TOWARD A HIERARCHICAL MODEL OF RHYTHM PRODUCTION: EVIDENCE FROM PHRASE STRESS DOMAINS IN BRAZILIAN PORTUGUESE

Plínio A. Barbosa* and Sandra Madureira†

**Universidade Estadual de Campinas and †Pontifícia Univerdade Católica de São Paulo, Brazil*

ABSTRACT

Acoustic final lengthening is addressed in this work by examining segment durations at three different phrase boundaries. Each prosodic boundary strength also orthogonally compares two kinds of Brazilian Portuguese lexical pattern (paroxyton and oxyton). Careful statistical analyses show that overall V#C lengthening distinguishes the two weaker phrase boundaries from each other. It is also shown that postboundary onset consonants are lengthened by the same process stretching the preboundary vowel. At utterance boundary, oxytons and paroxytons do not entrain distinct lengthenings in postboundary consonants, but they do so in the two other prosodic conditions. For them, the stress degree of the last vowel extends to the postboundary consonant. These results indicate that initial lengthening is not the only phenomenon cueing initial strengthening at phrase left edges. Based on these and earlier data, a dynamical, hierarchical model of rhythm production is presented.

1. INTRODUCTION

Acoustic final lengthening is well documented in the speech production literature. Research shows that segments undergo more than two degrees of lengthening, depending on the strength of the particular prosodic boundary (see [7][17][20], for English and [3], for French). These different strengths are usually considered as a result of the projection of specific nodes of a prosodic hierarchy onto the phonetic string of segments. Phonological words and phrases, intonational phrases and phonological utterances may constitute domains delimitating this string, depending on the particular node label.

At phrase boundaries, it is usually reported that the last vowel, the final consonant, the final rhyme or the final syllable are lengthened. What is pointed out in this work is, first, that the domain of final lengthening is not necessarily isomorphous to phonological units like syllables or their constituents but it extends across phrase boundaries and it includes at least one postboundary onset consonant. And second, that in Brazilian Portuguese (henceforth BP), this consonant lengthening cannot be characterized as initial lengthening alone but it is also a part of the very final lengthening that stretches the preboundary rhyme (here, a vowel).

2. RESEARCH BACKGROUND

By studying segmental acoustic durations in syllable-sized frames, we have presented evidence against phonological syllables as phrasal stress bearing units. Earlier works [2][4] have in fact shown that segments within VC sequences are homogeneously lengthened at phrasal stress. This result reinforces Vaissière's claim that rhymes (tautosyllabic VC) seem to be universal units for signalling prosodic strength [19]. Even phonological research confirms this trend by distinguishing heavy syllables from light syllables. The first ones are characterized by the branching structure of their rhymes (typically, VC) and are potential or obligatory stress attractors in many languages [10].

But our results have shown that the same is true for VC sequences across boundaries. That is to say, not only rhymes are lengthened at phrasal stress but also the postboundary onset consonant (if present). In other words, our findings suggest that VC_0 (#) C_0 units act as rhythmic programming units, i.e. potential candidates for bearing phrasal stress, depending on the particular prosodic boundary strength. This behavior is consistent with two related notions.

Marcus and colleagues [15] have coined the term P-center for designating acoustic anchor points in the speech signal which are used by listeners to perceive sequences of syllables as occurring isochronously. More ecological data [11] have confirmed Marcus's intuitions about the P-center location: the vowel onset. It seems to us that P-centers would delimitate the minimal programming units for the phonetic implementation of rhythm. Port and colleagues' experiments confirm this assumption by choosing vowel onsets as rhythmic beats in their speech cycling task [8].

Another related issue is the characterization of the speech chain as twofold: a continuous vowel flow (probably represented by the alternate movement of jaw openings and closings [18]) on which consonant gestures are superimposed [9][16]. We consider here that lengthened consonants in typical V(#)C sequences are by-products of greater displacements of jaw closing movements beginning during vowel production. Byrd and Saltzman's recent paper [6] seems to confirm this assumption for lip movement at five different prosodic boundaries.

The experiment we describe here shows that consonants after three different phrase stress boundaries are lengthened and that, at least in BP, this phenomenon can also be explained as final lengthening.

2.1. A Pilot Study

The starting point for the results shown here was the observation of a non-random variation of [p] occlusion durations in the phrase “para ele” (*to him*), embedded in the sentences “Digo *word* para ele.” (*(I) say __ to him*). At the variable place marked *word*, the three BP lexical stress patterns, proparoxytons (antepenultimate stress), paroxytons (penultimate stress) and oxytons (final stress) were contrasted in 126 sentences. Word length was three or four syllables long. Phrase stress falls on each contrasted word. This corpus was recorded with one subject at three speech rates (self-chosen normal and metronome-controlled fast and slow rates) in order to study BP rhythmic and intonational patterns.

We have measured [p] occlusions from the previous vowel offset to the onset of the first vowel in “para” [ˈparɐ]. We have decided to compare [p] durations according to the stress pattern of the previous word. Statistical analyses were carried out and no differences were found for the fast rate. But for the normal rate, occlusion intervals were 8 ms ($p < 0.0001$) shorter after paroxytons or proparoxytons in comparison to those after oxytons. An even greater difference in the same direction was found for the slow rate: 56 ms ($p < 0.0001$).

Proparoxytons and paroxytons have had the same effect on [p] durations. This can be explained by a tendency in BP for reducing non-final stressless vowels in proparoxytons (cf. “xícara” - *cup* -, pronounced [ˈʃikrɐ] or “veículo” - *car* -, pronounced [veˈiklʊ]). Thus, rhythmically, proparoxytons behave as the unmarked lexical stress paradigm: those of the paroxytons.

These figures indicate that phrasal stress seems to extend to the postboundary onset consonant. In order to evaluate this hypothesis a more controlled experiment was carried out.

3. EXPERIMENT DESCRIPTION AND RESULTS

Two other subjects (AJ and GR) recorded twelve repetitions of three-paired sentences. The three pairs of sentences contrasted prosodic boundaries of three different strengths. Very strong (or utterance boundary) s1: “Zé diz ___. Chapado também.” (*Joe says ___. Chapado too.*), strong s2: “Digo ___ chapado e baixo.” (*(I) say ___ unconsciously and softly.*) and weak s3: “José Paulo diz ‘___ chapado” (*Joseph-Paul says ‘___ unconsciously.’*). Each sentence pair orthogonally contrasts “se gaba” [sɪˈgabɐ] (*(he’s) boasting*) and “se gabá” [sɪgaˈba][1] (*to boast*), i.e. the paroxyton and the oxyton pattern.

The durations of three phonetic segments are shown in the tables below: the postboundary segment [ʃ], the first vowel [a], both in the word “chapado” [**ʃa**ˈpadʊ], and the sequence [a ʃ], in [sɪ gaˈb**a ʃ**aˈpadʊ] (only for oxytons). A t-test evaluated the null hypothesis for the differences in duration under two conditions: paroxyton or oxyton at phrase boundary. Significance was accepted for p-values lesser than 2%.

In table 1, speaker AJ produced eight repetitions (oxyton pattern) and four repetitions (paroxyton pattern) of the sentence s1 with additional silent pauses at phrase boundary. These sentences were discarded as it is preferable to preserve identical event sequences for all utterances. Results for those cases are 143 (16) ms (paroxyton) and 158 (21) ms (oxyton) and the difference between them is not significant. Speaker GR produced all sentences s1 with a silent pause at phrase boundary (that is why these results are not shown in table 1). For this speaker, postsilence mean (and standard deviation) [ʃ] durations are 132 (11) ms (paroxyton) and 135 (12) ms (oxyton). This difference is also not significant

sent.	speaker AJ		speaker GR	
type	paroxyton	oxyton	paroxyton	oxyton
s3	*112 (4)	*122(11)	124 (5)	127(12)
s2	*135 (16)	*155(14)	*150 (8)	*163(11)
s1	170 (18)	179 (11)	-	-

Table 1: Mean durations (and standard deviations) in ms of [ʃ] in the word “chapado” for the three prosodic boundary conditions. Stress pattern of previous word is indicated. Significant differences are starred. In s2, p is lesser than 1%. See text for additional explanations.

For speaker AJ, table 1 shows significant differences for [ʃ] duration after phrase boundary: [ʃ] following oxytons are longer than those following paroxytons in s3 (10 ms) and s2 (20 ms). Speaker GR shows the same trend: [ʃ] are 13 ms longer in average after oxytons in s2. In both subjects, these differences are no more distinctive at utterance boundary (s1).

Table 2 shows that this lengthening is restricted to the onset consonant [ʃ], it does not spread to the prestressed vowel in “chapado”. This vowel onset sets the beginning of the second stress group, which ends at utterance final stress.

sent.	speaker AJ		speaker GR	
type	paroxyton	oxyton	paroxyton	oxyton
s3	61 (6)	64 (9)	54 (7)	54 (5)
s2	60 (4)	58 (5)	65 (5)	63 (6)
s1	66 (6)	64 (3)	-	-

Table 2: Mean durations (and standard deviations) in ms of the first [a] in the word “chapado” for the three prosodic boundary conditions. Stress pattern of previous word is indicated. There are no significant differences between paroxyton/oxyton pairs.

Back to table 1, it is worth noting that differences between [ʃ] durations for paired sentences of different prosodic strengths (for each lexical pattern condition) are significant for subjects AJ (s1/s2, s2/s3 and s1/s3) and GR (s2/s3). But it is important to stress that the role of distinguishing between different phrase boundaries is mainly attributed to VC units (and also to individual segments, like the consonant [ʃ] in this work, as a by-product of lengthened VCs or as a consequence of phrase-initial phase adjustments: see section 5). If an utterance ends with a vowel, the boundary is cued by stretching this vowel alone. The same is valid for phrases ending with a vowel, if the postboundary segment is also a vowel. Degree of phrase stress can be inferred by observing durations of the V#C sequence [a ʃ] in table 3, for oxytons at phrase boundary. For sentence s1, only the vowel duration is shown.

The V#C durations distinguish phrase boundary conditions s3 and s2 in both subjects. But the vowel duration alone in s1 is not sufficient to separate this case from the others. Additional cues beside final lengthening are probably used to decode utterance boundary. The significantly different initial lengthening of [ʃ] seen in table 1 (speaker AJ, sentence s1) is certainly one of them but syntactic information is also certainly used (this seems to be the case for speaker GR).

sentence type	speaker AJ	speaker GR
s3	238 (24)	275 (34)
s2	343 (20)	417 (36)
s1	221 (14) [a]	-

Table 3: Mean durations (and standard deviations) in ms of the sequence [a ʃ] (oxytons) for the three prosodic boundary conditions. Differences between s3 and s2 are significant for both speakers. For s1 and s3, results are statistically indistinct.

4. DISCUSSION

We have shown elsewhere [2] that only V-to-V durations can reveal correct prosodic hierarchies underlying utterance production. Local perturbations to this general trend can nevertheless occur depending on syllable structure (in the sequence CVC#CV, weaker effects - than those shown here - on postboundary C are found), on particular language (for BP, the whole syllable is lengthened at lexical stress not phrasally prominent [2]). In this work, the duration of the sequence [a ʃ] distinguishes two phrase boundary strengths in both subjects. But, in BP, oxyton and paroxyton stress patterns allow us to go further.

If the lengthening of postboundary onset consonants were only due to the so-called initial lengthening, results would show no differences between the two lexical stress pattern conditions. But in fact they do. It seems easier to interpret this fact as being due to the extension of the stress degree of the previous vowel to the following consonant: in oxytons, the last vowel [a] is stressed and its stress degree extends to [ʃ] across the phrase boundary. In paroxytons, the last vowel [ɐ] is not stressed and its stresslessness extends to the next [ʃ]. Major boundaries (as in s1) seem to block this rule. Only for these boundaries lengthening would be restricted to the class of “initial lengthening” phenomena (cf. longer [ʃ] durations for s1, table 1, speaker AJ).

Another issue still needs to be treated here. Which time interval cues prosodic strength for paroxytons at phrase boundary. In this case, post-stressed VCs seem to be governed by the previous stressed vowel. The culminative interval for stress is likely to be bounded by the stressed vowel onset and the postboundary consonant across boundary with the inclusion of post-stressed segments.

Research on f0 patterns shows that post-stressed syllables constitute domains where f0 movements unfold [12]. For sentence s1, f0 contours do not change during all syllables in paroxytons but go down during post-stressed syllables in oxytons. For sentences s2 and s3 with paroxyton or oxyton patterns, f0 contours rise at stressed vowel onset, get higher during post-stressed syllables (in paroxytons) and level off during the postboundary syllable [ʃa]. If post-stressed syllables constitute a domain where stress culminates, the greater perceptual prominence of proparoxytons over the other two lexical stress patterns can also be explained.

Moraes [14] has in fact confirmed that listeners perceive proparoxytons as being more salient than paroxytons, which are more prominent than oxytons, *ceteris paribus*. In our hypothesis, this is due to the greater duration of the stressed interval: stressed VC plus two post-stressed units in proparoxytons, stressed VC plus one post-stressed unit in paroxytons and just the stressed VC in oxytons. In paroxytons, our results also show a slightly greater stability of the durations of the sequence [abɐ ʃ] (case 1) in comparison with those of the sequence [ɐ ʃ] (case 2) for sentence s3 (where segments are more coarticulated to each other at boundary). Coefficients of variation are 5%/7% (case 1/case2), for speaker AJ, and 8%/9% (case 1/case2), for speaker GR.

This discussion enables us to propose a model of rhythm production integrating Metrical Phonology insights but constrained by the dynamical apparatus of the speech production system.

5. TOWARD A HIERARCHICAL MODEL OF RHYTHM PRODUCTION

Articulatory Phonology (henceforth AP) [5] has shown the advantages and power of integrating the first-order dynamic constraints of a mass-spring system into their gestural framework. But AP fails to account for differences in vowel height among languages, easily explained by acoustic facts. It also fails to explain phenomena of reduction and elision in BP, usually characterized by a more constricted derived gesture. And it only has begun to show how the prosodic organization of utterances should be.

By adressing the first two questions, Albano [1] has tried to solve the AP drawbacks with the integration of acoustic information into a gestural framework and the introduction of other kinds of constraint in the paired tract variable equations associated with some gestures (as the glide gestures in the cited reference). We would like to address the third point by proposing a dynamical model of rhythm production.

This model uses the notion of period and phase entrainment reported in [13], but extends it to rhythm production. To entrain an oscillator means to synchronize it by means of external stimulation (from an input oscillator).

In order to explain acoustic duration data in BP and in French [2][3][4], we propose at least three hierarchically connected extrinsic, abstract clocks (or oscillators). The first clock stands for the vowel flow (section 2), the second one stands for lexical stress sequence and the third clock represents the eurhythmic phrasal stress sequence. These clocks act as the input oscillators in McAuley’s experiments.

The abstract vowel flow clock is related to a corresponding to-be-entrained clock representing, on the other hand, the level of rhythm production. The period of this clock explains, at the physical level, V-toV duration flow regularity. The onsets of the abstract vowel flow clock cycles would be anchored at vowel gesture onsets of AP-like gestural scores [1][5]. The corresponding physical clock would control the alternate jaw openings and closings. The phrasal stress clock, on the other hand, divides the utterance in evenly produced chunks by reinforcing the amplitudes of the previous clock at intervals of an integer number of syllable-sized periods. (Syntactic information would be used to avoid perturbing linguistic input by stress misplacement.) It also triggers a slow-down entrainment mechanism increasing the period of the vowel flow clock. These two clocks are supposed to be universal.

The (lexical) stressed vowel clock is supposed to be language specific. It acts as an input for the phrasal stress clock and as a stress beat which reinforces (and reduces eventual poststressed beats) the vowel flow clock amplitudes at lexical edges, in languages like BP. French would not have this kind of clock.

The perceptual sensation of syllable-timed or stress-timed languages would be explained by concentrating the attention either on the vowel flow clock or on the stressed-vowel clock (if it exists). French listeners would use the first kind of clock to judge the V-toV continuum and English listeners would use the second one to make expectancies about the lexical stress flow.

In this framework, segmental durations are only by products of this clock hierarchy controlling the mandible cycling and of the phasing of consonant gestures around vowel onsets [1][5]. Initial lengthening would be treated with this kind of phase readjustments. No clock is then needed at this microrhythmic level of implementation. The next figure shows the model currently being implemented with recurrent neural networks.

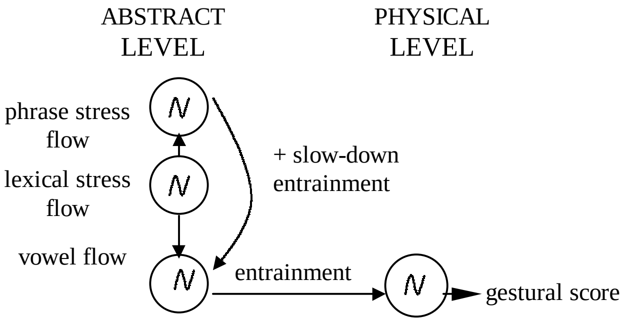

Figure: Clock levels of the hierarchical rhythm production model

ACKNOWLEDGMENTS

This work is financed in part by the project *Jovem Pesquisador em Centro Emergente* (proc. n° 95/09708-6) from Fundação de Amparo à Pesquisa do Estado de São Paulo-FAPESP and by a grant (proc. n° 524 110/96-4) from the Conselho Nacional de Desenvolvimento Científico e Tecnológico-CNPq. We also especially thank our subjects for their voices and Eleonora Albano for her very helpful suggestions.

NOTES

1. The deletion of the final infinitive marker /r/ is a common phenomenon in BP.

REFERENCES

[1] Albano, E.C. 1999. Gestural solutions for some glide epenthesis problems. *These Proceedings*.

[2] Barbosa, P.A. 1996. At least two macrorhythmic units are necessary for modeling Brazilian Portuguese duration. *Proceedings of the first ESCA Tutorial and Research Workshop on speech production modeling*, 85-88.

[3] Barbosa, P.A. 1994. *Caractérisation et génération automatique de la structuration rythmique du français*. Unpublished thèse de 3^e cycle, INPG/ICP, Grenoble, France.

[4] Barbosa, P. and Baillly, G. 1994. Characterisation of rhythmic patterns for text-to-speech synthesis. *Speech Communication*, 15 (1-2), 127-137.

[5] Browman, C. and Goldstein, L. 1990. Tiers in Articulatory Phonology with some implications for casual speech. In: Kingston, J. and Beckman, M.E. (Eds.) *Papers in Laboratory Phonology I.* Cambridge: Cambridge University Press, 341-376.

[6] Byrd, D. and Saltzman, E. 1998. Intragestural dynamics of multiple prosodic boundaries. *Journal of Phonetics*, 26, 173-199.

[7] Campbell, W.N. 1991. Phrase-level factor affecting timing in speech. *Proceedings of the 2nd European Conference on Speech Communication and Technology*, 2, 629-632.

[8] Cummins, F. and Port, R. 1998. Rhythmic constraints on stress timing in English. *Journal of Phonetics*,26, 145-171.

[9] Fujimura, O. 1995. Prosodic organization of speech based on syllables: the C/D model. *Proceedings of the $XIII^{th}$ International Congress of Phonetic Sciences,* 3, 10-17.

[10] Goldsmith, J. 1990. *Autosegmental and metrical phonology*. Oxford: Blackwell Publishers.

[11] Janker, P. 1995. On the influence of the internal structure of a syllable on the P-center perception. *Proceedings of the $XIII^{th}$ International Congress of Phonetic Sciences*,2, 510-513.

[12] Madureira, S., Barbosa, P.A., Fontes, M., Spina, D. and Crispim, K. 1999. Post-stressed syllables in Brazilian Portuguese as f0 markers. *These Proceedings*.

[13] McAuley, J.D. 1995. *Perception of time as phase: toward an adaptative-oscillator model of rhythmic pattern processing*. Unpublished PhD dissertation, Indiana University, USA.

[14] Moraes, J.A. 1986. Acentuação lexical e acentuação frasal em português. Um estudo acústico-perceptivo. *Conference presented at the II Encontro Nacional de Fonética e Fonologia,* Brasília, Brazil.

[15] Morton, J., Marcus, S. and Frankish, C. 1976. Perceptual centers (p-centers). *Psychological revue*, 83(5), 405-408.

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

[17] Oller, D.K. 1973. The effect of the position in utterance on speech segment duration in English. *J. Acoustic. Soc. Am*., 54, 1235-1247.

[18] Rhardisse, N. and Abry, C. 1995. Mandible as syllable organizer. *Proceedings of the $XIII^{th}$ International Congress of Phonetic Sciences*,3, 556-559.

[19] Vaissière, J. 1983. Language-independent prosodic features. In: *Prosody: models and measurements*. Cutler, A. and Ladd, D.R. (Eds.) Berlin: Springer-Verlag, 53-66.

[20] Wightman, C.W., Shattuck-Hufnagel, S., Ostendorf, M. and Price, P. 1992. Segmental durations in the vicinity of prosodic boundaries. *J. Acoustic. Soc. Am* ., 91, 1707-1717.