

Speech rate effects on speech rhythm

Alexsandro R. Meireles and Plínio A. Barbosa

Speech Prosody Studies Group, Department of Linguistics
State University of Campinas, Brazil

{meirelesalex@gmail.com; pabarbosa.unicampbr@gmail.com}

Abstract

This acoustic study investigates speech rhythmic restructurings due to speech rate increase in Brazilian Portuguese. Rhythmic restructuring is considered here as the reorganization of stress groups due to speech rate increase. The Dynamical Speech Rhythm model was used as a theoretical background for the acoustic analyses. Main results have shown that speech rate increase reorganizes speech rhythm and modifies some phonetic parameters in the following way: a) the standard deviation of vowel-to-vowel (VV) duration and stress group duration is smaller at faster rates; b) stress group duration tends to be constant with speech rate increase (rhythmic restructurings make VV units smaller, but with a greater number of VV units per stress group, what results in a statistically constant stress group duration); c) the number of VV units per stress group proportionally increases with speech rate increase due to rhythmic restructurings; and d) speech rate increase exacerbates the mixture character of Brazilian Portuguese rhythm, i.e., tendencies to syllable as to stress-timed rhythm.

1. Introduction

Meireles and Barbosa [9] have presented studies that show the influence of speech rate on speech rhythmic restructurings (see [2] for data in French, and [4], [1] for data in Brazilian Portuguese). Rhythmic restructuring is defined here as the reorganization of stress groups (henceforth SG) as speech rate increases. An example of this linguistic phenomenon can be seen in figure 1.

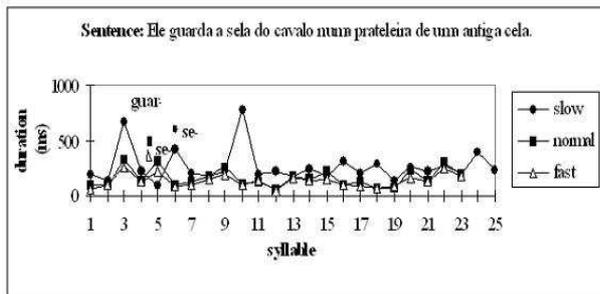


Figure 1: Duration contours of syllables in "Ele guarda a sela do cavalo numa prateleira de uma antiga cela" (He keeps the horse's saddle on a shelf in an ancient shell) at three speech rates (from [1]).

In figure 1, an utterance recorded at three statistically different speech rates is displayed. Considering syllable duration patterns as a descriptive framework, rhythmic restructurings can

be seen in the first three syllables of the stretch "e-le-guar". In this part there is a gradual change from 2 SG (slow rate) to 1 SG (cf. duration crescendo). It is important to highlight that, as Brazilian Portuguese is a right-headed language, peaks in duration contours represent the end of a stress group.

The speech rate influence on speech rhythmic reorganizations can be explained by the Dynamical Speech Rhythm model [3] [4] (henceforth DSR). In this model, speech rhythm is considered as a "consequence of the variation of perceived duration along the entire utterance" [3]. Intrinsic segmental duration is normalized through the use of abstract vowel-to-vowel (henceforth VV) duration (see [4] [8]). Entrained, abstract VV duration is used as the control parameter to generate prosodic variation. The maxima of VV duration delimitate produced SG as rhythmic units in Brazilian Portuguese.

Using the DSR model as a theoretical framework, Meireles and Barbosa have run acoustic and articulatory experiments that show the influence of speech rate on speech rhythm (see [8] for a broad explanation): (i) Meireles and Barbosa [10] investigated how speech rate acts to articulatorily reorganize lexical stress in Brazilian Portuguese; (ii) Meireles and Barbosa [9] ran acoustic and articulatory (EMMA) experiments, so as to study speech rhythm reorganizations due to speech rate increase in Brazilian Portuguese.

Meireles and Barbosa [9] have also shortly described an acoustic experiment in which several sentences were recorded at three different speech rates. In this paper, this study will be described in great detail.

2. Methods

As in our previous studies [8] [9] [10], a classical procedure in Dynamical Systems Theory [6] was used: a perturbation of the system caused by movement rate increase. Such procedure is used so as to reveal new stable patterns in a system. In our case, speech rate was increased in order to reveal new rhythmic patterns. Specifically, by the light of DSR, an acoustic experiment was run so as to observe speech rhythmic restructurings due to speech rate increase from the interaction of the syllabic and phrase stress oscillators [3] [4].

2.1. Data

A database of 11 sentences repeated ten times (sampling rate of 22.1 kHz) was recorded by a female (age 29), native speaker of Brazilian Portuguese, in a sound-treated room, at three speech rates (slow, normal, and fast). In this study, isolated sentences were used in order to avoid great variability in the delimitation of stress groups, as occurred in [8] [9].

The recorded corpus was composed by the following sentences: 1. Há três tipos de abóbora no centro de Belo Hori-

zonte; 2. Há três tipos de abóbora nanica em Belo Horizonte; 3. A análise de tantos dados parecia certa; 4. A análise computacional parecia certa; 5. Quando sinto cócegas, logo tenho falta de ar; 6. Dei um beijo de tirar o fôlego; 7. Fôlego de atleta foi exigido do competidor; 8. Meu fósforo terminou. Me empresta o fósforo; 9. Uso fósforo para acender o fogão; 10. Como carne de boi no sábado; 11. No sábado passado comi carne de boi.

2.2. Procedures

Semi-automatic procedures were used to observe different rhythmic structures with speech rate increase. Firstly, VV units were labeled in Praat [5]. Then, a Praat script (SGDetector) [4] was run, resulting in information, such as: (i) moment-to-moment VV duration; (ii) SG duration; and (iii) VV units per SG. Finally, in case of wrong attribution of phrasal stresses, we listened to the original recordings, so as to manually mark phrasal prominences.

2.3. Hypotheses

This corpus was used so as to deepen previous acoustic results [8] [9]. According to these results, it is expected in case of rhythmic restructurings that:

1. Standard deviation is generally smaller at fast rates, resulting in a greater sensation of isochronism at such rates;
2. SG duration tends to be kept constant with speech rate increase;
3. VV per SG proportionally increases with speech rate acceleration;
4. Speech rate increase exacerbates the mixture character of Brazilian Portuguese rhythm, i.e., tendencies to syllable as to stress-timed rhythm.

Besides, this database was also used in order to be observed the influence of syntactical factors in phrasal stresses [8]. For this reason, paired sentences (except 5) with close semantic content and some syntagmatic variation were chosen.

3. Results

One-way Anovas with VV duration as dependent variable and speech rate as independent variable were run in *Statistica* (www.statsoft.com) to evaluate if we had statistically different speech rates. Statistical analyses showed significative differences among rates for all sentences (see table 1 below). Nevertheless, *post-hoc* Scheffé tests showed that sentences 5,6, and 9 were different only between two rates with the following pattern: slow (S) \neq (normal (N) = fast (F)).

These results show that we have at least two statistically different speech rates for each sentence. In the following we will be investigating each one of the hypotheses mentioned above.

3.1. Standard deviation of VV duration

All sentences revealed a decreasing pattern of the standard deviation of VV duration with speech rate increase (see table 1). Yet, a One-way Anova showed that this pattern was statistically different only for sentences 1, 3, 4, 5, 10, and 11. Besides, a subsequent *post-hoc* Scheffé test also showed the following patterns: (i) S \neq N \neq F (1, 5); (ii) S \neq (N = F) (3, 11); and (iii) (S = N) \neq F (4, 10).

These statistical analyses considered all VV units in the SG. Nevertheless, considering that VV units in the DSR model grow

Table 1: VV duration (mean, standard deviation) in milliseconds, and significance level ($p <$) for duration (D) and standard deviation of VVs (SD) for all sentences at three speech rates: slow (S), normal (N), and fast (F). n.s. means non-significant.

	VV duration (mean and standard deviation)				
	S	N	F	D	SD
1	259 (118)	202 (99)	149 (81)	10^{-5}	3.10^{-5}
2	230 (102)	179 (99)	163 (92)	10^{-5}	n.s.
3	340 (146)	216 (97)	179 (83)	10^{-5}	3.10^{-5}
4	218 (84)	176 (76)	145 (63)	4.10^{-5}	0.006
5	277 (126)	218 (91)	205 (63)	10^{-5}	10^{-5}
6	263 (103)	212 (88)	198 (96)	10^{-5}	n.s.
7	249 (75)	173 (64)	148 (63)	10^{-5}	n.s.
8	220 (104)	179 (94)	164 (86)	10^{-5}	n.s.
9	206 (70)	162 (67)	178 (68)	4.10^{-4}	n.s.
10	215 (65)	181 (60)	143 (40)	10^{-5}	3.10^{-5}
11	233 (81)	189 (58)	161 (60)	10^{-5}	10^{-5}

exponentially up to a phrasal stress, if VV units at the end of a SG are eliminated, it will be easier to see the decrease of the standard deviation of VV units with speech rate increase. Moreover, unstressed VV units in the DSR model act as a reference for speech rate perception. Therefore, we ran One-way Anovas with standard deviation of unstressed VV duration as a function of speech rate. Results of this analysis are displayed in table 2.

Table 2: Unstressed VV duration (mean, standard deviation) in milliseconds, and significance level for all sentences at three speech rates: slow (S), normal (N), and fast (F). Bold items represent sentences with many gestural overlappings. n.s. means non-significant.

	Unstressed VV duration (mean and standard deviation)			
	S	N	F	$p <$
1	203 (67)	177 (70)	138 (40)	10^{-5}
2	218 (91)	157 (63)	141 (52)	10^{-5}
3	340 (146)	217 (97)	179 (83)	4.10^{-4}
4	199 (76)	158 (55)	131 (47)	0.001
5	251 (111)	194 (70)	182 (39)	10^{-5}
6	266 (119)	212 (88)	198 (96)	0.003
7	236 (81)	164 (61)	142 (58)	4.10^{-5}
8	180 (60)	151 (42)	141 (48)	0.04
9	190 (58)	146 (62)	173 (69)	n.s.
10	203 (67)	177 (63)	138 (40)	10^{-4}
11	210 (71)	176 (56)	142 (47)	2.10^{-5}

Statistical analyses have then shown a standard deviation decrease of unstressed VV duration with speech rate increase. All sentences but one (9) reached statistical significance. This only exception is explained by the fact that there were greater gestural overlappings for this sentence's gestures (as well as to the gestures in sentences 6 and 8) (see [8]). These gestural overlappings make the splitting of some acoustic VV impossible, resulting in greater VV duration compared to the ones without

these overlappings.

3.2. Standard deviation of SG duration

Differently from the standard deviation of VV duration, it was not possible to calculate the standard deviation of SG duration for all sentences at some rates. Due to their length, sentences 3 (fast rate), 6 (normal and fast rates), 9 (normal and fast rates), and 10 (normal and fast rates) had only 1 SG at faster rates (see table 4). Therefore, only the remaining sentences were analyzed, which resulted in (see table 3): (1) decreasing pattern from slow to fast rate (1, 4, 5, 11); (2) increasing pattern from slow to fast rate (2, 7, 8). To understand pattern (2) it is necessary to introduce a new term: DVV/SG_{ir} . This term represents the standard deviation of VV/SG intra rates, i.e., the standard deviation of VV/SG for each rate.

Table 3: Mean and standard deviation of stress groups (μ , σ) in milliseconds, and significance level ($p <$) for duration (D) and standard deviation of SG (SD) at three speech rates: slow (S), normal (N), and fast (F). Italic items (2, 7, 8) represent greater standard deviation with speech rate increase; black items represent the opposite pattern; and bold items (3, 6, 9, 10) represent only 1 SG per sentence.

	S	N	F	D	SD
1	1051 (300)	809 (174)	788 (169)	3.10^{-4}	0.003
2	858 (200)	1353 (320)	1226 (321)	10^{-5}	n.s.
3	1094 (365)	1133 (876)	1733 (90)	n.s.	n.s.
4	912 (368)	1026 (249)	845 (172)	n.s.	8.10^{-5}
5	1944 (901)	1091 (516)	924 (362)	10^{-4}	10^{-5}
6	897 (224)	1153 (43)	991 (79)	0.002	n.s.
7	1287 (42)	1199 (495)	939 (326)	0.005	10^{-5}
8	818 (357)	982 (881)	891 (758)	n.s.	10^{-4}
9	1220 (472)	1807 (36)	1542 (53)	0.002	n.s.
10	819 (498)	1206 (42)	973 (27)	0.003	n.s.
11	1201 (98)	962 (54)	806 (57)	10^{-4}	0.046

A non-parametric Kruskal-Wallis Anova (see table 4) showed a decrease (5) or statistical invariance (1, 4, 11) of DVV/SG_{ir} for the sentences with a decreasing pattern of standard deviation. On the other side, there was an increase of DVV/SG_{ir} for the remaining sentences (2, 7, 8). Therefore, in order to occur a decrease of the standard deviation of SG duration with speech rate increase, it is necessary that the number of VV/SG be not extremely different intra-rates. In other words, DVV/SG_{ir} should decrease or remain constant with speech rate increase so as to hypothesis (1) be valid for SG duration. Figure 2 is a real example of the standard deviation decrease of SG duration with speech rate increase.

3.3. Constant SG duration

Tables 3 and 4 show that 3 of the sentences (3,4,8) statistically confirm hypothesis (2) with a three-rate differentiation through a *post-hoc Scheffé* test. Before analyzing the remaining sentences, it is important to remind that rhythmic restructurings must necessarily occur between rates in order for this hypothesis to be verified. These restructurings make VV units smaller, but with a greater number of VV/SG, what results in a statistically constant SG duration. This fact explains why sentences 5

Table 4: DVV/SG_{ir} for all sentences in the corpus. Numbers like 3-3-7-3 represent the most common pattern of VV/SG for each rate, and the number in parenthesis represents the DVV/SG_{ir} . L represents slow rate; N, normal rate; F, fast rate; p represents the significance level of a non-parametric Kruskal-Wallis (K-W) analysis. Italic VVs (2, 7, 8) represent greater standard deviation with speech rate increase, black VVs represent the opposite pattern, and bold VVs (3, 6, 9, 10) represent sentences without VV/SG deviation at some rates.

	Rates			K-W
	S	N	F	$p <$
1	3-3-7-3 (2)	3-3-7-3 (2)	6-7-3 (2.1)	n.s.
2	3-3-6-3 (1.5)	6-9 (2.1)	6-9 (2.1)	10^{-5}
3	2-3-4 (1)	2-8 (4,2)	9	-
4	2-7-4 (2.5)	7-4 (2.1)	7-4 (2.1)	n.s.
5	5-9 (2.8)	3-7 (2.8)	3-6 (2.1)	0.001
6	2-5 (2.1)	6	5	-
7	5-5-5 (0)	9-5 (2.8)	9-5 (2.8)	0.003
8	2-5-4 (1.5)	2-9 (4.9)	2-9 (4.9)	10^{-4}
9	3-9 (4.2)	11	10	-
10	6-2 (2.8)	7	7	-
11	6-5 (0.7)	5-5 (0)	5-5 (0)	n.s.

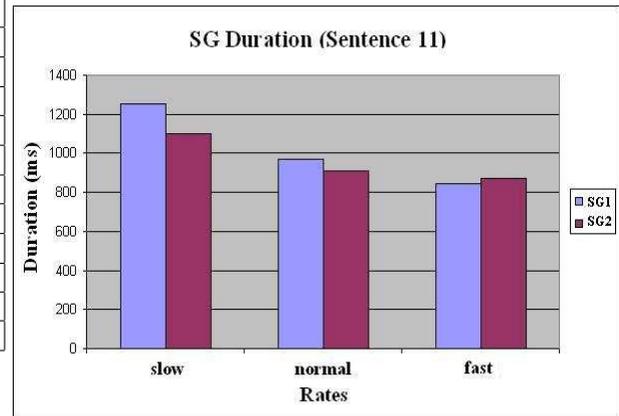


Figure 2: Example of a standard deviation decrease of SG duration with speech rate increase for sentence 11.

and 11 did not present such pattern (see table 4).

Through a *post-hoc Scheffé* test it can also be seen that the remaining sentences confirm hypothesis (2), but in a complex manner. Sentence 1 did not keep constant SG duration from slow to normal rate, since no rhythmic restructurings occurred. Yet, SG duration was constant from normal to fast rate. In this rate transition there was a change from 4 SG to 3 SG (see table 4). On the other hand, sentence 7 kept constant SG duration from slow (3 SG) to normal rate (2 SG), but not from normal to fast rate (see table 4). Sentences 6, 9, and 10 had a different behavior. Even though there were rhythmic restructurings from slow to normal rate (see table 4), SG duration increased for this rate change. Yet, a *post-hoc Scheffé* test showed that SG duration at the slow rate was not statistically different from SG duration at the fast rate. It implies that, in order to occur a constant SG duration between rates, it is necessary a greater speech

rate variation from one rate to the other. Also, for sentence 2, SG duration was different from slow to normal rate, but SG duration at the slow rate was statistically different from SG duration at the fast rate. This is explained by the smaller decrease of SG duration from normal to fast rate (see table 3) compared to sentences 6, 9, and 10. In this situation, a greater SG duration decrease from normal to fast rate is a necessary condition for a constant SG duration between slow and fast rate. Nonetheless, more data are needed to confirm this hypothesis, since there was no statistical difference of SG duration from normal to fast rate for sentences 2, 6, 9, and 10.

3.4. VV per SG increase

Table 5 corroborates the hypothesis (3) of VV per SG increase with speech rate acceleration due to rhythmic restructurings. All sentences showed such pattern (except 5 and 11, since no rhythmic restructurings occurred). Besides, VV per SG increase was statistically confirmed, through a Kruskal-Wallis Anova with VV per SG as a function of speech rate, for 5 (2, 3, 6, 9, and 10) out of 9 sentences.

Table 5: VV/SG (mean, standard deviation) at three speech rates (slow (S), normal (N), and fast (F)) for all sentences (1-11). p represents the significance level for a Kruskal-Wallis Anova (K-W) with VV/SG as a function of rate. Italic items (2,7,8) represent greater standard deviation with speech rate increase. Black items represent the opposite pattern. Bold items (3,6,9,10) represent only 1 SG per sentence repetition.

	S	N	F	p <
1	4 (2)	4 (2)	5.3 (2)	n.s.
2	<i>3.8 (1.5)</i>	<i>7.5 (2.1)</i>	<i>7.5 (2.1)</i>	10^{-4}
3	3 (1)	5 (4.2)	9 (0)	0.014
4	4.1 (2.1)	5.5 (2.1)	5.5 (2.1)	n.s.
5	7 (2.8)	5 (2.8)	4.5 (2.1)	n.s.
6	3.5 (2.1)	5.7 (0.7)	5 (0)	10^{-4}
7	<i>5.2 (0.3)</i>	<i>6.9 (2.9)</i>	<i>6.4 (2.9)</i>	n.s.
8	<i>3.7 (1.5)</i>	<i>5.5 (4.9)</i>	<i>5.2 (4.5)</i>	n.s.
9	5.9 (4.1)	11.2 (0.4)	9.8 (0.8)	5.10^{-4}
10	3.9 (2.7)	7 (0)	7 (0)	10^{-4}
11	5.2 (0.2)	5 (0)	5 (0)	n.s.

4. Discussion

In Dynamical Systems theory [7] the most common way of discovering new patterns is to cause a perturbation in the system and observe the emergence of regularities. Here we have used speech rate, one of the greatest causes of phonetic modification, in order to observe the emergence of new rhythmic structures.

Results have shown that DSR model is able to accurately explain the rhythmic variations due to speech rate increase. The interaction of this model's oscillators (syllabic and phrase-stress ones), through the manipulation of VV duration, can simulate what happens in real-speech when people talk fast.

The application of DSR model in our data revealed a gradual variation of rhythm with speech rate increase. This variation was observed through the duration contour of VV units along the sentences. In other words, our results show that speech rhythm is better explained if we take under consideration phonetic/quantitative aspects of languages.

5. Conclusion

The acoustic study described here corroborated previous acoustic results [8] [9] and improved previous hypotheses. The main results of our study, in case of rhythmic restructurings, are:

- the standard deviation of VV unit duration and stress group duration is smaller at faster rates;
- stress group duration tends to be constant with speech rate increase;
- the number of VV units (vowel-to-vowel) per stress group proportionally increases with speech rate increase;
- As standard deviation is consistently smaller at fast rates, speech rate increase exacerbates the mixture character of Brazilian Portuguese rhythm, i.e., tendencies to syllable as to stress-timed rhythm.

6. Acknowledgments

This work was supported by the following grant agencies: CAPES, CNPq (200199/2004-8), and FAPESP (03/09199-2).

7. References

- [1] Albano, E.; Barbosa, P. A.; Gama-Rossi, A.; Madureira, S.; Silva, A., 1998. A interface fonética-fonologia e a interação prosódia-segmentos.. In: *Estudos Lingüísticos XXVII, Anais do XLV Seminário do GEL - GEL 97*, Unicamp: Campinas, Brazil, [s.n.], 135-143.
- [2] Barbosa, P. A., 1994. *Caractérisation et génération automatique de la structuration rythmique du français*. PhD Thesis. Institut National Polytechnique de Grenoble, France.
- [3] Barbosa, P. A., 2002. Explaining Brazilian Portuguese resistance to stress shift with a coupled-oscillator model of speech rhythm production. *Cadernos de Estudos Lingüísticos* 43, 71-92. Unicamp: Campinas, Brazil.
- [4] Barbosa, P. A., 2007. From syntax to acoustic duration: A dynamical model of speech rhythm production. *Speech Communication* 49, 725-742.
- [5] Boersma, Paul; Weenink, David, 2007. *Praat: doing phonetics by computer* (Version 4.6.35) [Computer program]. Retrieved October 22, 2007, from <http://www.praat.org/>.
- [6] Kelso, J. A. S., 1995 *Dynamic Patterns : The Self-Organization of Brain and Behavior*. MIT Press, Cambridge, MA.
- [7] Kelso, J. A. S.; Saltzman, E.; Tuller, B. 1986. The dynamical perspective on speech production: data and theory. *Journal of Phonetics* 14, 29-59.
- [8] Meireles, A. R., 2007. *Reestruturações rítmicas da fala no português brasileiro*. PhD Thesis. Unicamp: Campinas, Brazil.
- [9] Meireles, A. R.; Barbosa, P. A., submitted. Articulatory-rhythmic reorganizations due to speech rate increase in Brazilian Portuguese. *Journal of the International Phonetic Association*.
- [10] Meireles, A. R.; Barbosa, P. A., submitted. Lexical reorganization in Brazilian Portuguese: an articulatory study. *Speech Communication: special issue on Iberian languages*.