Speech Prosody 2022
23-26 May 2022, Lisbon, Portugal

Temporal prosodic cues for COVID-19 in Brazilian Portuguese speakers

Flaviane Romani Fernandes-Svartman[1], *Larissa Cristina Berti*[2], *Marcus Vinícius Moreira Martins*[3], *Beatriz Raposo de Medeiros*[1], *Marcelo Queiroz*[1]

[1]University of São Paulo
[2] São Paulo State University
[3]Minas Gerais State University

flavianesvartman@usp.br, larissa.berti@unesp.br, marcus.martins@uemg.br, biarm@usp.br, mqz@ime.usp.br

Abstract

Temporal aspects of pause in COVID-19 patients' speech are investigated as part of the SPIRA project. Pause is presented as an important candidate to differentiate the speech of COVID-19 patients (target group, n=94) from the speech of healthy subjects (control group, n=99). In order to investigate pause duration and its distribution along the sentence as a prosodic cue, three hypotheses were raised: (1) patient speech includes more pauses than control speech; (2) patient pauses are longer than control pauses; (3) pause distribution is different between groups. Results show that patients' speech has more pauses (3.16 versus 0.85), which are also longer (0.53s versus 0.13s) with respect to control pauses (all differences with $p<0.001$, Mann-Whitney U-Test). A time series analysis was used to model pause distribution along the sentence, which is shown to be randomly spread in the target group, with pauses occurring at unexpected places, contrasting with predictable pauses for controls. Furthermore, a correct classification was obtained for 87-89% of both target and control groups. These findings, grounded in prosodic aspects, are promising and point out the important role of pause as a biomarker in the speech of COVID-19 patients.

Index Terms: prosodic cues, pause, COVID-19, Brazilian Portuguese

1. Introduction

This paper investigates temporal aspects of pause in uttered speech of COVID-19 patients as part of the studies developed within the SPIRA project [1]. Considering that respiratory failure is one of the aggravating symptoms of the disease, one goal of the SPIRA project is the development of an automatic patient monitoring system for early detection of pulmonary insufficiency, based on recorded speech.

To our knowledge, studies dedicated to developing non-invasive methods for detecting COVID-19 search for biomarkers [2, 3] in a broad range of vocal and linguistic aspects, nonetheless, not focusing on temporal prosodic elements in detail.

We consider pauses as prosodic elements that enable breathing during speech [4, 5, 6] and which, together with intonation and duration, indicate prosodic unit boundaries in Portuguese [7, 8, 9, 10, 11]. Studies on prosodic phrasing in subject-verb-object (SVO) sentences in different varieties of this language [7, 12, 13, 14, 15, 16, 17, 18, 19] reveal that pause is one of the most robust clues in identifying boundaries of prosodic constituents of the main phrasing patterns, namely (SVO) and (S)(VO). In a series of studies on the analysis of pauses in speech production of subjects with and without Parkinson's disease, [20] and [21] observed not only a larger variability (both inter- and intra-subjects) in the duration and type of pauses (silent, filled and mixed) in parkinsonian subjects, but also differences of frequency, duration and type of pauses in more advanced stages of the disease [22], with respect to pauses produced by healthy subjects [23].

Our work aims at showing that speech pause is an important candidate to differentiate the speech of patients affected by COVID-19 from the speech of healthy subjects (the control group). In order to investigate pause duration and its distribution along the spoken sentence as a prosodic cue, three hypothesis were raised: (1) there are more pauses in the speech of the patient group than in the control group; (2) pauses are longer in the patient group than in the control group; (3) pause distribution along the sentence is different between groups.

Considering that shortness of breath during speech increases the need to breathe in and, consequently, can affect the prosodic phrasing, we claim that the observation of pause production in the speech of both patients and healthy subjects may help identifying significant differences between these two groups. In order to verify the proposed hypotheses, we proceed with an analysis of silent pauses, defined as portions of silence, in 200 sentences uttered in Brazilian Portuguese by COVID-19-affected patients and healthy subjects. It is worth emphasizing that silent pauses are here understood as any portion of non-speech realization, implying that silent pauses may be filled with low energy noises related to breathing and other non-linguistic sounds produced at the vocal tract. Pause segments were obtained with a semiautomatic energy-thresholding segmentation tool, and were subject to statistical modeling and analysis to assess group differences of several pause segment parameters, including number, length and temporal distribution.

2. Dataset and pause extraction

2.1. SPIRA dataset

The SPIRA dataset corresponds to recordings of the sentence: *O amor ao próximo ajuda a enfrentar o coronavírus com a força que a gente precisa* ('Love of your neighbor helps to face the coronavirus with the strength we need'). This utterance has 31 syllables and branching prosodic and syntactic constituents. The subject, with seven syllables, is

10.21437/SpeechProsody.2022-43

formed by two prosodic words (PW) [24, 25], *o amor* (PW1) and *ao próximo* (PW2). The predicate, with 24 syllables, consists of seven prosodic words, *ajuda* (PW1) *a enfrentar* (PW2) *o corona* (PW3) *vírus* (PW4), *com a força* (PW5), *que a gente* (PW6) and *precisa* (PW7).

The subset selected for this study comprises two groups of 100 participants each: (i) the patient group, including COVID-19 patients with blood oxygenation level below 92%, indicating respiratory failure; (ii) and a control group, formed by healthy volunteer subjects. Each group is also gender- and age-balanced. Considering that this is an applied field research, recording conditions are bound to vary: recordings of the patient group were obtained in hospital wards using a mobile phone, whereas speech data for the control group was gathered through an online platform https://spira.ime.usp.br.html. Further details, including specifications and ethics committee approval, see [1]. After a careful visual and auditory inspection of these two hundred audio files, we inferred that they were reliable for temporal analyses of pause segments.

2.2. Data treatment and pause extraction

Data treatment involved the already mentioned visual and auditory inspection for adequacy of the samples, as well as manual segmentation for a small batch of files. A few samples (n=1 for controls, n=6 for patients) showed what we inferred to be literacy problems or vision impairment, since the collector's voice could be heard, and have been discarded. Afterwards, semi-automatic pause extraction was performed, observing the following steps: (1) automatic segmentation based on a signal energy thresholding, and (2) manual correction of the automatic segmenter output.

Automatic segmentation of the speech signals was implemented in Python using standard signal processing techniques. For each signal, an energy profile (dB scaled) was computed and the minimum and maximum dB values were used to obtain the background noise floor and the dynamic range. An energy threshold was optimized on a small batch of annotated recordings, in order to produce a binary classifier for each frame, indicating the presence of speech or background noise. The output of this binary classifier is expectedly jumpy (i.e., it produces transient false classifications due to rapid variations of the energy profile), and so a smoothing majority vote filter was applied (over 0.4 ms windows).

Manual correction of the automatic segmenter followed three criteria: (1) if the pause was located between two vowels, boundaries were inserted at the left vowel last pulse and at the right vowel first pulse; (2) before unvoiced plosives the burst was observed for the boundary location; and (3) before a fricative sound, the boundary was inserted at the beginning of the visible frication noise in the spectrogram. Regarding criterium (2), we observe that since pause boundaries are defined exclusively on the acoustic signal, the plosive silence portion was necessarily included in the overlapping silent pauses.

3. Results and discussion

3.1. Number of pauses and duration

In order to test hypotheses 1 and 2, the following variables were considered: (a) utterance total duration (UTD), (b) pause mean duration (PMD), (c) speech portion mean duration (SPMD), and (d) number of pauses per utterance (NPU). Variable (a) corresponds to the entire produced sentence, including pauses; variable (b) is the average duration of all pause portions; variable (c) represents the average duration of all portions of uttered speech, not including pauses; and variable (d) is the total number of pauses produced at each utterance.

A Kolmogorov-Smirnov test indicated that data distribution was not normal, leading us to choose a nonparametric hypothesis test, the Mann-Whitney U test. Statistical analyses were conducted with SPSS, version 22.

A Mann-Whitney U test of independent samples indicates that there are differences between the patient and control groups for all variables ($p<0.001$, for all comparisons in the sequel).

UTD mean is 7.96 ± 2.59 seconds (min=4.33, max=18.88) for patients and 5.34 ± 0.85s (min=3.77, max=8.01) for controls. It is observed that patients tend to produce longer utterances with larger variation, whereas controls tend to produce shorter utterances with smaller variation. A similar effect can be observed regarding the length of pauses: PMD in the patient group is 0.53 ± 0.19s (min=0.19, max=1.15), while for the control group it is 0.13 ± 0.16s (min=0, max=0.82).

SPMD is 1.72 ± 0.68s (min=0.77, max=5.20) for patients and 3.41 ± 1.45s (min=1.14, max=6.75) for the control group. The average NPU for the patient group is 3.16 ± 2.02 pauses per utterance (min=0, max=11) while the control group has 0.85 ± 0.94 pauses per utterance (min=0, max=4). The combined values of these variables indicate that patients tend to produce utterances with more pauses and, consequently, shorter phrases.

We proceed to analyze the pause/utterance ratio variable. A t-test of independent samples indicates that there is a statistical difference between groups for this variable ($p<0.001$, $t = -13.75$). The mean ratio is 0.21 (sd=0.11) for patients and 0.04 (sd=0.05) for controls. The results suggest that about 20% of the duration of patients' statements is filled with pauses, whereas in the control group this figure is 4%. Such results seem to indicate that, for the control group, the prevalent pattern is the non-insertion of pauses in the utterance.

UTD and PMD distributions can be seen in Figures 1 and 2. Note that, for both variables, the patient group has larger amplitude and higher median values. The large number of outliers, considering patients' UTD, may be associated with the particular clinical conditions of each patient. Regarding UTD outliers for the control group, it is possible to observe that, although there is some pattern for pause production (few and short), some individual characteristics also tend to remain (e.g., longer pauses).

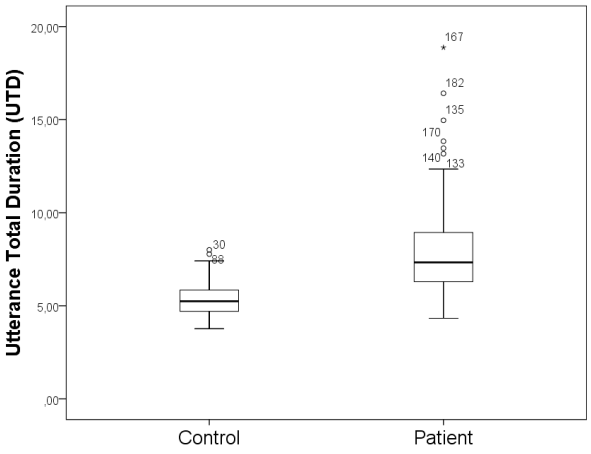

Figure1: *Box plots of UTD, in seconds, for the Control and the Patient group*

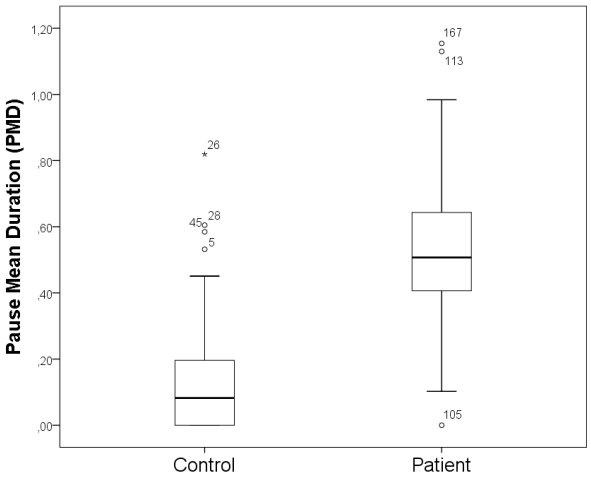

Figure 2: *Box plots of PMD, in seconds, for the Control and the Patient group*

In Figure 3, we see the NPU for each group. As pause insertion in the control group is rare, its distribution obviously differs from that of the patient group. It is important to emphasize that, according to grammatical patterns of the formation of larger prosodic units in Brazilian Portuguese [8, 26, 9, 27], two pauses may occur in the sentence "Love of your neighbor helps to face the coronavirus with the strength we need". The first pause would be placed between the subject "love of your neighbor" and the predicate "helps to face the coronavirus with the strength we need", and the second pause would appear within the predicate, between the object "the coronavirus" and the adjunct "with the strength we need". The vast majority of the control group follows this pattern (i.e., when they do produce pauses), while for patients the median NPU is 3, indicating that patients' pauses occur in unforeseen places in terms of the formation pattern of prosodic constituents.

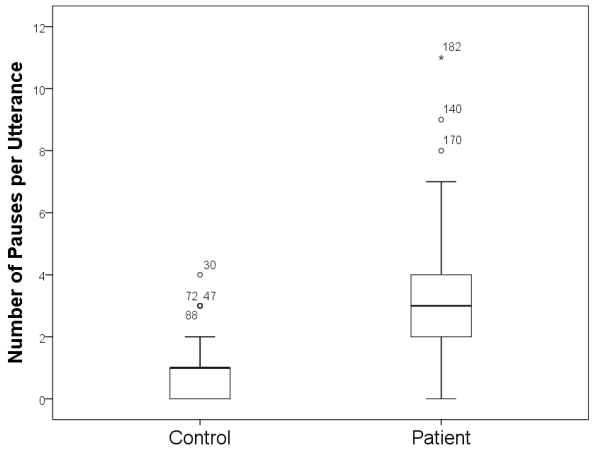

Figure 3: *Box plots of the number of pauses per utterance.*

We highlight that there are differences related to the first four variables (dependent variables) for both groups. This indicates that patients' speech prosody has its own pattern, which can be characterized by the following features: longer utterances with larger inter-subject variability, containing more, and longer, pauses. The observations above allow us to say that patients' speech sequences are organized with a larger number of prosodic units, formed by a smaller number of syllables. This echoes similar observations in the speech of Parkinson's disease patients [20,21,22,23].

3.2. Pause distribution

In order to address the third hypothesis, a time series model was built for both groups. Each sentence is indexed by a normalized time variable $t \in [0, 1]$, where $t = 0$ represents the beginning of the utterance and $t = 1$ its end. The statistical model corresponds to a series of random variables $\{X_G(t)\}$, where each $X_G(t)$ is the probability that the utterance of a subject in group G has a pause at time t. The models for the patient (P) and control (C) groups were obtained from all recordings of the corresponding group, by mapping each recording to a binary function $R(t) \in [0, 1]$, $\forall t \in [0, 1]$, with $R(t) = 1$ if (and only if) there is a pause in that recording at time t. We then compute $X_G(t)$ as the sum (over all recordings R) of all $R(t)$ divided by the number of recordings, separately in each group G. The temporal distributions obtained is shown in Figure 4 below, which was obtained by discretizing the normalized time t in steps of 1% (of the sentence duration).

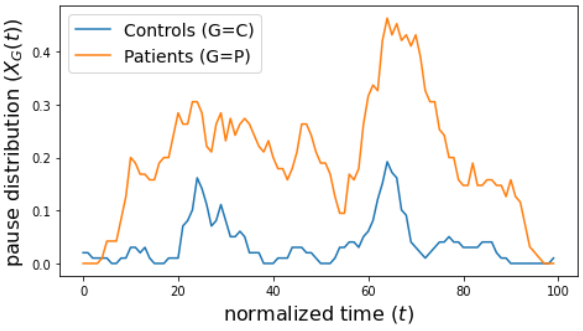

Figure 4: *Pause distribution along the sentence.*

Two immediate observations can be made about Figure 4: for each time t, fewer Controls have pauses with respect to Patients, and pauses are generally concentrated around two main regions ($[20 - 40]\%$ and $[60 - 70]\%$). These regions would correspond, under a steady-paced reading assumption, to the predictable pauses of the sentence "O amor ao próximo / ajuda a enfrentar o coronavirus / com a força que a gente precisa". The same concentrations are also apparent in the pause distribution profile for Patients, although the spread of the corresponding regions is considerably larger.

In order to validate the relevance of this time series model, a simple classifier was implemented to measure whether a recording with pause profile $R(t)$ would be most adherent to its own group distribution, i.e., whether we could tell apart a patient recording from a control recording based on its pause profile alone. Such a naive classifier obtained 88.9% correct classifications for control recordings and 87.4% correct classifications for patient recordings. These results are by no means meant to endorse such a classifier as a diagnosis screening procedure, but they provide evidence of the usefulness of this statistical model, alongside other clinically-validated criteria.

In order to deepen the analysis of the temporal distribution of pauses in these groups, we show in Figure 5 a scatter plot of the joint distribution (t, d), where t is the normalized central time (the temporal midpoint of each pause segment) and d is the normalized duration, for all pauses in both groups.

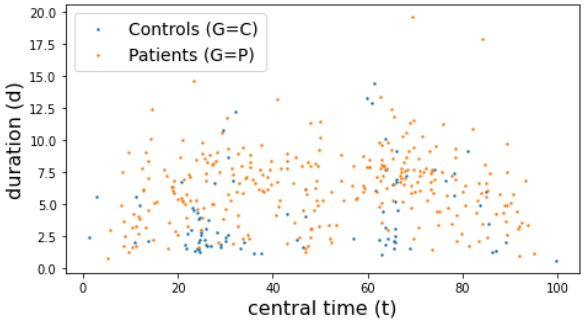

Figure 5: *Scatter plot of pause central times and durations.*

Figure 5 evinces that Patient pauses are widely spread in both central time and duration axes, whereas Control pauses are rather concentrated in the same temporal regions previously identified. Using an automated clustering procedure (Kmeans) we were able to cluster pause events for controls into 5 temporal regions (Figure 6), where regions 2 and 4 correspond roughly to the two regions previously identified (region 2 = 26.3±3.4 % and region 4 = 64.6±5.5 %), and account for 73% of control pauses.

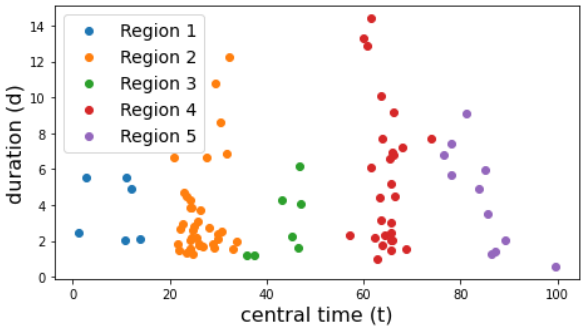

Figure 6: *Control pause clusters.*

The above observations corroborate hypothesis 3, in the sense that the temporal distribution of pauses is indeed very different between groups, allowing a successful classification of subjects into groups based on pause distribution alone. We also observed through temporal distribution analysis that patients frequently produce pauses at unexpected sentence places, when compared to the canonical speech of control subjects.

4. Conclusions

The prosodic time features with focus on pause behavior investigated in this study unveiled speech differences between healthy subjects and patients affected with COVID-19. Three hypotheses were raised regarding the (1) number, (2) duration, and (3) temporal distribution of pauses, and all were corroborated by the experimental data available: patients on average produce more pauses, which are also longer and unexpectedly-placed.

As predicted, in the control group, pauses occur in the S-VO boundary and before the adjunct, which is not related to the previous element. Although such pauses are also observed in patients, other pauses often occur at unexpected places and with longer duration with respect to controls. These temporal prosodic cues characterize patient' speech and may be used as biomarkers in COVID-19 detection.

We expect these contributions to have significance to studies related to phonology under pathological conditions. Within the scope of the SPIRA project, these findings may assist in the development of an application that enables the screening and monitoring of patients with COVID-19, aiming to reduce the high demand both on medical institutions and for health professionals.

5. Ethical procedures

This study was approved by the Research Ethics Committee of the Hospital das Clínicas da faculdade de Medicina da USP under protocol CAAE: 30918120.0.0000.0068.

6. Acknowledgements

We would like to thank the invaluable contribution of the students Ingrid G. G. da Silva, Letícia S. Ferreira, e Pedro L. Pereira, who contributed to data treatment and pause extraction. This research is part of the SPIRA Study, funded by FAPESP Grant 2020/06443-5. L. Berti acknowledges funding by CNPq, Grant 301735/2019-0. F. Fernandes-Svartman acknowledges funding by CNPq, Grant

313103/2018-6. M. Queiroz acknowledges funding by CNPq, Grant 307389/2019-7.

7. References

[1] M. Finger, S. M. Aluisio, E. A. Spazzapan, L. C. Berti, A. C. Camargo Neto, A. Candido Jr., E. Casanova, F. Fernandes-Svartman, R. Ferreira, R. Fernandes Jr., A. Goldman, L. R. Gris, P. L. Pereira, A. S. Levin, M. Martins, M. G. Queiroz, J. H. Quirino, B. R. Medeiros, E. C. Sabino, and D. Silva, "Detecting Respiratory Insufficiency by Voice Analysis: The SPIRA Project," in *Acoustic communication: an interdisciplinary approach*, E. Otta and P. F., Orgs., Universidade de São Paulo: São Paulo, pp. 164-180, 2021.

[2] T. F. Quatieri, T. Talkar, and J. S. Palmer, "A Framework for Biomarkers of COVID-19 Based on Coordination of Speech-Production Subsystems," in IEEE Open Journal of Engineering in Medicine and Biology, vol. 1, pp. 203-206, 2020. DOI: 10.1109/OJEMB.2020.2998051.

[3] G. Fagherazzi, A. Fischer, M. Ismael, and V. Despotovic, "Voice for Health: The Use of Vocal Biomarkers from Research to Clinical Practice," *Digital Biomarkers*, vol. 5, no.1, pp. 78–88, 2021. DOI: 10.1159/000515346

[4] L. C. Cagliari, "Prosódia: algumas funções dos supra-segmentos," *Cadernos de Estudos Linguísticos,* vol. 23, pp. 137-151, Jul./Dez. 1992.

[5] K. Stevens, *Acoustic Phonetics*. Cambridge: MIT Press, 2000.

[6] A. Marchal, and C. Reis, *Produção da Fala*. Belo Horizonte: Editora UFMG, 2012.

[7] S. Frota, *Prosody and focus in European Portuguese: Phonological phrasing and intonation*. New York: Garland Publishing, 2000.

[8] L. E. Tenani, *Domínios prosódicos no português: Implicações para a prosódia e para a aplicação de processos fonológicos*. Unpublished PhD dissertation. State University of Campinas, 2002. Available at: http://repositorio.unicamp.br/jspui/handle/REPOSIP/270935.

[9] C. Serra, *Realização e percepção de fronteiras prosódicas no português do Brasil: Fala espontânea e leitura*. Unpublished PhD dissertation. Federal University of Rio de Janeiro, 2009.

[10] P. Barbosa, and T. Raso, "Spontaneous speech segmentation: functional and prosodic aspects with applications for automatic segmentation," *Revista de Estudos da Linguagem*, vol. 26, no. 4, pp. 1361-1396, 2018.

[11] B. H. F. Teixeira, T. Raso, and P. A. Barbosa, "Detecção automática de fronteiras prosódicas entre unidades entonacionais," *Gradus - Revista Brasileira de Fonologia de Laboratório*, vol. 5, no. 1, pp. 17-46, 2020.

[12] G. Elordieta, S. Frota, P. Prieto, and M. Vigário, "Effects of constituent weight and syntactic branching on intonational phrasing in Ibero-Romance," in *Proceedings of the 15th International Congress of Phonetic Sciences*, Barcelona, Spain, 2003, pp. 487-490.

[13] G. Elordieta, S. Frota, and M. Vigário, "Subjects, objects and intonational phrasing in Spanish and Portuguese," *Studia Linguistica,* vol. 59, no. 2-3, pp. 110-143, 2005.

[14] M. D'Imperio, G. Elordieta, S. Frota, P. Prieto, and M. Vigário, "Intonational phrasing in Romance: The role of syntactic and prosodic structure," in *Prosodies*, S. Frota, M. Vigário, and M. J. Freitas, Eds., Mouton de Gruyter: Berlin/New York, pp. 59-97, 2005.

[15] S. Frota, M. D'Imperio, G. Elordieta, P. Prieto, and M. Vigário, "The phonetics and phonology of intonational phrasing in Romance," in *Segmental and Prosodic Issues in Romance Phonology*, P. Prieto, J. Mascaró, and M. J. Solé, Eds., John Benjamins: Amsterdam, pp. 131-153, 2007.

[16] S. Frota and M. Vigário, "Intonational phrasing in two varieties of European Portuguese," in *Tones and Tunes*, vol. 1, T. Riad and C. Gussenhoven, Eds., Mouton de Gruyter: Berlin, pp. 265-291, 2007.

[17] M. Cruz and S. Frota, "On the relation between intonational phrasing and pitch accent distribution: Evidence from European Portuguese varieties," in *Proceedings INTERSPEECH 2013 - 14th Annual Conference of the International Speech Communication Association*, Lyon, France, 2013, pp. 300-304.

[18] S. Frota, "The intonational phonology of European Portuguese," in *Prosodic Typology II*, S.-A. Jun, Ed., Oxford University Press: Oxford, pp. 6-42, 2014.

[19] F. R. Fernandes-Svartman, V. G. Santos, and G. Braga, "Fraseamento prosódico em português: semelhanças e diferenças entre variedades africanas e brasileiras," *Filologia e Linguística Portuguesa*, vol. 20, no. spe, pp. 119-138, 2018.

[20] L. Chacon and G. Schulz, "Duração de pausas em conversas espontâneas de parkinsonianos," *Cadernos de Estudos Lingüísticos*, vol. 39, pp. 51-70, Jul./Dez. 2000.

[21] L. Chacon, "Relação entre aspectos motores e cognitivos nas dificuldades de linguagem de parkinsonianos," *Veredas - Revista de Estudos Linguísticos*, vol. 6, no.1, pp. 141-152, 2002.

[22] E. C. Oliveira, *Um estudo comparativo do funcionamento das pausas na atividade verbal de sujeitos parkinsonianos*. Unpublished M.Sc dissertation. São Paulo State University, 2003. Available at: http://hdl.handle.net/11449/86602.

[23] L. F. Zaniboni, *Função das pausas na atividade discursiva de sujeitos com doença de Parkinson*. Unpublished M.Sc dissertation. São Paulo State University, 2002.

[24] M. Vigário, *The Prosodic Word in European Portuguese*. Berlin: Mouton de Gruyter, 2003.

[25] P. Toneli, *A palavra prosódica em português brasileiro*. Unpublished PhD dissertation. State University of Campinas, 2014. Available at: http://repositorio.unicamp.br/jspui/handle/REPOSIP/270939.

[26] F. R. Fernandes, "Tonal association in neutral and subject-narrow-focus sentences of Brazilian Portuguese: a comparison with European Portuguese," *Journal of Portuguese Linguistics*, vol. 5/6, pp. 91-115, 2007.

[27] S. Frota, M. Cruz, F. Fernandes-Svartman, G. Collischonn, A. Fonseca, C. Serra, P. Oliveira, and M. Vigário, "Intonational variation in Portuguese: European and Brazilian varieties," in *Intonation in Romance*, S. Frota and P. Prieto, Eds., Oxford University Press: Oxford, pp. 235-283, 2015.