Investigation of non-pitch-accented phrases in Brazilian Portuguese: no evidence favoring stress shift

Plínio A. Barbosa and **Pablo Arantes**

LAFAPE & Dep. of Linguistics/IEL/University of Campinas, Brazil

E-mail: {plinio,pablo}@iel.unicamp.br

ABSTRACT

The aim of this paper is twofold: (a) to investigate possible differences in durational patterns within phrase stress groups contrasting stress clash vs nonclash condition, and (b) to test the assumptions of a coupled-oscillator model of speech rhythm production. Measures of duration for relevant units in isolated read sentences uttered by four subjects were statistically analyzed and tonal events were annotated. The pattern that emerges from the analyses is that of a monotonous increasing of duration of syllable-sized units when approaching phrasal stress. These results contradict the Rhythm Rule and seem to indicate that no apparent, systematic, duration-related stress shift seems to take place in Brazilian Portuguese.

1. INTRODUCTION

In the 1990s, two studies presented in the Laboratory Phonology series tried to find acoustic correlates for the application of the Rhythm Rule (RR) [5,9], defined in the framework of Metrical Phonology as an optional rule favoring stress shift when stresses clash, in order to ensure a strong-weak prominence alternation [6]. Since stress is relational, two analyses for implementing the RR are logically possible for disyllabic, to-be-shifted (target) words: the Reversal Analysis (RA), by which the prominence of the first syllable of the target word is reinforced, and the Deletion Analysis (DA), by which the prominence of the second syllable of the target word is diminished.

Grabe and Warren's study [5] compared 42 pairs of sentences in Southern British English contrasting two-word sequences in clash vs nonclash conditions. In clash condition the two words (target followed by trigger) belong to the same phonological phrase (trigger word in nuclear position), whereas in nonclash condition there is a phonological phrase boundary between the two words, with the target word in nuclear position. All sequences were perceptually evaluated as regards the occurrence of stress shift. Since the authors only considered the possibility of the RA, their data failed to find strong acoustic evidence for the application of the RR in the perceptually shifted items, for all phonetic parameters analyzed. By choosing to contrast the target words in so distinct phonological environments (prenuclear vs nuclear positions) they were unable to separate the effects of this contrast from the contrast of interest (clash vs nonclash). Vogel et al.'s study [9] were more careful in this respect. By contrasting two-word sequences in Northeastern American English according to three distinct preceding contexts, clash vs nonclash condition, and number of syllables of target words, they showed that there are indeed duration and pitch effects on the second rhyme of the target word, which is less prominent than the first rhyme. Their results favor the DA and suggest that the effect would be restricted to the rhymes.

Based on previous descriptions of the English intonation system, Shattuck-Hufnagel [8] suggests that phenomena such as pitch placement alternation, early prominence, and the obligatory position of nuclear pitch accent would better explain the findings above. She proposed an intonational phonological view of early pitch accent placement in American English to explain accent placement in relation to the stress shift issue. She also suggested that prior shift of stress does not seem to be required for implementing pitch accent alternation, and that the investigation of stress shift in non-pitch-accented stretches of speech is crucial to shed new light on the matter.

Besides investigating the stress clash/shift issue in Brazilian Portuguese (BP), a language with (usually) lesser tonal events along isolated read utterances than English, we propose to raise the discussion to an explanatory level by simulating the stress patterns along utterances with a coupled-oscillator model of speech rhythm production.

2. CONTRASTING STRESS CLASH VS NONCLASH CONDITIONS IN BP

Two experiments were conducted in order to examine possible differences in duration for three units (rhymes, syllables and VV units – delimited by two consecutive vowel onsets) in the first words of a two-word sequence embedded in sentence pairs. First words (targets) in a pair are disyllabic oxytons followed either by a disyllabic paroxyton (thus creating a stress clash, and henceforth referred to clash condition) or by a disyllabic oxyton (not triggering a clash, and henceforth referred to nonclash condition). Second words in the sequence are referred to trigger words here, whether or not they create a clash. All sentences were prosodically annotated with respect to boundary strength and tonal events.

In the first experiment, one subject (São Paulo state, 25 years old) read ten repetitions of four isolated paired

sentences (sentences 1 to 4: target words in bold, lexical stress in trigger words indicated by a diacritic, A sentences represent the nonclash condition, whereas B sentences, the clash condition). The pairs differ in number of syllables, syntactic structure and location of nuclear accent. In sentences 1, 2 and 4 trigger words carry the nuclear pitch accent. In sentences 3, the two-word sequence is in post-focal position. Smoothed f_0 traces were obtained with Praat 4.0.7. The pattern for the eight sentences is the same, namely a start at the middle of the speaker register, followed by the onset of rise (see figures 1 and 2) within the target word– except for sentences 3 –, and a downdrift towards the end of the utterance to the bottom of the speaker register. With a possible exception of sentence 1 (cf. section 4), target words have no tonal events. Statistical analyses (2-way ANOVA with Scheffé post-hoc test) revealed no differences in f_0 values between first and second vowels of the target word, with the exception of sentence 1B ($p < 10^{-3}$).

Eu **comi** bolór sexta-feira à noite.	(1A)
Eu **comi** bólo sexta-feira à noite.	(1B)
O **bordeaux** chinês derramou-se pela mesa.	(2A)
O **bordeaux** xúcro derramou-se pela mesa.	(2B)
Um lindo **bebê** carmím.	(3A)
Um lindo **bebê** cálvo.	(3B)
Parece que **falou** 'baixóu', e não 'caiu'.	(4A)
Parece que **falou** 'báixo', e não 'caiu'.	(4B)

S	r			σ			VV		
	A	B	p <	A	B	p <	A	B	p <
1	34 /o/	30 /o/	ns	129 /ko/	137 /ko/	0.07	**89 /om/**	**101 /om/**	**0.02**
	59 /i/	59 /i/	ns	118 /mi/	126 /mi/	0.06	**148 /ib/**	**164 /ib/**	**10^{-2}**
2	96 /or/	93 /or/	ns	174 /bor/	166 /bor/	ns	165 /ord/	173 /ord/	ns
	81 /o/	87 /o/	ns	**151 /do/**	**167 /do/**	**0.02**	**195 /oʃ/**	**253 /oʃ/**	**10^{-7}**
3	66 /e/	67 /e/	ns	134 /be/	129 /be/	ns	135 /eb/	142 /eb/	ns
	75 /e/	77 /e/	ns	143 /be/	151 /be/	ns	**174 /ek/**	**200 /ek/**	**10^{-3}**
4	47 /a/	41 /a/	ns	163 /fa/	165 /fa/	ns	106 /al/	96 /al/	0.06
	83 /o:/	85 /o:/	ns	142 /lo:/	139 /lo:/	ns	**150 /o:b/**	**183 /o:b/**	**10^{-7}**

Table 1: Speaker 1 mean duration (in ms) and ANOVA p-values for rhymes (r), syllables (σ) and inter-vowel onset units (VV) in non-clash (A) and clash (B) conditions for sentences 1 to 4 for the two positions of the target words. Corresponding segments are indicated within slashes. Contrasted values with significant differences in bold.

All segmental durations were delimitated with Praat. Table 1 presents the mean durations of rhymes (r), syllables (σ) and VVs for the two positions (different rows) and the two conditions (different columns) of the target words. Since there is a great segmental diversity across sentences, ANOVA p-values are presented for each nonclash/clash contrast.

Results indicate that a difference in duration seem to be restricted to the second position of the target word, for syllable-sized units only (syllables in a minor sense, and VV more strongly). Three-way ANOVA using CONTEXT (clash vs nonclash), UNIT POSITION (first vs second in the target word) and SENTENCE (1 to 4) as factors confirm this first inspection. (The SENTENCE factor was highly significant for all dependent variables, $p = 0$, due to the clear segmental differences between the sentences. Scheffé post-hoc tests also revealed that the four sentences significantly differ from each other.)

When the dependent variable is the rhyme duration, the results exhibit significance for UNIT POSITION ($F_{1,158} = 135.39$, $p = 0$), while CONTEXT is not significant, even if only the second rhymes are considered for analysis. The same holds for syllable duration: significance for UNIT POSITION ($F_{1,147} = 15.7$, $p < 10^{-4}$), while CONTEXT is not significant ($F_{1,147} = 3.11$, $p < 0.08$). There is significant interaction between all combinations of factors. The interaction between UNIT POSITION and CONTEXT reveals particularly that significant differences can be restricted to one of the positions. Indeed, considering only the second syllables in the analysis, CONTEXT ($F_{1,147} = 8.84$, $p = 0.004$) becomes significant. The interaction between CONTEXT and SENTENCE is no longer significant, revealing the consistency of the clash/nonclash contrast across the sentences. When the dependent variable is the VV duration, the results exhibit significance for UNIT POSITION ($F_{1,147} = 921.73$, $p = 0$), and CONTEXT ($F_{1,147} = 96.56$, $p = 0$), with all interactions significant.

Results indicate that, for syllable-sized units, durations are longer in second position for clash and nonclash conditions, since this position is closer to phrasal stress (in trigger word). But more interestingly (because contrary to the Metrical Phonology prediction), syllable-sized durations in second position are longer for clash condition for the same reason: closer to the phrasally stressed syllable. The effect is more clear for VV units, since phrasally stressed consonants belong to them. In sentences 1, it seems that this effect affects the first VV marginally. A possible explanation for this result is discussed in section 4 in connection to early pitch accent placement.

The second experiment was designed in such a way as to examine a possible influence of the stress status of the left context, besides extending the analyses for three more subjects (2 to 4, São Paulo state, between 25 and 30 years old). The two pairs of sentences below have the same number of syllables (stressed syllables in bold, words whose durations were measured underlined) and were read five times. The target word is the oxyton "café" (coffee) followed either by a paroxyton ("frio", cold) or an oxyton ("Pilão", proper name), as in the first experiment. The word to the left of "café" is either a paroxyton (5) or an oxyton (6). Statistical analyses revealed no differences in

f_0 at the middle of the vowel for paired sentences (cf. figure 1). Trigger words carry nuclear pitch accent ("Pilão" or "frio"). Stress groups separated by '|'.

Ele e vo**cê** | não **to**mam ca**fé Pilão** | de **no**vo.   5A)
Ele e vo**cê** | não **to**mam ca**fé frio** | de **no**vo.   (5B)

Não pode**ri**a | to**mar** ca**fé Pilão** | de **no**vo.   (6A)
Não pode**ri**a | to**mar** ca**fé frio** | de **no**vo.   (6B)
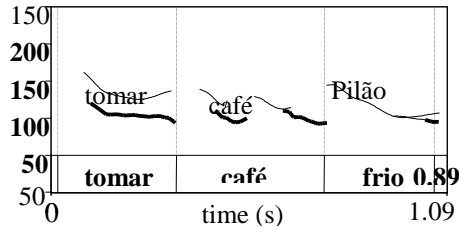

Figure 1: F_0 evolution of underlined sequence in clash (thick line) and non clash condition (plain line), sentences 6. Bold numbers refer to the scale for clash condition.

Tables 2, 3 and 4 present the results for durations of the three dependent variables in the target word (position in columns, condition in rows) for subjects 2, 3 and 4, respectively. Only the first position for VVs was measured since contrasted second VV have different segments.

S		r		σ		VV
	A	68	**69**	143	**160**	159
5	B	72	**87**	145	**178**	163
	p <	ns	10^{-3}	ns	**.003**	ns
	A	58	**73**	**133**	**167**	152
6	B	60	**92**	**148**	**187**	154
	p <	ns	10^{-5}	10^{-3}	10^{-3}	ns

Table 2: Speaker 2 mean duration (in ms) and ANOVA p-values for rhymes (r), syllables (σ) and VV units in non-clash (A) and clash (B) conditions for sentences 5 and 6. Contrasted values with significant differences in bold.

For a more consistent two-way ANOVA, means of durations for each subject instead of repeated observations were considered for analysis (which gives more realistic degrees of freedom). The two factors were CONTEXT (clash vs nonclash) and SENTENCE (5 vs 6) for the second position of rhymes and syllables. The dependent variable distributions were homocedastic across the three subjects.

When the dependent variable was the syllable duration, the results exhibit no significance for both factors and interaction. When the dependent variable was the rhyme duration, the results exhibit marginal significance for CONTEXT ($F_{1,8}$ = 5.24, p < 0.06) but not for SENTENCE ($F_{1,8}$ = 0.7, p = 0.43).

For this experiment the rhyme tends to exhibit the same pattern of result of the first experiment for the CONTEXT

factor (in second position). For these sentences, the left context has no influence on the rightward duration increasing towards phrasal stress. The results of both experiments will be validated in the light of a coupled-oscillator model of rhythm production [2,3].

S		r		σ		VV
	A	58	47	122	**123**	134
5	B	55	55	124	**137**	137
	p <	ns	ns	ns	**.03**	ns
	A	49	**50**	113	**123**	122
6	B	50	**63**	122	**149**	136
	p <	ns	**.004**	ns	**.02**	ns

Table 3: Speaker 3 mean duration (in ms) and ANOVA p-values. See legend of table 2 for details.

S		r		σ		VV
	A	51	**56**	127	**121**	115
5	B	48	**71**	130	**141**	118
	p <	ns	**.01**	ns	**.02**	ns
	A	48	**57**	131	**130**	122
6	B	42	**83**	130	**151**	110
	p <	ns	10^{-4}	ns	**.003**	ns

Table 4: Speaker 4 mean duration (in ms) and ANOVA p-values. See legend of table 2 for details.

3. SIMULATING THE RESULTS WITH THE MODEL OF RHYTHM PRODUCTION

A coupled-oscillator model composed by a phrase stress oscillator implemented by a train of pulses, and a syllabic oscillator implemented by a sinusoidal function was proposed elsewhere [2]. Since the pulses of the phrase stress oscillator are aligned with stressed vowel onsets, they characterize a linguistic rhythm in the traditional sense [6]. The phrase stress oscillator is related to the hypothesized universal property of stressing.

The syllabic oscillator implements an aspect of the so-called syllabicity, that one related to the vowel continuum [7]. The syllabic oscillator periods are nevertheless abstract, in the sense that their values do not specify overt durations, since the entrained durations delivered by the coupled-oscillator model stand for extrinsic timing. As to regard to intrinsic timing, the model is able to explain segmental duration as a consequence of a gestural perturbation upon the entrained syllabic oscillator. Gestures are specified in the lexicon [1].

Period coupling in the model is implemented by the set of equations (1) and (2), for 0 < n < N-1.

$$s\,(0) = w_0.\exp\,(-N + 2), \text{ and } s(N-1) = 0.05 \quad (1a)$$
$$s\,(n) = (1 - w_0).s(n-1) + w_0.\exp(-N + n + 1) \quad (1b)$$

$$\Delta T = \alpha.T.s(n).i(n) - \beta.(T-T_0).i(n-1) \quad (2)$$

The function s (.) is the synchrony function, which measures the degree of synchrony between the syllabic and phrasal oscillators. The parameters α, w_0, and T_0 are respectively the entrainment rate, the relative coupling strength, and the uncoupled syllabic-oscillator period. The function i (.) stands for the on-going phrase stress oscillator amplitude and T is the current syllabic oscillator period. The syllabic-oscillator period resetting, implemented by the second term of (2), operates after each phrasal stress with a decay rate specified by β during two cycles of the syllabic oscillator. This term is only present if the current VV unit is dominated by an on-going phrase stress. The current period T is updated only at syllabic oscillator maxima by iterating (2).

The values of w_0, α, and β were optimized in order to minimize the error between the evolution of syllabic oscillator periods, and the mean of VV-duration evolution along 3-VV and 4-VV stress groups in an ad hoc corpus. The choice of 3- and 4-VV stress groups was made in order to ensure statistical reliability. More details in [3]. By simulating the stress groups for the sentences of the first experiment, with (α, β, w_0, T_0, i) = (0.38, 1.13, 0.78, 138 ms, 1.0), the following, non-overt values are obtained for the two conditions (nonclash/clash) for the second VV of the target words: (163 ms, 212 ms), sentences 1 and 2; (161 ms, 204 ms), sentence 3, and (165 ms, 218 ms), sentence 4. This gives a 30 % average difference between the two conditions. Overt, measured differences from table 1 give differences of 11 % (s. 1), 30 % (s. 2), 15 % (s. 3), and 22 % (s. 4).

Despite the non coincidence of figures, the general pattern is the same: the model predicts that the closer the phrasally stressed VV is to a preceding unit, the longer is the duration of the latter. This is so because the model is insensitive to stress clash inside a stress group (delimitated by two consecutive phrasal stresses). As regards patterns of duration, the only possibility is a rightward increase of duration inside the group, thus predicting the data presented here. Data and model predictions are in complete opposition to metrical-phonological-based predictions for BP.

4. FINAL REMARKS

For sentence 1B, the duration of the first VV is also lengthened in clash condition (see table 1). A possible explanation is a pitch-related duration lengthening, since this sentence is the only one with statistically distinct f_0 pattern in the target word, as can be seen in figure 2. In clash condition, the phrasally stressed (and accented) VV is one position to the left in comparison to nonclash condition, which could explain the early f_0 peak during the target word. This f_0 peak is audible only paradigmatically, by careful listening of the target words in sequence.
This means that, if the durational patterns do not favor stress shift, there is room for an optional, possibly limited to particular speaking styles, accent shift.
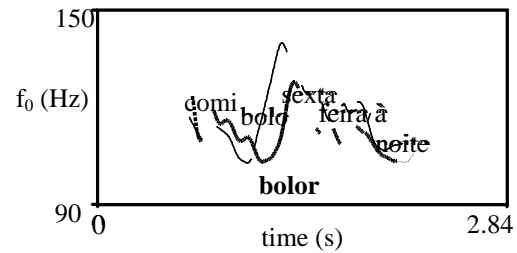

Figure 2: Contrast between f0 excursions in clash (plain) and nonclash conditions (thick line) for sentences 1.

REFERENCES

[1] E. C Albano, *O Gesto e suas bordas: esboço de fonologia acústico-articulatória do português brasileiro*. Campinas, Brazil: Mercado de Letras, 2001.

[2] P. A. Barbosa, "Explaining cross-linguistic rhythmic variability via a coupled-oscillator model of rhythm production", *Proceedings of Speech Prosody 2002*, Aix-en-Provence, France, pp. 163-166, 2002.

[3] P. A. Barbosa, "Explaining Brazilian Portuguese resistance to stress shift with a coupled-oscillator model of speech rhythm production", *Cadernos de Estudos Lingüísticos*, 43, forthcoming.

[4] B. Connell and A. Arvaniti (eds.), *Phonology and phonetic evidence: papers in Laboratory Phonology IV*. Cambridge University Press: Cambridge.

[5] E. Grabe and P. Warren, "Stress shift: do speakers do it or do listeners hear it?", In: [4], pp. 95-110, 1995.

[6] M. Liberman and A. Prince, "On stress and linguistic rhythm, *Linguistic Inquiry,* 8 (2), pp. 249-336, 1977.

[7] S. Öhman, "Coarticulation in VCV utterances: spectrographic measurements", *J. Acoustic. Soc. Am.*, 39, pp. 151-168, 1966.

[8] S. Shattuck-Hufnagel, "The importance of prosodic transcription in empirical approaches to 'stress shift' versus 'early accent': comments on Grabe and Warren, and Vogel, Bunnell and Hoskins", In [4], pp. 128-140, 1995.

[9] I. Vogel, H. T. Bunnell and S. Hoskins, "The phonology and phonetics of the Rhythm Rule", In [4], pp. 111-127, 1995.