

Rhythm Typology of Brazilian Portuguese dialects

Alexsandro Rodrigues Meireles¹, Viviany de Paula Gambarini¹

¹Phonetics Laboratory, Federal University of Espírito Santo, Brazil

meirelesalex@gmail.com

Abstract

This paper analyzes the speech rhythm of three Brazilian Portuguese dialects under the light of a coupled oscillator approach to rhythm typology. The data showed a scale from a less stress-timed to a more stressed-timed dialect as follows: Minas Gerais state (*coupling strength* = 1.83, *speech rate* = 6.3) > São Paulo state (1.66, 6.6) > Bahia state (1.26, 5.9) > Espírito Santo state (0.95, 4.8). Also, the results show that stress-timing seems to be associated to faster rates. Finally, the theoretical background used here, Barbosa's speech rhythm model, is capable of being sensible to linguistic variations such as speaking styles, speech rates, syntactic structures, and text knowledge.

Index Terms: speech rhythm, speech rate, rhythm typology, stress and syllable timing.

1. Introduction

According to Fraise [13, p.28], speech rhythm, as well as other rhythms, is composed of two essential elements: periodicity and structure, which are interdependent: the structure is always present in the periodicity, and the periodicity is always the organization of structures (see also [14]). Thus, because rhythm is composed of two indivisible elements — periodicity and structure, they have been included in the speech rhythm model (henceforth, SRM, see [5]). In this model, periodicity is represented by the syllabic oscillator and the phrase-stress oscillator, and the structuring of the syllabic periodicity is realized by the phrase-stress oscillator. Besides, because they are indissociable, both are bidirectionally coupled.

As the SRM works with speech rhythm, the phrase-stress oscillator specifically signals, from time to time, positions with relevant linguistic information, when perturbing syllabic regularity. This makes the sequence of vowel onsets non-isochronous, and, therefore, perceptually salient [4, p. 100]. Furthermore, as in the syllabic oscillator, this oscillator has a cyclic pattern, which is influenced by higher linguistic levels such as syntax, semantics, and pragmatics.

As can be inferred from this discussion, the most important advancement of the SRM in relation to other models of speech production is that it not only accounts for rhythm production, but, most importantly, it incorporates in a single model information from other “higher level” linguistic levels. Thus, it contributes significantly towards a scientific modeling of language structure.

Therefore, in order to describe the rhythm of the languages of the world one should take into consideration the essential aspects of the constitution of rhythm presented above. For this reason, Barbosa [6] maintains that advances related to speech rhythm typology should consider a key aspect of rhythm production (and perception): “the interplay between regularity and structuring constraints that takes place between syllable and higher-level units”.

The speech rhythm literature shows two main approaches to speech rhythm research: a) the coupled oscillators approach [3, 11, 22]; and the b) descriptive approach [23, 16, 20, 12, 8].

The descriptive approach measures rhythm typology according to parameters such as %V (vocalic interval durations [23]), ΔC (consonantal interval durations [23]), PVI (pairwise variability index [20, 12]) and CCI (control/compensation index [8]). Although the two last parameters have considerably improved the measurements of speech rhythm, by including information on local differences of duration, none of these models deal with Fraise's essential properties of rhythm production. Also, according to Kohler, neither the vocalic and consonantal interval durations nor the PVI is an explanatory model of rhythm in speech and language, it is only a form of “data sorting on the basis of consonantal and vocalic, i.e. local segmental, durations in their point-to-point variability, not with global rhythmical patterns” [15].

On the other hand, the coupled oscillator approach, which deals with the interaction between a syllable oscillator and a stress group oscillator, is able to explain “both universal and language-specific properties of rhythm by means of general principles applicable to all languages” [6]. Moreover, a comparison between a syllable-sized PVI index and the coupling strength between the two oscillators of the SRM has shown that the latter parameter is “more tuned with the subjects' rhythmic performance as it reflects differences in speaking styles” [6].

Generally, the coupling oscillator approaches classify languages' speech rhythm through the coupling strength of syllable and stress group oscillators, which is measured by the ratio (relative strength, henceforth r) between the intercept coefficient (a) and the inclination coefficient (b) in the linear regression equation $I = a + bn$, in which I is the stress group duration and n the number of syllable-sized units. Thus, stress-timed languages would have a greater influence of the stress group oscillator ($r > 1$), and syllable-timed languages would have a greater influence of the syllabic oscillator ($r \leq 1$). Nevertheless, Barbosa and colleagues [7] have demonstrated that the inclusion of the level of prominence (henceforth p) of the stress group oscillator, though presenting similar results, better reflects the variance of stress group durations. So, p will be included in our analysis, which is based on the concepts of the SRM model [5].

Barbosa [2, 6] and Meireles [19], however, have shown that caution should be observed when analyzing speech rhythms, since r may exhibit great variation as a function of speech styles, speech rates and dialects. A brief review of this variation follows below.

2. Timing in Brazilian Portuguese dialects

Despite the tradition of classifying languages according to one dimension or another (stress- vs. syllable-timing), Brazilian linguists always considered Brazilian Portuguese (henceforth BP) to display both sorts of timing as a function of different linguistic and/or speech contexts. These authors have

shown that BP rhythm varies according to: a) the size of stress groups (henceforth SG): syllable-timing tends to occur in smaller SGs (4 to 8 syllables) and stress timing in greater SGs (above 8 syllables) [21]; b) dialectal variation [10, 1]; c) speech rate variation [1, 2, 19]; d) speech style variation [6, 7]; e) the number of vowel-to-vowel units (henceforth VV) per SG: the greater the number of VVs per SG, the greater the occurrence of stress-timing [2]; and f) syntactic structure variation [2].

Meireles' results [17, 18] have shown that the standard deviation of VV duration as well as SG duration is smaller at faster rates. Consequently, both stress group and VV duration tend to be constant with speech rate increase. Thus, speech rate increase exacerbates the mixture character of BP rhythm, i.e., the tendency to use both syllable as well as stress-timed rhythm. The rhythmic pattern adopted will depend on the coupling strength between the syllable and the stress group oscillators.

Based on this assumption, Meireles [19] carried out a cross-dialectal study to observe speech rhythm variation in two BP dialects (Minas Gerais state (MG) vs. Bahia state (BA)), as follows: syllable-timed languages are expected to have smaller standard deviation of VV duration, as well as higher standard deviation of SG duration; b) stress-timed languages are expected to have greater standard deviation of VV duration, as well as smaller standard deviation of SG duration.

The data suggests that the MG dialect is more stress-timed than the BA dialect. Not only does the BA dialect have smaller variability of VV duration, but also the standard deviation of SG duration is greater at the fast rate. This evidence was found by running a factorial ANOVA with SG standard deviation as a function of rate and dialect ($F(1, 682) = 17.439, p < .00003$). Using this statistical analysis, SG standard deviation decreased with speech rate increase for the MG dialect and increased with speech rate increase for the BA dialect.

Meireles' data used a different approach to analyze SR typology, based on descriptive measurements of acoustic data. Therefore, it lacks the interaction between the stress group and the syllabic oscillators. The present paper, on the other hand, couched in the SRM model, compares Meireles' previous results regarding these two dialects using the common approach of classifying speech rhythms according to the relative strength parameter. Also, the ES dialect (from Espírito Santo state) will be included in the comparison.

3. Methods

3.1. Corpora

A 110-word excerpt of a well-known Brazilian children's book [17] was read by 12 native speakers of BP, aged 15 to 25 years, at three speech rates (slow, normal, and fast). These speakers are from Minas Gerais (henceforth MG) state (2 men and 2 women), Bahia (henceforth BA) state (2 men and 2 women), and Espírito Santo (henceforth ES) state (2 men and 2 women). The MG speakers were born and raised in Belo Horizonte, the capital of MG. The BA speakers were born and raised in Conceição do Jacuípe, a small town in BA. The ES speakers were born and raised in Vila Velha, a metropolitan region of the capital of the state. They are all considered typical representatives of their dialects.

The distinct speech rates were obtained according to the following instructions and order: (1) normal: speak in a comfortable way; (2) slow: speak as slowly as you can, whilst preserving the prosodic structure of the sentences; (3) speak as fast as you can without introducing distortions in your speech.

3.2. Procedures

Semi-automatic procedures were used to observe different rhythmic structures with speech rate increase. First, VV units were labeled in Praat [9] using the BeatExtractor script [4], followed by manual correction. The interval from the beginning of a vowel up to the next vowel defines the so-called VV unit. For example, "em seguid(a)" was segmented as: /eNs/, /eg/ e /id/, and then the duration of each VV unit was extracted. Then, a Praat script (Stress Group Detector, [5]) was run, resulting in information such as: (i) moment-to-moment VV duration; (ii) SG duration; and (iii) VV units per SG.

Stress groups were automatically delimited through the following steps: a) z-score transform to eliminate the effects of the segment's intrinsic duration in each VV duration [6]; b) smoothing of z-score evolution of z-values using the formula:

$$z_{smoothed} = \frac{5z^j + 3z^{j-1} + 3z^{j+1} + z^{j-2} + z^{j+2}}{13}$$

This procedure minimizes local oscillation effects (non-phrasally marked lexical stresses and the remaining effects of intrinsic duration) and determines SG group boundaries through points of maxima in the smoothing duration curve; c) computation of SG duration and number of VV units per SG to compare distributions in three speech rates — slow, normal and fast — for each speaker (the distribution of duration and number of VV units). The number of phonological syllables per stress group was counted manually.

3.3. The coupling strength between oscillators

As previously seen, the coupling strength between the syllable and the stress group oscillators is calculated by the ratio between the intercept coefficient (a) and the inclination coefficient (b) in the regression equation. Moreover, as the addition of the level of prominence (p) better reflects the variance of stress group durations, this measure will be taken into account using multiple linear regression. Thus, the equation used to analyze the rhythm typology of BP dialects and their variability is:

$$I = a + bn + cp \quad (1)$$

in which I is the stress group duration, n is the number of syllable-sized units in the stress group, and p is the level of prominence of the phrase stress. The last measure is simply the smoothed z of each phrase stress, which is returned by the SGDetector script. As Barbosa [6] points out, "due to the techniques applied, this value is not a subpart of the stress group duration", but it does improve the correlation coefficients of the data.

4. Results

Similar to the results produced by Barbosa et al. [6, 7], the number of VV units did not produce significant intercept coefficients for most of the speakers and conditions (6 out of 32 coefficients, 2 dialects x 4 speakers x 3 rates, and BA dialect x 4 speakers x 2 rates (normal, fast)). That is why we also included the number of phonological syllables (PS) as an independent variable in our analysis. Nevertheless, differently from Barbosa's studies, the use of PS did not result in a greater number of significant intercept coefficients (5 out of 32 coefficients). As can be noticed, the slow rate was excluded from the analysis of the BA dialect, since these speakers did not produce the slow rate as instructed. All six BA speakers (we included 2 extra ones) introduced pauses in between the words.

It is important to highlight the fact that of the 6 significant intercept coefficients using VV units, 5 were found at the fast rate. Also, all significant coefficients using PS were found at

the fast rate. This finding may be related to Meireles' results [17, 18, 19] regarding the decrease of standard deviation of VV duration and SG duration at fast rates. Considering this standard deviation decrease, the probability of finding significant results in linear regressions is greater at fast rates.

As speech rate increase may be one of the factors that change BP speech rhythm, it is necessary to check whether the qualitative rates (slow, normal, fast) corresponded to statistically different quantitative rates (measured by the rate influence on VV duration). Therefore, one-way ANOVAs with VV duration as a function of rate were run. The results have shown that all speakers presented a decreasing pattern of VV duration from slow to fast rate. So, we assured the analyses were not influenced by the non-statistical significance of rates.

Also, a one-way ANOVA with speech rate as a function of the dialects, all speakers included, has shown a significant difference among dialects only for the normal rate ($F(2, 2068)=8.6228, p < 0.00019$). This result may be explained by the instructions for rate acquisition. The speakers spoke as slowly as they could for the slow rate and as fast as they could for the fast rate, but they spoke comfortably for the normal rate, i.e., this last rate better reflects a natural way of pronouncing sentences in their dialects. Moreover, it reflects previous results regarding a comparison between southeastern (MG, ES) and northeastern (BA) BP dialects. Meireles and colleagues [19] have proved that southeastern speakers speak faster than northeastern speakers. So, our data corroborates this result since the two statistically equal southeastern dialects (MG, ES) were significantly different from the northeastern dialect (BA) at the normal rate (MG mean (x) = 4.6, and standard deviation (sd) = 0.5; ES x = 4.4, and sd = 0.4; BA x = 3.9 mean, and sd = 0.4).

Nevertheless, only one instance of a marginally significant intercept coefficient was found at the normal rate. On the other hand, all the other significant intercepts, either using the VV unit or the PS, were found at the fast rate. Thus, all rhythm comparisons among dialects were based on this rate. Also, texts with no significance for this parameter were excluded from the dialectal comparison. In the end, we had 2 speakers representing the BA dialect (JU, RO), 2 speakers representing the MG dialect (LZ, VH), and 1 speaker representing the ES dialect (BM). As can be seen, these speakers were grouped together in table 1. Besides, the regression equations for the slow and the normal rates, all speakers grouped together, were added in table 1 as a matter of illustration. Bold items are meaningless values, since no significance was found for the intercept coefficient.

Table 1 shows the linear regression equations at the fast rate according to dialect (MG, ES, BA) and type of syllable-sized unit (VV or PS). The coupling strength c is the ratio between the intercept (a) and the inclination (b) coefficients of the equations. All correlation coefficients are highly significant ($R^2 > 0.82, p < 10^{-4}$). The significance of the intercept coefficients is indicated in parentheses. c is considered undefined for a non-significant value of the intercept coefficient or a negative a . Speech rate (sr) is given in VV/s or PS/s. These results are plotted against speech rate in figure 1.

Table 1. Regression equations at the fast rate for 3 BP dialects (MG, ES, BA) considering the number of VV units (nVV) and the number of phonological syllables ($nSyl$). Coupling strength (c), speech rate (sr), and the phrase stress magnitude (p) are given. S = slow; N = normal, and F = fast. In the ' c ' column, bold items are meaningless values, and ' u ' is an undefined value.

Dialect	Regression equation	c	sr
MG (S)	$I = -121 + 236nVV + 73p$ (n.s.)	u	3.6
MG (S)	$I = -72 + 181nSyl + 78p$ (n.s.)	u	3.6
MG (N)	$I = 13 + 186nVV + 72p$ (n.s.)	0.07	4.6
MG (N)	$I = 15 + 148nSyl + 70p$ (n.s.)	0.10	4.6
MG (F)	$I = 222 + 132nVV + 106p$ ($6.4e^{-5}$)	1.68	6.3
MG (F)	$I = 187 + 102nSyl + 86p$ ($7.1e^{-5}$)	1.83	6.3
ES (S)	$I = 34 + 207nVV + 52p$ (n.s.)	0.16	3.8
ES (S)	$I = 88 + 156nSyl + 50p$ (n.s.)	0.56	3.8
ES (N)	$I = -9 + 197nVV + 58p$ (n.s.)	u	4.4
ES (N)	$I = 59 + 150nSyl + 51p$ (n.s.)	0.39	4.4
ES (F)	$I = -232 + 272nVV + 60p$ (0.01)	u	4.8
ES (F)	$I = 121 + 128nSyl + 49p$ (0.02)	0.95	4.8
BA (N)	$I = 36 + 205nVV + 49p$ (n.s.)	0.17	3.9
BA (N)	$I = 63 + 158nSyl + 48p$ (n.s.)	0.39	3.9
BA (F)	$I = 125 + 153nVV + 63p$ (0.002)	0.81	5.9
BA (F)	$I = 143 + 113nSyl + 63p$ ($6.4e^{-3}$)	1.26	5.9

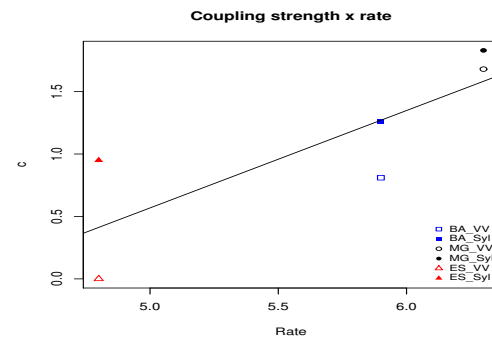


Figure 1. Relationship between speech rate (PS or VV units) and coupling strength for 3 BP dialects.

In addition, ANCOVAs with SG duration as a function of the continuous variables nVV (or $nSyl$) and p , and the categorical variable *dialect* were run, so as to evaluate whether the regression equations were statistically different among the dialects at the fast rate. As before, only the subjects previously selected for the dialectal comparison were used in the analysis. The results have shown that the regression equations were statistically different among each other either with the nVV ($F(4,188) = 265.8, p < 2.2e^{-16}, R^2 = 0.85$) or the $nSyl$ ($F(4,188) = 373.8, p < 2.2e^{-16}, R^2 = 0.89$) parameters. See figure 2 for a graphical display of the regression equations with the nVV parameter for the three dialects.

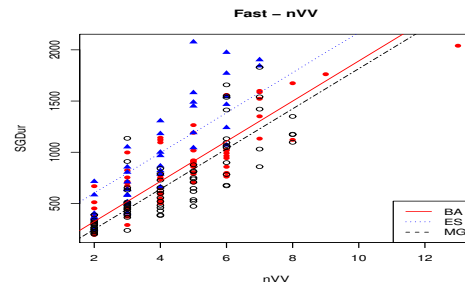


Figure 2. Regression equation of the BP dialects.

Considering the number of PS, results show that the most stress-timed dialect is MG (1.83) followed by BA (1.26) and ES (0.95), which is agreement with Meireles' findings [19] who followed a different methodological approach. At least for two dialects, this pattern is the same if we consider the number of VVs (MG (1.68) > BA (0.81)). Moreover, although no significance for rate as a function of dialects was found, the greatest rate was found in the MG dialect, which is also according to previous results produced by Meireles.

Even though our data suggests that the BA dialect is more stress-timed than the ES dialect, extra data should be collected in order to clarify this point, since only one significant a was found for the ES dialect. It may even be the case that this ES speaker is a naturally slow speaker who does not follow the major rhythmical tendencies of his dialectal area.

5. Discussion

The above explanation supports the use of a dynamical systems approach to explain rhythm variability in the languages of the world. Not only is the SRM able to represent linguistic rhythms taking into consideration all important aspects of rhythm production (basically periodicity and structuring), but it is also sensible to linguistic variations such as speaking styles, speech rates, syntactic structures, and text knowledge.

As an example of rhythm typology application, we can extend our analysis to include the data of São Paulo state (SP). This data will be extracted from Barbosa [6], who used the same speech corpus of this paper. Considering similar nominal speech rates, we may infer a scale from less stress-timed to more stressed-timed as follows: MG ($c=1.83$, $sr=6.3$) > SP (1.66, 6.6) > BA (1.26, 5.9) > ES (0.95, 4.8).

Another interesting point presented here is the occurrence of significant intercept coefficients mostly at the fast rate. This fact may be supported by Meireles' studies [17, 18, 19], which show a decrease of standard deviation at fast rates, i.e., less duration variability is expected at extreme rates. As a consequence, it may be easier to find significant regression equations at these rates. In addition, this rate was associated to higher levels of stress timing patterns.

On the other hand, Barbosa [6] argues that speech rate increase favors syllable-timing. This argument seems to be reflected in his data (see table 2). Nevertheless, we are not sure this is true for BP in general, since in our data the most stress-timed dialects were the fastest ones (see table 1 above). Meireles and colleagues [19] have discussed this point within the BP literature (see [1, 2]) and concluded that fast rates may be associated with stress-timing. So, further studies with more speakers and dialects need to be run to investigate speech rate effects on speech rhythm more deeply.

6. Acknowledgements

The authors would like to thank Plínio Barbosa, Karen Currie, and Pablo Arantes for suggestions, and the CNPq for financing the research.

7. References

- [1] Abaurre-Gnerre, M. B. Processos fonológicos segmentais como índices de padrões prosódicos diversos nos estilos formal e casual do português do Brasil. *Caderno de Estudos Lingüísticos*, 2: 23-34, 1981.
- [2] Barbosa, P. A. Syllable-timing in Brazilian Portuguese": uma crítica a Roy Major. *D.E.L.T.A.*, 16 (2), 369-402, 2000.
- [3] Barbosa, P. A., "Explaining Cross-Linguistic Rhythmic Variability via a Coupled-Oscillator Model of Rhythm

- Production", Proc. Speech Prosody 2002 Conf. [CD], Aix-en-Provence, 163-166, 2002.
- [4] Barbosa, P. A., *Incursões em torno do ritmo da fala*, Campinas, Brazil: Pontes/Fapesp, 2006.
- [5] Barbosa, P. A., "From syntax to acoustic duration: a dynamical model of speech rhythm production", *Speech Communication*, 49:725-742, 2007.
- [6] Barbosa, P. A. "Measuring speech rhythm variation in a modelbased framework". In.: Proc. Interspeech 2009, Brighton, UK, 1527-1530, 2009.
- [7] Barbosa, P. A, Viana, M. C., Trancoso, I. "Cross-variety Rhythm Typology in Portuguese". In.: Proc. Interspeech 2009, Brighton, UK, 1011-1014, 2009.
- [8] Bertinetto, P. M., Bertini, C., "Towards a unified predictive model of Natural Language Rhythm", *Quaderni del Laboratorio di Linguistica della SNS*, 7, 2007/08.
- [9] Boersma, P., Weenink, D., "Praat: doing phonetics by computer" (Version 5.0.35) [Computer program], Online: <http://www.praat.org>, accessed in 2008.
- [10] Cagliari, L. C. and Abaurre, M. B. M. Elementos para uma investigação instrumental das relações entre padrões rítmicos e processos fonológicos no português brasileiro. *Caderno de Estudos Lingüísticos*, 10: 39-57, 1986.
- [11] Cummins, F., Port, R., "Rhythmic constraints on stress timing in English", *J. Phon.*, 26:145-171, 1998.
- [12] Deterding, D., "Letter to the Editor. The measurement of rhythm: a comparison of Singapore and British English", *J. Phon.*, 29:217-230, 2001.
- [13] Fraisse, P. Les rythmes. *Journal Français d'Oto-Rhinolaryngologie*, n. Supplément numéro 7, p. 23-33, 1968.
- [14] Fraisse, P. *La Psychologie du Rythme*. Paris: Presses Universitaires de France, 1974.
- [15] Kohler, K. J. "Rhythm in speech and language: a new research paradigm", *Phonetica*, 66:29-45, 2009.
- [16] Low, E. L., Grabe, E., Nolan, F., "Quantitative characterizations of speech rhythm: Syllable-timing in Singapore English", *Language and Speech*, 43:377-401, 2000.
- [17] Meireles, A. R. *Self-organizing rhythms in Brazilian Portuguese: speech rate as a system perturbation*. Germany: VDM Verlag, 2009.
- [18] Meireles, A. R. and Barbosa, P. A. Speech rate effects on speech rhythm. In: *Speech Prosody 2008 Conference*, 2008, Campinas. *Proceedings of the Speech Prosody 2008 Conference*. Campinas: RG. v.1. p.327 - 330, 2008.
- [19] Meireles, A. R., Tozetti, J. P., Borges, R. R. Speech rate and rhythmic variation in Brazilian Portuguese. In: *Speech Prosody 2010 Conference*, 2010, Chicago. *Proceedings of the Speech Prosody 2010 Conference*. Chicago: RG. v.1. p.1 - 4, 2010.
- [20] Mok, P. P. K., Dellwo, V., "Comparing native and nonnative speech rhythm using acoustic rhythmic measures: Cantonese, Beijing Mandarin and English", *Proc. Speech Prosody 2008*, Campinas, 423-426, 2008.
- [21] Moraes, J. A. and Leite, Y. F. Ritmo e velocidade da fala na estratégia do discurso: uma proposta de trabalho. In: R. Ilari (org.) *Gramática do Português falado, v. II: níveis de análise lingüística*. Campinas: Editora da Unicamp, 1992.
- [22] O'Dell, M., Nieminen, T., "Coupled Oscillator Model of Speech Rhythm", *Proc. XIVth ICPHS*, San Francisco, 1075-1078, 1999.
- [23] Ramus, F., Nespors, M., Mehler, J., "Correlates of linguistic rhythm in the speech signal", *Cognition*, 73:265-292, 1999.