



Statistical modeling of prosodic contours of four speech acts in Brazilian Portuguese

Luma da Silva Miranda^{1,2}, João Antônio de Moraes^{2,3}, Albert Rilliard^{2,4}

¹Eötvös Loránd University, Hungary

²Federal University of Rio de Janeiro, Brazil

³CNPq, Brazil

⁴LIMSI, CNRS, Université Paris-Saclay, France

lumah.miranda@gmail.com, jamoraes3@gmail.com, albert.rilliard@limsi.fr

Abstract

In this paper, the prosody of four speech acts (assertions, echo questions, wh-questions and wh-exclamations) in Brazilian Portuguese was acoustically described and their variations analyzed using a non-linear statistical model. The aim of the paper is to formally present the relevant changes in prosody linked to these four types of speech acts. The sentence *Como você sabe* was recorded by ten speakers (5 male) from Rio de Janeiro that produced the four speech acts ten times, resulting in a total of 400 utterances. The corpus was analyzed in terms of F0, intensity and duration. Automatic extraction of acoustic measures was made and submitted to normalization procedures. A Generative Additive Mixed Model (GAMM) was applied to the acoustic measures. Results showed that the four speech acts presented significant differences in their F0 patterns, with four different types of contours differentiated either on the initial high F0 peak or on the final tonic syllable; the wh-questions were also performed with a higher mean pitch. Differences in the intensity of some syllables among speech acts were found, but not in their mean levels. No significant differences in duration were verified among the speech acts.

Index Terms: prosody, speech acts, Brazilian Portuguese.

1. Introduction

In Brazilian Portuguese (BP), the prosodic features of speech acts [1] other than assertions and yes-no questions are still under-described. For instance, the prosody of echo questions, wh-questions and wh-exclamations remain underexplored compared to assertions and yes-no questions. Regarding the pragmatic context of these speech acts [2], echo questions are a type of yes-no question produced to signal a lack of understanding of the preceding speech interaction that triggers a yes-or-no response, similarly to the neutral yes-no question, whereas assertions are produced to communicate a state of affairs. Wh-questions are information requests by the interlocutor, and wh-exclamations are produced to express the speaker's reaction about either an event or what is said by the interlocutor [3]. In two previous studies [4, 5], in which the same production dataset of the present study was perceptually analyzed, we verified that BP listeners identify the intonation patterns of these four speech acts based on their prosody.

In terms of speech production, several works in different languages have shown that prosody conveys the meaning of speech acts [6, 7, 8]. For instance, questions are spoken in a faster speech rate than statements in many languages [9]. The

findings of BP modal intonation [10] indicate that these four speech acts are encoded in specific fundamental frequency (F0) contours, carried either in the first or in the last stressed syllables of the utterance, where the phonological contrast between assertion, echo question, wh-question and wh-exclamation is performed. Assertions and echo questions in BP are distinguished by pitch accents on the nucleus [2], with assertion contours presenting a falling F0 nuclear configuration, and interrogative contours a rising one. Conversely, previous works in BP [3, 10, 11, 12] have shown that wh-questions and wh-exclamations present distinctive F0 contours on the pre-nucleus: wh-questions have a higher initial F0 peak than wh-exclamations. The nuclear region of both intonation patterns shows a similar falling F0 contour, having a distinct F0 inclination across Brazilian varieties, which can be steeper [10, 12] or smoother [3, 11].

Although the intonation patterns of these four speech acts have been described in BP, the use of other prosodic parameters (intensity and duration) still requires attention. A comparative study analyzing three BP varieties [13] showed the role of these two parameters for declaratives and yes-no questions: whereas the intensity of the last stressed syllable was higher in the interrogative contour compared to the declarative for all varieties, the last stressed syllable of the interrogative contour was shorter than the declarative in Salvador variety and longer in Rio de Janeiro and Fortaleza varieties. Apart from the results of this study, these parameters in BP have not been explored for echo questions, wh-questions and wh-exclamations, to the best of our knowledge.

This study investigates the phonetic information distinguishing these four speech acts in BP: assertions, echo questions, wh-questions and wh-exclamations. Based on the literature, these speech acts are expected to present recurrent differences in their F0 configurations and also for the other prosodic parameters. A production experiment was conducted to quantitatively describe these acoustic correlates, the results of which are reported here.

2. Method

2.1. Structure of the corpus

The corpus is based on one sentence starting with a WH-word: *Como você sabe*. Ten speakers produced it with distinctive prosodic patterns, which change the sentence's meaning, for assertion (*As you know.*), echo question (*"As you know?"*, meaning *"Was that what you said?"*), wh-question (*How do you know?*) or wh-exclamation (*How clever you are!*).

2.2. Participants

The ten participants (five male) were speakers of the Rio de Janeiro dialect and the mean age was 28.5 years old. All of them were university-educated.

2.3. Recording procedure

Recordings were made with a Zoom H4 at UFRJ’s Laboratory of Acoustic Phonetics in an acoustically isolated room. The pragmatic contexts of the four speech acts were presented to the speakers. They were asked to read the contexts and imagine themselves in these situations prior to producing the speech acts ten times (the corpus is composed of 400 utterances). During the recording sessions, the experimenter instructed the speech act to be produced by the speaker.

3. Acoustic analysis

3.1. Extraction of acoustic measures

With the use of Praat [14], the 400 utterances were manually segmented at the phone level. Next, an estimate of the fundamental frequency (F0) was made with the Praat default algorithm. A Praat script [15] was used to extract from each sound regularly spaced F0 estimates values in order to produce a time-normalized F0 contour: from each vowel, ten F0 samples were collected and five from each consonant. The same script was later modified to extract time-normalized intensity. This way, temporally normalized curves of intonation and intensity were obtained. The script SG-detector [16] was used to extract the phonemes’ duration, transforming their raw values in smoothed Z-score values to remove intrinsic and cointrinsic duration differences. The information of segmental duration was based on the measurement of six V-to-V units [17].

The statistical analysis of the set of F0 (34.000 measures), intensity (34.000 measures) and duration (2.400 measures) was made with the R software [18]. The fundamental frequency was expressed in semitones (ST) with reference to 1 Hz [19]. The intensity was measured with A-weighted decibel (dB) [20] and corrected for the mean observed recording level for each speaker (i.e., the intensity mean observed over all sentences produced by the same speakers was removed from raw values), so as to compensate for unknown differences of recording levels. This prevents the comparison of potential differences between speakers but allows for the measuring of differences of intensity among the four types of utterances. The duration was analyzed in V-to-V units expressed in smoothed Z-score. The results of this analysis are presented in section 4.

3.2. Statistical modeling: GAMM

For each dependent prosodic variable, five independent factors were considered: the normalized time of the sentence (x-axis of the graphics of Figures 1 to 3), the modality of the sentence (four levels: assertion, echo question, wh-question, wh-exclamation), the speaker’s gender (two levels: male and female), the repetition (ten levels) and randomly selected individual speakers (ten levels). The dependent variables of this model are the F0 measures, the intensity values in dBA and the V-to-V duration values. A Generalized Additive Mixed Model (GAMM—analyzed with the libraries “mgcv” and “gamm4” of the R software, [21, 22]) was applied to the three measures. This type of regression allows the modeling of

nonlinear variations along time which is typical of prosodic changes. The GAMM models used spline functions to approximate the discrete acoustic measures in the form of a continuous curve (see [23] for a similar use of spline for intonation stylization). A specific model was fitted for each prosodic parameter, described hereafter.

4. Results

4.1. Fundamental frequency – F0

F0 was fitted using equation (1), with two fixed intercepts for the mean levels of each speech act and gender and a smooth curve for each speech act, plus random smooths to model the individual difference among speakers, for each type of speech act (cf. details in [22]). The high autocorrelation in residuals (0.97) was dealt with by adding an autoregressive AR (1) model and their heavy tailed distribution by using a “scaled-t” distribution (following [22]).

$$F0 \sim \text{Act} + \text{Gender} + s(\text{Time}, k=12) + s(\text{Time}, \text{by}=\text{Act}, k=12) + s(\text{Time}, \text{Speaker}, \text{by}=\text{Act}, \text{bs}=\text{"fs"}, m=1) \quad (1)$$

The speaker’s gender had a significant (and foreseen) effect on the speakers’ mean F0 level (men having voices 10 ST lower than females). The speech act factor also had a significant effect, linked to a higher mean F0 used for wh-questions (5.6 ST higher than assertion—cf. Table 1 for details). The smooth curves fitted for each speech act also presented significant deviation from the assertion’s smooth, which shows differences in the F0 contours that are observed for each speech act. These contours are plotted on Figure 1 (the absolute level used in Fig. 1 is the one estimated for the female group). The details for the smooth functions in the GAMM are presented in Table 2.

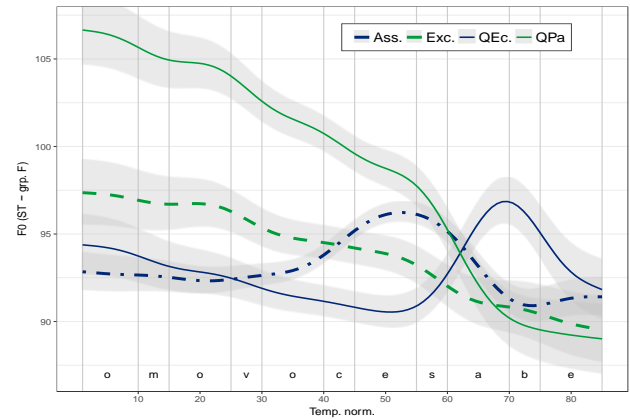


Figure 1: F0 contour modeled by the GAMM model for each speech act: assertion (Ass.), echo question (QEc.), wh-exclamation (Exc.) and wh-question (QPa); all speakers together.

Table 1: Parametric coefficients of the GAMM model adjusted over the F0 values for the four speech acts.

	Estimate	Std. Error	t value	Pr(> t)
Intercept	93.13	0.76	122.4	< 0.05
Act (Ex.)	0.79	1.15	0.7	0.49
Act (EcQ.)	-0.25	1.02	-0.2	0.81
Act (WhQ.)	5.64	1.04	5.4	< 0.05
Gender (Male)	-10.56	0.71	-14.8	< 0.05

Table 2: Smooth function terms for the four speech acts for F0. The first line shows the reference level (assertion), lines two to four represent difference smooth functions comparing other levels to the reference. The four final lines show the random-effects structure. edf: estimated degrees of freedom. The p-value assesses if the smooth significantly differs from straight line.

Smooth functions	edf	F	p-value
s(sample): Act Ass.	10.7	35.0	< 0.05
s(sample): Act Exc.	10.4	20.1	< 0.05
s(sample): Act QEc.	10.9	81.9	< 0.05
s(sample): Act QPa.	10.2	20.7	< 0.05
s(sample, Speaker): Act Ass.	79.9	31.7	< 0.05
s(sample, Speaker): Act Exc.	81.4	67.7	< 0.05
s(sample, Speaker): Act QEc.	80.4	46.1	< 0.05
s(sample, Speaker): Act QPa.	81.7	66.7	< 0.05

In Fig. 1, it can be observed that: (i) the assertion contour is characterized by a rising F0 movement over the final prestressed syllable “cê” and a falling F0 over the last stressed syllable “sa” (confirming the literature descriptions [2, 10]); (ii) the echo question shows an F0 fall down to the final pretonic “cê” followed by rising along the final tonic “sa” (also confirming earlier descriptions in the literature [2]); (iii) wh-exclamation starts at a medial F0 that gradually falls down until the last tonic “sa”, marked by an F0 fall (see [3, 10, 11, 12]); and (iv) the wh-question starts at a high F0 level, then falls down to the tonic syllable “sa” (see [10, 11, 12]), which is marked by a 5 ST steep fall.

In summary, the four speech acts presented different F0 configurations, mainly in the prenuclear region for the wh-questions and wh-exclamations, and in the nuclear region, for assertions and echo questions.

4.2. Intensity

Intensity was fitted using equation (2), which is similar to (1), but for the smoothing details (smoother curves are required so to avoid fitting phonemic details). As for F0, the autocorrelation in residuals (0.8) was dealt with by adding an autoregressive AR (1) model and their heavy tail by using a “scaled-t” distribution.

$$\text{Intensity} \sim \text{Act} + \text{Gender} + s(\text{Time}, k=30) + s(\text{Time}, \text{by}=\text{Act}, k=15) + s(\text{Time}, \text{Speaker}, \text{by}=\text{Act}, \text{bs}=\text{"fs"}, m=1) \quad (2)$$

As expected, speaker’s gender factor had no effect on mean intensity level (as measures were normalized by speaker), nor did the speech acts (cf. Table 3 for details). The smooth curves fitted for each speech act present significant deviations from the assertion’s smooth, showing intensity differences that are observed between these speech acts. The intensity smooths are plotted on Fig. 2 and the details are presented in Table 4.

Table 3: Parametric coefficients of the GAMM model adjusted over intensity values for the four speech acts.

	Estimate	Std. Error	t value	Pr(> t)
Intercept	-4.22	0.93	-4.5	< 0.05
Act (Exc.)	1.00	1.31	0.8	0.44
Act (EcQ.)	0.96	1.13	0.9	0.39
Act (WhQ.)	0.86	1.16	0.7	0.45
Gender (Male)	-0.16	0.48	-0.3	0.73

Table 4: Smooth function terms for the four speech acts for intensity. First line shows the reference level, lines two to four the difference smooth functions. The four final lines show the random-effects structure. edf: estimated degrees of freedom. The p-value assesses if the smooth significantly differs from straight line.

Smooth functions	edf	F	p-value
s(Sample): Act Ass.	28.9	524.1	< 0.05
s(sample): Act Exc.	10.6	3.7	< 0.05
s(sample): Act QEc.	12.3	11.4	< 0.05
s(sample): Act QPa.	12.8	7.1	< 0.05
s(sample, Speaker): Act Ass.	70.7	7.8	< 0.05
s(sample, Speaker): Act Exc.	73.3	12.6	< 0.05
s(sample, Speaker): Act QEc.	68.1	6.0	< 0.05
s(sample, Speaker): Act QPa.	69.0	6.6	< 0.05

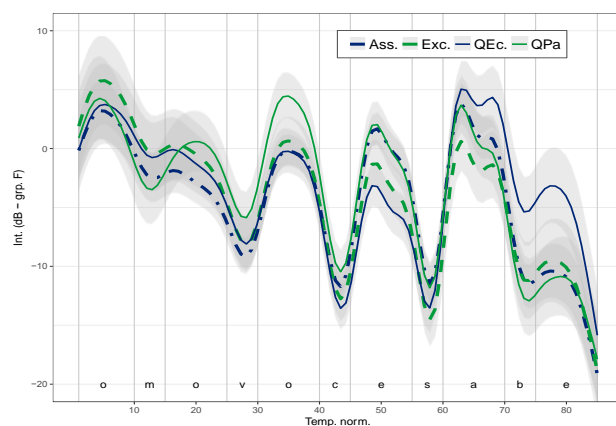


Figure 2: Intensity contour modeled by the GAMM model for each speech act: assertion (Ass.), echo question (QEc.), wh-exclamation (Exc.) and wh-question (QPa).

Fig. 2 shows the clear increase of intensity for each vowel (the bumps on the curves) and the declination line across the sentence. The same figure also shows differences among speech acts: (i) for assertions, the two strongest syllables are the final tonic “sa” and the pretonic “cê”, if one disregards the declination line; (ii) a drop of intensity over the final pretonic “cê” followed by an increased intensity on the tonic “sa” was observed for echo questions, followed by a comparatively high level for the posttonic “be”; (iii) wh-exclamations start with a highest intensity level (compared to assertions) for the two first syllables “co” and “mo”, then follow the assertion pattern; and (iv) wh-questions start at the assertion level and reach a peak on syllables 2 “mo” and 3 “vo”, then follow the assertion pattern.

In sum, the intensity patterns for the wh-questions and wh-exclamations differentiate the prenuclear region of the contours, while for assertions and echo-questions, it is the nuclear region that bears the main differences.

4.3. Duration

Duration was fitted using equation (3), with two fixed factors for speech act and gender, plus smooth of the V-to-V unit durations over its position in the sentence, for each speech act, and random smooth for each speaker over act. Simpler smooth was used as only a few duration points are considered for each sentence. No autocorrelation was observed in the residuals,

but their heavy tail was dealt with using a “scaled-t” distribution.

$$\text{Duration} \sim \text{Act} + \text{Gender} + s(\text{Position}, k=6) + s(\text{Position}, \text{by}=\text{Act}, \text{bs}=\text{"tp"}, k=6) + s(\text{Position}, \text{Speaker}, \text{by}=\text{Act}, \text{bs}=\text{"fs"}, m=1, k=6) \quad (3)$$

Table 5 shows that there is a significant gender difference for V-to-V units’ duration: males have a 0.9 z-score shorter units, compared to females. No effect of speech act was observed on duration. The smooth curves fitted for each speech act also showed no significant deviations from the assertion’s smooth. There is thus no compelling evidence that the duration of each V-to-V unit is used to distinguish between these speech acts. The duration smooths are plotted on Fig. 3 and the details for being presented in Table 6.

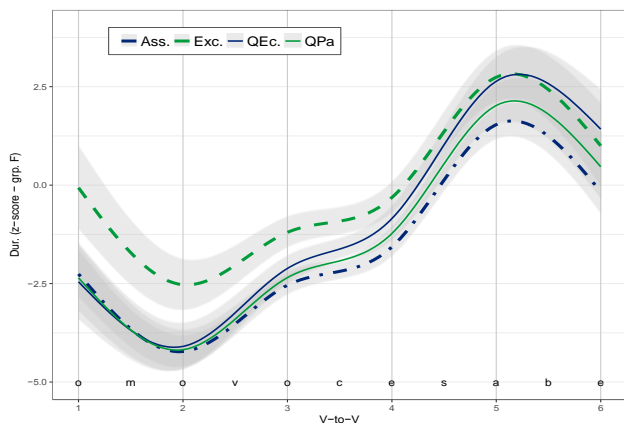


Figure 3: Duration contour modeled by the GAMM model for each speech act: assertion (Ass.), echo question (QEc.), wh-exclamation (Exc.) and wh-question (QP).

Table 5: Parametric coefficients of the GAMM model adjusted over V-to-V durations for the four speech acts.

	Estimate	Std. Error	t value	Pr(> t)
Intercept	-1.54	0.41	-3.8	< 0.05
Act (Exc.)	1.47	0.79	1.9	0.06
Act (EcQ.)	0.63	0.57	1.1	0.27
Act (WhQ.)	0.27	0.52	0.5	0.61
Gender (Male)	-0.90	0.36	-2.5	<0.05

Table 6: Smooth function terms for the four speech acts for duration. First line shows the reference level, lines two to four the difference smooth functions. The four final lines show the random-effects structure. edf: estimated degrees of freedom. The p-value assesses if the smooth significantly differs from straight line.

Smooth functions	edf	F	p-value
s(position): Act Ass.	4.98	149.3	< 0.05
s(position): Act Exc.	2.05	1.3	0.31
s(position): Act QEc.	1.61	2.6	0.13
s(position): Act QPa.	1.00	0.8	0.38
s(position, Speaker): Act Ass.	50.66	20.6	< 0.05
s(position, Speaker): Act Exc.	53.50	58.7	< 0.05
s(position, Speaker): Act QEc.	53.02	28.8	< 0.05
s(position, Speaker): Act QPa.	51.59	27.0	< 0.05

Fig. 3 shows that the duration units are shorter in the beginning of the sentences and shows a progressive temporal lengthening until the last tonic unit. This shape is similar across speech acts. There is thus no differing strategy

involving duration. Meanwhile, the curve of wh-exclamation presents longer units for the three first units, when combining the mean levels of this speech act and the slight difference in the smooth. One may thus suppose that there could be relatively lengthened syllables for the first part of this speech act, which could be coherent with the higher intensity levels found in these syllables, with this difference being relatively small as compared with differences observed for other parameters.

In summary, the four speech acts did not show significant distinction in relation to the duration patterns verified in our production data.

5. Conclusion

The aim of this study was to describe the prosodic parameters distinguishing four speech acts (assertions, echo questions, wh-questions and wh-exclamations) in Brazilian Portuguese. The description of the acoustic parameters of F0, intensity and duration provided by the GAMM models showed that the four speech acts are performed with systematic differences in their melodic contours and that these differences are supported by coherent increases of intensity in critical vowels. Note that the model included a random factor for each individual speaker, but our data did not allow testing a potential effect of age.

The main conclusion of this modeling study thus supports the descriptions found in the literature. The difference between assertions and echo questions is in the nuclear region of the contours, with the F0 configurations of the final prestressed and stressed syllables having a distinctive role [2]. Additionally, the wh-questions and wh-exclamations presented a falling F0 configuration throughout the contour [3, 10, 11, 12], with differences in the initial F0 peak, which is higher in the wh-questions, and in the inclination of the nuclear F0 fall, which is steeper (5 ST) in the wh-questions [10, 12]. The differences observed for intensity between the assertive and interrogative contours were already reported in another study for the Rio de Janeiro dialect [13]. We thus confirm that the last pretonic of the nucleus presents a higher intensity level in assertive utterances compared to the interrogative ones, while the intensity on the nucleus’ tonic and posttonic syllables is higher for interrogatives. This may be described as a later F0 peak for interrogatives, accompanied by changes in intensity.

The results, on the one hand, confirm our hypothesis regarding the distinctive intonation contour for each speech act; on the other hand, they partially confirm the second hypothesis that expected the intensity and duration parameters to also distinguish these four speech acts, since the duration parameter did not show compelling differences, but did show a slight lengthening for initial syllables of wh-exclamations. We conclude that the combination of the F0 movements along with the intensity patterns are the acoustic parameters that best distinguish the four speech acts and that systematic variations could be observed and related to these speech acts, over ten speakers, and independently of individual strategies.

6. Acknowledgments

This research has been funded by the scholarship 88882.331896/2015-01 (awarded by the Brazilian Federal Agency for Support and Evaluation of Graduate Education–CAPES).

7. References

- [1] J. R. Searle, *Speech Acts*. Cambridge: Cambridge University Press, 1969.
- [2] S. Frota, M. Cruz, F. R. F. Svartman, G. Collischonn, A. Fonseca, C. Serra, P. Oliveira, M. Vigarito. “Intonational variation in Portuguese: European and Brazilian varieties”, in S. Frota, P. Prieto (Org.), *Intonation in Romance*. Oxford: Oxford University Press, pp. 235-283, 2015.
- [3] K. Zendron da Cunha, *Sentenças exclamativas em português brasileiro: um estudo experimental de interface*, PhD dissertation, Santa Catarina, Florianópolis, 2016, retrieved from: http://fonapli.paginas.ufsc.br/files/2017/06/Tese_KarinaZendrondaCunha1.pdf.
- [4] L. Miranda, M. Swerts, J. Moraes, A. Rilliard, “The role of the auditory and visual modalities in the perceptual identification of Brazilian Portuguese statements and echo questions”, *Language and Speech*, 2020. <https://doi.org/10.1177/0023830919898886>
- [5] L. Miranda, J. Moraes, A. Rilliard, “Audiovisual perception of wh-questions and wh-exclamations in Brazilian Portuguese”, in *ICPHS 2019 – 19th International Congress of Phonetic Sciences, August 5-9, Melbourne, Australia, 2019*, pp. 2941–2945.
- [6] N. Hellbernd, D. Sammler, “Prosody conveys speaker’s intentions: Acoustic cues for speech act perception”, *Journal of Memory and Language*, vol. 88, pp. 70-86, 2016.
- [7] J. Moraes, A. Rilliard, “Describing the intonation of speech acts in Brazilian Portuguese: Methodological aspects”. In: I. Feldhausen, J. Fliessbach, M. M. Vanrell (Eds.). *Methods in prosody: A Romance language perspective* (Studies in Laboratory Phonology 4). Berlin: Language Science Press, pp. 229-262, 2018.
- [8] R. Tommaso, H. Mello, A. Jesus, L. Deus, “Uma aplicação da teoria da Língua em Ato ao português do Brasil”, *Revista de Estudos da Linguagem*, v. 15, n. 2, pp. 147-166, 2007.
- [9] F. Torreira, E. Valtersson, “Phonetic and visual cues to questionhood in French”, *Phonetica*, 72, pp. 20-42, 2015.
- [10] J. Moraes. “The pitch accents in Brazilian Portuguese: Analysis by synthesis”, in *SPEECH PROSODY 2008 – 4th Conference on Speech Prosody, May 6-9, Campinas, Brazil, Proceedings*, 2008, pp. 389-397.
- [11] J. Oliveira, V. Pacheco, M. Oliveira, “Análise perceptual das frases exclamativas e interrogativas realizadas por falantes de Vitória da Conquista/BA”, *Signum: Estudos Linguísticos*, vol. 17, no. 2, pp. 354-388, 2014.
- [12] L. S. Miranda, *Análise da entoação do português do Brasil segundo o modelo IPO*, Master thesis, Rio de Janeiro, Rio de Janeiro, 2015, retrieved from: <http://www.posvernaculas.letas.ufrj.br/images/Posvernaculas/3-mestrado/dissertacoes/2015/16-MirandaLS.pdf>.
- [13] L. R. Couto, C. G. Silva, L. S. Miranda, “Prosódia dos enunciados declarativos e interrogativos totais nas variedades de Salvador, Fortaleza e Rio de Janeiro”, *Revista de Estudos da Linguagem*, vol. 25, no. 3, pp. 1105–1142, 2017.
- [14] P. Boersma, D. Weenink, *Praat: doing phonetics by computer*, 2016, [Computer program], (Version 5.1.05), URL <<http://www.praat.org/>>.
- [15] P. Arantes, “Time-normalization of fundamental frequency contours: a hands-on tutorial”, in A. Meireles (Org.), *Courses on Speech Prosody*. Newcastle upon Tyne: Cambridge Scholars Publishing, pp. 98-123, 2015.
- [16] P. A. Barbosa, “Semi-automatic and automatic tools for generating prosodic descriptors for prosody research”, in *TRASP 2013 – 1st Tools and Resources for the Analysis of Speech Prosody, August 30, Aix-en-Provence, France, Proceedings*, 2013, pp. 86-89.
- [17] P. A. Barbosa, *Incurções em Torno do Ritmo da Fala*. Campinas: Pontes, 2006.
- [18] R Core Team. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria, 2018, URL <<https://www.R-project.org/>>.
- [19] F. Nolan, “Intonational equivalence: An experimental evaluation of pitch scales”, in *ICPhS 2003 – 15th International Congress of Phonetic Sciences, August 3-9, Barcelona, Spain, Proceedings*, 2003, pp. 771-774.
- [20] J. S. Liénard, C. Barras, “Fine-grain voice strength estimation from vowel spectral cues”, in *INTERSPEECH 2013 – 14th Annual Conference of the International Speech Communication Association, August 25-29, Lyon, France, Proceedings*, 2013, pp. 128–132.
- [21] S. N. Wood, *Generalized Additive Models: an introduction with R*, Chapman and Hall/CRC, 2017.
- [22] M. Wieling, “Analyzing dynamic phonetic data using generalized additive mixed modeling: a tutorial focusing on articulatory differences between L1 and L2 speakers of English”, *Journal of Phonetics*, vol. 70, pp. 86-116, 2018.
- [23] D. Hirst, R. Espesser, “Automatic modelling of fundamental frequency using a quadratic spline function”, *Travaux de l’Institut de Phonétique d’Aix*, vol. 15, pp. 75-85, 1993.