Macro- and Micro-prosodic Changes in the Duration of Brazilian Portuguese Vowels

João Antônio de Moraes[1,2] Albert Rilliard[1,3]

[1] Federal University of Rio de Janeiro, Brazil; [2] CNPq; [3] Université Paris Saclay, CNRS, LISN, France
jamoraes3@gmail.com; albert.rilliard@lisn.upsaclay.fr

ABSTRACT

This accounts for the effect of intrinsic segmental characteristics on the duration of Brazilian Portuguese oral vowels, compared to durational changes that can be attributed to macro-prosodic features, such as stress location, sentence modality, and syntactic structure. A multi-speakers corpus containing logatomes with systematic variations of vowel quality (seven vowels) embedded in sentences with controlled characteristics was recorded. A semi-automatic procedure allowed to segment and measure the duration of 3036 vowels; their duration was fitted to a series of factors, representing the controlled micro- and macro-prosodic characteristics. This modeling process showed the importance of micro-prosodic changes mainly in the case of stressed vowels (explained by jaw aperture and tongue posteriorization), which decreases with the vocalic reduction on prestress and post-stress positions, depending on macro-prosodic factors. The main macro-prosodic effects are linked to stress, the syllable position, final lengthening, and modality.

Keywords: Vocalic duration, Micro-prosody, Macro-prosody, Brazilian Portuguese.

1. INTRODUCTION

Segmental effects on the fundamental frequency (F0), intensity, and duration are referred to as micro-prosodic effects and may influence the description of macro-prosodic changes; thus the choice often made by prosodists to work on lab speech based on controlled sentences. This paper aims to describe quantitatively the variations induced in the duration of Brazilian Portuguese (BP) vowels by both micro-prosodic factors (i.e., the segmental, intrinsic characteristics of these vowels) compared to macro-prosodic ones (stress position in the word, the word's place in the sentence and the sentence modality).

The vocalic system of BP has seven oral vowels, /i e ɛ a ɔ o u/, that form minimal pairs at stress positions, but it reduces to five opposable qualities at pre-tonic positions with the neutralization of the middle vowels /e/ ~ /ɛ/ and /ɔ/ ~ /o/ [1], although the seven qualities are still produced mainly due to vocalic harmony processes [2]. Post-tonic positions show only three qualities in the word-final position, /i a u/, but may have four or even five qualities (/i e a o u/) in non-final post-tonic syllables (e.g., pen-ultimate syllables of proparoxytonic words) [3].

The effect of vowel quality on segmental duration has been evidenced for Dutch or English [4], [5], or as part of a larger set of factors (e.g., [6]), as an intrinsic duration factor [7], with jaw aperture being the main explication to intrinsic duration changes. Several works have dealt with vocalic duration for Brazilian Portuguese. [8] measured the duration of BP vowels of ten speakers from Rio de Janeiro at different sentence positions and confirmed a lengthening linked to aperture until mid-open vowels – /a/ being shorter than /ɛ/ and /ɔ/; differences along the anteroposterior dimension were observed, but not consistently. Similar results were replicated for another BP dialect, São Paulo [9]. [10] observed the duration of BP oral vowels in 20 speakers from São Paulo for stress position of disyllabic paroxytones. Differing from the two previous studies, they showed an increased duration linked to aperture even for the /a/; they also observed some effect on the duration of the front-back articulatory dimension.

The present article targets the duration of BP oral vowels and is part of a wider project aiming at describing micro-prosodic effects (intrinsic and co-intrinsic) in relation to macro-prosodic changes; to that aim, known factors of variation were controlled while creating the corpus. After presenting the corpus construction and recording, the output of the fit of syllable durations to the controlled factors is given and discussed.

2. METHODS

2.1. Corpus construction

The corpus is based on trisyllabic $CV_1CV_2CV_3$ logatomes (with C being /p/, and V_x on the seven oral vowels of BP /i, e, ɛ, a, ɔ, o, u/) inserted in the carrier sentence "*Ele dirá X [de novo], mas parecerá Y [de novo]*" ("*He will say X [again], but it will sound like Y [again]*"), where X and Y are two different logatomes, and [*de novo*] an optional extension. The sentence may or not propose an extension of the nominal group containing the logatome, and it was produced with two modalities – as an assertion or an interrogation.

The logatomes had the stress at each possible position (oxytone, paroxytone, proparoxytone). They were constructed with the same vowel at each position (thus, 21 logatomes are considered here: 7 vowels x 3 stress positions), with a limitation of the vowel inventory at unstressed syllables (see introduction).

2.2. Recording procedure

Six speakers (3 females and 3 males, mean age 46 years) were recorded in a sound attenuated room, using an omnidirectional measurement DBX RTA PHM919 microphone placed 30cm from their mouth, plugged into a Zoom H5 recorder, with the recording level calibrated by a Brüel & Kjær 4230 sound level calibrator. Speakers were presented with the sentences in random order via a computer interface and had to read them once – unless the experimenter judged the reading unsatisfactory.

2.3. Segmentation and measurements

A forced alignment procedure was applied to the recorded sentences based on their orthographic transcription supplied to the Montreal Forced Aligner [11]. Ad hoc phonetization of logatomes was provided to the program. The segmentation was then hand-corrected using Praat [12] by two experienced phoneticians. Considering two sentence modalities, two logatome positions in the sentence, the eventual extension, three vowels per each of 21 logatomes, and six speakers, it ended with 3036 individual vowels. From this dataset, a Praat script extracted the duration of each segment and other measures not presented here.

2.4. Statistical analysis

The following factors were controlled: the *Logatome position* (**LogP**) – at the syntagm end or not (Boundary or No boundary); the syntagm was at the initial part of the sentence or the end (*Syntagm position* (**SynP**): initial or final); the sentences were produced with a given *Modality* (**Mod**: Assertive or Interrogative). The logatomes also bear a specific *Stress pattern* (**Stress**: Oxitonic, Paroxitonic, Proparoxitonic), and its syllables are at three positions (*Syllable position* (**SylP**): Antepenultimate, penultimate, and ultimate). These five controlled factors are linked with macro-prosodic characteristics.

Then, the set of phonological vowels constitutes the micro-prosodic controlled factor (Vowel quality – **VowQ**). As existing vowels in PB vary according to the syllable position and the word tonal pattern (with syllables bearing the stress or being at pre- or post-stressed positions), different linear mixed models were fitted to the data. A first model was fitted on the cardinal extreme vowels /i a u/, found in all possible levels of the controlled factors (even if reduced at post-stressed positions) to test the effect of all macro-prosodic factors. Then, different mixed linear models were fitted on subsets of syllables, according to their position relative to stress (which determines the number of possible vocalic timbres): one model was made on stressed syllables (7 qualities), one on prestressed vowels (7 qualities with 5 having phonological oppositions), one on post-stress syllables at word-final position (3 qualities), and one on non-final post-stress syllables of proparoxytones (5 qualities).

All models were fitted using the same process (following [13] and using the lme4 library [14] of the R software [15]): we started by fitting a maximal model, with, as fixed factors, the micro-prosodic factor (having a varying number of levels as a function of the selected vocalic qualities, from 3 to 7) and the available macro-prosodic factors, with all interactions, and a random structure with an intercept for the six speakers (**Spk**) and for each produced sentence (**Snt**) nested in the speakers. This maximal model was simplified until reaching a minimal model adequate for describing the change in vocalic duration (**Dur**, in ms), used as the dependent variable of these models. The following section describes these different models and the main effects observed in each case; ANOVA tables are presented in Tables 1 & 2 for the first two models, with the effect size estimated using the partial ω^2 given by the *effectsize* library [6].

3. RESULTS

3.1. Models of duration

The exact formulas of the minimal models are given in the endnotes for the sake of clarity.

3.1.1. Duration of cardinal vowels

The minimal model[1] was based on all main fixed factors, implied in two four-way interactions linked to macro-prosodic effects (*Stress:SylP: SynP:LogP* and *SylP:SynP:LogP:Mod*), and three-way and two-way interactions mixing micro- and macro-prosodic factors (*VowQ:SynP* and *VowQ:Stress:SylP*). The random structure kept only the *Spk* effect. This model explained about 81% of the variance, with fixed factors responsible for 75% ($R^2_{marginal} = 0.75$; $R^2_{conditional} = 0.81$). From this complex structure, most of the changes in duration are explained by the *Stress*:*SylP* interaction, i.e., if the syllable bears or not the stress. In Figure 1, the durations of stressed vowels are systematically longer than the unstressed ones and above 100ms; most unstressed syllables have a duration between 50 and 75ms. Factors that induce lengthening are syllable position at word end (ultimate) when a syntactic boundary occurs and syntagm at the sentence-final position.

The *SylP:SynP:LogP:Mod* interaction is essentially linked to a further lengthening of sentence-final vowels (i.e., ultimate syllable of the word, for logatome at syntagm end and syntagm at sentence end), for interrogative modality compared to assertive.

The micro-prosodic effect showed longer /a/ and shorter /i/, with /u/ in between (thus an effect of vowels height and backness dimensions) – but this figure is affected by macro-prosodic characteristics: this holds for the longest vowels (i.e., the stressed ones). Conversely, most unstressed vowels, with centered qualities, showed few to no micro-prosodic differences (see latter models for details).

	χ^2	df	p	ω^2
VowQ	85.2	2	0.000	0.06
Stress	73.7	2	0.000	0.06
SylP	184.6	2	0.000	0.13
LogP	319.3	1	0.000	0.21
VowQ: SynP	6.1	2	0.048	0.00
SynP:SylP	57.7	2	0.000	0.04
Stress:SylP	3994.6	4	0.000	0.76
SynP:LogP	4.2	1	0.041	0.00
SylP:LogP	149.8	2	0.000	0.11
Mod:LogP	6.2	1	0.013	0.00
VowQ:Stress:SylP	35.0	8	0.000	0.02
SynP:Stress:SylP	12.4	4	0.014	0.00
SynP:SylP: LogP	31.1	2	0.000	0.02
Stress:SylP: LogP	112.3	4	0.000	0.08
SynP:Mod: LogP	7.4	1	0.006	0.00
Stress:SylP:SynP:LogP	11.9	4	0.018	0.00
SylP:SynP:LogP:Mod	13.6	2	0.001	0.00

Table 1: Significant factors for the model fit on cardinal vowels. For each factor or interaction: corresponding χ^2, degree of freedom, p-value, and partial ω^2. Only significant effects are reported here.

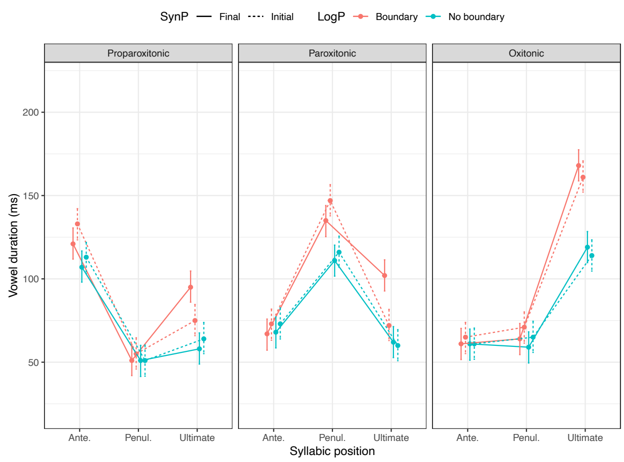

Figure 1: mean fitted vocalic duration for the *Stress:SylP:SynP:LogP* interaction

3.1.2. Duration of stressed vowels

The model on stressed vowels was based on *VowQ* plus the four-way interaction between *SylP*, *LogP*, *SynP*, and *Mod*, keeping the complete random structure. There was thus no interaction between the micro- and macro-prosodic levels: the *VowQ* factor, with seven qualities, was replaced by two factors – for the vowel's "***Height***" (high, mid, low) and "***Backness***" (front, back)– to reach the minimal model[2] (see Table 2); the interaction *Height:*

Backness was not significant. The model had a lower (72%) explanatory power, with fixed factors explaining only 37% of the variance: this shows the importance of stress lengthening, which is not accounted for here (considering only stressed vowels), leaving more importance to random factors.

	χ^2	df	p	ω^2
Height	181.5	2	0.000	0.20
Backness	36.4	1	0.000	0.03
SynP	37.7	1	0.000	0.07
SylP	93.7	2	0.000	0.16
Mode	0.4	1	0.505	0.00
LogP	295.5	1	0.000	0.38
SynP: SylP	15.8	2	0.000	0.03
SynP:Mode	13.7	1	0.000	0.03
SylP:Mode	0.4	2	0.833	0.00
SynP:LogP	9.3	1	0.002	0.02
SylP:LogP	43.3	2	0.000	0.08
Mode:LogP	7.7	1	0.005	0.01
SynP:SylP:Mode	2.1	2	0.346	0.00
SynP:SylP:LogP	1.8	2	0.402	0.00
SynP:Mode:LogP	22.6	1	0.000	0.04
SylP:Mode:LogP	1.6	2	0.451	0.00
SynP:SylP:Mode:LogP	6.2	2	0.046	0.00

Table 2: Significant factors for the model fit on stressed vowels. For each factor or interaction: corresponding χ^2, degree of freedom, p-value, and partial ω^2.

The 4-way interaction explains classical macro-prosodic lengthening at phrase final boundary, an interaction with modality; details cannot fit here. The two micro-prosodic factors, vowel *Height* and *Backness,* show the importance of both characteristics on the duration change, with a relatively important effect size for *Heigh* (ω^2=0.2; with a 9ms lengthening from high to mid vowels, and a 12ms one from mid to "low" ones, that groups /a, ɛ, ɔ/), and reduced effect size for *Backness* (ω^2=0.03; with a 9ms lengthening for back vowels over front ones). We did not observe a duration difference between low /a/ and low-mid vowels – thus, a three-level *Height* factor was sufficient (and /a/ was processed as a front vowel for *Backness*).

3.1.3. Duration of unstressed vowels

Three models were fitted for the duration of unstressed vowels, which descriptions do not fit here (see formulas in footnotes): one for pretonic vowels[3], one for final post-tonic vowel[4], and one for non-final post-tonic[5]. An interaction between micro- and macro-prosodic factors was observed for pretonic vowels, showing that the micro-prosodic differences no longer apply for macro-prosodic factors that most reduce the vocalic duration. This was also observed for non-final post-tonic vowels, the position showing the shortest vowels, where the five vowel qualities have mostly comparable durations. In the case of final post-tonic vowels lengthened because of their position, the three qualities show the same hierarchy described for cardinal vowels (/a/ > /u/ > /i/).

4. DISCUSSION & CONCLUSION

This work gives a first and rapid analysis of the micro- and macro-prosodic factors that influence oral vowel duration in BP in this dataset. As only intrinsic duration phenomena were considered here, this shall be extended in the future to report on co-intrinsic phenomena and include nasal vowels and diphthongs.

The results mostly back what was already described in the literature on micro- and macro-prosodic effects. It replicates the same results as [8, 9], with /a/ not being the longest vowel but comparable in aperture effect to /ɛ/ and /ɔ/, a finding that differs from [10] results (in our data, /ɔ/, not /a/, was the longest stressed vowel: as a "low" –i.e., within the category of longest vowels in our data, for the vowel Height factor– and back vowel). The effect of posterior articulation was systematic in our data for stressed vowels. This differs from the three previous studies, which do not report systematic lengthening with posterior articulation.

Another difference was linked to using a wide set of macro-prosodic factors, the effect of which has been shown here to interact with the micro-prosodic level. Further investigation may be required to understand when and why this interaction happens and propose an interpretation of the reasons for such behavior, but it seems there is a link between the duration of a vowel and the strength of the micro-prosodic effects, thus they are more obvious on stressed vowels. Vowels with shorter durations tend to be centralized, and this affects the importance of articulatory gestures – thus the reduced micro-prosodic changes in these cases.

5. ACKNOWLEDGMENTS

The authors are deeply grateful to the six speakers that agreed to a considerable effort to record the corpus.

6. REFERENCES

[1] J. M. Câmara, Para o estudo da fonêmica portuguesa. Padrão, 1977.
[2] S. H. Lee, 'Sobre as vogais pré-tônicas no Português Brasileiro', Estud. Lingüíst., vol. XXXV, pp. 166–175, 2006.
[3] L. Bisol, 'Neutralização das átonas', DELTA Doc. Estud. Em Lingüíst. Teórica E Apl., vol. 19, no. 2, pp. 267–276, 2003, doi: 10.1590/S0102-44502003000200002.
[4] S. G. Nooteboom, 'Production and perception of vowel duration in Dutch: A Study of Durational Properties of Vowels in Dutch', University of Utrecht, 1972.
[5] D. H. Klatt, 'Interaction between two factors that influence vowel duration', J. Acoust. Soc. Am., vol. 54, no. 4, pp. 1102–1104, Oct. 1973, doi: 10.1121/1.1914322.
[6] J. P. H. van Santen, 'Contextual effects on vowel duration', Speech Commun., vol. 11, no. 6, pp. 513–546, Dec. 1992, doi: 10.1016/0167-6393(92)90027-5.
[7] A. Di Cristo, De la microprosodie à l'intonosyntaxe, 2 vols. Aix-en-Provence : Marseille: Publications, Université de Provence ; Diffusion, J. Laffitte, 1985.
[8] J. A. De Moraes, 'Durée Intrinsèque des Voyelles du Portugais Brésilien', in Mélanges Léon: Phonetique, Phonostylistique, Linguistique et Litterature, Mélodie., P. Martin, Ed. Toronto, 1992, pp. 367–377.
[9] P. A. Barbosa, 'Revelar a estrutura rítmica de uma língua construindo máquinas falantes: pela integração de ciência e tecnologia de fala', in Estudos de prosódia, E. M. Scarpa, Ed. Campinas, SP, Brasil: Editora da UNICAMP, 1999, pp. 21–52.
[10] P. Escudero, P. Boersma, A. S. Rauber, and R. A. H. Bion, 'A cross-dialect acoustic description of vowels: Brazilian and European Portuguese', J. Acoust. Soc. Am., vol. 126, no. 3, pp. 1379–1393, Sep. 2009, doi: 10.1121/1.3180321.
[11] M. McAuliffe et al., 'MontrealCorpusTools/Montreal-Forced-Aligner: Version 2.0.6'. Zenodo, Aug. 08, 2022. doi: 10.5281/ZENODO.6972621.
[12] P. Boersma and D. Weenink, 'Praat: doing phonetics by computer [Computer program]. Version 6.2.08'. Feb. 05, 2022. [Online]. Available: http://www.praat.org/
[13] S. T. Gries, Statistics for linguistics with R: a practical introduction, 3rd revised edition. Berlin Boston: de Gruyter Mouton, 2021.
[14] D. Bates, M. Mächler, B. Bolker, and S. Walker, 'Fitting Linear Mixed-Effects Models Using lme4', J. Stat. Softw., vol. 67, no. 1, 2015, doi: 10.18637/jss.v067.i01.
[15] R Core Team, R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing, 2022. [Online]. Available: https://www.R-project.org/
[16] M. S. Ben-Shachar, D. Lüdecke, and D. Makowski, 'effectsize: Estimation of Effect Size Indices and Standardized Parameters', J. Open Source Softw., vol. 5, no. 56, p. 2815, 2020, doi: 10.21105/joss.02815.

---

[1] Formula for the model on cardinal vowels:
Dur ~ VowQ + TonP + SylP + SynP + LogP + Mod + VowQ:TonP:SylP + VowQ:SynP + TonP:SylP:SynP:LogP + SylP:SynP:LogP:Mod + (1 | Spk)

[2] Formula for the model on stressed vowels:
Dur ~ Height + Backness + SylP * LogP * SynP * Mod + (1 | Spk/Snt)

[3] Formula for pretonic vowels:
Dur ~ VowQ + SynP + SylP + Mod + LogP + VowQ: SylP + SynP: SylP + SylP: LogP + SynP:Mod + SynP: LogP + Mod: LogP + SynP:Mod: LogP + (1 | Spk/Snt)

[4] Formula for final post-tonic vowels:
Dur ~ VowQ + SynP + Stress + Mod + LogP + SynP:LogP + SynP:Stress + Stress:LogP + SynP:Stress:LogP + SynP:Mod + Mode:LogP + SynP:Mod:LogP + (1 | Spk)

[5] Formula for non-final post-tonic vowels:
Dur ~ VowQ + SynP + LogP + SynP:LogP + (1 | Spk/Snt)