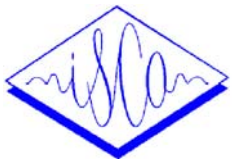

Speech Prosody 2002
Aix-en-Provence, France
April 11-13, 2002

Explaining Cross-Linguistic Rhythmic Variability via a Coupled-Oscillator Model of Rhythm Production

Plínio A. Barbosa

Lab. of Phonetics and Psycholinguistics & Dep. of Linguistics, IEL/UNICAMP, Brazil
plinio@iel.unicamp.br

Abstract

A recent work by Ramus and colleagues has renewed the interest in rhythm typology connected to the evaluation of speech production data. They proposed that differences in rhythm type could be accounted for by a segmental set of variables derived from the acoustic duration of consonants and vowels. However, rhythm typology can be more interestingly characterized and understood by modeling speech rhythm production. The purpose of our study is then threefold: (a) showing why a deeper understanding of languages' rhythmic types can only be achieved by modeling their underlying rhythmic systems; (b) presenting the results of the implementation of a coupled-oscillator rhythmic system which simulates language-specific continuous patterns of syllable-sized durations; (c) suggesting that only a rhythmic system integrated to a gestural framework can account for the complexity of rhythm phenomena.

10.21437/SpeechProsody.2002-26

1. Introduction

Based on a notion of speech isochrony as the regular occurrence of a sequence of beats along the speech chain, Pike [18] has coined the terms of syllable-timed and stress-timed languages. The former would have the sequence of all syllables in an utterance as regular beat carriers and, the latter, the sequence of the stressed syllables. In 1967, Abercrombie [1] asserted that "every language in the world is spoken with one kind of rhythm or with the other." (p. 97), setting the basis for the emergence of a dichotomy. Undeniably, his assumption influenced the search for absolute isochrony on speech production which turned out to be troublesome no matter Kelly's [11] attempt to minimize Abercrombie's influence by arguing that the latter was only concerned with the phonological properties of languages when establishing a rhythm typology. Some surveys on the history of rhythm typology can be found in [7], [13] and, more recently, in [4], with respect to a reanalysis of the case of Brazilian Portuguese (BP), which completely rejects Major's [14] findings.

Since absolute isochrony in stress-timed languages such as English [12] and in syllable-timed languages such as French [20] has been denied, a weaker notion of isochrony has emerged from perceptual [3][13] and more elaborated linguistic studies [7][10]. These studies revealed that the explanation for rhythm type recognition is certainly more complex than originally thought: several variables seem to play a role in speech rhythm perception and production.

Recently, a renewal of interest in speech production acoustic data aimed at evaluating rhythm typology is found in [19]. Ramus and colleagues proposed that differences in rhythm type could be accounted for by a set of variables derived from the acoustic duration of consonantal and vocalic acoustic intervals, which is very similar to Dauer's considerations on the role of cross-linguistic differences in syllable structure and vowel reduction (among others).

2. A segmentally-oriented technique for describing rhythm typology

The acoustic variables used by Ramus and colleagues were the percentage of vocalic interval durations (%V) and the standard-deviation of consonantal interval durations (ΔC) within the same utterance. (Vocalic and consonantal intervals are uninterrupted strings of vowels and consonants, respectively.) According to perceptual tests carried out by the authors, newborns seem to distinguish languages on the sole basis of percentage of vocalic interval durations.

These two segmental variables were able to discriminate data on eight languages. Polish, English and Dutch constituted a first class (stress-timed), Spanish, Catalan, French and Italian, a second class (syllable-timed) and Japanese, a third one (mora-timed). Interestingly, the two variables, %V and ΔC, seem to be strongly correlated.

Unfortunately, there is no reference in their data to speech rate, which could indicate that it was not controlled at all. If this is so, very different results for position of languages' data onto the (%V, ΔC) plane can be expected.

Another technique for characterizing rhythm types can be used with comparable success as to descriptive power, but with more chance of explanatory success. This technique, explained in the next section, has been recently used [5] to compare European Portuguese with BP. Besides demonstrating the impossibility of any kind of cross-linguistic comparison for corpora uttered at distinct rates, the results also suggest the only way to advance towards the understanding of rhythm typology is taking into account both segmental and prosodic factors in speech rhythm modeling.

3. Coupled-oscillators as long-term, qualitative descriptors of speech rhythm production

O'Dell and Nieminen [16] used a mathematical technique to obtain long-term, qualitative descriptions of coupled oscillator models. By suggesting the coupling of a syllable and a stress group oscillator, they were able to explain the durational patterns of early research on speech isochrony based entirely on strict considerations of timing and coupling strength.

They showed that the coupling strength of the stress group oscillator on the syllable oscillator is equivalent to the ratio between the point of interception and the inclination of the regression line computed with the variables: "duration of stress group" and "number of syllables within the stress group". By doing so they were able to restate Dauer's data [10] on isochrony in continuous terms: the higher the coupling strength, the more stress-timed a language is. According to the authors, the coupling strength seems to exhibit a relative

stability across speakers and speech styles [17]. By using the same technique, we confirmed in a recent paper [4] the interest of such a technique to reevaluate BP rhythmic typology.

Besides showing that Major's five arguments for characterizing BP as stress-timed are all misleading, this work shows that BP has a great amount of syllable timing, and that the comparability of speech rate is crucial in cross-linguistic studies. In fact, as speech rate increases in BP, the coupling strength also increases. By comparing utterances of a BP corpus pronounced at three distinct speech rates with other languages, BP presents less stress timing than Thai and British English, and more stress timing than Italian and Greek. In comparison to European Spanish and Finnish, the three languages seem to exhibit the same degree of stress timing. (The correlation coefficients for all regression analysis carried out on BP utterances were significant and greater than 80 %.)

By means of a qualitative, long-term analysis it is not possible to track what the syllable oscillator is undergoing at every cycle. In order to do that, a model taking into account the moment-to-moment consequences of the syllable and stress group oscillator coupling is needed.

4. Coupled-oscillators as moment-to-moment descriptors of speech rhythm perception

In 1995 McAuley proposed an Entrainment Model in order to account for human rhythm perception processing [15]. Entraining means modifying the parameters of an oscillator according to predefined laws of adaptation operating on the adaptive system. It is clear that such a system is a class of coupled-oscillators systems.

In McAuley's model, a train of pulses, considered as the input, entrains phase resetting and period coupling onto a cosinoidal oscillator. The phase resetting is triggered by the addition of a coupling strength function $w_i.i(t)$ (where i(t) represents the input stimulus) to the cosinoidal oscillator amplitude.

In the latter oscillator, period coupling is controlled by phase resetting, direction of coupling (with a reset-phase sign), and an output signal o (.) measuring the current degree of synchronicity between the oscillator being adapted and the input. The amount of 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 period, T is the current period, P (.) is the reset phase sign, and M (.) is the impulse response function. The function M (.), which has the value of 1 during entrainment and of 0 during decay, allows to take into account decay and entrainment in a single formula, by separating their respective contributions. (In fact, a simultaneity of entrainment and decay is possible during transitional periods of time.)

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

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

5. A moment-to-moment coupled-oscillator model of speech rhythm production

In our model, the input signal i (n) is a train of pulses, representing a stress group oscillator and the to-be-entrained oscillator is a syllable-sized oscillator, as in the O'Dell-Nieminen model (the syllable-sized oscillator maxima coincide with vowels onsets, based on the importance of CV transitions as rhythmic anchor points [3][4].)

Stress beats in the stress group oscillator are not necessarily periodic (although they can be) and are given by higher level linguistic input (semantic, syntactic, and lexical components), as well as eurhythmic constraints (see figure 4, below).

The measure of synchronicity is given by the empirically determined function s(.), in (2). This function was obtained by non-linear regression analysis on BP stress groups. The relative change of V-V durations along each stress group was computed and transformed by means of a given function. The best results were obtained with an exponential. N is the number of syllables in the stress group, and w_0 is a language-specific parameter representing the degree of coupling strength (the coupling strength is in fact given by $w_0.i$).

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

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

Phase resetting is considered to be achieved at each syllable-sized oscillator maximum. Period coupling is obtained with the finite-differences equation (3).

In this equation, period resetting is achieved after each stress beat by the second term of (3), where β is the decay rate (provided the utterance bears an on-going stress beat, otherwise the entrainment mechanism represented by the first term of the equation is used instead. This is exactly the case at utterance-final position, where stressed and post-stressed V(.)C are lengthened). This period resetting mechanism is active during the two first periods of the syllable-sized oscillator. In the first term of (3), s'(n) is the synchronicity signal s(.) with w_0 factored out, and α is the entrainment rate. The period is up-dated only at the maxima of the syllable-sized oscillator.

$$\Delta T = \alpha.T.s'(n).w_0.i(n) - \beta.(T-T_0).i(n) \quad (3)$$

This model is able to simulate classes of language-specific syllable-sized durational patterns in accordance to the notion of stress timing. These patterns account for the prosodic aspects of rhythm production (lower-level aspects of rhythm production will be considered in section 7).

6. Generating language-specific syllable-sized durational patterns

The entrainment of the syllable-sized oscillator was implemented with the help of MatLab®, version 6.0 (release 12). In the following examples of simulations, the period coupling was simulated with three degrees of strength by changing the parameter w_0 in (3). All the other parameters were kept unchanged during the simulations, that is, the values for the entrainment and decay rates (0.5 and 0.8, respectively), the amplitude of the stress group oscillator (1.0), and the initial period of the syllable-sized oscillator (100 ms). The number of syllables in the stress group is either four or eight (chosen as such to make the comparisons with the stress timing literature easier).

The next three figures show the entrained syllable-sized oscillators for three values of w_0 (0.3, 0.7, and 1.3) standing for distinct degrees of stress timing for three hypothetical right-headed languages.

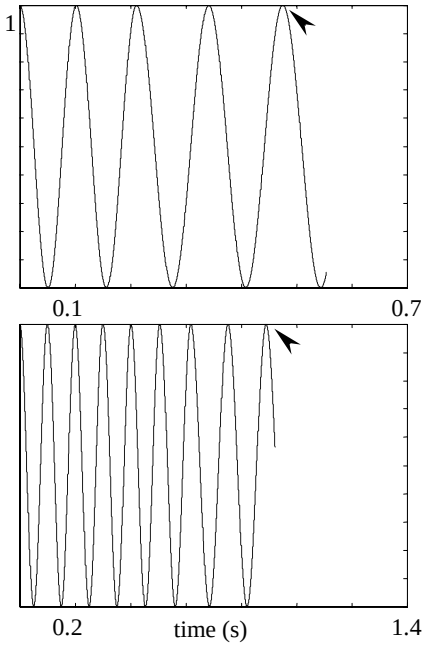

Figure 1: Entrained syllable-sized oscillators for four- (top) and eight-syllable (bottom) stress groups with $w_0 = 0.3$. The arrows point to the end of the fourth and eighth cycle, respectively.

In figure 1, the time position of the last peak of the eight-syllable stress group (0.89 s) is very close (but not coincides) to twice the time position of the last peak of the four-syllable stress group (0.48 s), which is an alleged characteristic of syllable-timed languages. The period increasing in both stress groups is slow, and it is more effective only during the last three cycles in both groups, which creates the syllable-timed pattern (note however that there is no such a thing as isochronous syllables: all cycles have different durations).

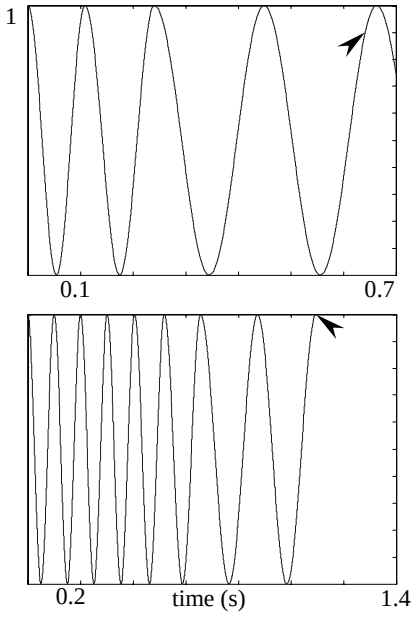

Figure 2: Entrained syllable-sized oscillators for four- (top) and eight-syllable (bottom) stress groups with $w_0 = 1.3$.

In figure 2, the position of the peak of the last cycle of the eight-syllable stress group (1.0 s) is closer to the position of the peak of the last cycle of the four-syllable stress group (0.67 s) than in the case of figure 1, which is an alleged characteristic of stress-timed languages (note however that there is no equality of durations between 4-syllable and 8-syllable stress groups).

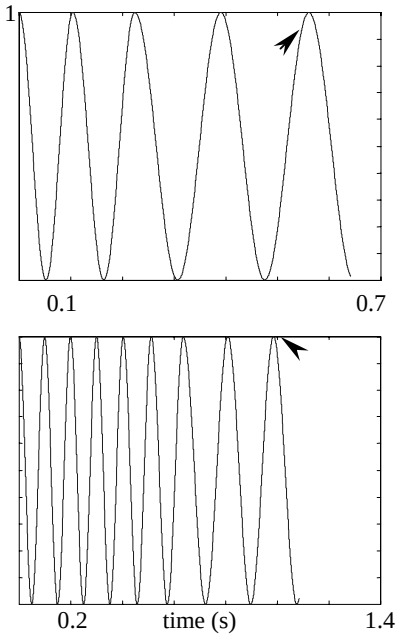

Figure 3: Entrained syllable-sized oscillators for four (top) and eight syllables (bottom), with $w_0 = 0.7$.

Figure 3 shows an in-between situation: the last peak of the eight-syllable stress group is at 0.99 s and that of the last peak of the four-syllable stress group is 0.56 s.

By using the durations of both stress groups in each pair to compute the coupling strength according to the O'Dell-Nieminen technique, we found respectively 0.6 (coupling strength less than 1, a tendency to syllable timing), 2.1 (coupling strength greater than 1, a tendency to stress timing), and 1.3 (coupling strength around 1, an in-between stress timing characteristic). These results confirm what has just been stated about the simulations for the entrained syllable-sized oscillators. BP would correspond to the situation in figure 3.

It is important to remark that despite the use of a period expressed in milliseconds, the entrained syllable-sized oscillator is an abstract oscillator, that is, it does not deliver overt values for V-V durations. The first step in order to obtain actual durations (either articulatory or acoustic) is to consider the inclusion of the interaction of the coupled-oscillator system (the rhythmic system) with the gestural score for an entire utterance. This gestural score is however essentially different from the Articulatory Phonology (AP) framework [8], because the positions at the left edges of Tongue Body gestures representing vowels (not closed, non-critic constraint degree) are set extrinsically by the entrained syllable-sized oscillator peaks (in AP, with the exception of the rearranging of consonants according to the vowel flow, and to a rhythmic tier [9], there are no detailed rules for the coordination of vowels). The rules for C-to-V coordination are given by the AP framework [9].

The last step to obtain actual durations is done by modeling prosody-gestures interaction using a recurrent neural network, just as shown in figure 4 (reproduced from [6]).

At the left of figure 4, the coupled-oscillator system represents the rhythmic system, with the entrained syllable-sized oscillator acting as a pacemaker for the gestural score (built by reorganizing the lexical gestural scores). The connectionist network is recurrent in order to take time into account. By achieving the interaction of the gestures *g* of the gestural score with the entrained syllable-sized oscillator, it delivers silent pause *p* and segment durations *dur (s)*.

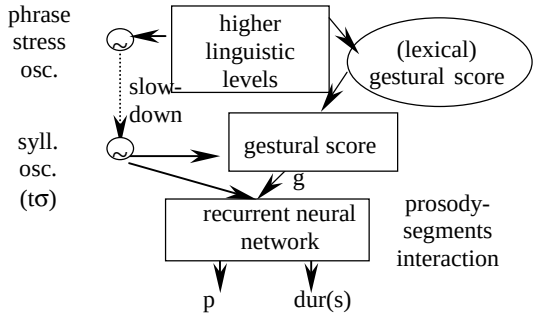

Figure 4: The dynamical model of rhythm production

In this framework, (phrase) stress and syllabic oscillators are considered universal properties of language because they are hypothesized to be present as coupled oscillators in our cognitive system. The coupling strength between the two oscillators, the lexical gesture score, and the higher linguistic levels are considered language-specific properties.

7. Perspectives within a dynamical framework

The simulations shown here demonstrate the computational power of coupled oscillator systems: their capacity for reproducing distinct patterns of duration variation as a result of changing a single parameter in equation 3 (w_0).

It is also possible to simulate the effect of distinct stress beat magnitudes onto the duration pattern: the greater the value of the amplitude of i (.), the greater the changing in the period. In this way, increases in duration can be obtained in a single language by controlling the magnitude of stress beats.

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 [2]).

By simulating a second interaction with the recurrent neural network, and by exploring in greater detail the way the first interaction takes place, it will be possible to attain a deeper understanding of the complexity of languages' rhythmic types.

8. Acknowledgments

We thank Eleonora Albano and Sandra Madureira for their helpful suggestions. This work was partially financed by a grant from a FAPESP project (n° 95/09708-6) and by a research grant (n° 350382/98-0) from CNPq, associated with the project n° 524110/96-4. It is integrated to the FAPESP project number 01/00136-2: "Integrating Continuity and Discreteness in Modeling Phonic and Lexical Knowledge".

9. References

[1] Abercrombie, D., 1967. *Elements of General Phonetics*. Edinburgh: Edinburgh University Press.

[2] Albano, E. C., 2001. *O Gesto e suas Bordas: Esboço de Fonologia Acústico-Articulatória do Português Brasileiro*. Mercado de Letras: Campinas, Brazil.

[3] Allen, G. D., 1972. The Location of Rhythmic Stress Beats in English I & II. *Language & Speech*, 15, 72-100, 179-195.

[4] Barbosa, P. A., 2000. "Syllable-Timing in Brazilian Portuguese": uma Crítica a Roy Major. *D.E.L.T.A.*, 16 (2), 369-402.

[5] Barbosa, P. A., 2000. Illuminating some Methodological Issues Concerning Speech Timing Research from a Comparison between European and Brazilian Portuguese. *Cadernos de Estudos Lingüísticos*, 39, 41-50.

[6] Barbosa, P. A., 2001. Generating Duration from a Cognitively Plausible Model of Rhythm Production. *Proceedings of the Eurospeech 2001*, Ålborg, Denmark. v. 2, 967-970.

[7] Bertinetto, P. M., 1989. Reflections on the dichotomy "stress-" vs "syllable-timing". *Revue de Phonétique Appliquée*, 91-92-93, 99-130.

[8] Browman, C.; Goldstein, L., 1989. Articulatory Gestures as Phonological Units. *Phonology*, 6, 201-251.

[9] Browman, C.; Goldstein, L., 1990. Tiers in Articulatory Phonology with some Implications for Casual Speech. In *Papers in Laboratory Phonology I*, Kingston, J. and Beckman, M. E., eds. Cambridge: Cambridge University Press, 341-376.

[10] Dauer, R. M., 1983. Stress-Timing and Syllable-Timing Re-Analysed. *Journal of Phonetics*, 11, 51-62.

[11] Kelly, J., 1993. David Abercrombie (Obituary). *Phonetica*, 50, 68-71.

[12] Lea, W. A., 1974. Prosodic Aids to Speech Recognition: IV. A General Strategy for Prosodically-Guided Speech Understanding. *Univac Report PX10791*, *Sperry Univac, DSD*, St. Paul, Minnesota, USA.

[13] Lehiste, I., 1977. Isochrony reconsidered. *Journal of Phonetics* 5, 253-263.

[14] Major, R. C., 1981. Stress-timing in Brazilian Portuguese. *Journal of Phonetics*, 9, 343-351.

[15] McAuley, J.D., 1995. *Perception of Time as Phase: Toward an Adaptative-Oscillator Model of Rhythmic Pattern Processing*. Unpublished PhD dissertation, Indiana University, USA.

[16] 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, 1075-1078.

[17] O'Dell, M.; Nieminen, T., 2001. Speech Rhythms as Cyclical Activity. In *21. Fonetiikan päivät Turku 4.-5.1.2001*, S. Ojala; J. Tuomainen (eds.). *Publications of the Department of Finnish and General Linguistics of the University of Turku*, 67, 159-168.

[18] Pike, K., 1945. *The Intonation of American English*. Ann Arbor: University of Michigan Press.

[19] Ramus, F., Nespor, M. & Mehler, J.,1999. Correlates of linguistic rhythm in the speech signal. *Cognition*, 73, 265-292.

[20] Wenk, B. J.; Wioland, F., 1982. Is French really syllable-timed? *Journal of Phonetics*, 10, 193-216.