F_1 and Spectral Correlates of Secondary Stress in Brazilian Portuguese

Pablo Arantes and Plinio A. Barbosa

Speech Prosody Studies Group
State University of Campinas, Brazil
parantes@fastmail.fm

Abstract

Vowel opening and spectral emphasis are investigated as possible acoustic correlates of secondary stress in Brazilian Portuguese. F_1 and energy difference between frequency bands were measured for vowel /a/ in five- and six-syllable stress groups. Results for F_1 indicate greater opening for vowels in initial and stressed positions. As for the spectral feature, relative emphasis in bands $B2$ (0.5 to 1 kHz) and $B4$ (2 to 4 kHz) has some relevance to phrase stress signaling, but not to secondary stress if it is to be equated to binary alternations.

1. Introduction

Most experimental work done in Brazilian Portuguese (henceforth BP) gives little support to the idea that binary alternation is at the core of secondary stress (henceforth SS) attribution rule (see the bibliography in [1] and [2] for references about SS theoretical and experimental accounts). Evidence gathered by the the authors of the present paper in earlier works [1][2] suggest that at least for read speech, SS in BP may be better described as an initial prominence signaled by gradient lengthening of a stress group's first V-to-V unit (a V-to-V unit comprises segments between two consecutive vowel onsets. It has been established [3] that V-to-V durations is the main acoustic correlate of stress in the phrasal domain in BP) and also by a high tone (H*) the first syllable of the stress group's first content word. Duration data seems to be well accounted for by a coupled oscillator-based speech rhythm model [3].

In order to further refine the phonetic description of SS in BP we set out to investigate the role of not so commonly studied acoustic correlates of stress in speech production, namely vocalic quality and spectral emphasis. Although duration and f_0 failed to support the traditional intuition about SS placement being binary in nature, both vowel reduction and spectral emphasis can be applied independently from duration and f_0. So there is the possibility that the alleged binarity could be anchored in the presence of reduced vowels along the chain of pre-stressed vowels in polysyllabic words. The binarity effect could also be caused by some pre-stressed vowels being uttered with more vocal effort than others. There is also the possibility that vowel reduction and spectral emphasis fall in a gradient pattern similar to what has been found for duration.

There are just a few broadly acknowledged studies investigating the role of spectral emphasis as an acoustic correlate of stress and accentuation, specially in what concerns Romance languages. For the most part, existent work tends to limit its focus to the role spectral emphasis might play in distinguishing stress and accent. Therefore, it seems worthwhile gathering experimental data that may help to: (i) access the relevance that spectral emphasis might have in the description and modeling of prosodic phenomena in Romance languages; (ii) broaden the range of prosodic phenomena that can be investigated through spectral emphasis analysis and (iii) improve and sharpen the description of SS in BP.

2. Methodology

The corpus we have been using to study SS in BP is composed of carrier sentences where a target word is embedded. The carrier sentences have the following structure:

"A *target-word* [∅|**ru**de|ru**ral**|bu**dis**ta|bico**lor**] parece menor"

Phrase stress falls on the target word's lexically stressed syllable when there is no modifier and shifts to the modifier's stressed syllable (printed in boldface) when it is present. This manipulation creates four levels of distance between the target's stressed syllable and the one bearing phrasal stress — zero when both coincide, 2 (*rude*), 3 (*rural* and *budista*) and 4 syllables (*bicolor*). This independent variable will be referred to as DISTANCE from now on. The greater the DISTANCE factor, the longer the chain of phrasally unstressed syllables, making it easier for the alleged binary tendencies to manifest itself. A naive male speaker of the southeastern BP variety read ten repetitions of each carrier sentence in a sound-attenuated booth.

From a total of 17 target words the present paper will only focus on two of them: *pata**ra**ta* and *jarata**ca**ca* (lexical stress in boldface, alleged secondary stress underlined). The first word has four syllables and the second has five syllables. Both vowel reduction and spectral emphasis will be measured for each vowel in the selected target words plus the definite article "a" which starts the stress group. Each syllable is a level of the independent variable POSITION.

Acoustic analysis were carried out using Praat.

2.1. Vowel Reduction

LPC analysis was used to extract formant values. FFT spectrums were extracted in the vicinity of the duration midpoint of the vowels in the target words and then plotted against the LPC spectrum obtained at the same point to check the goodness of fit. The LPC spectrum was recalculated with a different filter order until a satisfactory fit was achieved. The procedure was manually repeated for nearly 400 individual vowels. Although time costly, this methodology, inspired by [4], was used in order to prevent possible spurious values generated by unsupervised automatic analysis.

Because we are interested in vowel reduction, only F_1, generally accepted as a reliable acoustical correlate of jaw height [5], was analyzed here.

2.2. Spectral Emphasis

Spectral emphasis was extracted following the methodology outlined in [6], one of the most cited sources in the literature on the subject. In that paper, the authors have chosen words and pseudowords with the vowel /a/ in their syllable nuclei and this decision affected the details of the spectral emphasis extraction. The choice of target words from our own corpus was influenced by theirs.

In [6] the choice for the vowel /a/ is justified by the possibility of isolating f_0, F_1, F_2 and $F_3 + F_4$ in four contiguous non-overlapping frequency bands so that it becomes possible to single out what formant or group of formants receives emphasis in each different circumstance. The four frequency bands defined in [6] are: B1: 0 to 0.5 kHz, B2: 0.5 to 1 kHz, B3: 1 to 2 kHz and B4: 2 to 4 kHz.

A Praat query function that calculates energy differences between user-defined frequency bands was employed to obtain a measure of spectral emphasis out of the FFT spectrums generated for the formant analysis described above. Differences were calculated between each two consecutive bands in the set of four bands defined above. B2–1, B3–2 and B4–3 will be used to refer respectively to the difference (expressed in dB) between B2 and B1, B3 and B2 and, finally, B4 and B3.

2.3. Statistical Analysis

F_1 values and energy difference between bands were the dependent variables in separate two-way Anovas having DISTANCE (with four levels) and POSITION as independent variables. POSITION has five levels when the target word is *patarata* (the definite article “a” plus the noun’s four syllables) and six levels when the word is *jaratacaca* (the definite article “a” plus the noun’s five syllables). When multiple comparisons where necessary, p-values were corrected using the Bonferroni method. A 5% α level was fixed for all the tests. R statistical environment [8] was used for all statistical analysis.

3. Results

3.1. Vowel Reduction

Figure 1 shows mean F_1 value for vowel /a/ in the four syllables of the target word *patarata*.

POSITION ($F(4, 210) = 219,\ p < 10^{-16}$) and DISTANCE ($F(4, 210) = 3.1,\ p < 0.02$) factors yielded significance, but not their interaction. Pairwise comparisons shows that all position means differ from each other ($p < 10^{-3}$ or less), except for syllables “pa” and “ra”. No significant difference between levels emerged for the DISTANCE factor in the pairwise comparisons.

Figure 2 shows mean F_1 value for vowel /a/ in the five syllables of the target word *jaratacaca*.

POSITION ($F(5, 232) = 148.3,\ p < 10^{-16}$) and DISTANCE (($F(4, 193) = 4.4,\ p < 0.002$)) factors yielded significance, but not their interaction. Pairwise comparisons shows that all position means differ from each other ($p < 10^{-5}$ or less) except for the definite article when compared to noun syllable “ta”. No significant difference between levels emerged for the DISTANCE factor in the pairwise comparisons.

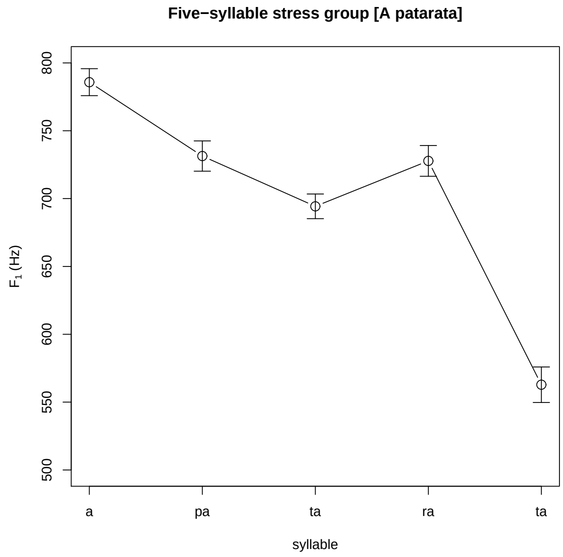

Figure 1: *Mean F_1 value for vowel /a/ in each syllable in the stress group. Whiskers indicate 95% confidence interval around the mean.*

3.2. Spectral Emphasis

3.2.1. Five-syllable stress group

Figure 3 shows mean energy difference (in dB) between frequency bands for vowel /a/ in the five syllables of the stress group [A patarata]. Statistical analysis won’t be discussed for band B3-2 because that band showed no prosodically meaningful pattern. Values for that band will be omitted in the figure as well.

B2–1: POSITION ($F(4, 171) = 70,\ p < 10^{-16}$) is the only factor yielding significance. Pairwise comparisons shows that the means for the definite article and the post-stressed syllable differ significantly ($p < 10^{-5}$ or less in all cases) from all others. The pre-stressed and the stressed syllable do not differ.

B4–3: POSITION ($F(4, 210) = 57.7,\ p < 10^{-16}$) is the only factor yielding significance. Pairwise comparisons shows that means for pre-stressed “ta” and post-stressed “ta” don’t differ significantly as well as means for article “a” and “pa”. All other comparisons yield significance ($p < 10^{-6}$ or less).

3.2.2. Six-syllable target word

Figure 4 shows mean energy difference (in dB) between frequency bands for vowel /a/ in the five syllables of the stress group [A jaratacaca]. Statistical analysis won’t be discussed for band B3-2 because that band showed no prosodically meaningful pattern. Values for that band will be omitted in the figure as well.

B2–1: POSITION ($F(5, 232) = 93.8,\ p < 10^{-16}$) and DISTANCE ($F(4, 193) = 2.6,\ p < 0.005$) factors yielded significance, but not their interaction. Pairwise comparisons shows that all means differ from the others ($p < 0.003$ or less), except for syllable “ra” compared to “ta” as well as “ta” compared to stressed “ca”. No significant difference between levels emerged for the DISTANCE factor in the pairwise comparisons.

B4–3: POSITION ($F(5, 232) = 48.5,\ p < 10^{-16}$ is the

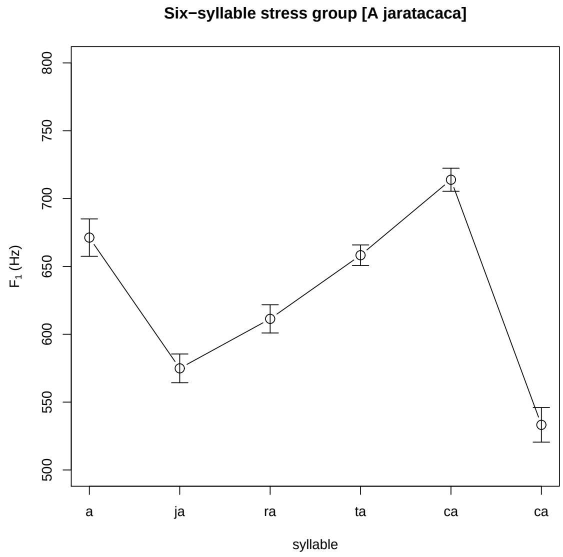

Figure 2: *Mean F_1 value for vowel /a/in each syllable in the stress group. Whiskers indicate 95% confidence interval around the mean.*

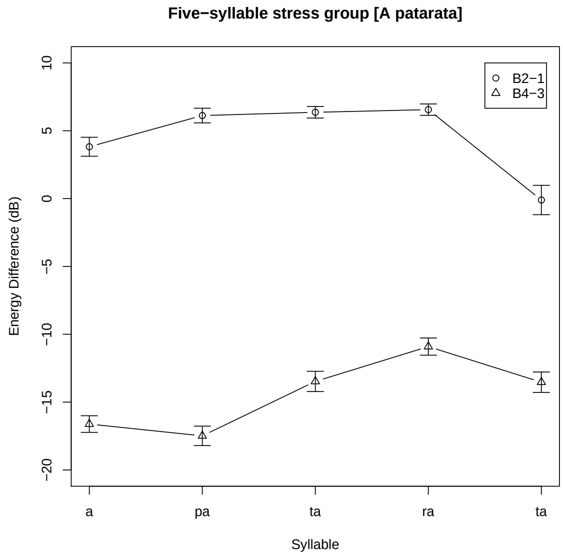

Figure 3: *Mean energy difference between frequency bands for vowel /a/ in each syllable in the stress group. Whiskers indicate 95% confidence interval around the mean.*

only factor yielding significance. Interaction between POSITION and DISTANCE is marginally significant $(F(20, 232) = 1.5,\ p = 0.07)$. Pairwise comparisons shows significance for every pair of means ($p < 0.005$ or less), except for definite article "a" compared to post-stressed syllable "ca", syllable "ja" compared to "ra" and stressed "ca" compared to post-stressed "ca".

4. Discussion

Results concerning F_1 data indicate prosodically-driven gradience in vowel opening. Data from stress groups containing the target word *patarata* suggest the first syllable in the stress group (the definite article "a") receives relative prominence, indicated by an open vowel articulation that wears off in the two subsequent vowels, corresponding to the noun's pre-stressed syllables. A relative opening takes place in the stressed vowel of stressed syllable "ra" and then a huge fall follows in the post-stressed vowel as it would be expected. The observed result could be interpreted under the assumption that the noun's first two syllables, *pa* and *ta*, are being parsed as an iambic binary metrical foot. This explanation, however, would leave unaccounted the significant openness observed in the stress group's first syllable which would not be left out unparsed according to one of the algorithms proposed in the literature [7].

Data from the analysis of the stress groups containing the target word *jaratacaca* can also be interpreted as a case of initial prominence signaled by a relative vowel opening in the stress group's first syllable. But unlike the case discussed in the previous paragraph, the relative openness wears off promptly so that a drop exceeding 100 Hz in F_1 is observed in the next syllable's vowel. A continuous opening movement then starts and peaks at the noun's stressed syllable. These data leave no room for an interpretation based on binary tree parsing, since there is an steady opening among the noun's pre-stressed syllable.

The kind of culminative gradience exhibited by F_1 here can be compared to V-to-V duration patterns observed in our earlier works [1][2]. This resemblance is no surprise, since acoustic duration for vowel segments can be regarded at least in part as a byproduct of jaw excursion. This special relation between articulation and timing patterns is interesting from the modeling point of view because if a model is able to correctly predict duration patterns (as the one taken as reference here [3]) it is also capable of drawing inferences about the openness of vocalic segments.

F_1 results may also be pointing to a more general phenomenon of vowel space enlargement. In order to further investigate this possibility, the front–back dimension should also be taken in consideration (through F_2 measures). Also high and mid vowels should be included in this study.

Spectral emphasis results are also different for the two target words. For stress groups containing the target word *patarata*, B2–1 seems to draw a line separating the noun's post-stressed syllable and the definite article from the other vowels as in a figure-ground pattern that enhances the pre-stressed and stressed syllables (the figure) against a less prominent ground. B4–3, on the other hand, shows the unfolding of a culminative phrasal stress that peaks at the noun's lexically stressed syllable. It means that the vowel spectrum gradually becomes less steep (more energy is present around F_3–F_4 region) under the influence of the upcoming stressed syllable, be it phrasally stressed or not. B3–2 has no apparent prosodically motivated variation.

When it comes to the stress groups containing the target word *jaratacaca*, B2–1 jointly plays the part of distinguishing the post-stressed vowel from the others and at the same time also reflecting the unfolding of the phrasal stress. It seems that a gradual energy boost around F_1 can grant the stressed syllable more prominence. The other band differences have no easy explanation, except that for B4–3 the noun's first two syllables seem to receive some kind of prominence when compared to the other positions. An initial prominence pattern, the stress group's first syllable excluded, is at odds with the smoothly in-

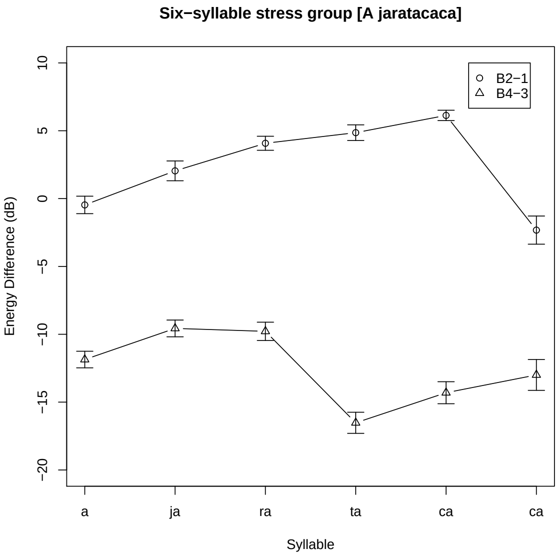

Figure 4: *Mean energy difference between frequency bands for vowel /a/ in each syllable in the stress group. Whiskers indicate 95% confidence interval around the mean.*

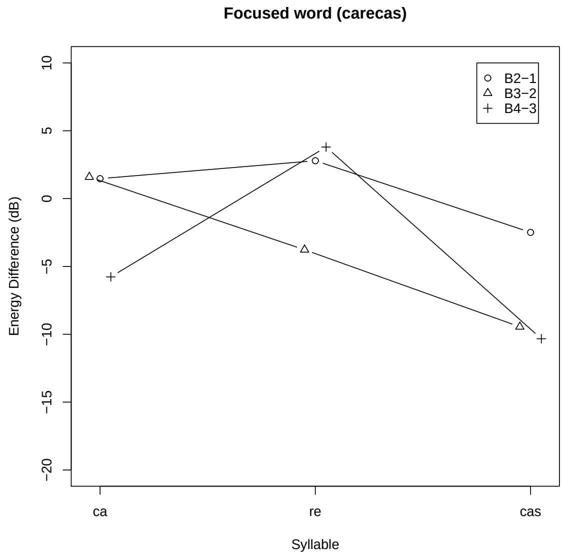

Figure 5: *Energy difference between frequency bands for vowel /a/ in each syllable of "carecas".*

creasing behavior of B2–1, the relatively low F_1 values for the noun's first two syllables and also with the behavior of B4–3 in *patarata*.

The present study is a preliminary work and no strong statements can be drawn about the role of spectral emphasis in SS or prosody in general based on the results presented here. It is possible to say, though, that both vowel opening and spectral emphasis variation seem to be following the same culminative pattern observed for acoustic duration, what brings more evidence to support the reference model of speech rhythm we adopt [3]. Additional work must be done to determine more accurately if the frequency bands that receive more emphasis change when the word is in or out of focus, for instance. Some initial observations of spectral emphasis measures taken from a corpus of radio broadcasted speech seem to point to an influence of focus in spectral emphasis. In a sentence taken from this corpus ("aquilo que eles estão carecas de saber fazer") the word *ca**re**cas* (stressed syllable in boldface) receives narrow focus that surfaces as an early pitch accent in the pre-stressed syllable. Spectral analysis measures taken from the stressed and pre-stressed vowels of *carecas* in this single case show that in the focus situation both vowels have an even much less steep spectrum than those seen in the corpus analyzed here (see figure 5). Although this is a single case analysis, it is interesting to notice how the presence of focus narrows down band energy differences which may be seen as an indication of less steep spectral damping. A less steep vowel spectrum in focused syllables in comparison to non-focused ones were found by [6] even though they only looked at read speech. The most striking feature in figure 5 is that B4–3 value in the stressed syllable position is even higher than B2–1 is, indicating a sizeable amount of emphasis in the high frequency domain for vowels in focused syllables. B2–1 can be said to have a similar pattern as that found for the target word *patarata*, i.e., higher values in pre-stressed and stressed syllables and a lower one in the post-stressed. In this single word, B3-2, as in the other cases discussed earlier, also doesn't seem to be of much use as far as prosody is concerned. More complete studies are necessary to corroborate these initial observations, though.

5. Acknowledgements

This work was supported by Fapesp grant 06/50166-9.

6. References

[1] Barbosa, P. .A; Arantes, P.; Silveira, L. S., 2004. Unifying stress shift and secondary stress phenomena with a dynamical systems rhythm rule. In *Proceeding of the Speech Prosody 2006 Conference*, Nara, Japan, 49–52.

[2] Arantes, P.; Barbosa, P. A., 2006. Secondary stress in Brazilian Portuguese: the interplay between production and perception studies. In *Proceeding of the Speech Prosody 2006 Conference*, Dresden, Germany.

[3] Barbosa, P. A., 2007. From syntax to acoustic duration: A dynamical speech rhythm production. *Speech Communication*, 49, 725–742.

[4] Vallabha, G. K.; Tuller, B., 2002. em Systematic errors in formant analysis of steady-state vowels, *Speech Communication*, 38, 141–160.

[5] Lindblom, B.; Sundberg, J., 1971. Acoustical consequences of lip, tongue, jaw and larynx movement, *J. Acoust. Soc. Am.*, 50(4), 1166–1179.

[6] Sluijter, A. M. C.; Heuven, V. J. van, 1996. Spectral balance as an acoustic correlate of linguistic stress. *J. Acoust. Soc. Am.*, 100, 2471–2485.

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

[8] R Development Core Team, 2007. R: A language and environment for statistical computing. *R Foundation for Statistical Computing*, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.