# Articulatory Reorganizations of Speech Rhythm due to Speech Rate Increase in Brazilian Portuguese

*Alexsandro R. Meireles* [1], *Plínio A. Barbosa* [2]

[1] Phonetics Laboratory, Federal University of Espirito Santo, FAPES, Brazil
[2] Speech Prosody Studies Group, State University of Campinas, Brazil
meirelesalex@gmail.com, pabarbosa.unicampbr@gmail.com

## Abstract

This paper examines how speech rate increase acts to change speech rhythm at the articulatory level. Main results show that speech rate increase worked to change articulatory parameters in the following way: a) decrease of acceleration duration; b) decrease of y-extremum; c) decrease of constriction displacement; d) decrease in modulus of peak and/or valley velocity; e) decrease of gestural duration; and f) constant proportional time-to-peak (or valley) velocity. Besides, results have shown that speech rate tends to affect all gestures in an utterance independently of their phrasal position. Nevertheless, there was evidence that some articulatory parameters could, if properly manipulated, provide cues for rhythmic restructurings in speech. Finally, results show that the dynamical speech rhythm model (Barbosa, 2007) is more appropriate to deal with Brazilian Portuguese acoustical data than the pi-gesture model (Byrd & Saltzman, 2003), and that both models could explain articulatory reorganizations due to speech rate increase.

**Index Terms**: speech prosody, speech rhythm, speech rate, EMA, articulatory study.

## 1. Introduction

The speech rate influence on speech rhythmic reorganization can be explained by two recently proposed prosodic models: (i) the speech rhythm model [1, 2, 3] (henceforth SRM); and (ii) the π-gesture model [5] (henceforth PG). A rhythmic reorganization is considered here as a change in the temporal prosodic structure of articulatory gestures [9]. Both models account for prosodic structuring using a dynamical systems approach (cf. [7, 8]).

Byrd and Saltzman [5] state the main features of prosodic gestures: (i) prosodic gestures have a temporal extension and overlap with constriction gestures; (ii) prosodic gestures' gestural scores represent the activity of a set of abstract point attractors, in order to make the model as abstract as possible; (iii) prosodic gestures do not have an independent articulatory realization. Therefore, they are only indirectly realized by its effect on the articulatory dynamics.

Barbosa [2, 3] highlights the linguistic characteristics of SRM, as follows: (i) the linguistic rhythm is a consequence of the way the phrase-stress oscillator pulses align with the onset of lexically stressed vowels; (ii) the hierarchy between the magnitude of phrase-stress oscillator pulses generates a dynamical metrics; (iii) the relative coupling force is language dependent; (iv) the prosody-segments interaction is different among languages; (v) syntactic information is crucial, though not alone, to explain the variability of placement and magnitude of phrase stresses; (vi) phrasal stresses are a superficial consequence of the prominences expressed by peaks of abstract duration of the syllabic oscillator; (vii) lexical stress is defined in the abstract gestural score.

These two prosodic models (SRM and PG) predict similar phonetic consequences to syllable-sized gestures, namely: (i) gestures adjacent to prosodic boundaries will be lengthened; (ii) degree of slowing will be greatest as the gesture approaches a prosodic boundary; (iii) boundaries of different strengths are only expected to be distinct in degree of effect. Nevertheless, the scope of action in the SRM model is global, i.e., the phrase stress oscillator acts throughout the whole utterance. The longer vowel-to-vowel (from the beginning of one vowel up to the beginning of the next vowel, henceforth VV) duration at the end of a stress group (Brazilian Portuguese is a right-headed language at this level) is a cumulative function of previous VV durations from the beginning of this group, while the scope of action in the PG model is local: "the π-gesture locally slows the clock that controls the timeflow of an utterance" [5, p. 160]. Another difference between them regards their units of action. While in the SRM these units are articulatory gestures on the lexical level whose stiffness is modified by the effect of the phrase stress oscillator pulses at the properly rhythmic level, in the PG model, these units are gestures spanning a phrasal boundary. Finally, π-gesture action is limited to phrase edges, as can be noticed in the passage: "effects will be limited to gestures near the domain edge and will not occur at gestures quite remote from it" [5, p. 162].

Despite these differences, it is worthwhile to highlight that both models fall into the category of the so-called intrinsic timing, for in the π-gesture model "the activation level dynamics of the clock and the constriction level dynamics of the gestural units are bidirectionally coupled and hence form a single higher-order dynamical collective" [5, p. 156]. Barbosa's dynamical model of speech production [2, 3] works likewise, for the two coupled oscillators, through prosody-segments interaction, is bidirectionally coupled with the gestural score and thereby also form a single higher-order dynamical collective. Furthermore, prosodic timing is explicitly controlled for utterance production in the SRM, that is, the coupled oscillators in interaction with the other levels of grammar control speech rhythm, so that a separate abstract executor is not needed to do this control.

Based on the assumptions of the speech rhythm model, this paper intends (i) to study how speech rate acts to change the temporal structure of articulatory gestures, as related to linguistic rhythm, and (ii) to compare the predictions of both dynamical prosodic models (SRM and PG). To study speech rhythm at the articulatory side, we used the jaw as the basis for articulatory rhythm as proposed by Erickson [6].

Pursuing Erickson's idea of the jaw as a rhythm articulator, we conducted an experiment to examine how speech rate acts to change the phrasal prominences throughout the utterance. The novelty of the present study is to make a fine articulatory description of how speech rate variation works to change the durational patterns of articulatory gestures (defined in [4]).

10.21437/SpeechProsody.2014-40

Our main hypothesis is that phrasal prominences along the utterance are restructured with speech rate increase, and, as a consequence, stressed vowels under a phrasal boundary are realized with lesser jaw opening at fast rates. Besides, it is hypothesized that at fast rates there is no jaw movement reset after some minor prosodic boundaries. This hypothesis is based on Barbosa's studies [2, 3], which show that Brazilian Portuguese (henceforth BP) VV durations exponentially increases up to a phrasal boundary and then a reset of VV duration values occurs, i.e., after reaching its maximal duration, VV duration decreases and starts increasing all over again up to the next phrasal boundary.

## 2. Methods

A female native speaker of BP (age 28-30) was recorded acoustically (sampling rate: 22.5 kHz) and articulatorily at the USC Phonetics Laboratory. The speaker was paid for the participation in the experiment and signed an approved informed consent form explaining the purpose of the experiment. A 2-D Articulograph AG-200 (www.articulograph.de) (cf. [12], EMA magnetometer system) was used for tracking jaw movement. The movement data was sampled at 200 Hz, head-corrected, rotated to the occlusal plane, and low-pass filtered at 25 Hz. Pellets were attached to the following articulators: tongue (close to the palatal region), lower lips, jaw (at the lower incisors). Two other pellets were used as reference for the signal acquisition system: one at the nose bridge, and one at the center anterior surface of the maxillary incisors. Only the y-movement of the jaw was measured. As in Erickson [6], jaw opening was measured "in terms of the lowest vertical position of the mandibular pellet in the syllable from the maxillary occlusal plane" [6, p. c-134].

The recorded sentences are displayed in Table 1. Ten repetitions of each sentence (randomized within blocks) at three speech rates were recorded. This results in a total of 120 utterances for analysis (4 sentences x 10 repetitions x 3 speech rates).

To obtain three distinct speech rates, the subject was asked to read the sentences according to the following instructions and order: (1) normal: speak in a comfortable way; (2) slow: speak as slow as you can preserving the sentence's meaning and without introducing pauses between words; (3) fast: speak as fast as you can without introducing distortions in speech.

Table 1. *Sentences used in the experiment with their respective translation (TR) and VV phonetic transcription (PT). Bolded words represent where phrasal prominence is expected to fall (no such markings appeared in the stimuli for reading.*

| Sentence 1 | Ela diz mão de **máfia** no carro da moça do **papai**. |
|---|---|
| PT | [ɛl.ɐdʒ.izm.ãʊdʒ.ɪm.af.ɪɐn.ʊk.aɤ.ʊd.am.os.ɐd.ʊp.ap.aɪ] |
| TR | She says mafia's hand in the car of my father's girl. |
| Sentence 2 | Foi bão gostar demais de **papai**, mas de **máfia**!? |
| PT | [f.oɪb.ɐʊg.ost.aɤdʒ.ɪm.aɪzdʒ.ɪp.ap.aɪm.azdʒ.ɪm.af.ɪɐ] |
| TR | It were good to love too much my dad, but not mafia... ["were" instead of "was" implies a very informal style] |
| Sentence 3 | Mamãe não quer mais qu'eu **babe**, mas qu'eu **pape**. |
| PT | [m.ãm.ãɪn.ãʊk.ɛɤm.aɪsk.ɪeʊb.ab.ɪm.ask.ɪeʊp.ap.ɪ] |
| TR | My mother doesn't want that I dribble anymore, but that I eat [child language]. |
| Sentence 4 | Vou lá levar o **pavê** pra filha do **papai**. |
| PT | [v.oʊl.al.ev.aɾ.ʊp.av.epɾ.af.iʎ.ɐd.ʊp.ap.aɪ] |
| TR | I am going there to deliver the "pavê [kind of dessert]" to dad's daughter. |

For transcription and labeling, MAVIS software [14], modified at University of Southern California, was employed to measure the jaw sensor movement in the vertical dimension (y-axis). The following articulatory variables were measured (see [9, p. 139] for details): (i) jaw maximum extension: measured at y-velocity zero-crossing at maximum opening; (ii) jaw constriction displacement: measured as the difference between the zero crossings at constriction onset and extremum; (iii) jaw gesture (related to acoustic) duration: measured as the interval between maximum peak velocity[1] and maximum (modulus) valley velocity; (iv) jaw gesture duration: measured between y-velocity zero-crossings at constriction onset and maximum; (v) constriction jaw gesture peak/valley velocity (y); (vi) jaw acceleration duration: measured from the time of zero-crossing at constriction onset up to the time of constriction peak (or valley) velocity; (vii) proportional time-to-peak (or valley) velocity: measured from the ratio of y acceleration duration to total constriction formation duration.

## 3. Results

As objective means of automatically detecting stress group boundaries is not yet available for articulatory data, acoustic analyses were run to detect the stress groups' boundaries following the procedures presented in Meireles' papers [9, 10, 11]. Also, subsequent statistical analyses have shown that the duration of the articulatory VVs were not significantly different from the duration of the acoustic VVs.

Our results suggest that speech rate increase tends to strengthen the right-headness characteristic of BP, i.e., the greatest phrasal prominences occur to the right of the sentence at fast rates. These greatest phrasal prominences are considered here as the greater duration of a VV unit in comparison with the other VV unit. For sentence 1, with VV duration as a function of the articulatory VVs (related to the acoustic one) [afɪ] and [aɪ], the greatest phrasal prominence occurred at [afɪ] for slow ($F(1,4) = 15.858$, $p < 0.017$) and normal rates ($F(1,10) = 72.681$, $p < 0.0002$), and at [aɪ] for fast rate (n.s.). For sentence 2, with VV duration as a function of the articulatory VVs [ap] and [afɪ], the greatest phrasal prominence occurred at [ap] for slow ($F(1,8) = 85.689$, $p < 0.00003$) and normal rates ($F(1,6) = 22.611$, $p < 0.0032$), and at [afɪ] for fast rate (n.s.). For sentence 3, with VV duration as a function of the articulatory VVs [ab] and [ap], no statistical difference was found for slow rate, and for normal ($F(1,6) = 6.0367$, $p < 0.0494$) and fast rates ($F(1,12) = 15.042$, $p < 0.003$) the greatest phrasal prominence occurred at [ap]. For sentence 4, with VV duration as a function of the articulatory VVs [epɾ] and [aɪ], there was no statistical difference among rates. Nevertheless, as these one-way ANOVAs confirmed this tendency only for sentence 3, further studies are necessary to corroborate this hypothesis.

A statistical comparison of the sentences' vowels (cf. table 2) as a function of rate indicates a significant general tendency for smaller displacements (y-extremum and constriction

[1] As we are using velocity in modulus, all general references to peak velocity also applies for valley velocity. So, the hypothesis in fig. 1 for peak velocity can also be extended to valley velocity.

[2] VVs with 2 vowels mean that the second vowel was acoustically produced with a voiceless pattern and, thus, included in this syllable-like unit.

displacement) with speech rate increase. High and mid-high vowels tend to be less high, and low vowels tend to be less low from slow to fast rates (see table 3 and [9, p. 148-149] for details).

A comparison of the peak velocity (consonants) or valley velocity (vowels) as a function of rate for all gestures suggests a decreasing of maximum velocity (modulus) from slow to fast rate. Although this decreasing pattern was statistically found for some gestures within the sentences (SENTENCE 1: 2 out of 15, SENTENCE 2: 5 out of 12, SENTENCE 4: 7 out of 12), only a general pattern of decreasing maximum velocity was found for the gestures in sentence 3 (F(2,191) = 4.7334, p < .0099, using maximum velocities with their absolute values). Therefore, future studies are needed in order to support this hypothesis of gesture's maximum velocity decrease with speech rate increase.

Table 2. *Vowels (bold type) used in a one-way ANOVA with y-extremum and constriction displacement as a function of rate. Some vowels were not analyzed, because they occurred at the same jaw movement of the preceding consonant/vowel.*

| | |
|---|---|
| Sentence 1 | Ela diz mão de m**á**fi**a** no c**a**rro d**a** m**o**ç**a** do p**a**p**ai**. |
| Sentence 2 | Foi bão gostar dem**ai**s de p**a**p**ai**, m**a**s d**e** m**á**fi**a**!? |
| Sentence 3 | Mamãe não quer mais qu'**eu** b**a**b**e**, m**a**s qu'**eu** p**a**p**e**. |
| Sentence 4 | Vou lá lev**a**r o p**a**v**ê** pr**a** f**i**lh**a** do p**a**p**ai**. |

Rate effects on both articulatory and acceleration duration revealed a decreasing duration pattern from slow to fast rate (ARTICULATORY DURATION: SENTENCE 1, $F$ (2,183) = 13.080, p < $10^{-5}$, SENTENCE 2, $F$ (2,141) = 9.1624, p < .0002, SENTENCE 3, $F$ (2, 177) = 52.992, p < $10^{-4}$, SENTENCE 4, $F$ (2,273) = 76.072, p < $10^{-4}$; ACCELERATION DURATION: SENTENCE 1, $F$ (2,183) = 12.549, p < .00002, SENTENCE 2, (2,141) = 12.382, p < .00002, SENTENCE 3, $F$ (2,176) = 34.077, p < $10^{-5}$, SENTENCE 4, $F$ (2,273) = 30.905, p < $10^{-5}$).

Analysis of the proportional time-to-peak/valley-velocity as a function of rate indicates no common pattern for the gestures. Some gestures had a downward pattern and others an upward pattern from slow to fast rate. Thereby, stressed and unstressed vowels were grouped together. This grouping still indicates no rate effects on proportional time-to-peak/valley-velocity, but it indicates that stressed vowels have smaller proportional time-to-peak/valley velocity than unstressed vowels (SENTENCE 1, $F$ (1,91) = 9.6706, p < .003; SENTENCE 2, $F$ (1,70) = 5.4489, p < .023; SENTENCE 3, n.s.; SENTENCE 4, $F$ (1,136) = 10.051, p < .002).

In summary, the data indicate a stiffness increase from slow to fast rate, and, consequently, a shrinking of the gestures and smaller spatial movements. Therefore, rhythmic restructurings are more likely to occur at fast rates, since stress groups (henceforth SG) ending at weak boundaries in slower rates may delete due to stiffness increase in fast rates. It is worth to recall that longer durations are necessary to delimit phrasal boundaries in BP.

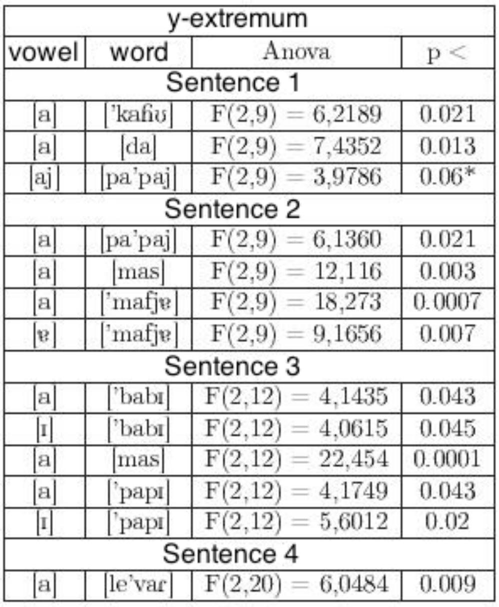
### 3.1. Kinematic sources of rhythmic restructurings

Based on the articulatory results outlined above, we proposed the following kinematic sources of rhythmic restructurings with speech rate increase (cf. figure 1): a) decrease of acceleration duration; b) decrease of y-extremum; c) decrease of constriction gesture displacement; d) decrease of peak/valley velocity (modulus); e) decrease of articulatory duration; and f) constant time-to-peak/valley-velocity.

Another kinematic source of rhythmic restructuring not related to speech rate change is the diminishment of the proportional time-to-peak/valley-velocity on stressed vowels compared to unstressed ones.

According to these sources of articulatory patterning, a general decrease of articulatory parameters from slow to fast rate is expected. Figure 1a represents an articulatory shortening caused by a stiffness change (cf. [13]), which indicates a shorter acceleration duration. Figure 1b exhibits a smaller y-extremum caused by a less open jaw position (low vowels) or less closed jaw position (high mid-high vowels), which can be understood as a gesture undershoot. Closely related to figure 1b is figure 1c that represents a smaller difference of y displacement at the initial and final positions of the gesture. If a smaller y-extremum is expected, a smaller constriction displacement at fast rates is also expected. Figure 1d shows a peak velocity decrease from slow to fast rate. This hypothesis can be explained by the fact that to reach a greater peak velocity followed by a change in movement direction at zero velocity (jaw goes up and down), a greater distance is needed, which can be found at slow rates, not fast ones. Figure 1e is merely a consequence of the previous kinematic sources, i.e., diminishment of articulatory duration. Finally, figure 1f shows that the time-to-peak/valley-velocity keeps constant with speech rate increase.

Table 3. *Significance of y-extremum as a function of speech rate. This table represents vowels with their respective word, ANOVA and significance for all sentences in the corpus. *means marginally significant.*



| y-extremum | | | |
|---|---|---|---|
| vowel | word | Anova | p < |
| Sentence 1 | | | |
| [a] | ['kaɦʊ] | F(2,9) = 6,2189 | 0.021 |
| [a] | [da] | F(2,9) = 7,4352 | 0.013 |
| [aj] | [pa'paj] | F(2,9) = 3,9786 | 0.06* |
| Sentence 2 | | | |
| [a] | [pa'paj] | F(2,9) = 6,1360 | 0.021 |
| [a] | [mas] | F(2,9) = 12,116 | 0.003 |
| [a] | ['mafjɐ] | F(2,9) = 18,273 | 0.0007 |
| [ɐ] | ['mafjɐ] | F(2,9) = 9,1656 | 0.007 |
| Sentence 3 | | | |
| [a] | ['babɪ] | F(2,12) = 4,1435 | 0.043 |
| [ɪ] | ['babɪ] | F(2,12) = 4,0615 | 0.045 |
| [a] | [mas] | F(2,12) = 22,454 | 0.0001 |
| [a] | ['papɪ] | F(2,12) = 4,1749 | 0.043 |
| [ɪ] | ['papɪ] | F(2,12) = 5,6012 | 0.02 |
| Sentence 4 | | | |
| [a] | [le'var] | F(2,20) = 6,0484 | 0.009 |

## 4. Discussion

As seen in the last session, main results have shown that speech rate increase seems to change the articulatory parameters in a uniform way, disregarding their phrasal position. Despite that, there is evidence that some factors may be able to explain the rhythmic restructurings found at the articulatory level in BP. More data and speakers are needed to confirm these trends, however. Normalization in the articulatory signals is needed in order to be able to obtain a consistent methodology to work with them. Recall that acoustic stress group boundaries were found based on normalization procedures applied to the acoustical signal only (cf. [3, 9]). Furthermore, a comparison of the PG and SRM models have led us to different conclusions on the acoustic and the articulatory side

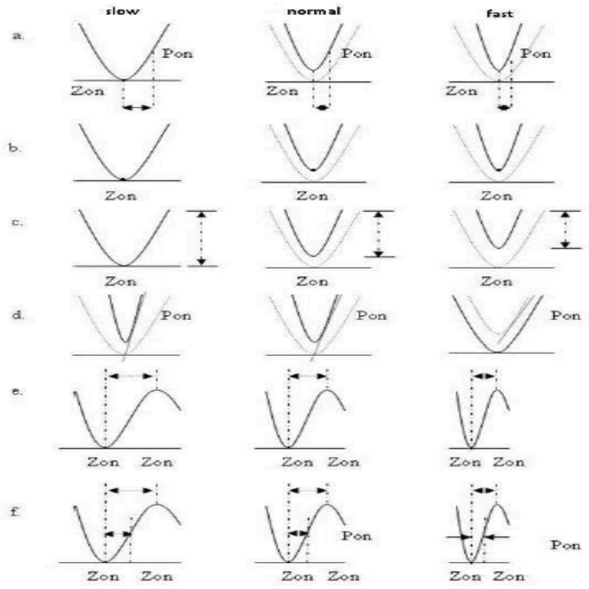


Figure 1. *Kinematic sources of rhythmic restructurings with speech rate increase: a. diminishment of acceleration duration (stiffness increase); b. y-extremum decrease; c. constriction gesture displacement decrease; d. peak velocity decrease; e. articulatory duration decrease; f. constant proportional time-to-peak velocity.*

On the acoustic side, according to Barbosa [1, 2, 3] and Meireles's acoustical data [9, 10, 11], since in BP the greater duration at the end of a SG is a consequence of a exponential increase from the beginning of this SG, SRM model better explains the rhythmic variations found in this paper. Recall that PG model only explains longer durations at (or near) a phrasal boundary.

Yet, on the articulatory side, since we only found some possible influence of articulatory parameters exactly at the places where rhythmic restructurings occurred, both models would work to explain the results found up to this point. However, we remind that because PG researchers have specifically worked with articulatory data, they are more advanced at the methodological side to work with articulatory gestures influenced by prosodic structure. Despite that, we have showed here some evidence that SRM model could perfectly deal with such data once more information about BP gestures is provided and new methods applied to articulatory data are designed.

## 5. Conclusions

The results of this articulatory study suggest that speech rate increase affects all gestures in a sentence disregarding their phrasal position. Nevertheless, future work is needed since acceleration duration, constriction displacement and articulatory duration seem, though not conclusive, to reflect the rhythmic restructurings found on the articulatory side. Also, results showed that the SRM model is more appropriate to deal with BP acoustical data than the PG model, and that both models could explain the articulatory variations due to speech rate increase.

Finally, this paper's main results have shown that rhythmic structure variation is modified gradually with speech rate increase, i.e., quantitative aspects of speech are acting to modify speech rhythm, showing how a dynamical systems approach to language perfectly suits to linguistic descriptions

## 6. Acknowledgements

This work was supported by FAPES (54652499/2011), and NIH (DC03172, D. Byrd). The authors thank Elliot Saltzman, Louis Goldstein, and Hosung Nam for their insightful comments on this research. Many thanks are due to USC Phonetics Laboratory, Haskins Laboratories, Dani Byrd and Sungbok Lee, for help with the articulatory design and support, and James Mah, for help in the articulatory recordings. Also, we would like to thank the anonymous reviewers for the remarkable comments.

## 7. References

[1] 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.

[2] Barbosa, P. A., Incursões em torno do ritmo da fala, Campinas, Brazil: Pontes/Fapesp, 2006.

[3] Barbosa, P. A., "From syntax to acoustic duration: a dynamical model of speech rhythm production", Speech Communication, 49:725–742, 2007.

[4] Browman, C. and Goldstein, L., Articulatory gestures as phonological units. Phonology, v. 6, p. 201–251, 1989.

[5] Byrd, D. and Saltzman, E., The elastic phrase: modeling the dynamics of boundary-adjacent lengthening. Journal of Phonetics, v. 31, p. 149-180, 2003.

[6] Erickson, D., On phrasal organization and jaw opening. In: From Sound to Sense: 50+ Years of Discoveries in Speech Communication, MIT, June 11-13. [S.l.: s.n.], 2004. p. S15.

[7] Kelso, J. A. S., Dynamic patterns: the self-organization of brain and behavior. Cambridge, USA: MIT Press, 1995.

[8] Kelso, J. A. S., Saltzman, E. L. and Tuller, B., The dynamical perspective on speech production: data and theory. Journal of Phonetics, v. 14, p. 29–59, 1986.

[9] Meireles, A. R., Self-organizing rhythms in Brazilian Portuguese: speech rate as a system perturbation. Germany: VDM Verlag, 2009.

[10] 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.

[11] Meireles, A. R., Tozetti, J. P. and 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.

[12] Perkell, J., Cohen, M., Svirsky, M., Matthies, M., Garabieta, I. and Jackson, M., Eletromagnetic midsagittal articulometer (emma) systems for transducing speech articulatory movements. J. Acoust. Soc. Am., p. 3078–3096, 1992.

[13] Saltzman, E. and Munhall, K. G., A dynamical approach to gestural patterningin speech production. Ecological Psychology, v. 1, n. 4, p. 333–382, 1989.

[14] Tiede, M. K., Vatikiotis-Bateson, E., Hoole, P. and Yehia, H., Magnetometer data acquisition and analysis software for speech production research. ATR Technical Report TR-H 1999. Kyoto, Japan:ATR Human Information Processing Labs, 1999.