SPEECH RHYTHM OF ENGLISH AS L2: THE INFLUENCE OF DURATION AND F_0 ON FOREIGN ACCENT INVESTIGATION

O ritmo da fala em inglês como L2: a influência da duração e F_0 na investigação do sotaque estrangeiro

SILVA JR., Leônidas J. [1]*
BARBOSA, Plínio A. [2]

[1] University of Paraíba - University of Campinas
[2] University of Campinas

Abstract: *When speaking a foreign language (L2), non-native speakers' (NNS) speech contains some variable degree of foreign accent (FA) that is perceivable by the native speakers (NS) of that language based on the production of phonetic gestures characteristic of their mother tongue (L1), and that differ from those of the the foreign language (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, for example, are 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 acoustic correlates such as duration and fundamental frequency (F0) influence FA degree. 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 of analysis, phonetic data from 20 BP and 04 American speakers were collected. Next the data were segmented into diferent unit procedures for the purpose of carrying out acoustic, perceptual and statistical analysis. Results pointed out to a significant difference between L1 and L2 rhythms.*
Keywords: *speech rhythm; foreign accent; acoustic correlates; English as L2.*

Resumo: *Ao falar uma língua estrangeira (L2), a fala de falantes não nativos contém um grau variável de sotaque estrangeiro (SE) que é percebido pelos falantes nativos da língua em questão com base na produção de gestos fonéticos característicos de sua língua materna, e que diferem da língua estrangeira em termos de segmentos (vogais e consoantes) e prosódia (acento, ritmo e entonação). Alega-se que causas como neuro-plasticidade e tempo de residência, por exemplo, interferem na produção da L2. Este trabalho tem como objetivo analisar como o ritmo da fala do inglês como L2 é produzido pelos falantes do português brasileiro (PB) e como correlatos acústicos (duração e frequência fundamental (F0)), influenciam o grau de SE. Esta pesquisa é baseada em Barbosa (2006) para a determinação dinâmica do ritmo da fala, além de Ramus et. al. (1999) dentre outros, para escolha de métricas e procedimentos de segmentação. Quanto aos métodos de análise, foram coletados dados fonéticos de 20 falantes do PB e 04 falantes americanos. Em seguida, os dados foram segmentados em diferentes procedimentos unitários, com o objetivo de realizar análises acústicas, perceptuais e estatísticas. Os resultados apontaram uma diferença significante entre os ritmos L1 e L2.*
Palavras-chave: *ritmo da fala; sotaque estrangeiro; correlatos acústicos; ingles como L2.*

1 Introduction

When speaking a foreign language (L2), non-native speakers' (NNS) speech contains some variable degree of foreign accent (FA) that is perceivable by the native speakers (NS) of that certain language based on the production of phonetic gestures characteristic of their mother tongue (L1) and that differ from those of the L2 in terms of the segmental (vowels and consonants) and prosodic (stress, rhythm and intonation) features (see Flege, 1995).

In fact, the literature of L2 acquisition has sustained several probable causes to explain the maintenance of FA such as neurological maturation of NNS followed by decreased neuro-plasticity and neuromotricity during L2 sound production, inaccurate perception of L2 sounds and awkward phonetic input from an early age in addition to the amount of input, length of residence, aptitude, proficiency level, motivation, etc.

The diversity of causes above cited gives a great dimension of the complexity of FA, especially in the prosodic domain (see Flege op. cit.), that is, they occur at the suprasegmental levels and can substantially hinder NNS speech productions' intelligibility. Such prosodic

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

phenomena are laid from the syllable structure to higher constituent levels and signal 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, the current research aims to analyze how English/L2 is produced by Brazilian Portuguese (BP) speakers from (semi) spontaneous speech and how acoustic correlates such as duration and fundamental frequency (F0) influence FA degree.

2 Theoretical framework

Regarding speech rhythm, 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.

With respect to L2 speech rhythm studies, phonetic literature has laid on different mathematical parameters for its characterization. Nooteboom (1997), Ramus, et al. (1999), Grabe & Low (2002), Barbosa (2006, 2012) and others proposed the so-called "rhythm metrics." Silva Jr. and Barbosa (2019) coded these metrics and the ones established over the last 25 years of speech rhythm research in addition to new melodic parameters based on F0 centrality, dispersion and derivative.

With reference to the reliability of these metrics, Gut (2012) mentions that each metric generates different results in different studies. This inconsistency observed for the metrics is 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. To test multiple metrical possibilities, for the metrical parameters, Silva Jr and Barbosa (in press) segmented speech into phonetic syllable-sized units besides vowels and consonants suggested by the literature so far and for the F0 parameters, they segmented speech into chunks following syntactic and prosodic criteria.

3 Methods

We collected phonetic data from 20 BP speakers of English/L2 (experimental group) and four American English (AmE) native speakers (control group). For this study, two categorical proficiency levels of English for the BP speakers (10 subjects per group) were established: *high-intermediate* and *advanced,* as determined through the Oxford Online Placement Test (OOPT) (<https://www.oxfordenglishtesting.com/>).

Speakers were told that their task would be to read aloud (an Aesop's fable) text and this would be recorded. They were shown the text in advance to be familiarized with and to avoid anxiety while reading. Speakers were recorded at a radio station studio from a Boss Br 1600 Recorder, a sampling rate at 48 kHz and 16-bit quantization to ensure high quality of dataset that would be later used for acoustic analysis.

3.1 Acoustic, statistical and discriminant analysis

To perform the acoustic analysis, data were segmented into vowel (V), consonant (C) and phonetic syllable-sized (VV) units for the extraction of the metrics; sentences (S) and syntactic-prosodic larger units (chunks - CH) for the F0 parameters. For the metrics and melodic values we ran a script for Praat (*Metrics&AcousticsExtractor* - Silva Jr. and Barbosa, 2019). The script

extracts parameters such as: (*metric*: %V..., ΔC..., varcoC..., simple, raw and normalized pairwise variability indexes and rhythm ratio; *acoustic*: regular and derivative F0 peak, minimum, median, st.dev, skewness, rate, spectral emphasis and speech rate), in the case of the metrics for all units: V, C, VV, and, in the case of acoustic parameters in the domains of S and CH for each Sound/TextGrid pair, manually segmented into the above units, and generates an output file with all parameters. The segmentation and script application were run in Praat (Boersma and Weenink, 2019).

With the metric and acoustic measures returned from the script, statistical training of the models under Analysis of Covariance (ANCOVA) was done in which we extracted the significant parameters that did not fail to meet the three statistical conditions for the conventional ANOVA (normality of the residuals, homoscedasticity of variances, and independence of the samples). The metric significant parameters that showed robustness under analysis were: for VV: **Δ**, varco, VI, r-PVI; for C: r-PVI and YARD and the F0-related parameters: F0min, standard-deviation of F0deriv., mean of F0deriv. and speech rate.

Besides the acoustic analysis, a perception experiment with 10 AmE/L1 speakers was run to rank the degree of FA between the speakers' production from a six-point scale: the higher the grade, the higher the degree of FA. From the perception test results, we ran a linear discriminant analysis (LDA) in order to classify the speakers into one of the two clearly defined language groups (AmE or BP). All steps of the statistical analysis (for the metric/acoustic parameters and discriminant grades) were performed in R environment (R CORE TEAM, 2019).

4 Results

For the comparison between the two groups, the best metric parameters have a determination coefficient of 55% ($R^2 = 0.55$) and the best acoustic ones have a determination coefficient of about 60% ($R^2 \cong 0.6$). It was also observed that both consonant and phonetic syllable (especially the latter, not mentioned so far by L2 speech rhythm research) were the most reliable segments in determining and discriminating duration-related values between AmE/L1 and BP/L2 speech rhythm of English as seen in the boxplots in figure 1:

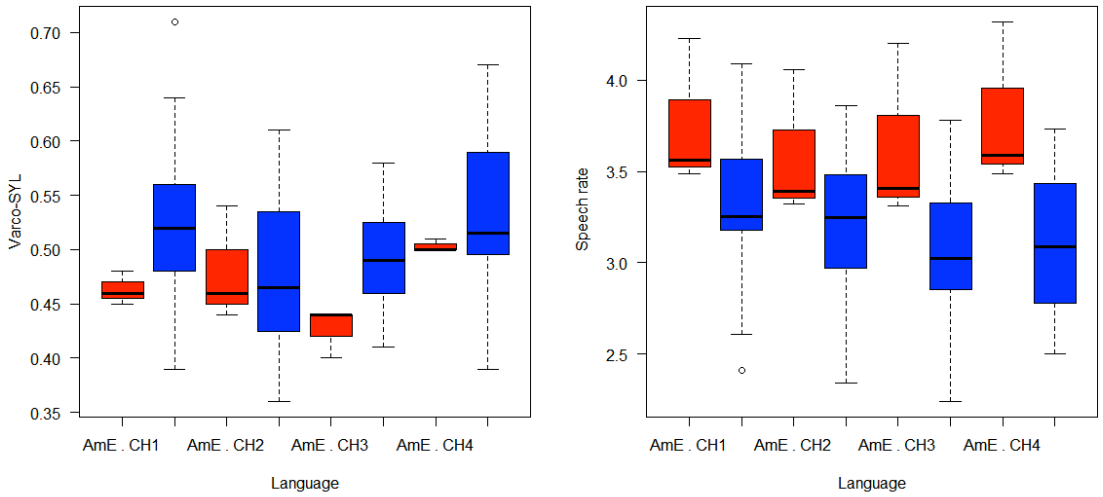

Figure 1: Boxplots of the *varco-syl* (left portion) and for *speech rate* (right portion) for the NS (red boxes) and BP speaker (blue boxes) productions.

It is also worth noting that there is a significant opposite variation in the F0 trajectory towards the phrasal stress between both languages; a left-to-right decaying shift (HL) by the AmE speaker and a left-to-right rising shift (LH) for the BP one as seen in figure 2. This brings up the discussion of how important it is to take into account F0 parameters, to say, F0min, standard-deviation of F0deriv., mean of F0deriv and the speech rate (the most important herein) as far as speech rhythm is concerned.

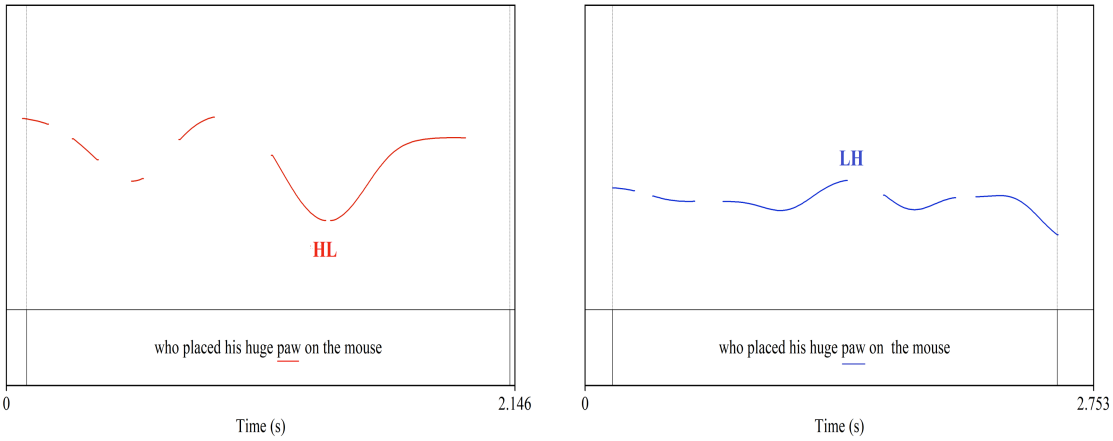

Figure 2: F0 curve and phrasal stress over time for the utterance: [*who placed his huge PAW on the mouse*]U spoken by a NS of AmE (red curve on the left portion) and by a BP speaker (blue curve on the right portion).

5 Conclusion

We may conclude from these observations that it is possible to establish a reliable acoustic and perceptually-oriented framework for the F0 parameters in the research of L2 speech rhythm rather than only the traditionally metric-based one.

REFERENCES

1. Barbosa, P. Panorama of Experimental Prosody Research. In: Proc. of the GSCP Workshop, Belo Horizonte Firenze University Press, 33-42, 2012.

2. Barbosa, P. *Incursões em torno do ritmo da Fala*. Campinas: Pontes Editores, 2006.

3. 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.

4. Barbosa, P.; Bailly, G. Characterisation of rhythmic patterns for text-to-speech synthesis, *Speech Communication* 15, 127-137, 1994.

5. Boersma, P.; Weenink, D. *Praat: doing phonetics by computer*. (Version 6.0). http://www.praat.org, 2019.

6. 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.

7. 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.

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

9. Nooteboom, S. The Prosody of Speech: Melody and Rhythm. In. *The handbook of phonetic sciences*, Utrecht, UtdallasEdu, 640–673, 1997.

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

11. Silva Jr., L; Barbosa, P. *Metrics&AcousticsExtractor* version 1.0. Script for Praat. 2019.

12. Silva Jr., L; Barbosa, P. *The Production and Perception of English Speech Rhythm as L1 and L2: investigating the degree of foreign accent*. (in preparation).