is the coefficient of determination, which represents the effect size, i.e., how much duration and F0 are explained in terms of variance.

In the present study, R² explains 77% (R² = 0.77) of the variation regarding the duration of the V-V units between both BRA and NAT groups. At this point, we propose that, in fact, it is likely that Brazilian L2 (English) speakers, even on their way to being advanced proficient speakers (see Pollit, 2019 based on CEFR) are liable to prosodic transfers from their L1 to the target-L2. As for the adjusted R² for F0 between the groups, the model explains 51% (R² = 0.51) of the variation. Furthermore,

Based on the R² herein presented, we conclude, on a preliminary basis, that both z-scored duration and F0 showed a strong effect when comparing L1 and L2 prosody. We may also account for a slight advantage for the duration-based model over the F0-based one. However, we emphasize that the statistical literature indicates that the coefficient of determination depends on the quantity of the samples. In addition, if R² is equal or greater than to 50%, with a sample number above 100 events (in this case, we had 120 tokens for the present study), it is possible to trust the model fit (Rietveld & Hout, 2005; Anderson, et al. (2005; Triola, 2014) From the statistical studies abovementioned, both of the presented models (Eq.1 and Eq. 2) are reliable in terms of z-scored duration on F0 from L1- and L2-English speech production.

5 FINAL REMARKS

In this study we proposed that L2 prosody should be present in pronunciation teaching to enhance oral communication by promoting greater intelligibility and comprehensibility in foreign speech. We understand and claim that it is through prosody – interacting to segments - that the speaker’s pronunciation must be shaped, since elements such as reduced vowels/consonants are observed throughout the utterances, such as those hereon evidenced.

The teaching of prosody in L2 classes based on prosodic-acoustic features, just like duration and F0 presented in the current paper is still a
challenge, which includes speech variability, different foreign accents, contexts, language complexity, among others. It is important to keep in mind that the use of such features by the teachers has a chance to succeed in terms of prosodic awareness. Furthermore, once the student is prosodically aware of the L2, s/he will create oral communication instances in different context that go beyond vocalic and consonantal productions.

In future studies, as an attempt to minimize prosodic problems and optimize the teaching of L2 prosodic dimension in pronunciation classes, we suggest the use of metacognitive strategies involving L2 stress, rhythm and intonation were used in order to obtain from learners a more accurate, intelligible and understandable pronunciation, taking into account processes that occur in the suprasegmental domain, however, without forgetting segmental aspects internal to the syllabic structure that directly affect L2 speech. We also intend to apply perceptual tests based on prosodic-acoustic parameters herein conducted, bearing in mind the interdependence of duration and F0 to the perception of L2 speech prosody.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the financial support from the National Council for Scientific and Technological Development - CNPq, grant no. 307010/2022-8 for the first author, and grant no. 302194/2019-3 for the second author. Finally, the authors acknowledge the valuable contributions, and kind collaboration of the participants in this research.
REFERENCES


