ISCA Archive
http://www.isca-speech.org/archive

Speech Prosody 2004
Nara, Japan
March 23-26, 2004

Unifying Stress Shift and Secondary Stress Phenomena with a Dynamical Systems Rhythm Rule

Plínio A. Barbosa, P. Arantes & L. S. Silveira

LAFAPE, Instituto de Estudos da Linguagem, Unicamp, Brazil
plinio@iel.unicamp.br

10.21437/SpeechProsody.2004-11

Abstract

Rhythmic patterns related to Brazilian Portuguese adjacent stresses and secondary stress were investigated under a dynamical systems perspective. Paired utterances contrasting alleged stress clash vs non clash reveal a duration pattern difference in the opposite direction of the Rhythm Rule: the closer a syllable-sized unit is to phrase stress, the longer its duration. Polysyllabic words may exhibit initial lengthening as a possible indication of an initial secondary stress. The two phenomena are simulated by using the same parameters with the coupled-oscillators model of speech rhythm production.

1. Introduction

During the last Speech Prosody Conference, a model of speech rhythm production able of explaining cross-linguistic differences in syllable-sized patterns of duration was presented [4]. The model, based on previous proposals [3][7], is couched on dynamical systems theory, particularly coupled-oscillators theory ([16] applies this model in several domains of the human performance, specially motor and perceptual activities, and [24] uses a functional (task-specific), dynamical model for converting Articulatory-Phonology – henceforth AP – abstract gestural scores [9] into actual articulatory trajectories).

Our model aims at generating segmental duration as a by-product of higher prosodic domains for languages which control duration as a prosodic parameter, as it is the case of Brazilian Portuguese (henceforth BP), language studied here. In a first step, a sequence of vowel onset-to-vowel onset (henceforth V-V) durations is delivered by a coupled-oscillators model constituted by two oscillators: a syllabic oscillator and a phrase stress oscillator (left side of Fig. 1). By the interaction between the two oscillators, complex patterns of duration can be generated.

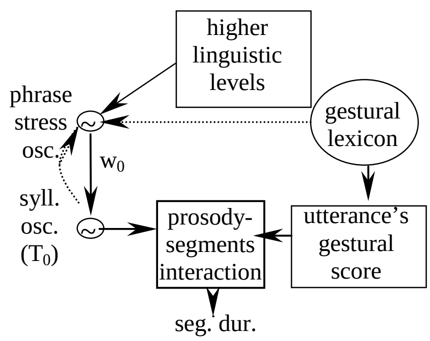

Figure 1: *The dynamical model of rhythm production*

Indeed, in [4], it was shown that language-specific rhythmic patterns within a stress-timing continuum can be explained by this model. After being entrained by the phrase stress oscillator, the syllabic oscillator acts as a pacemaker for a gestural score of the whole utterance.

The phrase stress oscillator is currently implemented by a train of pulses, and the syllabic oscillator implemented by a sinusoidal function. The bi-directional influence of one oscillator onto the other has its degree specified by the relative coupling strength parameter (w_0), which is a real number between 0 and 1. Only the patterns generated by the syllabic oscillator are highlighted here.

The location and magnitude of the phrase stress oscillator pulses, which respectively specify 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 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 [17]. Moreover, the relative coupling strength parameter is language-specific, although it may vary across individuals of a same language with less variability than across languages. Both the phrase stress and the syllabic oscillators are related to universal properties, respectively phrasal prominence and syllabicity.

The syllabic oscillator cycles are abstract in the sense that their duration values do not directly specify overt duration values: they constitute a holistic, gestaltic pattern to be subsequently modified by the perturbational effect of the intrinsic durations of gestures specified in a lexicon. Period coupling in the syllabic oscillator is implemented by the formula (1) below.

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

Where the synchronicity function s(.), given by (2), measures the degree of closeness between the current cycle of the syllabic oscillator and the on-going phrase stress pulse dominating the current stress group. (Note that BP is a right-headed language.)

$$\begin{aligned} s(0) &= w_0.\exp(-N+2) \\ s(N-1) &= \exp(-5.81+0.016.T_0) \qquad (2) \\ s(n) &= (1-w_0).s(n-1) + w_0.\exp(-N+n+2), \text{ for } 0< n < N-1 \end{aligned}$$

In the formulae (1) and (2), N is the number of V-V units in the current stress group, n is the V-V unit index (starting from 0, for the first V-V unit in the stress group), w_0 is the relative coupling strength, and T_0 is the uncoupled syllabic-oscillator period. The parameters α, and β, are respectively the entrainment rate, and the decay rate. The variable i (m) stands for the magnitude of the pulse of the phrase stress oscillator dominating the current stress group, which is the m-th pulse of the train of phrase stress pulses specifying the eurhythmic and syntactic-semantic relations in the utterance.

The current syllabic oscillator period (the V-V abstract duration) is specified by T, and the following one, obtained by adding ΔT to T. The syllabic-oscillator period resetting, implemented by the second term of (1), operates after each phrase stress with a decay rate specified by β. This term is only present if the current V-V unit is dominated by a following, on-going phrase stress. According to this, in utterance-final position (for which there is no such an on-going phrase stress) both stressed and post-stressed V-Vs would be lengthened, simulating the so-called final lengthening. Period resetting is only activated during the first two periods of the syllabic oscillator. The model was optimised in [5] in order to reproduce the stress patterns of a corpus of 36 BP sentences read in three speech rates. The procedure gave the values of 118 ms (fast rate), 138 ms (normal rate), and 205 ms (slow rate) for T_0 (each one of these values can further vary to reflect actual, continuous speech rate control); and the following values for: $w_0 = 0.78$, $(\alpha,\beta) = (0.31, 1.13)$ for the fast rate, $(\alpha,\beta) = (0.38, 1.13)$ for the normal rate, and $(\alpha,\beta) = (0.43, 1.04)$ for the slow rate (same observation as for T_0. Note the smallest value of the entrainment rate for the fast rate, which explains lesser differentiated V-V periods for fast rate, that is, lesser variable V-V durations, which is in accordance with behavioural data).

As regards intrinsic timing, the model is conceived in such a way as to explain segmental cross-linguistic contrasts, such as full vs reduced vowels and consonants by specifying full and reduced gestures in a lexicon, as predicted by [9] and recently argued in [1], from analyses of BP acoustic phonetic data. In [1: 141-179] the interaction of metrical and gestural timing was already suggested. This interaction is crucial to generate overt durations (see [3] for an application in speech synthesis, where the prosody-segments interaction is realised by a recurrent neural network). This part of the model as well as the interaction with the higher linguistic levels are currently under implementation.

The coupled-oscillators model is able to generate global stress patterns for BP utterances, as those related to stress clash and secondary stress phenomena (provided stress group delimitation and number of V-V in each group are given).

2. Stress clash in (Brazilian) Portuguese

In [6], statistical analyses of minimal-paired two-word sequences (embedded in identical carrier phrases) contrasting a stress clash condition and a non-stress clash condition, recorded by four BP male subjects, revealed that lexical stress clash does not trigger the rhythm rule [17]. On the contrary, the second (and not the first) syllable-related unit of the first, stress-shiftable word in the sequence was systematically more prominent in the clash condition. These results are seemingly contradictory to those of [14], and [15], but it is important to note that, in a language such as English, stress shift is apparently a matter of pitch accent placement, which seems to support the notions of early pitch accent placement and pitch accent alternation, proposed in [26]. The results in [19] and, more recently, in [22], seem to confirm the alternation account, since the probability of stress shift increases in direct proportion to the increasing of the distance between potential candidates for pitch accent placement. A related phenomenon, reported in [22], is the occurrence of stress shift in utterances where stresses do not clash. Only the results for V-V duration patterns are shown here, since the main goal of this paper is to compare the behavioural data with the model simulations.

2.1. Stress clash analysis and simulation

In the paired sentences in (3), from a larger corpus analysed in [6], the two-word sequences are italicised, and the two stress groups are separated by squared brackets. Lexically stressed syllables are underlined. In sentence (3a), two lexically stressed units clash (stress clash condition), but not in (3b) (non stress clash condition).

[O *bordeaux xucro*] [derramou-se pela mesa.] (3a)
[O *bordeaux chinês*] [derramou-se pela mesa.] (3b)

The results and significance presented here are based on ten repetitions of one speaker [6]. The V-V actual mean durations for the first two units in the shiftable word "bordeaux" are (173 ms; 253 ms) in (3a), and (165 ms; 195 ms) in (3b), that is, the second V-V, [o ʃ], is 30 % longer in the stress class condition.

This result was simulated with our speech rhythm production model, by using the parameters in section 1 for the normal rate, which gives the abstract (non-perturbed by intrinsic duration specification) durations for the first and second V-V: (146 ms; 163 ms) for a simulation of (3a), and (161 ms; 212 ms) for a simulation of (3b). The second simulated V-V duration is then 30 % longer than the first one in the stress clash condition, that is, the accordance between model prediction and behavioural data is exact!

In [6], we concluded that there is no evidence for duration-related stress shift in BP. As regards f_0, [18] showed, with one notable exception, for a subset of the corpus presented here, no evidence for accent shift. The exception is related to the left-edge marking of an utterance under stress clash condition with a slight rise-fall f_0 movement (and longer durations for both V-V of the shiftable word), which can be explained as an early pitch accent placement [26] signalling the left edge of the first stress group.

In order to investigate if the phenomenon described here would be found as well as could be simulated in another Portuguese variety, a fifty-year old, female speaker of European Portuguese (EP) was recorded. The speaker read the same corpus of five paired sentences, which was read by the BP male speaker analysed here.

The results for V-V durations in EP conform to those of BP: V-V duration differences across the two conditions are non-significant in the first V-V of the target word, and highly significant in the second V-V ($p < 0.001$). For the pair of sentences in (3), the V-V mean durations in the second position are 246 ms (clash condition) and 218 ms (non-clash condition), a 13 % difference. This smaller difference was simulated with the model with $(\alpha, \beta, T_0, w_0) = (0.15, 1.00, 0.200 \text{ s}, 0.95)$ producing the contrast 242 ms (second V-V, clash cond.) vs 215 ms. No differences in f_0 excursion were found for this subject across the two conditions, with the exception of an expected difference in tonal alignment within the trigger word (earlier in 'xucro'), as in BP.

3. Secondary stress in Brazilian Portuguese

The majority of the generative and post-generative rule-based treatment of Secondary Stress (henceforth SS) assignment in Romance are based on lexical items, compounds, or two-word phrases (see [23] for Spanish, [27] Italian; and [12] BP). These studies also share another aspect, that of impressionistically assigning SS position. Even though the rule of initial prominence predominates in [27] for monomorphemic words, phonological rules state a binary alternation algorithm for SS assignment, which includes BP [12].

Phonetic studies of SS in Romance (see [21] for Spanish, [8] Italian, and [13][20] BP) on the other hand, fail to find evidence in the classic stress-related prosodic parameters (duration, f_0, and intensity) for a rule of binary alternation. They advance the notion of one initial secondary prominence in the items studied. These results nevertheless contrast with those presented in [25], for which no articulatory correlates distinct from those of unstressed syllables were found in Spanish, nor in Polish. In the following, we try to reconcile all these findings.

In [2], we started a pilot-study aiming at investigating the (possible) acoustic phonetic correlates of SS in BP. Our results to date depict a more complex panorama than that presented in the literature. Since duration is the most important acoustic parameter for the implementation of stress in BP, we started our investigation with the analysis of duration patterns for V-V units and syllables.

Seventeen polysyllabic, penultimately-stressed words with different number of pre-stressed units (from two – e.g., *horroroso*, to five – e.g., *dirigibilidade*) were chosen for analysis. They were inserted in four distinct carrier sentences, in order to vary distance (d_σ) of phrase stress from the stressed syllable of the word in four ways: phrase stress falls on the stressed syllable of the word, $d_\sigma = 0$ ("A __ parece menor hoje.", The ___ seems shorter today.), or from two to four syllables to the right, $d_\sigma = 2, 3, 4$, ("A __ rude/rural/bicolor parece menor." The rude/rural/bicolor __ seems shorter.). If the word begins with a vowel, the plural form of the article was used instead, with all plural flexions in the sentence changed accordingly. Ten repetitions of the sentences were read by another BP male speaker.

The next figures show mean patterns of normalized V-V durations (z-scores) from the beginning of the stress group for all nine words with two pre-stressed syllables (Fig. 2), and all six words with three pre-stressed syllables (Fig. 3). Different patterns in the same figure stand for the four conditions of phrase stress placement. For the statistical analyses, position of V-V in the stress group (POSITION) and distance from the stressed syllable of the word to phrase stress (DISTANCE) were taken as the two categorical factors.

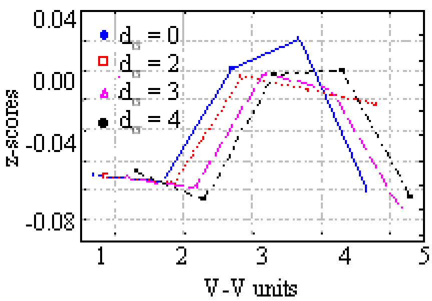

Figure 2: *V-V z-score patterns within words with two pre-stressed units (for d_σ from 0 to 4, rightwards).*

For words with two pre-stressed syllables (Fig. 2), POSITION was highly significant ($F_{2,1305} = 456.06$, $p<10^{-4}$), but not DISTANCE neither their interaction. However, a Scheffé test reveals that positions 1 and 2 do not differ significantly from each other. The phrasally stressed V-V is, however, significantly different from the other phrasally stressed V-V in the stress group with $d_\sigma = 2$ to 4 (Scheffé test with $p<10^{-2}$). For words with three pre-stressed syllables (Fig. 3), POSITION was also highly significant ($F_{3,1144} = 124.3$, $p<10^{-4}$), but not DISTANCE neither their interaction. The lexically stressed V-V which coincides with phrase stress is also significantly different from the other stressed V-V (Scheffé test with $p<10^{-2}$). Taking only the two first V-V durations for analysis, positions 1 and 2 are significantly different for $d_\sigma = 2$ (p=0.002) and $d_\sigma = 4$ (p=0.03).

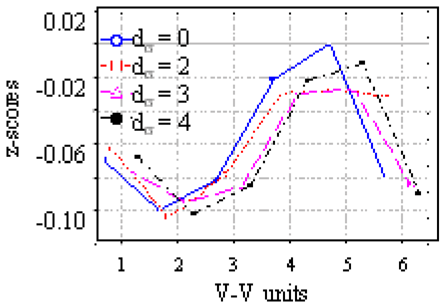

Figure 3: *V-V z-score patterns within words with three pre-stressed units (for d_σ from 0 to 4, rightwards).*

These results indicate that, as the number of V-V in the stress group increases (due to an increase of word length), there is clear signalling of initial prominence in the stress group containing this word (this first V-V is indeed formed by the vowel of the article, and a following consonant, which can be either the consonant of the first syllable of the word, or the /s/ of the article plural form). This prominence seems then to be related to the rhythmic patterns' structuring for the whole utterance. No binary alternation seems to take place. The data in [25], commented above, can be understood by considering the small significance when comparing the first two V-V durations within the stress groups.

This general pattern was also simulated with the rhythm production model, giving the patterns presented in Fig. 4, for the parameters $(\alpha,\beta,T_0,w_0) = (0.5, 0.8, 0.158\ s, 0.7)$. Note the strengthening of the initial prominence as the number of V-V within the stress group increases (from left to right: 5 to 8).

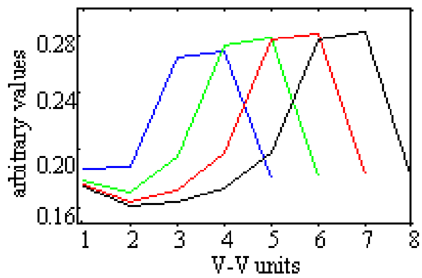

Figure 4: *Simulated V-V durations.*

4. Unifying the two phenomena

Rhythmic patterns related to secondary stress implementation as initial prominence were simulated by the dynamical systems model proposed in [4], [5]. The results shown here also indicate a distinction in the dynamics of phrase stress implementation according to the distance of phrase stress from the beginning of the stress group. This kind of distinction also explains the durational differences in sentences contrasting a condition of seeming (lexical) stress clash against a control condition with no clash. The only case of left-edge mark in this respect could be simulated via duration (f_0 movement implementation requires a model integrating the generation of intonational patterns) in the same way as for the simulation of the patterns for polysyllabic words embedded in carrier sentences. If this left-marking was really triggered by the (lexical) stress clash, it is a matter for further investigation.

Thus, according to the model, both phenomena behave exactly in the same way: as mechanisms of left- and right-edge marking of the stress group domain. As regards clash, there is no lexical stress clash if both words are in the same stress group. A clash would be possible for adjacent words in two different stress groups. But this clash would be probably undone by changing phrase stress hierarchy or position before utterance implementation (in the model, this is done by the higher linguistic levels).

5. Final remarks

The speech rhythm production model presented here is able to simulate and relating two phonetic/phonological processes in BP: duration patterning under lexical stress clash and duration patterning for polysyllabic words. This was possible by implementing extrinsic timing with a coupled-oscillators model, which suggests that this kind of model is a candidate for supplementing gestural phonologies [10] with rhythmic information distributed over the whole utterance, as opposed to the local solution posited by the π–gestures in [11].

6. Acknowledgements

This work was partially financed by research grants from CNPq. It is integrated to the FAPESP project number 01/00136-2: “Integrating Continuity and Discreteness in Modeling Phonic and Lexical Knowledge”.

7. References

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

[2] Arantes, P.; Barbosa, P. A., in press. Acentuação secundária em português brasileiro à luz de modelo dinâmico do ritmo: um estudo piloto. *I Congresso Internacional de Fonética e Fonologia.* Belo Horizonte, Brazil (October 2002).

[3] Barbosa, P. A., 2001. Generating duration from a cognitively plausible model of rhythm production. *Eurospeech 2001*. Ålborg, Denmark, 2, 967-970.

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

[5] Barbosa P. A., 2002b. Explaining Brazilian Portuguese resistance to stress shift with a coupled-oscillator model of speech rhythm production, *Cadernos de Estudos Lingüísticos* 43, 71-92.

[6] Barbosa, P. A.; Arantes, P., 2003. Investigation of non-pitch-accented phrases in Brazilian Portuguese: no evidence favoring stress shift. *XVth ICPhS.* Barcelona, Spain, 475-478.

[7] Barbosa, P. A.; Madureira, S., 1999. Toward a hierarchical model of rhythm production: evidence from phrase stress domains in Brazilian Portuguese. *XIVth ICPhS*. San Francisco, USA, 1, 297-300.

[8] Bertinetto, P. M., 1976. L’accento secondario nella fonologia italiana. Analisi teorica e sperimentale. In *Studi di fonetica e fonologia*. Roma: Bulzoni, 189-236.

[9] Browman, C. P.; Goldstein, L., 1986. Towards an articulatory phonology. *Phonology Yearbook* 3, 219-252.

[10] Browman, C. P.; Goldstein, L., 1990. Tiers in Articulatory Phonology, with some implications for casual speech. In *Papers in Laboratory Phonology I,* J. Kingston; M. E. Beckman (eds.). Cambridge: C. U. P., 341-376.

[11] Byrd, D.; Saltzman, E., 2003. The elastic phrase: modeling the dynamics of boundary-adjacent lengthening. *J. Phon.* 31, 149-180.

[12] Collischonn, G., 1994. Acento secundário em português. *Letras de Hoje* 29, 43–53.

[13] Gama Rossi, A. J., 1998. Qual é a natureza do acento secundário no português brasileiro? *Cadernos Centro Universitário São Camilo* 4, 77–92.

[14] Grabe, E.; Warren, P., 1995. Stress shift: do speakers do it or do listeners hear it? In *Phonology and phonetic evidence: papers in Laboratory Phonology IV*, B. Connell; A. Arvaniti (eds.). Cambridge: C. U. P., 95-110.

[15] Gussenhoven, C., 1991. The English rhythm rule as an accent deletion rule. *Phonology* 8, 1-35.

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

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

[18] Madureira, S., 2002. An acoustic study of phonological phrases containing sequences of words with adjacent primary-stressed syllables: does stress shift occur in Brazilian Portuguese? *Cadernos de Estudos Lingüísticos* 43, 109-125.

[19] McCormack, P. F.; Ingram, J. C., 1995. The rhythm rule in speech production: the effect of inter-stress syllables. *XIIIth ICPhS*. Stockholm, 3, 242-245.

[20] Moraes, J. A., 2003. Secondary stress in Brazilian Portuguese: perceptual and acoustical evidence. *XVth ICPhS.* Barcelona, 2063–2066.

[21] Prieto, P.; van Santen, J., 1999. Secondary stress in Spanish: some experimental evidence. In *Aspects of Romance Linguistics*. C. Parodi; C. Quicoli; M. Saltarelli; M. L. Zubizarreta (eds.). Washington: Georgetown Univ. Press, 337-356.

[22] Quené, H.; Port, R. F., 2003. Produced speech rhythm depends on predictability of stress patterns. *XVth ICPhS.* Barcelona, Spain, 2445-2448.

[23] Roca, I., 1986. Secondary stress and metrical rhythm. *Phonology Yearbook* 3, 341-370.

[24] Saltzman, E. L.; Munhall. K. G., 1989. A dynamical approach to gestural patterning in speech production. *Ecological Psychology* 1, 333-382.

[25] Scharf, G.; Hertrich, I.; Roca, I.; Dogil, G., 1995. Articulatory correlates of secondary stress in Polish and Spanish. *XIIIth ICPhS*. Stockholm, 4, 634-637.

[26] Shattuck-Hufnagel, S.; Ostendorf, M.; Ross, K., 1994. Stress shift and early pitch accent placement in lexical items in American English. *J. Phon.* 22, 357-388.

[27] Vogel, I.; Scalise, S., 1982. Secondary stress in Italian. *Lingua* 58, 213-242.