EXPLAINING BRAZILIAN PORTUGUESE RESISTANCE TO STRESS SHIFT WITH A COUPLED-OSCILLATOR MODEL OF SPEECH RHYTHM PRODUCTION

PLÍNIO ALMEIDA BARBOSA
(LAFAPE/IEL/UNICAMP)

RESUMO Um modelo de osciladores acoplados da produção do ritmo da fala é apresentado, modelo este capaz de reproduzir via simulação os padrões de duração de unidades do tamanho da sílaba em português brasileiro, incluindo casos de encontro acentual. Para tanto, os parâmetros do modelo são otimizados com relação a frases naturais e manipulados convenientemente a fim de reproduzir os dados comportamentais. Ao comparar padrões de duração para três tipos de unidades prosódicas mínimas (sílaba, GIPC e rima) em condição ou não de choque acentual, não se encontrou evidência clara para o deslocamento acentual: os padrões nas duas condições exibem um aumento da duração das unidades em questão em direção ao acento frasal à direita. Discute-se uma possível relação do deslocamento acentual com a colocação do *pitch accent* bem como com a alternância de proeminências ao longo do enunciado.

1. INTRODUCTION

In this paper, the coupled-oscillator model of speech rhythm production we are using for explaining acoustic duration in Brazilian Portuguese is discussed in great detail (cf. Barbosa, 2001; Barbosa, 2002 for previous, shorter presentations). Particular aspects of parameter optimization and control, as well as the linguistic phonetic consequences of the way duration is specified in the model with particular reference to the stress clash/shift phenomenon, are also considered.

In the model framework, speech rhythm is understood as a consequence of the variation of perceived duration along the entire utterance. In order to produce such an effect, the great majority of scholars consider the measured duration as the most relevant parameter of control. (Despite well-known studies showing that parameters such as fundamental frequency and intensity also play a role when perceiving duration. See, for instance, Lehiste, 1978).

Besides adopting duration as the control parameter in the speech rhythm production model, two levels of duration encoding are hypothesized here: an intrinsic, lexical level of duration specification and an extrinsic, properly rhythmic level of duration control, the latter implemented by a model of two coupled oscillators. This model takes into account the moment-to-moment consequences of the coupling between a syllabic and a phrase stress oscillator. Even if the proposal of two oscillators (instead of three, as in Barbosa & Madureira, 1999) is a direct consequence of speech rhythm research with

Brazilian Portuguese, particularly in connection to the results of Albano's work (2001), it was also inspired by O'Dell and Nieminen's work (1999, 2001). More information on the latter authors' research can be found in this very volume. Yet the starting point of our own model was the moment-to-moment implementation of a model of adaptive oscillators for perception of timing by Devin McAuley.

2. A MOMENT-TO-MOMENT MODEL OF TIME INTERVALS PERCEPTION

McAuley (1995) proposed an Entrainment Model in order to account for human perception of time processing. In his terms, entraining means to modify the parameters of an oscillator according to predefined laws of adaptation operating on the adaptive system, which is a class of coupled-oscillators systems.

In McAuley's adaptive-oscillator model, a train of pulses, considered as the input, $i(t)$, entrains phase resetting and period coupling onto a sinusoidal oscillator. Both the input and the oscillator constitute the adaptive system. Phase resetting is triggered by the addition, to the sinusoidal oscillator amplitude, of a product coupling strength function $w_i.i(t)$. Resetting occurs every time the sum of the oscillator amplitude and the corresponding value of the coupling strength function is greater than 1.0, at any time.

Period coupling between external input and the oscillator is controlled by three characteristics of this particular adaptive system: (a) phase resetting, (b) reset phase sign (positive if the external input is advanced in time in relation to the oscillator and negative otherwise), and (c) an output signal $o(.)$ measuring the current degree of synchronicity between the oscillator being adapted and the input. The amount of oscillator period subject to change is given by formula (1), where α is the entrainment rate, β is the decay rate, T_0 is the initial oscillator period, T is the current oscillator period, $P(.)$ is the reset phase sign function, and $M(.)$ is the impulse response function.

$$\Delta T = \alpha.T.P (\Phi^r, o). M (i^r) - \beta.(T-T_0). (1 - M(i^r)) \tag{1}$$

The function $M(.)$, which has the value of 1 during entrainment and 0 during decay, allows to take into account decay and entrainment in a single formula, by separating their respective contributions. Yet a simultaneity of entrainment and decay is possible during transitional periods of time and this very interaction revealed to be critical to modeling human performance in tempo discrimination.

McAuley's adaptive oscillator revealed to be an adequate model of the human perception of time intervals. His adaptive oscillator is able to be entrained by a great variety of complex temporal patterns including situations of simulated degraded environments up to 10% added noise (see especially McAuley, 1995, chapter 4). But because he focussed on single adaptive oscillators, his model does not directly address the complex problem of perception of meter.

The main ideas in McAuley's time-processing perception model were used in our speech production model, but several modifications were introduced.

3. A MOMENT-TO-MOMENT COUPLED-OSCILLATOR MODEL OF SPEECH RHYTHM PRODUCTION

In the model presented in figure 1, the two oscillators to the left stand for a phrase stress oscillator implemented by a train of pulses (upper oscillator), and a syllabic oscillator (from 'syllabicity', and not 'syllable'), implemented by a sinusoidal function. They are coupled with each other, which means that the patterns of oscillation of one oscillator can influence those of the other and vice-versa. The bi-directional influence has its degree expressed by the coupling strength between both oscillators. Yet only the consequences onto the syllabic oscillator are highlighted in this paper, since the focus of this work is on stress patterns expressed by syllable-sized units along the utterance. When our model is compared with that in O'Dell-Nieminen's work, two important differences need to be signaled: (a) in their model, the stress group oscillator is a sinusoidal function, and (b) in their model, the coupled oscillators are not directly subject to linguistic control.
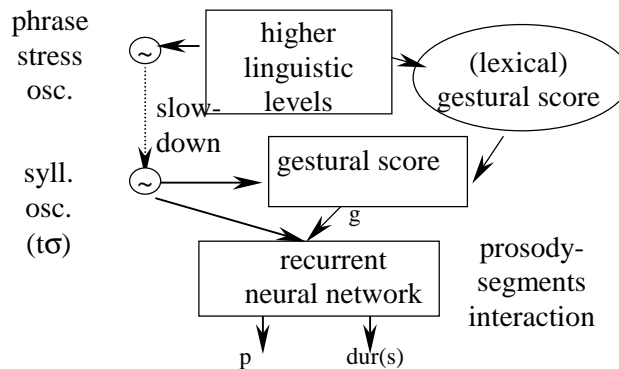

Figure 1: The dynamical model of speech rhythm production

The location and magnitude of the phrase stress oscillator pulses, which specifies respectively position and degree of prominence of phrase stresses along the utterance, are determined by the interplay between higher-level linguistic knowledge and eurhythmic constraints. In the present state of the model, only position of word primary stresses and very simple eurhythmic constraints are considered. Because the pulses of the phrase stress oscillator are aligned with stressed vowel onsets, they characterize a linguistic rhythm in the traditional sense (Liberman & Prince, 1983). Another sense of linguistic, with respect to the coupled-oscillators model refers to the parameterization of the coupling strength between them, as explained below. The phrase stress oscillator is related to a universal property of phrasal prominence (all languages signal prominence at one or more positions along the utterances).

The syllabic oscillator implements an aspect of the so-called syllabicity, that one related to the vowel continuum (see Öhman, 1966 for the vowel continuum hypothesis, and its more concrete development in the use of Fujimura's *base function*: Fujimura, 1995, and

this volume). The syllabic oscillator maxima are related to the sequence of vowels onsets, since CV transitions are anchor points for speech rhythm production and perception (Allen 1972; Dogil & Braun, 1988; Barbosa, 2000, and the literature on the perceptual-center: Morton, Marcus & Frankish, 1976; Pompino-Marschall, 1989; Janker, 1995). The syllabic oscillator periods (henceforth VV periods) are nevertheless abstract, in the sense that their values do not specify directly the acoustic duration between two consecutive vowel onsets (henceforth VV durations). The entrained VV periods delivered by the coupled-oscillator model stand for **extrinsic timing** in this framework.

As regards **intrinsic timing**, the model is able to explain alleged segmental cross-linguistic differences, such as differences in vowel and consonant reduction, by constrasting full and reduced gestures in a (gestural) lexicon (Albano 2001). A language such as English would have frequent specification of vowel reduction, in comparison to Castillan Spanish, for instance. Contrasted gestures can differ both on magnitude specification and intrinsic time, as proposed by Browman & Goldstein (1986, 1989). The interaction with the right components of the model in figure 1 (lexical gestural score, gestural score and recurrent neural network) is then crucial to generate overt durations. (Indeed the model was used to generate acoustic segmental durations for a concatenative TTS synthesis system in Barbosa, 2001). This implies that, in this framework, acoustic duration is the surface manifestation of two levels of timing: an extrinsic level (properly rhythmic) subject to independent, prosodic control, implemented by the coupled-oscillators model, and an intrinsic level, specified in a gestural lexicon. The model is mathematically implemented by two functions similar to those in McAuley's work, with the differences signaled below.

3.1. Implementing the coupled-oscillators model of speech rhythm production

The measure of synchronicity between the phrase stress oscillator and the syllabic oscillator is implemented by the empirically determined function s(.), given in (2). Note that this synchronicity function is a crucial component of the speech rhythm production model and that, differently from McAuley's output signal above, it was empirically obtained from non-linear regression analysis of VV durations along the stress groups in a corpus of 36 Brazilian Portuguese (henceforth BP) read sentences recorded at three speech rates by a native speaker. Fast and slow speech rates were recorded by using a light metronome. The evolution of the relative change of VV duration ($\Delta VV_i/VV_i$) along each stress group was computed and transformed by means of some predefined functions before doing the regression analysis. The best result ($R^2 = 0.999$) was obtained with the exponential function.

$$s(0) = w_0.\exp(-N + 2), \text{ and } s(N-1) = \exp(-5.81 + 0.016.T_0) \qquad (2a)$$
$$s(n) = (1 - w_0).s(n-1) + w_0.\exp(-N + n + 2), \text{ for } 0 < n < N-1 \qquad (2b)$$

In formula 2, N is the number of VV units in the stress group, and n is the VV unit index (starting from 0 for the first VV unit in the stress group). The relative coupling strength

74

w_0 is a continuous, language-specific parameter representing the degree of coupling between the phrase stress and the syllabic oscillators, and T_0 is the uncoupled syllabic-oscillator period (the same as in formula 3).

The relative coupling strength presents another, new sense of rhythm-related linguistic knowledge, since rhythm-related language-specificity is represented by a real number (and not a discrete one). It is expected that w_0 will vary for different speakers, but that this variability would be much smaller than the variability between these same speakers and those of a different language. We intend to carry out the optimization procedure described in the following section with other speakers of BP and possibly with European Portuguese, in order to assess this assumption.

Phase resetting within this particular coupled-oscillator model is considered achieved at each syllabic oscillator maximum, that is, at positions isomorphic to vowel onsets. Period coupling is obtained with the finite-differences equation (3).

$$\Delta T = \alpha.T.s(n).i(m) - \beta.(T-T_0).i(m-1) \qquad (3)$$

The parameters α, β, and T_0 are respectively the entrainment rate, the decay rate, and the uncoupled syllabic-oscillator period. The functions $s(n)$ and $i(m)$ are respectively the synchronicity function (given by formula 2) and the on-going phrase stress oscillator amplitude. T is the current syllabic oscillator period. The syllabic-oscillator period resetting, implemented by the second term of (3), operates after each phrasal stress with a decay rate specified by β. This term is only present if the current VV unit is dominated by a next, on-going phrase stress. According to this, at utterance-final position, where there is no such an on-going phrase stress, both stressed and post-stressed VV would be lengthened (since there is no resetting). This simulates what is observed in BP natural data (for an example, Barbosa, 1996). This period resetting is only active during the two first periods of the syllabic oscillator. The current period T is up-dated only at syllabic oscillator maxima by the iteration in 4.

$$T(k) = T(k-1) + \Delta T \qquad (4)$$

The values of w_0, α, and β were optimized in order to minimize the error between the evolution of syllabic oscillator periods given by formulae 2 and 3, and the mean of VV-duration evolution along 3-VV and 4-VV stress groups in the corpus (given in table 1). The choice of 3- and 4-VV stress groups only was made in order to ensure statistical reliability (more examples). The first and last stress groups of the sentences were not considered for the computation of the means in table 1. These mean VV durations are considered an approximation of the more abstract VV periods values.

Table 1: Evolution of VV duration mean (standard deviation) along 3-VV to 6-VV stress groups in milliseconds for three speech rates. n is the number of stress groups for a particular rate and size in the corpus.

	slow rate	normal rate	fast rate
3 VV	n = 48	n = 41	n = 42
	201 (76); 289 (58); 320 (78)	150 (41); 190 (35); 209 (47)	118 (38); 160 (31); 164 (45)
4 VV	n = 28	n = 26	n = 26
	215 (88); 200 (81); 289 (75); 339 (133)	125 (59);140 (58); 211 (23); 202 (87)	122 (40); 114 (31); 155 (49); 165 (42)
5 VV	n = 17	n = 8	n = 10
	202 (97); 181 (122); 221 (67) 247 (64); 338 (91)	144 (54); 107 (70); 131 (52) 207 (44); 194 (52)	105 (48); 88 (51); 105 (39) 182 (33); 155 (62)
6 VV	n = 8	n = 7	n = 7
	240 (100); 185 (74); 207 (61) 183 (39); 223 (85); 282 (119)	119 (39); 87 (71); 156 (33) 148 (14); 152 (43); 166 (30)	101 (42); 84 (54); 109 (43) 150 (36); 133 (58); 139 (29)

3.2. Optimizing the parameters of the model

As a first step, we tried to match only the values of table 1 corresponding to normal rate with the entrained syllabic oscillator periods, by a trial-and-error approach with parameters w_0, α, and β. This resulted in the values of 0.70 for w_0, 0.50 for α, and 0.80 for β (errors lesser than 25 ms between VV period and VV duration). These parameter values were used by Barbosa (2001) to obtain the segmental acoustic durations for a TTS system.

It can be hypothesized that the values of the pair (α,β) may be different for distinct speech rates. On the other hand, based on the language-specificity hypothesized above, w_0 must be the same for data obtained from one speaker. These considerations motivated a second step in the optimization procedure, which implemented a function for minimizing the quadratic errors between the mean VV durations in table 1 and the values for the entrained VV periods delivered by the coupled-oscillator model. The optimization function was implemented with MatLab®, version 6.0 (release 12). The two criteria used for terminating the recurrent algorithm were either maximum number of iteration steps (350) or maximum error between the modeled VV periods and the mean VV durations lesser than 20 %. All three parameters were allowed to iterate with the classic gradient-descent algorithm with a learning rate of 0.03 for α and β, and of 0.01, for w_0. The uncoupled value for the syllabic oscillator period (T_0) was estimated from the mean of three VV durations: first position in 3-VV stress groups as well as first and second position in 4-VV stress groups. These estimations gave the values of 118 ms for fast rate, 138 ms for normal rate and 205 ms for slow rate.

The result of the optimization procedure gave the following values: w_0 = 0.78, with (α,β) = (0.31, 1.13) for the fast rate, (α,β) = (0.38, 1.13) for the normal rate, and (α,β) = (0.43, 1.04) for the slow rate. It can be seen that the relative coupling strength (w_0) is very close to the trial-and-error value obtained previously for the normal rate only. As speech rate increases, the decay rate increases (from 1.04 to 1.13), but the effect of this changing (analyzed more closely below) is not important, though possibly ensuring a

certain amount of initial lengthening for slow rates. On the other hand, as speech rate increases, the entrainment rate decreases, indicating possibly that perfect coupling is more difficult to be achieved for fast rates (the difference between the phrase stress and the syllabic oscillator periods being the greatest), which would explain a slightly lesser important differentiation among the syllabic oscillator periods for fast rates.

The values of the optimized parameters gave the following results for the entrained syllabic oscillator periods. In the 3-VV stress group: (127 ms, 150 ms, 151 ms) for the fast rate, with a correlation coefficient R = 1.00, (152 ms, 185 ms, 187 ms) for the normal rate (R = 0.96), and (241 ms, 290 ms, 300 ms) for the slow rate (R = 1.00). In the 4-VV stress group: (117 ms, 129 ms, 163 ms, 164 ms) for the fast rate (R = 0.93), (137 ms, 154 ms, 204 ms, 206 ms) for the normal rate (R = 0.99), and (215 ms, 233 ms, 318 ms, 329 ms) for the slow rate (R = 0.95). All correlation coefficients are higher than 0.9. Comparing with the values in table 1, the maximum error was less than 20 % for the slow rate, less than 11 % for normal rate, and lesser than 14 % for fast rate. At phrase stress, the maximum error is lesser than 11 %, for normal rate, for a 3 VV stress group, and lesser than 3 %, for slow rate, for a 4 VV stress group.

It is important to check the effect of changing decay rate, entrainment rate and the amplitude of the on-going phrasal stress beat, in order to evaluate the predictive power of the model. This is done in the next three sections for a 5 VV stress group at slow rate (greater number of examples for this size of stress group in comparison to the other speech rates), where an interesting phenomenon of alternation at the beginning of the stress group was found (this can also be obtained for any stress group with more than four VV units).

3.3. Changing the decay rate (β) and the uncoupled period (t_0)

The result for the evolution of the syllabic oscillator periods for a 5-VV stress group is (206 ms, 215 ms, 244 ms, 332 ms, 344 ms), with a correlation coefficient R = 0.87, obtained for the slow rate with (α, β, i, w_0, T_0) = (0.43, 1.04, 1.0, 0.78, 205 ms). Even if the errors are higher (maximum error lesser than 35 %) when compared to the ones in 3- and 4-VV stress groups, the values are nevertheless inside the (mean + standard-deviation) limit. Of course the means for 5-VV stress groups cannot be taken as having strong statistical reliability, since the number of examples is restricted to seventeen. But when the decay rate is changed to 0.9 and subsequently, to 0.8, these values change to (**223 ms, 218 ms**, 247 ms, 337 ms, 348 ms), R = 0.87, and (**235 ms, 223 ms**, 252 ms, 344 ms, 356 ms), R = 0.88, respectively. It can be seen (first two values in bold) that an alternation at the beginning of the stress group emerges, which is attested in all cases of 5- and 6-VV stress groups for the three speech rates. The fall from the first to the second VV for the 5 VV stress group if of 10 % in the corpus, and in the model, of 2 % for β= 0.9, and of 5 % for β = 0.8. Note that the effect of changing β from 1.04 to 0.8 at phrasal stress is of only 3 % (from 344 to 356 ms), the more important change occurring at the first VV (206 to 235 ms, or 14 %). An apparent initial lengthening seem then to be a consequence of the interaction between entrainment and decay rate, possibly in long stress groups (more than 4 VV units).

If, besides β being changed to 0.8, the uncoupled period for the syllabic oscillator is changed from 205 ms to 165 ms, we obtain the following values for the VV periods: (204 ms, 183 ms, 208 ms, 283 ms, 289 ms), R = 0.87, which represents errors lesser than 16 %, with the same percentage of first-to-second VV fall in comparison to the one in the corpus. Thus, the interaction between the decay rate and the uncoupled syllabic-oscillator period seems to be a crucial aspect of the model in order to reproduce rhythmically acceptable patterns. What happens with the changing of the entrainment rate?

3.4. Changing the entrainment rate (α) and the uncoupled period (t_0)

The result for the evolution of the syllabic oscillator periods for a 5-VV stress group is (204 ms, 183 ms, 208 ms, 283 ms, 289 ms), R = 0.87, obtained for the slow rate with (α, β, i, w_0, T_0) = (0.43, 0.8, 1.0, 0.78, 165 ms). When the entrainment rate is changed to 0.5, and after to 0.6, the values change to (204 ms, 185 ms, 213 ms, 304 ms, 310 ms), R = 0.86, and (206 ms, 187 ms, 222 ms, 335 ms, 344 ms), R = 0.87, which increases the maximum error (at penultimate phrasal stress position) to values lesser than 24 % and than 36 %, respectively, while maintaining equivalent correlation coefficients. Maintaining α = 0.6 and changing the uncoupled period to 140 ms has the effect of reducing the maximum error to a value inferior to 18 % (186 ms, 162 ms, 192 ms, 290 ms, 295 ms), R = 0.86, with an almost similar first-to-second-VV falling: 11 %.
By manipulating α, β, and T_0 is then possible to get close to the natural values for VV period evolution. Changing the amplitude of the on-going phrasal stress beat will simulate, for fixed values of α, β, w_0, and T_0, the amount of phrasal stressing (which has in BP the perceptual effect of increasing prominence).

3.5. Changing the amplitude of the on-going phrase stress beat (i)

Starting with the values obtained with (α, β, i, w_0, T_0) = (0.6, 0.8, 1.0, 0.78, 165 ms), that is, (204 ms, 183 ms, 208 ms, 283 ms, 289 ms), R = 0.87, a change in the amplitude of the on-going phrasal stress beat (i) to 1.2 and 0.9, produces respectively: (187 ms, 165 ms, 202 ms, 325 ms, 332 ms), R = 0.86, and (185 ms, 160 ms, 187 ms, 273 ms, 277 ms), R= 0.86. The maximum error for the first series is lesser than 32 %, at penultimate phrasal stress position, and lesser than 19 %, at phrasal stress position. It can be seen that as the amplitude of the on-going phrasal stress beat increases , the VV period at phrasal stress increases (from 277 ms, passing by 289 ms, to 332 ms). The initial VV periods for i = 0.9 and 1.2 are similar, due to the effect of resetting (amplified by i in formula 3) at the two first positions. Nevertheless, changes in the amplitude of the on-going phrasal stress beat are almost equivalent to changes in entrainment rate (they appear as a product in the first parcel of formula 3). The coincidence is not perfect since this amplitude also appears in the second half of the same formula, without the entrainment rate.
After having simulated the effects of changing the most relevant model parameters, it is interesting to assess the results of the model for some sentences.

78

3.6. Obtaining the syllabic oscillator vv periods for particular sentences

It is important to note that the VV periods for the sentences derived in this section are not the predicted acoustic durations for the VV units (the VV durations). This is due to the fact that the intrinsic timing level was not considered until now. This level corresponds to the specification of the number and nature of the corresponding gestures associated with each VV cycle. These particular phonological gestures modify, perturb the entrained VV periods given by the coupled-oscillators model and produce the overt, segmental acoustic durations. The entrained syllabic oscillator is then a kind of guiding grid for the superposing segmental gestures. A first implementation of this perturbation model of speech gestures was made with a recurrent neural network (Barbosa, 2001). Since the interest here is in taking into account the linguistic rhythm by evaluating global patterns of stressing in syllable-sized domains along the utterances, the overt acoustic durations will not be considered in this paper.

Six sentences were chosen for this study as follows. Phrasally stressed syllables are those just before the vertical bars. Words in bold are explained below.

Ele guar|da a sela do cava|lo numa pratelei|ra de uma antiga se|la.   .1.
(He keeps the horse's shaddle on a shelf in an ancient cell.)

Cerzi | uma man|ga.   .2.
((I) darned a sleeve.)

Ele e você | não tomam **café fri|o** de no|vo.   .3.
(He and you do not take cold coffee again.)

Ele e você | não tomam **café Pilão** | de no|vo.   .4.
(He and you do not take Pilão coffee again.)

O **bordeaux xu|cro** derramou-se pela me|sa.   .5.
(The bad-quality Bordeaux poured out onto the table.)

O **bordeaux chinês** | derramou-se pela me|sa.   .6.
(The Chinese Bordeaux poured out onto the table.)

These sentences were produced by BP native speakers. The first two sentences were extracted from the corpus used to optimize the model (one single utterance each, normal rate). Sentences 3 and 4 were produced five times by another native speaker and sentences 5 and 6 repeated ten times by a third native speaker. Sentences 3 to 6 contrast cases of stress clash condition (3 and 5) against non-clash condition (4 and 6) in such a way as to evaluate the results of the statistical analyses carried out on the natural sentences presented in section 4. The stress groups are indicated in the sentences with vertical bars. This parsing is based on syntactic as well as eurhythmic criteria, and corresponds to the perceptual impression we get when listening to the three speakers. The single values for the VV durations of sentences 1 and 2 and the mean VV durations

for sentences 3 to 6 were computed after segmentation of the utterances and given in table 2.

In order to simulate the stressing patterns for the six sentences, the parameters values of the model used were $(\alpha, \beta, w_0, T_0, i) = (0.38, 1.13, 0.78, 138\ ms, 1.0)$. The duration of a final catalectic VV in the final stress groups (since all final words are paroxytons) was modeled by maintaining the last value of the synchronicity function s (.) and by iterating formula 3 one step further. In order to compute, for each sentence, the VV periods from the second stress group to the final one, the first current period value used (see formula 3) was the last VV period of the previous stress group. The predicted values given in table 2, second column, were obtained after simulation. For sentences 3 to 6, only the words signaled in bold above were predicted and compared to the corresponding natural VV durations. The numbers in bold in table 2 refer to phrase stress position. For sentence 1, the correlation coefficient R = 0.71, and for sentence 2, R = 0.49. (Since there are only two points for correlation in sentences 3 to 6, the correlation coefficient is 1.)

Table 2: Natural VV durations and predicted VV periods for six sentences in miliseconds. Phrase stresses are given in bold. See text for details.

Sentence	Natural	Predicted
1	147, 149, **274**, 261, 170, 116, 129, 219, **199**, 124, 61, 167, 146, 138, **168**, 153, 38, 74, 215, 130, 279, **265**, 105	153, 202, **204**, 130+140, 142, 148, 166, 219, **222**, 138, 141, 146, 164, 217, **219**, 138, 138, 139, 141+148, 165, 218, **221**, 223
2	334, **135**, 110, 179, **240**, 99	179, **181**, 152, 185, **187**, 189
3	163, 238	161, 218
4	159, 182	148, 166
5	173, 253	161, 212
6	165, 195	146, 163

In comparison to the natural VV durations of sentence 1, where the first and final phrase stresses seem to be the strongest ones, the three final predicted VV periods are very similar to each other (222, 219 and 221 ms). Since the on-going phrase stress amplitudes were fixed in 1.0, the VV periods at phrase stress depend only on the size of the group. If syntactic and eurhythmic criteria[1] are used to give the values of 0.6, 1.0, 0.3, and 1.2 for the four respective phrase stress amplitudes of this sentence, the four final values obtained for the VV periods are: 176 ms, 227 ms, 160 ms, 245 ms, R = 0.74, which shows, in a more differentiated way and with a slightly higher R, that the weakest phrase stress is the third one. For sentence 2, the predicted VV periods at phrase stress are very close to each other. If the same syntactic and eurhythmic criteria are used (phrase stress

---

[1] Basically, the eurhythmic criteria give a higher value for phrase stresses closer to the middle or the end of the sentence (in the example we could have 1.5, 2, 1.5, 2.0). The syntactic criteria are based on the projection of a dependent-grammar surface tree (see Bailly, 1986; Tesnière, 1965). In this example the values would be 4, between V and NP, 5, between NP and PP, 2, between the two PP, where the second is nested inside the first one. At the end of the sentence the value is 6. After that, the values from both criteria are multiplied and divided by 10.

oscillator amplitudes respectively equal to 0.8 and 1.2), the values of (172 ms, 198 ms), R = 0.29, are obtained at phrase stress, which exhibit the same trend of the natural values but with a smaller R. This can indicate that the real prominence for the two phrase stresses are inverted in relation to the criteria used. Indeed this is exactly the perceptual impression one gets. If the phrase stress oscillator amplitudes are reversed, the values of (189 ms, 176 ms) are obtained at phrase stress, with R = 0.54. Of course, the non-coincidence is not possible, since the VV period account only for the extrinsic timing level. The two pairs of VV periods summed up in table 1 for sentence 1 give the possibility of using the VV periods for predicting VV durations when natural phonic processes such as sândhi seem to elide some segments.

For sentences 3 and 4, the figures show that the natural VV duration corresponding to the vowel /ɛ/ (in "caf**é**") plus the next consonant, is 31 % greater in the sequence "café frio" than in the sequence "café Pilão". Exactly the same percent difference is found in the predicted VV periods. For sentences 5 and 6, the natural VV duration corresponding to the vowel /o/ (in "bord**eaux**") is 30 % greater in the sequence "bordeaux xucro" in comparison to the sequence "bordeaux chinês". Exactly the same percentage of difference is found for the predicted VV periods. The higher values for the VV periods or durations are a consequence of the closeness of the phrase stress beat to the previous word. Note that at the phrase stress oscillator level, the model do not provide any expedient for taking into account a possible stress shift. Let us remind that the model gives duration patterns only in the case duration is the control variable in the particular language (which is the case of BP). These results will be reminded in the following section, where data for analyzing the possibility of stress shift in BP are examined.

4. STRESS CLASH IN BRAZILIAN PORTUGUESE

Liberman and Prince (1983) defined   stress clash as an adjacency of a mark of prominence ('X') at one level of the metrical grid, without an intervening mark at the immediately lower level. For instance,

```
       X    level 3              X                X
     *X X   level 2            *X     X             X    X
     X X X  level 1            X X     X           X X X X
     X X X X level 0           X X X   X   X       X X X X
     ca fé fri o      .7a.     ca be ça quen te  .7b.    ca fé Pilão        .7c.
```

In the examples 7a to 7c (stressed syllables in bold), sequences 7a and 7b clash at level 2 (without an intervening 'X' at level 1), while in 7c there is no clash, since at level 1 there is an intervening 'X'. The Rhythm Rule undo the clash of 7a and 7b with the rule of Iambic Reversal, which results in a stress shift of the first words (the first word in 7a would be heard more paroxyton-like, instead of an oxyton. The first word in 7b would be heard as more proparoxyton-like, instead of paroxyton).

```
        X                   X
X   X           X           X
X X X           X X     X
X X X X         X X X   X   X
café frio       .7d. cabeça quente       .7e.
```

Two studies presented at the Laboratory Phonology series focussed on the phenomenon of stress shift (Grabe & Warren, 1995; Vogel, Bunnell & Hoskins, 1995). Both studies tried to find phonetic evidence for the application of the Rhythm Rule. Both authors agree that the perception of stress shift may be explained by the relational character of stress: besides reinforcing the stressness of the first word in sequences such as *thirtèen mén* (clash condition), which would allow reinforced phonetic parameters for the first syllable of the first word, the alternative decrease in prominence in the second syllable of the first word would also favor the perception of an iamb. Vogel *et a.l* (1995) show that the duration effect is mainly restricted to the rhymes (the universal properties of rhymes as carriers of the consequences of distinct prosodic strengths between constituents of a sentence is very well accepted. See Vaissière, 1983 for a survey). Nevertheless, the perception tests carried out by the authors have also shown that a great number of the first words in the sequences are perceived as shifted even in non-clash conditions.

Based on previous descriptions of the English intonation system, Shattuck-Hufnagel and colleagues claim (Shattuck-Hufnagel, 1995; Shattuck-Hufnagel, Ostendorf & Ross, 1994) that phenomena such as pitch placement alternation, early prominence, and the obligatory position of nuclear pitch accent would better explain the findings above. They proposed a metrical-intonational phonological view of early pitch accent placement in American English (Shattuck-Hufnagel *et al.*, 1994) to explain accent placement in relation to the stress shift issue. They 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. BP is adequate for this investigation, since f0 modulations in this language are correlates of phrasal accents, and syllable-sized duration has the main role to cue both phrasal and lexical stresses (Massini, 1991; Barbosa, 1996). Two experiments were conducted in order to assess the possible phonetic consequences of stress clash in BP (duration is analyzed in this paper, f_0 is analyzed by Madureira, this volume).

4.1. Description of experiment 1

In experiment 1, one speaker (native of the State of São Paulo, male, 25 years old) recorded ten repetitions of the five pairs of sentences shown in table 3. During the recording session, the sentences were randomized within and across repetitions. In order to avoid that two contrasted sentences were uttered one just after the other, the order was randomly changed if this happened to be the case. Acoustic duration was measured for the segments of the underlined sequences in table 3, where words in bold are the target words, that is, those in whose domain the phonetic manifestation of stress shift would

82

occur. Mean duration for three types of unit of the target word are presented with statistical analyses in table 4. In these sentences, a disyllabic oxyton/oxyton sequence represents the non-clash condition (sentences A in the pair), and a disyllabic oxyton/paroxyton sequence represents the contrasted clash condition (sentence B in the pair). A focussed version of utterance 8 was also prepared as a response to the question "Você bebeu chá sexta-feira à noite?" (Did you drink tea Friday night?), in order to restrict narrow focus within the sequences "comi bolo" and "comi bolor". This utterance was prepared in such a way as to assess if the greater, pitch-related prominence within the sequence under investigation would trigger stress shift with consequences to acoustic duration.

Table 3: Sentences of experiment 1. Target words are in bold. Phrase stress in the sequence indicated with acute diacritic. Segments that were measured are underlined.

Eu **comi** bolór sexta-feira à noite.   .8A.
(I ate mould Friday night.)

Eu **comi** bólo sexta-feira à noite.   .8B.
(I ate cake Friday night.)

O **bordeaux** chinês derramou-se pela mesa.   .9A.
(The Chinese Bordeaux poured out onto the table.)

O **bordeaux** xúcro derramou-se pela mesa.   .9B.
(The bad-quality Bordeaux poured out onto the table.)

Foi à **TV** francésa e fez a denúncia.   .10A.
((He) went to the French TV and made the accusation.)

Foi à **TV** Fránça e fez a denúncia.   .10B.
((He) went to France TV and made the accusation.)

Parece que **falou** 'baixóu', e não 'caiu'   .11A.
(It seems that (he) said 'it dropped' and not 'it fell'.)

Parece que **falou** 'báixo', e não 'caiu'   .11B.
(It seems that (he) said 'I drop' and not 'it fell'.)

Um lindo **bebê** carmím.   .12A.
(A beautiful crimson baby.)

Um lindo **bebê** cálvo.   .12B.
(A beautiful bald baby.)

4.2. Results of experiment 1

In table 4 the mean duration for three types of units, rhymes (r), syllables (σ) and inter-vowel onset units (VV), for the two positions of the target words are shown. Sentences

of type A (non-clash condition) are compared with sentences of type B (clash condition) by using one-way analysis of variance (ANOVA) with a significance level of 5 %.

With the exception of the sentence under focus, for which all results are non significant, VV units for all target words exhibit significant differences when non-clash are compared with clash condition. The second VV has higher duration than the first VV in clash condition (this second VV in clash condition is closer to a stressed syllable to its right than the second VV in non-clash condition). Even though this effect is mainly due to the higher duration of the next onset consonant of the following disyllabic paroxyton, it is important to note that the same effect is significant in sentence 9 and marginally significant in sentence 8, both for the syllable as an unit.

Table 4: Mean duration (in milliseconds) 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 8 to 12 for the two positions of the target words. Corresponding segments are indicated within slashes. Contrasted values with significant differences in bold.

	r			σ			VV		
	A	B	p <	A	B	p <	A	B	p <
sentence 8	34	30	ns	129 /ko/	137	0.07	**89**	**101**	**0.02**
	/o/	/o/			/ko/		**/om/**	**/om/**	
	59	59	ns	118	126	0.06	**148**	**164**	**0.004**
	/i/	/i/		/mi/	/mi/		**/ib/**	**/ib/**	
sentence 9	96	93	ns	174	166	ns	165	173	ns
	/or/	/or/		/bor/	/bor/		/ord/	/ord/	
	81	87	ns	**151**	**167**	**0.02**	195	253	10^{-7}
	/o/	/o/		**/do/**	**/do/**		/oʃ/	/oʃ/	
sentence 10	45	48	ns	**133**	**144**	**0.004**	101	101	ns
	/e/	/e/		**/te/**	**/te/**		/ev/	/ev/	
	64	**72**	**0.05**	120	125	ns	**205**	**245**	10^{-8}
	/e/	**/e/**		/ve/	/ve/		**/ef/**	**/ef/**	
sentence 11	47	41	ns	163	165	ns	106	96	0.06
	/a/	/a/		/fa/	/fa/		/al/	/al/	
	83	85	ns	142	139	ns	**150**	**183**	10^{-7}
	/ow/	/ow/		/low/	/low/		**/owb/**	**/owb/**	
sentence 12	66	67	ns	134	129	ns	135	142	ns
	/e/	/e/		/be/	/be/		/eb/	/eb/	
	75	77	ns	143	151	ns	**174**	**200**	10^{-3}
	/e/	/e/		/be/	/be/		**/ek/**	**/ek/**	
sentence 8 under focus	30	34	ns	125	132	ns	82	89	ns
	/o/	/o/		/ko/	/ko/		/om/	/om/	
	57	58	ns	109	113	ns	145	164	ns
	/i/	/i/		/mi/	/mi/		/ib/	/ib/	

With regard to the syllable, only sentence 10 seem to exhibit stress shift, since durations are higher at first position for this sentence (this difference is due to a stronger attack of /t/ in clash condition). Sentence 10 is also the only one exhibiting significant difference in duration when rhymes are the corresponding units: the second rhyme has higher duration in clash condition. Both patterns (higher duration for the first syllable as well as higher duration for the second rhyme) can be related to the two strong heteromorphemic

84

syllables of the acronym TV (see also Shattuck-Hufnagel et al. 1994's data) and would consist in a special case of the phenomenon under analysis.
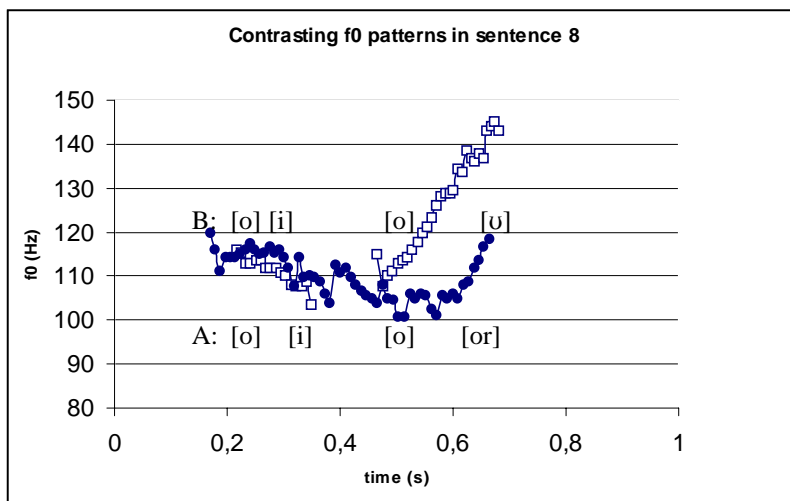

Figure 2: Differences in f0 pattern for sentence 8. Open squares: "comi bolo" (B condition), and closed circles: "comi bolor" (A condition).

As regards sentence 8, for the two positions of the target word, the marginal significant difference for syllables as well as the significant difference for VV merit some closer examination, since it can signal a different production strategy in comparison to the other utterances. In fact, if f0 patterns for both paired sentences are compared, the following mean values, at the middle of the vowels /o/ e /i/, are obtained: 114 Hz and 112 Hz, respectively for the two positions, in non-clash condition, difference non-significant. And 115 Hz followed by 107 Hz ($p < 10^{-3}$), in clash condition. This seems to indicate a two-fold strategy: duration is signaling a continuous increase of stress towards the phrasally stressed syllable in the second word, and f_0 is the only parameter related to a prominence shift, in such a way as to avoid two consecutive pitch accented units. Note, in figure 2, the earlier prominence in the word "bolo" in comparison with that of the word "bolor", as well as the decreasing f_0 pattern during "comi" in clash condition as opposed to the f_0 flat pattern in non-clash condition.

No other f_0 differences for the target word across the paired sentences were found (see Madureira, this volume).

4.3. Description of experiment 2

In experiment 2, three native speakers (all from the State of São Paulo, male, between 25 and 30 years old) recorded five repetitions of four paired sentences containing candidate domains for stress shift. The sentences were recorded with fifteen other sentences (non analyzed here) following the same procedure as in experiment 1. The corpus, shown in table 5, was set to investigate stress shift with respect to the entire

utterance. They allow to orthogonally compare non-clash (sentences A) to clash conditions (sentences B), as well as another parameter of control: two different kinds of stress placement for the word at the left of the target word: oxyton (sentences 15, word "tomar") and paroxyton (sentences 14, word "tomam").

Table 5: Sentences of experiment 2. All lexically stressed syllable are in bold. Measured segments are underlined.

O **Mes**tre To**más** Cris**pa**do pre**gou** por pa**rá**bolas.	.13A.
O **Mes**tre To**más** **Cris**pa pre**gou** por pa**rá**bolas.	.13B.
Ele e vo**cê** não **to**mam ca**fé** Pi**lão** de **no**vo.	.14A.
Ele e vo**cê** não **to**mam ca**fé** **fri**o de **no**vo.	.14B.
Não pode**ri**a to**mar** ca**fé** Pi**lão** de **no**vo.	.15A.
Não pode**ri**a to**mar** ca**fé** **fri**o de **no**vo.	.15B.

The need for the same phonetic environment forced the sentences in experiment 1 to be a little bit unusual in terms of meaning. That is why, in experiment 2, besides using the same kind of paradigm of experiment 1 with the paired sentences 13, a more natural contrast was included with sentences 14 and 15. In these sentences, the control of the duration-related consequences of distinct phonetic environments was minimized by using in both contrasted sentences voiceless labial obstruents followed by the same vowel /i/ ("Pilão" and "frio"). In this experiment, long non-pitch-accented stretches of speech were found, which allow to investigate more closely the use of duration for signaling phrasal prominence in both conditions for the three speakers.

4.4. Results of experiment 2

Results are presented in tables 6a to 6c, one for each speaker. Since the sequence of segments in the second VV of the target word in sentences 14 and 15 is not the same when comparing clash with non-clash conditions, the corresponding VV durations are not presented for these sentences. Unless when signaled, no differences in duration for the three units considered here were found for the left-context words ("mestre", "tomam", "tomar"). In all contrasted utterances no differences in f_0 were found.

Table 6a: Speaker 1 mean duration (in milliseconds) 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 8 to 12. Corresponding segments are indicated in slashes. Contrasted values with significant differences in bold.

		r		σ		VV	
	A	53	**151**	144	212	114	**271**
sentence 13	B	60	**172**	149	231	119	**316**
	p <	ns	**.02**	ns	.08	ns	**.004**
	A	68	**69**	143	**160**	159	
sentence 14	B	72	**87**	145	**178**	163	
	p <	ns	10^{-3}	ns	**.003**	ns	
	A	58	**73**	**133**	**167**	152	
sentence 15	B	60	**92**	**148**	**187**	154	
	p <	ns	10^{-5}	10^{-3}	10^{-3}	ns	

86

For this subject, differences in duration at second position for rhymes in clash vs non-clash conditions are all significant. This is not the case at first position. The same holds for syllables in sentence 14 and for VV, in sentence 13. In all these cases, duration is higher for clash condition at second position, which is exactly the opposite as predicted by the Rhythm Rule (in clash condition, either the stressness at first position in the target word would be higher, or there would be a decrease in stressness at second position in the same word), as illustrated by the analyses of Grabe & Warren (1995) and Vogel, Bunnell & Hoskins (1995)'s data. The contrasted first positions for VV in sentences 14 and 15 are not significant either. But the first syllable in sentence 15 has a significant higher duration at clash condition, which could indicate a stress shift. The difference is due to a stronger attack of the consonant /k/ (mean 89 ms for clash condition and 75 ms for non-clash condition, $p < .003$). This difference can not be associated to differences in f0 patterns, since there is no differences in f0 pattern for the sequence "tomar café" in both conditions (in both cases the f_0 peak of "tomar" is placed earlier, in its first syllable. The second f_0 onset of rising is placed in the stressed vowel of the next, phrasally stressed word). Furthermore, there is a flat (non-significant differences) f_0 pattern between the two syllables of the target word. For sentences 14 and 15, the percentage of the total variance explained by the clash/non-clash contrast is greater than 75% (for sentence 13, greater than 62%).

An example of the pattern of duration for a part of the utterance can be seen in figure 3 for sentence 13. Since they are more accepted units for signaling prominence, rhymes are presented here. The part shown is "-tre To**más Cris**pa" (stressed syllables in bold, phrasal stress underlined) for condition B, and "-tre To**más** Cris**pa**do" for condition A. The first part of the sentence was suppressed due to the impossibility of separating segment [s] from [t] in the word "mestre". Note that increasing patterns of duration are found for both contrasted conditions toward the rhyme /as/. At this position the rhyme in clash condition has a higher duration than that in non-clash condition. It is possible that a duration-related focus is associated with this rhyme, but a normalized duration must be used before concluding, due to the intrinsically higher duration of /a/ compared with that of /i/ in the following rhyme.
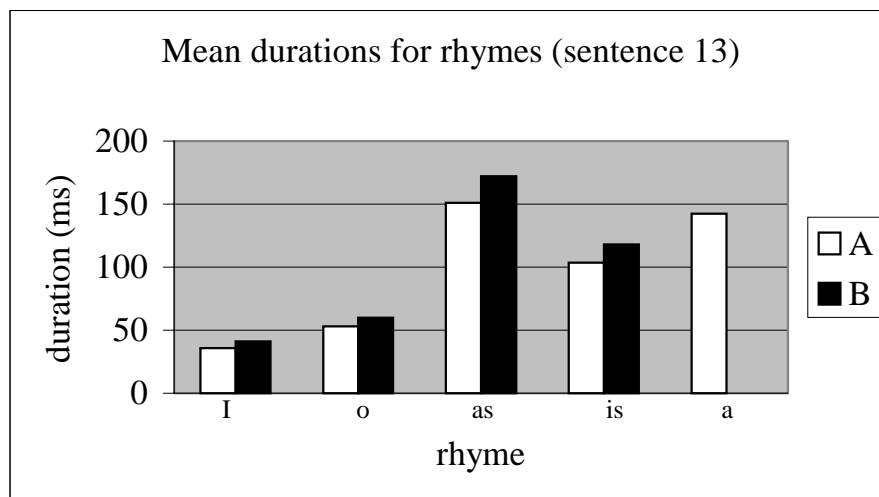
Figure 3: Mean durations for rhymes for part of the sentence 13. A: non-clash condition, and B: clash condition

Table 6b: Speaker 2 mean duration (in milliseconds) 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 8 to 12. Corresponding segments are indicated in slashes. Contrasted values with significant differences in bold.

		r		σ		VV	
sentence 13	A	75	232	164	291	134	362
	B	68	229	169	293	132	382
	p <	ns	ns	ns	ns	ns	ns
sentence 14	A	58	47	122	**123**	134	
	B	55	55	124	**137**	137	
	p <	ns	ns	ns	**.03**	ns	
sentence 15	A	49	**50**	113	**123**	122	
	B	50	**63**	122	**149**	136	
	p <	ns	**.004**	ns	**.02**	ns	

For speaker 2, the effects just described for speaker 1 are somewhat weaker. They are not observed in sentence 13, observed for syllables in sentence 14 and for syllables and rhymes in sentence 15. At first position of target words no differences are found. However, the pattern that emerges is exactly the same: duration is higher for units closer to the phrasally stressed syllable to its right (in clash condition).

88

Table 6c: Speaker 3 mean duration (in milliseconds) 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 8 to 12. Corresponding segments are indicated in slashes. Contrasted values with significant differences in bold.

		r		σ		VV	
sentence 13	A	51	**140**	136	**186**	**96**	-
	B	56	**170**	157	**228**	**115**	-
	p <	ns	**.004**	ns	**.003**	**.006**	-
sentence 14	A	51	**56**	127	**121**	115	
	B	48	**71**	130	**141**	118	
	p <	ns	**.01**	ns	**.02**	ns	
sentence 15	A	48	**57**	131	**130**	122	
	B	42	**83**	130	**151**	110	
	p <	ns	**10^{-4}**	ns	**.003**	ns	

The pattern that emerges for speaker 3 is similar to those of the other two speakers. For this subject, differences in duration for all rhymes and syllables are significant at second position. No significance is found at first position. The lack of a measure for the second VV in sentence 13 for this speaker was due to the impossibility of separating /ɾ/ from /i/ in the sequence /kɾisp/ (the tap was realized as an approximant). In this very sentence, there is significance at first position for VV (higher values in clash condition): the only difference between the contrasted utterances is the greater duration of /m/ (word "mestre") in clash condition: mean value of 59 ms against 45 ms (p < .008).

5. DISCUSSION

What emerges from both experiments is that, as regards VV units, there is an increase of duration when approaching phrasal stress in non-clash as well as in clash condition. Since stress groups in clash condition have minus one VV in comparison to those in non-clash condition, the rate of increasing is greater for the former, which explains the higher duration at second position in the target word. There is a significant difference between the two positions of the target word for VV in sentence 13 for subjects 1 and 2. This general pattern is present even for syllables and rhymes for the three subjects of experiment 2 (see Figure 3 for the increasing pattern of rhymes within the target word). This reinforcement of the iambic pattern of the target word is completely opposite to what has been described for American English. It follows, however, the patterns predicted by the speech rhythm production model presented in this paper as shown in table 2 (section 3.6) with two paired sentences of section 4 as examples: when a stressed syllable follows a word such as "café", the duration of its rhyme (and that of the following consonant) is significantly longer than when a pre-stressed syllable follows the same word. This would be true for right-headed languages using duration as independent variables signaling phrasal stress. It is likely, however, in situations related to pitch accent placement, and in two intriguing cases discussed in the following that stress shift seems to occur in BP.

In experiment 1, sentence 8, stress shift seems to be related to differences in pitch alignment, which means that it could be the case that it does not occur without an accompanying accent shift. Interestingly, when this very sentence is realized under focus the pattern disappears. This very fact is seemingly related to the realization of only one pitch accent in the phrasally stressed syllable of the focussed utterance (which means that there is no accent clash). In experiment 2, two cases of stress shift seem to occur: one for sentence 15 (subject 1), and another for sentence 13 (subject 3). For these two cases, the first unit of the target word (syllable for subject 1, VV for subject 3) has higher duration in clash condition, which is in accordance to one of the solutions of the Rhythm Rule. It is important to note, however, that these patterns do not occur without a higher duration at second position (see tables 6a and 6c).

The higher duration in the first syllable of the target word for subject 1 is due to a stronger attack of /k/ (in the word "café") in clash condition, which is likely to be related to the previous stressed syllable of "tomar" (see Barbosa & Madureira, 1999, where we claim that the heterosyllabic VC is the domain for phrasal stress culmination). Besides following the general duration pattern predicted by the model, this subject seems to signal a stress alternation by making the first onset consonant in the target word more prominent in the sequence "tom**ar c**afé fr**i**o". This general pattern is in accordance with the model in another sense: since phrasal stress is closer to /k/ than in non-clash condition, the rate of duration increasing is higher in that case, which explains higher durations for the VV /ark/.

For subject 3, the only difference between the paired utterances besides the higher duration of the VV at first position in the target word is a stronger attack of /m/ (in the word "mestre") in clash condition, which could be related to an alternation of strong and weaker VV in the whole sentence, reinforcing the impression of the succession of a trochaic pattern (paroxytons, the non-marked paradigm in BP): "O m**estr**e **Tom**ás **Crisp**a preg**ou p**or par**áb**olas". For the first VV, the explanation that unifies both patterns is the same as before: as the rate of duration increasing predicted by the model is higher in clash condition, higher VV durations are expected from the beginning of the stress group to the end.

Thus these intriguing cases do not contradict the model, since the duration increasing pattern towards phrasal stress is higher for clash conditions, as can be seen for the predicted VV periods in table 2: sentence 3 against 4, and sentence 5 against 6. All results indicate, otherwise, that stress shift seems to be optional and related to the organization of the whole utterance, including the control of f_0 for signaling pitch accents.

How to account for these seemingly segmental manifestations of stress shift in the model is still matter of further investigation. As regards accent shift, an integration of a model of intonation (in strict sense) within the grid given by the speech rhythm coupled-oscillator model need to be developed.

As regards the applications of the model for other languages, we have shown (Barbosa 2002), that the manipulation of the relative coupling strength parameter w_0 can indeed simulate data on rhythm typology: the greater the value of w_0, the closer the pattern of VV periods to that of a stress-timed language; the lesser the value of w_0, the closer the

pattern of VV periods to that of a syllable-timed language. Languages can also vary depending on the way their underlying rhythmic systems interact with the higher-level components of the grammar and with the gestures in the lexicon (where gestural coordination is found, including the one representing lexical stress, as proposed by Albano, op. cit.).

The results presented here seem to indicate that no apparent, systematic, duration-related stress shift seems to take place in BP.

_____

REFERENCES

ALBANO, E. C. (2001). *O Gesto e suas Bordas: Esboço de Fonologia Acústico-Articulatória do Português Brasileiro*. Mercado de Letras: Campinas, Brazil.

ALLEN, G.D. (1972). The location of rhythmic stress beats in English I & II. *Language & Speech*. 15: 72-100, 179-195.

BAILLY, G. (1986). Un modèle de congruence relationnel pour la synthèse de la prosodie du français. *Actes des 15es Journées d' Étude sur la Parole*. Aix-en-Provence, France. pp. 75-78.

BARBOSA, P.A. (1996). At least two macrorhythmic units are necessary for modeling Brazilian Portuguese duration: emphasis on segmental duration generation. *Cadernos de Estudos Lingüísticos*. 31: 33-53.
_____. (2000). "Syllable-Timing in Brazilian Portuguese": uma Crítica a Roy Major. *D.E.L.T.A.,* 16 (2): 369-402.
_____. (2001). Generating Duration from a Cognitively Plausible Model of Rhythm Production. *Proceedings of the Eurospeech 2001*. Ålborg, Denmark. v. 2, pp. 967-970.
_____. (2002) Explaining Cross-Linguistic Rhythmic Variability via a Coupled-Oscillator Model of Rhythm Production. In: Bernard Bel & Isabelle Marlien (eds.) *Proceedings of the Speech Prosody 2002 Conference*. 11-13 April, Aix-en-Provence, France. pp. 163-166.

BARBOSA, P.A. & MADUREIRA, S. (1999). Toward a hierarchical model of rhythm production: evidence from phrase stress domains in Brazilian Portuguese. *Proceedings of the XIVth International Congress of Phonetic Sciences*. v. 1. pp. 297-300.

BROWMAN, C.P. & GOLDSTEIN, L. (1986). Towards an articulatory phonology. *Phonology Yearbook*. 3: 219-252.
_____. (1989). Articulatory Gestures as Phonological Units. *Phonology*. 6: 201-251.
_____. (1990). Tiers in Articulatory Phonology with some Implications for Casual Speech. In: KINGSTON, J.; BECKMAN, M. E. (eds.) *Papers in Laboratory Phonology I,* Cambridge: Cambridge University Press. pp. 341-376.

CONNELL, B.; ARVANITI, A., eds. (1995). *Phonology and Phonetic Evidence: Papers in Laboratory Phonology IV*. Cambridge: Cambridge University Press.

DOGIL, G.; BRAUN, G. (1988). *The PIVOT model of speech parsing*. Wien: Verlag.

FUJIMURA, O. (1995). Prosodic organization of speech based on syllables: the C/D model. *Proceedings of the XIIIth International Congress of Phonetic Sciences*. v. 3,. pp. 10-17.

GRABE, E.; WARREN, P. (1995). Stress Shift: do speakers do it or do listeners hear it? In: CONNELL, B. & ARVANITI, A. pp. 95-110.

JANKER, P. (1995). On the influence of the internal structure of a syllable on the P-center perception. *Proceedings of the XIIIth International Congress of Phonetic Sciences. v.* 2. pp. 510-513.

LEHISTE, I, (1978). Temporal organization and prosody. Perceptual aspects. *Joint Meeting of A.S.A. and A.S.J.* v.1. Honolulu. pp. 1-17.

LIBERMAN, M. & PRINCE, A. (1983). On stress and linguistic rhythm. *Linguistic Inquiry*. 8 (2): 249-336.

MASSINI, G., (1991). *A Duração no estudo do acento e do ritmo em português.* Master's thesis. Campinas: University of Campinas.

MCAULEY, J.D. (1995). *Perception of Time as Phase: Toward an Adaptative-Oscillator Model of Rhythmic Pattern Processing*. Unpublished PhD dissertation, Indiana University, USA.

MORTON, J.; MARCUS, S. & FRANKISH, C. (1976). Perceptual centers (p-centers). *Psychological revue*. 83 (5): 405-408.

O'DELL, M. & NIEMINEN, T. (1999). Coupled Oscillator Model of Speech Rhythm. *Proceedings of the XIVth International Congress of Phonetic Sciences.* San Francisco, USA. v. 2. pp. 1075-1078.
_____. (2001). Speech Rhythms as Cyclical Activity. In: S. OJALA; J. TUOMAINEN (eds.). *21. Fonetiikan päivät Turku 4.-5.1.2001*, *Publications of the Department of Finnish and General Linguistics of the University of Turku*. 67. pp. 159-168.

ÖHMAN, S. (1966). Coarticulation in VCV utterances: spectrographic measurements. *J. Acoustic. Soc. Am.* 39: 151-168.

POMPINO-MARSCHALL, B. (1989). On the psychoacoustic nature of the p-center phenomenon. *Journal of Phonetics*. 17: 175-192.

SHATTUCK-HUFNAGEL, S. (1995). 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: CONNELL, B.; ARVANITI, A. (eds.). pp. 128-140.

SHATTUCK-HUFNAGEL, S.; OSTENDORF, M. & ROSS, K. (1994). Stress shift and early pitch accent placement in lexical items in American English. *Journal of Phonetics*. 22: 357-388.

TESNIÈRE, L. (1965). *Éléments de syntaxe structurale*. Paris: Klincksieck.

VAISSIÈRE, J. (1983). Language-independent prosodic features. In: CUTLER, A.; LADD, D.R. (eds.) *Prosody: models and measurements*. Berlin: Springer-Verlag. pp. 53-66.

VOGEL, I.; BUNNELL, H.T. & HOSKINS, S. (1995). The Phonology and Phonetics of the Rhythm Rule. In: CONNELL, B.; ARVANITI, A. pp. 111-127.