ILLUMINATING SOME METHODOLOGICAL ISSUES CONCERNING SPEECH TIMING RESEARCH FROM A COMPARISON BETWEEN EUROPEAN AND BRAZILIAN PORTUGUESE

PLÍNIO A. BARBOSA
(UNICAMP)

RESUMO Algumas questões metodológicas referentes à pesquisa sobre a organização temporal da fala são apresentadas e discutidas. A partir de uma comparação de dois *corpora* de frases lidas por dois locutores brasileiros e uma locutora portuguesa, avalia-se duas técnicas que visam à determinação do tipo rítmico das línguas: uma técnica dinâmica voltada para aspectos prosódicos e uma técnica voltada para aspectos segmentais. Discutem-se a importância da condição ceteris paribus e os critérios de mensuração nos experimentos em Fonética Acústica, no intuito de iluminar o uso adequado de ambas as técnicas. A primeira parece, no entanto, dar melhor conta daquilo que se conhece da tipologia das línguas do mundo. A segunda pode ser melhorada se alguns cuidados metodológicos básicos são levados em conta. O uso conjunto das duas técnicas poderá ser de grande utilidade para os estudos de tipologia rítmica.

0. INTRODUCTION

Research on speech timing is a typical example of how important it is to maintain under severe control all but one factor that could affect the measure of any phonetic parameter. Let us take a linguistic variable such as lexical stress and a phonetic parameter such as acoustic duration of vowels as examples. If the study of how different lexical stress patterns affect vowel duration is carried out, all extra-linguistic and all the other linguistic variables must be, as far as it is possible, exactly the same across speech materials. This requirement is called the *ceteris paribus* (CP) condition and it is absolutely necessary for deriving correct conclusions from data. In fact, if one or more variables were not controlled, the variability exhibited by the vowel duration could be attributed to the variability of anyone of the uncontrolled variables.

In this respect, the design of an experiment where duration is the parameter to be measured is particularly difficult, so many factors can affect its value, even for a single segment. These factors can be linguistic, such as the segment's own nature (e.g., open vowels usually take more time to be produced than close vowels), the nature of the surrounding segments (in Brazilian Portuguese – henceforth BP –, and in European Portuguese, henceforth EP, vowels are longer when followed by voiced obstruents than when followed by the voiceless homorganic obstruents), the segment's position within the syllable, within the word (stressed, pre-stressed or post-stressed), within the stress group and the sentence, and extra-linguistic factors such as speaking style, including speech rate. (A more exhaustive list can be found in Lehiste 1970 and Klatt 1987 for English, and in O'Shaughnessy 1984 and Bartkova & Sorin 1987 for French.)

Among the referred factors, speech rate is crucial. As an utterance speeds up, phenomena such as coarticulation can drastically alter the citation-form phonetic segments and their acoustic correlates (see Kohler 1995, for instance). Furthermore, as vowels and consonants are subject to different compressibility laws, the relation between consonant and vowel durations also alter significantly. That is why every experiment set to study speech timing should at least ensure that speech rate is absolutely under control, after all, to ignore the CP condition is a short path to derive wrong conclusions from data.

Cross-linguistic speech timing studies are maybe the area where it is more common to get the CP condition violated, so many are the ways two languages can vary.

This work presents an attempt to compare BP and EP rhythms from a dynamical, prosodic point of view. By trying to do so, some methodological issues on speech timing research are discussed in fine detail vis-à-vis a segmentally-oriented technique also able to evaluate rhythm typology.

1. RESEARCH BACKGROUND

On the basis of some hypotheses as the universality of notions such as *stress group* and *syllable-size unit* (these ones related, on the other hand, to the notions of *stressing* and *syllabicity*), I recently proposed (Barbosa, 2000) to evaluate the usefulness of the linear regression between "stress group duration" and "number of syllables within the stress group" for shedding some light on the stress-timing vs syllable-timing dichotomy and for reopening the discussion about stress-timing in speech production.

This reevaluation was possible by considering the equilibrium state of a coupled-oscillator system constituted by a syllable and a stress-group oscillator, as proposed by O'Dell & Nieminen (1999)[1].

The O'Dell-Nieminen system is able to take into account the consequences of the mutual effect between the stress-group and the syllable oscillators at the equilibrium state. The effect of the stress-group oscillator on the syllable oscillator is represented by a real number, the coupling strength r, which they showed to be equivalent to the ratio between the interception point and the inclination of the regression line for the variables "stress group duration" (I, plotted at the y-axis) and "number of syllables within the stress group" (n, plotted at the x-axis). From the equation of the regression line in (1), the coupling strength (r) is obtained by doing r = a/b.

$$I = a + b.n \qquad (1)$$

The greater the value of r, the more stress-timed a language is. Thus, in order to make intra- and cross-rhythmic comparisons by computing coupling strengths, the following methodological steps must be observed:

---

[1] We proposed a similar system on the same occasion (Barbosa & Madureira 1999), with two differences: the addition of a third oscillator (lexical stress) and the possibility of generating moment-to-moment V-V and segment durations. Recently, we have modified this rhythm production model (Barbosa in press) in order to integrate phonological gesture scores by attributing lexical stress to differences in gesture magnitude and intergestural phasing (Albano 2001).

1. Delimit the stress groups by identifying stresses for each utterance (this can be done by performing a perceptual test and by using production criteria: e.g, V-V duration greater than twice the averaged V-V duration). If the language is left-headed (as English, German and EP), the stress position starts the group; if the language is right-headed (as BP and French), the stress position ends the group;

2. Compute the duration I of each stress group by previously setting its boundaries at vowel onsets, instead of syllable edges (see Barbosa 2000 for details);

3. Count the number n of V-V units for each stress group;

4. Compute the linear regression parameters a and b (and also the correlation coefficient R) for the pair of variables (n, I);

5. Obtain the coupling strength r by computing the ratio a/b. Retain the parameter b, as it is a measure of speech rate.

By recording a corpus of 36 sentences read by a BP male speaker under three speech rates, and rigorously following the five steps above, we were able to obtain the coupling strengths according to speech rate and size of stress groups (this was done in order to guarantee the CP condition for the comparison of BP with the languages in Dauer 1983, where stress groups have up to four syllables).

Table 1: Strength of coupling (r), regression line equations and correlation coefficient (R) for three speech rates. Size of stress groups (SG) is also presented as a superior limit for number of V-V units. BP is considered as right-headed. Adapted from table 2 in Barbosa (2000).

Speech Rate	size SG < 5 V-V	size SG < 8 V-V
Slow	I = 267 + 186.n R = 0.73 **r = 1.44**	I = 270 + 185.n R = 0.84 **r = 1.45**
Normal	I = 114 + 143.n R = 0.80 **r = 0.80**	I = 169 + 122.n R = 0.82 **r = 1.39**
Fast	I = 152 + 99.n R = 0.78 **r = 1.53**	I = 172 + 92.n R = 0.81 **r = 1.86**

Table 1 shows that coupling strengths vary with speech rate and also with size of stress groups. It can be seen that slow and fast rates favor stress-timing, as indicated by their greater coupling strengths in comparison with that in normal rate. Size of stress groups also favors stress-timing, as the inclusion of groups with more than four V-V units has the effect of increasing r. It is worth noting that correlation coefficients are greater than 70% for all cases.

It is not surprising that fast rate and slow rates exhibit different behaviors, since they are associated with very different underlying production strategies, as studied by Kohno (1992, 1995). Fast rates are produced as a holistic (gestaltic) sequence of syllable-size units. A consequence of this holistic production is that fast-rate unstressed syllables are very stable units that exhibit little variability. On the other hand, normal

and slow rates are analytical sequences of syllable-size units (they are not produced as a whole) which much more variable units, especially at stressed positions (very slow rates would be over-analytical in this sense, that is, very slow utterances would have very long stressed V-V, which would yield a great value for the interception point a, and consequently for r, as it seems to be the case in table 1 for the slow rate). According to Kohno, the threshold between the two phenomena is about a V-V interval of 300 ms.

As stress groups of more than four V-V units are considered for the linear regression analysis, the amount of duration added in the respective group is lesser than the rate of duration increasing in the shorter groups. This has the effect of turning the regression line clockwise, which simultaneously increases the a parameter and decreases the b one (producing greater values of coupling strength).

These findings reinforce my contention in the introduction: speech rate (and size of stress groups) must be the same across two speech materials in order to correctly attribute stress-timing tendencies as due to effective rhythmic differences.

2. CONTROLLING SPEECH RATE OR WHY CROSS-LINGUISTIC COMPARISONS ARE NOT ALWAYS POSSIBLE

The same procedure was employed to analyze 20 sentences produced by two Portuguese-speaking female subjects: one of the Brazilian variety (São Paulo) and the other of the European variety (Lisbon)[2].

Results are presented in table 2. Number of V-V units within stress groups differs a little in both varieties due to differences in vowel reduction or vowel epenthesis. The word "objetivo" (goal) has four syllables in EP and five in BP (due to the epenthesis of [ɪ]: [o̞biʒe'ʧivu]). Words like "descrever" (to describe), which in EP are presented with two elided vowels, have two or three syllables, because in this variety, /d/ and/or /skɾ/ may function as syllabic nuclei (Delgado-Martins, personal communication). Other examples of words with divergent number of syllables across the two varieties were "ofereceu" ((he/she) offered) and one instance of "governador", respectively pronounced [ofɾe'seʊ] and [gvɐrnɐ'doʁ] in EP and with all canonic vowels in BP. (We agree with Moraes & Leite 1992, p. 74, when they say that the measurements must consider the syllables effectively pronounced, but this does not mean that consonant syllabicity is not possible. Besides, EP listeners hear words like "descrever" as having two or three syllables, but never one: Mateus & Delgado-Martins 1982.)

Table 2: Coupling strength (r) and regression equations (R is the correlation coefficient) according to size of stress groups (SG). Outliers were excluded.

BP		EP[3]	
size SG < 5 V-V	**size SG < 8 V-V**	**size SG < 5 V-V**	**size SG < 8 V-V**
I = 78 + 148.n	I = 110 + 143.n	I = 86 + 126.n	I = 125 + 115.n
R = 0.84	R = 0.89	R = 0.82	R = 0.90
r = 0.52	**r = 0.77**	**r = 0.68**	**r = 1.09**

The results indicate, for BP and EP, an increase in coupling strength as stress group size increases. This phenomenon has already been mentioned. It is also worth noting that all correlation coefficients are greater than 80 %. The comparison of the BP-speaking subjects with each other (tables 1 and 2) and of the BP-speaking subjects with the EP-speaking one might reveal interesting aspects of rhythm typology. We will show that, unfortunately, this comparison cannot be completely achieved with these data.

As discussed above, speech rates must be comparable across speech materials; that is, their value must be relatively close according to some predetermined criterion. By choosing in table 1 the closest value for the b parameter of the table-2 BP subject (it is the value for normal rate, both for size of stress groups lesser than five: 143 ms/V-V unit), it can be said that BP coupling strength is smaller for this female speaker. This means that her production is more syllable-timed than the production of the male subject.

The situation is more complex for EP, because this subject has very different b parameters in comparison with the BP-speaking female subject (126 and 115 ms/V-V unit, and 143 and 148 ms/V-V unit respectively). This indicates that the sentences were uttered at distinct speech rates (EP is faster) and a comparison here is not possible, because the CP condition is not fulfilled. Thus, the only thing that can be said about the BP and EP sentences of this corpus is that they were uttered at different speech rates[4].

If the fast-rate production of the male subject is chosen for comparison with the EP-speaking subject instead (see table 1), the difference between the b parameters for BP and EP is greater than 22 % (for size of stress groups up to five or eight V-V units). It is a matter of further investigation whether this difference can be considered negligible.

If the normal-rate production of the male subject is chosen for comparison instead, the closest value for b is that of 122 ms/V-V unit (for size of stress groups up to eight V-V units). This comparison would indicate a BP production more stress-timed than the EP production, a result that needs to be considered with care. This care is related to the evaluation of the comparability of other aspects of speaking style across both corpora, which so far were not considered here. Some of these aspects are related to the degree of lexical-item novelty in both corpora for the corresponding subjects, to the degree of

---

[3] Considered as left-headed. If EP were considered as right-headed the values of *r* would be 0.63, for sizes less than five V-V units and 0.86, for sizes less than eight V-V units.

[4] If other aspects of speaking style can be said to be the same, this result would advance a possibly relevant conclusion: EP speakers normally speak faster than BP ones.

informality the subjects take when uttering, the subjects' systematic use of focus, just to cite a few. (An interesting comparison can nevertheless be done with these EP utterances with those of other languages, as it was done in Barbosa (2000) with Dauer's data. This comparison revealed that this instance of EP is more stress-timed than Cuban Spanish, and more syllable-timed than American English and Swedish. These results are in accordance with knowledge on rhythm typology for these languages.)

It must be clear by now that the coupling-strength technique, which we have considered prosodically-oriented[5], can be satisfactorily used to draw interesting conclusions about rhythm typology. All that it is necessary is to plot coupling strengths (r) against inclination of regression line (b), and compare languages within the same range of b. It is also crucial to previously separate language varieties (e.g. American from British English) because the putative sameness of their rhythmic type should only be decided after experimental work. Both varieties could also have been uttered at distinct speech rates (Dauer 1983's data on American and British English are clearly not comparable due to the non-fulfillment of the latter condition).

Recently, another technique was presented in the literature and has been used by some prosodists in order to guide their investigation on rhythmic typology (see, for instance, Frota & Vigário 1999).

3. COMPARING THE PROSODICALLY-ORIENTED WITH A SEGMENTALLY-ORIENTED TECHNIQUE FOR EVALUATING RHYTHM TYPOLOGY

Ramus, Nespor & Mehler (1999) introduced a technique for evaluating cross-rhythmic typology, which also uses real-valued variables. These variables are 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 or 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. Due to the exclusive use of acoustic segmental durations (and no prosodic variables), I will refer to this technique as segmentally-oriented.

The two segmental variables above were able to discriminate data on eight languages (ibidem, fig. 1, p. 273). 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 (if this is so, both variables are giving the same information, and one could be used instead of the other). Even if this technique can indeed be useful for rhythm typology research, the corresponding experimental work presents some methodological problems.

---

[5] In fact, only prosodic variables as syllable-size units and stress-groups are taken into account.

Ramus *et al.* (1999)'s speech material (four speakers per language, five 15- to 19-syllable sentences per speaker) has a small number of stress groups for each language, which is problematic to obtain statistically stable means and standard-deviations. Besides, there is no reference 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 could be obtained if the speech rate were controlled across languages. This technique presents other disadvantages, which will be discussed later in reference to Portuguese.

Frota & Vigário (1999) used the Ramus-Nespor-Mehler technique for analyzing the same 20-sentence corpus presented above. After computing %V and ΔC for each one of the 20 utterances, Frota and Vigário built a scatter-plot with these data. The regression lines obtained seem to show significant differences between the two varieties[6]: BP seems to have more relative vocalic durations and less consonant dispersion than EP.

The position of BP in the (%V, ΔC) plane would reveal a tendency for syllable-timing for this language, as signaled by Frota & Vigário (ibidem) when comparing Portuguese (both varieties) against Ramus *et al.* 1999's results for the eight languages cited above. Instead, their analyses put BP very close to Japanese (traditionally considered mora-timed). EP, on the other hand, seems to exhibit some syllable-timing (it has more percentage of vocalic duration than Spanish and Italian), which is contrary to expectations (recall what was said above when comparing EP with Cuban Spanish). With respect to BP, EP seems to have comparable %V but ΔC between that for English/Dutch/Polish and that for Italian/Catalan/Spanish/French.

All these results really seem to indicate that these EP utterances are simply too syllabified for typical EP utterances. This conclusion is also in accordance with the data obtained with the coupled-oscillator technique: coupling strengths smaller than one indicate a tendency to syllable-timing and coupling strengths around one are an indication of a mixed rhythmic type (see Barbosa 2000 for details). It is also striking how the syllables of this subject sound distinctively pronounced when listening to the acoustic signal.

A comparison between these EP and BP subjects could yield more interesting results if speech rate were previously controlled (see section 2 for discussion: speech rate in both subjects seems to be distinct). The rather surprising result of the closeness between Japanese and BP may also be due to cross-linguistic data at distinct speech rates.

There are other methodological issues related to a technique that computes vocalic interval durations. Let us examine two examples from Portuguese and discuss a problem of perceptual accuracy.

In Portuguese, there are sound sequences, as obstruent-tap (or tap- obstruent), which exhibit a vocalic sound between the obstruent and the tap (or between tap and obstruent), as in "prato", pronounced ['pratu] which, in a narrower transcription, would

---

[6] About the pros and cons of the putative sameness of BP and EP, see Galves, Nunes & Raposo (2000).

be something as ['pᵉratʊ][7]. In this example, the short sound emerges after the plosive release and before the tap constriction, as a consequence of the on-going movement of the tongue body position toward the following [a]. When a tap is produced in the frame of a complex onset, the articulatory vowel begins around the plosive release and ends around the second plosive onset. A problem arises when making acoustic measurements, because the multidimensional articulatory consequences onto the acoustic medium are linear and an ad hoc decision must be done. One such a decision could be to consider the short vocalic duration adjacent to the tap as part of the duration of the latter (this increase in the consonant interval duration by including the tap's, would change the overall value of the ΔC variable in an ad hoc way). Another possible decision is to consider the short vocalic duration as part of the duration of [a], discarding the tap duration itself as a part of the vowel's (this solution would modify the value of the %V variable). For the coupling-strength technique this kind of decision does not need to be taken: "prato" has two V-V units and that is all (this computation is needed in order to obtain n, the number of V-V units within each stress group).

Another problematic case is the palatalization of the alveolo-dental plosive, which occurs before /i/ in the majority of the BP dialects. In words like "ritmo" (where /t/ is pronounced as the affricate $[\widehat{t\int}]$), there are simultaneous [t] and [ɪ] gestures (otherwise there would be no palatalization) that may overlap either partially or completely. If the complete overlap takes place, how to measure the vocalic interval duration? In terms of number of V-V units, this kind of decision does not need to be taken: as other cases of epenthesis of /i/, this word has three V-V units and that is all.

Still another issue is the perceptual relevance of short vocalic durations. If these measures serve as a basis for early developmentally judgments of C-V alternation, a crucial point is the possibility of newborns to perceive short durations as those in the examples just mentioned as effective vocalic intervals (Portuguese-speaking adults, for instance, judge the short vocalic duration before the tap in "prato" as an integral part of [ɾ]). If, just as adults, infants are not able to judge short vocalic intervals as such, sequences as [pᵉɾ] should instead be considered as a consonantal interval (in this case, [pɾ]). The same reasoning applies to post-stressed epenthetic vowels as those in the BP words "apto" or "ritmo" and the like.

Even if these segmental issues do not constitute any problem for a prosodically-oriented technique as the coupling-strength one, some related methodological aspects must not be ignored. They concern cases where decisions about number of V-V units and stressed positions along the utterances are not so clear. The first case was already mentioned with the EP examples of "descrever" and "governador". The determination of stressed position also deserves some attention. A previous perceptual test is absolutely necessary. From the results of this test, the more systematic hearers' choices can be considered as stressed positions. An additional production criterion can be used in order to consider as stressed the V-V units with at least twice the duration of the averaged V-V duration for each sentence.

---

[7] Silva (1996) has described this phenomenon for BP in acoustic terms.

Despite the existence of methodological issues for both the prosodically-oriented and the segmentally-oriented techniques, the high correlation coefficients, the clear control of speech rate, and the coherence of cross-linguistic comparisons with traditional knowledge on rhythm typology, seem to give advantage to the first one.

4. CONCLUSIONS

The dynamical method advocated here by following O'Dell & Nieminen 1999's suggestions, that is, the prosodically-oriented coupling strength technique, seems to give a more precise account of linguistic rhythm typology.

The Ramus et al. 1999 segmentally-oriented technique seems to be promising but the experimental work needed to derive typological conclusions must be carried out with much more care.

As prosodic and segmental aspects considered together usually give a better account of linguistic typology, my suggestion is the use of both techniques. In this respect, planes of vocalic interval percentage (%V) against coupling strengths (r) variables, parameterized according to speech rate (b is a good measure) are very likely to yield interesting results.

ACKNOWLEDGMENTS

REFERENCES

ALBANO, E.C. (2001) *O Gesto e suas Bordas: esboço de fonologia acústico-articulatória do português brasileiro*. Campinas, Brazil: Mercado de Letras.

BARBOSA, P.A. (in press) O Universal e o Específico a Língua em um Modelo Dinâmico de Produção do Ritmo (Fonético e Fonológico) da Fala. To appear in *Anais do II Congresso Internacional da Abralin*, 14 a 16 de março, Fortaleza, Ceará.

_____. (2000) "Syllable-timing in Brazilian Portuguese": uma crítica a Roy Major. *D.E.L.T.A.*, 16 (2): 369-402.

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*, 1: 297-300.

BARTKOVA, K. & SORIN, C. (1987) A model of segmental duration for speech synthesis in French. *Speech Communication* 6: 245-260.

DAUER, R.M. (1983) Stress-timing and syllable-timing re-analysed. *Journal of Phonetics*, 11: 51-62.

FANT, G., KRUCKENBERG, A. & NORD, L. (1991) Durational correlates of stress in Swedish, French and English. *Journal of Phonetics* 19: 351-365.

FROTA, S. & VIGÁRIO, M. (1999) *Aspectos de Prosódia Comparada: Ritmo e Entoação no PE e no PB*. Presented in the Congress of the Associação Portuguesa de Lingüística in September.

GALVES, C., NUNES, J. & RAPOSO, E. (Orgs.) (2000) Novos Estudos em Gramática Gerativa. Homenagem a Mary Kato. D.E.LT.A. 16 (nº especial).

KLATT, D. (1987) Review of text-to-speech conversion for English, *J. Acoust. Soc. Am.* 82: 737-793.

KOHLER, K.J. (1995) Articulatory reduction in different speaking styles. *Proceedings of the XIIIth International Congress of Phonetic Sciences*, 2: 12-19.

KOHNO, M. (1995) Two different systems for rhythm processing and their hierarchical relation. *Proceedings of the XIIIth International Congress of Phonetic Sciences*, 1: 94-97.
_____. (1992) Two mechanisms of processing sound sequences. In: Tohkura, Y., Vatikiotis-Bateson, E. & Sagisaka, Y. (Eds.), *Speech perception, Production and linguistic structure*. IOS Press. 287-293.

LEHISTE, I. (1970) *Suprasegmentals*. Cambridge, Mass.: MIT Press.

MAJOR, R.C. (1981) Stress-timing in Brazilian Portuguese. *Journal of Phonetics*, 9: 343-351.

MATEUS, M.H.M. & DELGADO-MARTINS, M. R. (1982) Contribuição para o estudo das vogais átonas [ə] e [u] no português europeu. BIBLOS, LVIII, 111-128.

MORAES, J.A. & LEITE, Y.F. (1992) Ritmo e velocidade da fala na estratégia do discurso: uma proposta de trabalho. In: Ilari, R. (Org.) *Gramática do Português falado, v. II: Níveis de análise lingüística*. Campinas: Editora da Unicamp.

O'DELL, M. & NIEMINEN, T. (1999) *Coupled oscillator model of speech rhythm*. Proceedings of the XIVth International Congress of Phonetic Sciences. August 1-7, San Francisco, USA, v. 2, 1075-1078.

O'SHAUGHNESSY, D. (1984) A multispeaker analysis of durations in read French paragraphs. *J. Acoust. Soc. Am.* 76 (6): 1664-1672.

RAMUS, F., NESPOR, M. & MEHLER, J. (1999) Correlates of linguistic rhythm in the speech signal. *Cognition* 73: 265-292.

SILVA, A.H.P. (1996) Para a descrição fonético-acústica das líquidas no Português Brasileiro: dados de um informante paulistano. Master's thesis. Campinas, Brazil: Universidade Estadual de Campinas.