On the Robustness of some Acoustic Parameters for Signalling Word Stress across Styles in Brazilian Portuguese

Plínio A. Barbosa[1], *Anders Eriksson*[2], *and Joel Åkesson*[2]

[1]Speech Prosody Studies Group, Dep. of Linguistics, State Univ. of Campinas, Brazil
[2]Dep. of Linguistics, Philosophy and Theory of Science, University of Gothenburg, Sweden
pabarbosa.unicampbr@gmail.com, anders.eriksson.gu@gmail.com, joelakesson@gmail.com

Abstract

This work aims at examining three classes of acoustic correlates of lexical stress in Brazilian Portuguese (BP) in three speaking styles: informal interview, phrase reading and word list reading. In the framework of an international collaboration, a parallel corpus was recorded in the three speaking styles with 10 subjects so far in each one of the following languages: Swedish, English, French, Italian, Estonian and BP. In BP, duration, F0 standard-deviation and spectral emphasis values for stressed vowels tend to be higher in comparison with vowel acoustic parameters in unstressed position. These three parameters are robust across styles, especially vowel duration, for which circa 50 % of the variance is explained by stress and speaking style factors. The parameters pattern according to stress level is very similar between interview and phrase reading styles, which points to a similar effectiveness of reading and spontaneous styles in uncovering the word stress acoustic correlates in BP.

Index Terms: word stress, speaking style, Brazilian Portuguese

1. Introduction

This work is part of the international project "A typology for word stress and speech rhythm based on acoustic and perceptual considerations" coordinated by the second author and aiming at uncovering the production mechanisms and the perceptual cues signalling lexical stress as a function of speaking style in genetically related and unrelated languages such as Swedish, English, French, Estonian and Brazilian Portuguese (henceforth BP). The project proposes the recording of a parallel corpus across languages consisting of three speaking styles: word list reading (WR), phrase reading (PR) and informal interview (SI), the latter two in comparable phonic contexts. Another goal of this project is the investigation of the robustness of the acoustic parameters that signal lexical stress across speaking styles. This is one of the differences between this study and previous studies on BP lexical stress acoustic correlates.

In BP words, stress can fall in one out of three positions from the right edge (stressed syllables in bold, secondary stressed syllables distinguished from the other syllables): final stress, e.g., *repli**car*** (to reply); 2^{nd} syllable from the end, e.g., *re**pli**ca* (s/he replies); 3^{rd} syllable from the end, e.g., ***ré**plica* (replica). Secondary stress occurs in compounds such as guar*da*-***chu**va* (umbrella) and derivatives such as ho*men***zi***nho* (little man). In the latter cases, the primary stress of the original, leftward words *guarda* and *homem* are downgraded to secondary when forming the new word. Vowels after primary stress are largely weakened in comparison with vowels before primary stress in BP, justifying different IPA symbols for them, e.g., *abata* ([aˈbatɐ]), down (verb, imperative). The vowel quality of pre-stressed vowels is thus similar to that of stressed vowels.

Fernandes [1] and Moraes [2] have shown that syllable duration followed by total intensity are the main correlates of lexical stress in BP isolated utterances. By using a more controlled experimental setting, Massini [3] showed, on the basis of a corpus of 20 isolated, declarative sentences containing two target words repeated 8 times, that the syllable is the domain for the acoustic realisation of lexical stress in BP. The acoustic correlates for lexical stress found in her study were duration, total intensity and vowel quality in decreasing order of importance. As far as isolated utterances are concerned, these three studies also pointed out that fundamental frequency (F0) has a role for signalling lexical stress only at intonational phrase boundaries.

By using a large corpus of declarative sentences with a structure similar to newspaper headlines, Barbosa [4] confirmed the leading role of syllable-sized duration for signalling lexical stress. This study also showed that phrase stress is signalled by peaks of normalised V-to-V durations in BP. In a later study, Barbosa [5] confirmed that F0 has a role in signalling both prominence and boundary at strong syntactic boundaries also in spontaneous speech.

Despite these studies attesting the relevance of up to four acoustic parameters for signalling lexical stress, no studies to our knowledge investigated lexical stress correlates in spontaneous speech. Another important contribution of the work presented here for the understanding of the phonetics of BP lexical stress is the use of a more appropriate method for assessing the role of intensity by computing spectral emphasis.

2. Methodology

2.1. A parallel corpus

For ensuring a comparison across the languages under study in the aforementionned project, the BP corpus was obtained by using the same procedure for recording as in the other languages, as outlined below. Results for Swedish were also obtained [6].

Accordingly, 5 male and 5 female speakers were recorded at a sample rate of 16 kHz. Male speakers' age ranges between 21 and 30 years (mean of 25.4 yrs), whereas female speakers' age ranges between 18 and 26 years (mean of 22 years). All males and females but one have complete undergraduate courses. As for females, the 18-year-old subject was an undergraduate student at the time of recording.

The SI-style was recorded in the first place during a 15-20-minute, informal interview with the ten speakers carried out by two interviewers. In each case, interviewer and interviewee were close friends, which ensured long monologues from the part of the interviewee. All recordings were then orthographically transcribed. The PR-style was obtained by the reading of suitable chunks selected from the SI-transcriptions. Each subject read the transcription of his/her own interview up to two weeks after the first recording. The selected material was read three times in random order. Finally, the WR-style is the reading by the same respective speakers of a list of 15 content words taken from the phrases. This list was also read three times in random order. Speakers used long pauses between words and list intonation. For the sake of comparison, the same 15 words had their acoustic parameters computed in the three styles phonewise. The choice of words in BP was guided by the distribution of lexical stress patterns according to Cintra [7] for polyssylabic words (circa 70 % of penultimate stress, 20 % of final stress and 10 % of antepenultimate stress). The number of syllables varied from 2 to 6-syllable words with 84 % of 3- and 4-syllable words. The majority of words (62 %) are in medial position in the phrase. The majority of the selected words in the PR- and SI-styles are prominent at the utterance level (82 % in males and 78 % in females, no less than 60 % for any single speaker). In the WR-style, due to the nature of the task, all words are prominent.

All BP vowels are represented per gender in each stress position. The words selected for the analyses were synchronously annotated with the audio files by using the Praat software [8] by creating a 4-tier TextGrid. The first tier contains the boundaries and orthographic labels either of the chunk from which the words were selected (SI and PR-styles) or of the selected words (WR-style). The second tier contains the boundaries and orthographic labels of the selected words in the three styles. The third tier delimits and transcribes the word phones. Finally, the fourth tier indicates the stress level of each vowel of the word. All these annotations were done manually by the first author.

2.2. The acoustic analyses

A Praat script, LexStressProcessing, was implemented in order to generate statistical descriptors for four acoustic parameters extracted for each phone: duration in ms, F0, SPL intensity and spectral emphasis. The descriptors of F0 in the phone domain were mean, median, standard deviation, relative standard deviation and baseline in Hertz (Hz) as well as median, standard deviation and relative standard deviation in semitones. Relative F0 standard deviation, either in Hz or in semitones, was defined as the ratio between standard deviation and F0 median. Baseline was defined according to Traunmüller and Eriksson [9] by expression 1 with all values in Hertz

$$baseline = F0_{median} - 1.43 \times F0_{SD} \qquad (1)$$

The descriptors of SPL intensity were mean and standard deviation in dB, whereas spectral emphasis was computed according to Eriksson et al. [10] by equation 2, where L the intensity of the whole spectrum of the segment up to the Nyquist frequency and $L0$ is the spectrum band up to 1.43 times the mean F0 in the phone.

$$\text{Spectral Emphasis } (dB) = L - L0 \qquad (2)$$

F0 traces were obtained by using a window containing the word interval added by 100 ms at both word margins to ensure continuity of values.

The LexStressProcessing script generates a table with the phonewise extracted parameters above with the corresponding categorical values specifying subject, speaking style, word, phone identity, phone type (V or C) and stress level. Stress level received one of the four values: 0 for consonants and post-stressed (PsStr) segments; 1 for pre-stressed vowels (PrStr); 2 for secondary stressed vowels; and 3 for primary stressed vowels (Str). For the study presented here, only the acoustic parameters for vowels were taken into account due to their importance to signal stress in BP. Secondary stressed words were excluded from the analyses due to insufficient statistical representativity.

3. Results

The analyses showed that F0 and SPL measures of centrality do not signal lexical stress. This is because the former is ruled by intonation requirements and the latter does not differentiate across stress status in the three styles. On the other hand, vowel spectral emphasis, duration and F0 standard-deviation are distinct across at least two stress levels in the three speaking styles. The general patterns of variation for the three acoustic parameters above are shown as boxplots with SPEAKING STYLE and STRESS level as factors.

Due to the non-normality and heterocedasticity of the original residuals, the statistical technique used here is the Scheirer-Ray-Hare (SHR) extension [11, p. 175] to Kruskal-Wallis test, which allows a 2-Way non-parametric alternative to 2-Way ANOVA. For doing so, a formula proposed by Poisot [14] was used. Significance levels both for the general model and for post hoc tests were set to 0.01 for the six models (2 genders x 3 parameters). Non-parametric, post-hoc comparisons [12, p. 213-214] after SHR test were used to assess the significance of paired levels. All tests were carried out using the R package [13]. Corrected effect sizes (ω^2) for the two factors and their interaction were manually computed from the table generated by each SHR model. Results revealed that the two factors and their interaction were highly significant for the six models. In the following, post hoc results, effect sizes and mean values are examined for each parameter.

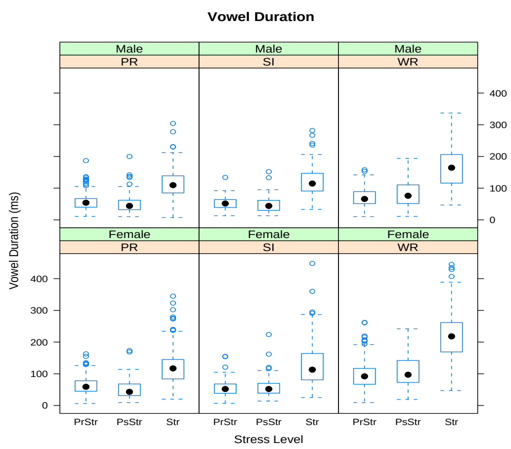

Figure 1: *Vowel duration in ms for male (top) and female (bottom) speakers according to stress level and speaking style. See text for more details.*

As can be seen in Figure 1, vowels are longer and vary more in duration in stressed position, irrespective of speaking style and gender. There is no distinction in duration between pre-stressed and post-stressed vowels in both genders (mean 63 ms for males and 76 ms for females), but stressed vowels are significantly longer (mean 137 ms for males and 164 ms for females). There is no difference in duration between the PR and SI-styles for both genders (mean 74 ms for males and 79 ms for females), whereas vowels in the WR-style are longer (mean 102 ms for males and 136 ms for females), all stress levels pooled together. The results for the WR-style reveal that hyperarticulation of the words produces longer vowels for all stress positions. Also in WR-style post-stressed vowels are longer than pre-stressed vowels for males (95 ms against 80 ms) due to final lengthening, but not for females (mean 79 ms). In the same style, stressed vowels are from 2.2 (females) to 2.4 (males) times longer than unstressed ones. Stress and speaking style explain 51 % of duration variance in females, whereas they explain 48 % in males. Effect size is higher for stress (30 % in females and 39 % in males) than for speaking style (20 % in females and 8 % in males). The interaction effect size is low (around 1 %) for both genders. Overall the results are very similar in both genders: stress prevails in explaining the duration variance in comparison with speaking style and stressed vowels are more than twice longer than unstressed vowels.

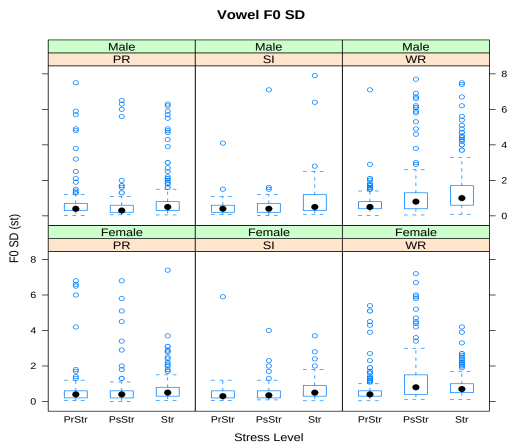

Figure 2: *Vowel F0 standard deviation in semitones re 1 Hz for male (top) and female (bottom) speakers according to stress level and speaking style. See text for more details.*

As for F0 standard-deviation in st, data in Figure 2 show slightly higher values in stressed positions in comparison with unstressed positions in the three styles, but post hoc analyses reveal that there is no significant difference between post-stressed and stressed positions in both genders (mean 0.8 st in females and mean 1.0 st in males), with pre-stressed vowels exhibiting significantly lesser values (0.6 st in both genders). There is no statistical distinction between the SI and PR-styles in both genders (mean 0.6 st in females and mean 0.7 st in males), but the WR-style has significantly higher values (0.8 st in females and 1.1 st in males). Stress and speaking style explain 9 % of F0 standard-deviation variance in both genders. Effect size is low both for stress (5 % in females and 2 % in males) and for speaking style (3 % in females and 6 % in males). The interaction effect size is low (around 1 %) for both genders. Overall the results point to a small, but significant effect of both factors in explaining the F0 standard-deviation variance.

As for spectral emphasis, Figure 3 shows that post-stressed vowels have significantly less spectral emphasis than the other two levels, which are indistinct in this respect. WR and PR-styles have the same mean spectral emphasis (5 dB for females

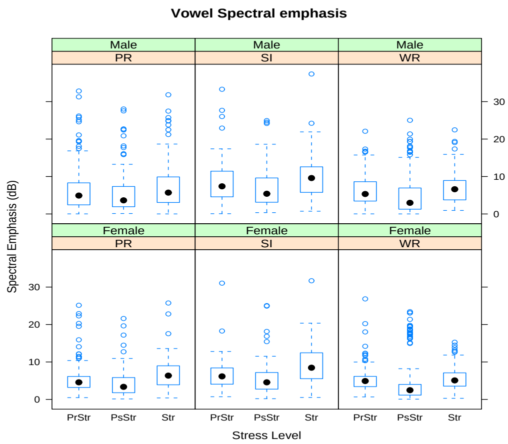

Figure 3: *Vowel spectral emphasis in dB for male (top) and female (bottom) speakers according to stress level and speaking style. See text for more details.*

and 6 dB for males), whereas the SI-style has a higher spectral emphasis (7 dB for females and 9 dB for males), which suggests more vocal effort in this style, which is somewhat contrary to expectations. This higher vocal effort in BP can be related to the higher pitch values in the SI-style for both genders. Stressed vowels (7 dB for both genders) have more spectral emphasis than pre-stressed vowels (5 dB for females and 6 dB for males), which in turn have more spectral emphasis than post-stressed vowels (4 dB for females and 5 dB for males). Stress and speaking style explain 7 % of spectral emphasis variance in males, and 13 % in females. Effect size is low both for stress (10 % in females and 4 % in males) and for speaking style (3 % in both genders). The interaction effect size is very low (around 0.3 %) for both genders. Overall the results point to a higher effect of stress for females in explaining the spectral emphasis variance. This effect is higher than for F0 standard deviation, which signals a difference in the importance of spectral emphasis in comparison with the male subjects, at least as far as the number of 5 subjects per gender is concerned.

Because both genders behave in a similar way for duration and F0 standard deviation, a pooled analysis considering both genders together was carried out. The pooled data was examined according to the three lexical stress patterns (final, penultimate and antepenultimate stress words) in order to check for differences in their behavior in terms of effect size and mean parameter value differences. As for duration, both factors explain only 32 % of the duration variance in antepenultimate stress words, whereas circa 50 % of the variance is explained in penultimate and final stress words. In penultimate stress words, the PR- and SI-styles mean duration is 55 ms for pre-stressed and post-stressed vowels pooled together, and 124 ms for stressed vowels, revealing that stress doubles the duration of unstressed syllables. As for the WR-style, vowel duration is 1.5 times the duration of the other styles in pre-stressed (84 ms) and stressed (192 ms) vowels, and the double in post-stressed (99 ms) vowels. The general picture for antepenultimate stress words is the same as for penultimate stress words, although in the former lexical stress pattern mean vowel duration is less affected by stress in the WR-style (136 ms against 78 ms for unstressed vowels), probably due to a greater distance of stressed vowels from the end of the word in comparison with the other two lexical stress patterns. Finally, in final stress words figures are also similar to the other stress patterns in the PR- and SI-styles, but in the WR-style, duration mean in stressed vowels is 3 times that in pre-stressed vowels (235 ms against 77 ms). This longer duration is related to the additional condition that stressed vowels are in absolute final position in the WR-style, which adds the effect of final lengthening.

As for F0 standard deviation in st, both factors explain 18 % of the F0 standard deviation variance in final stress words, 7 % in penultimate stress words (from which 5 % by speaking style) and 4 % in antepenultimate stress words for the stress factor only. In this picture, antepenultimate stress words appear to be largely affected by other factors not considered here in terms of vowel duration. Also for F0 standard deviation, the variance explained in antepenultimate stress words is lesser than for the other lexical stress patterns. As presented above, pre-stressed vowels have the lowest mean values of F0 standard deviation across styles, which is 0.6 st. This difference is enlarged in the WR-style with post-stressed vowels with the highest values in antepenultimate (0.9 st against 0.7 st in stressed vowels) and penultimate (1.3 st against 1.1 st in stressed vowels) stress words. In final stress words, F0 standard deviation mean value is 2.0 st in stressed vowels in the WR-style, two times the mean value in the other two styles.

By separately analysing, according to lexical stress patterns, the two genders as for the spectral emphasis parameter, it appears that stress is the factor that explains a higher percentage of the parameter variance in females for the most frequent lexical stress patterns (11 % against 3 % for speaking style in penultimate stress words and 6 % against 3 % for speaking style in final stress words) but not in males, where both factors equally explain spectral emphasis variance (a total of 9 % in penultimate stress words, 6 % in final stress words, and 3 % in antepenultimate stress words). Both in antepenultimate and penultimate stress words post-stressed vowels have the lowest value of spectral emphasis (4 dB against 7 dB for the other two stress levels) irrespective of style in females. For males the picture is the same in penultimate stress words (6 dB against 8 dB), whereas in antepenultimate stress words stressed and post-stressed vowels have the same mean values (6 dB) irrespective of style. In both genders the SI-style has the highest values of mean spectral emphasis irrespective of stress position and lexical stress pattern with one exception: pre-stressed vowels have the highest spectral emphasis in the PR-style for antepenultimate stress words (9 dB against 7 dB for post-stressed and stressed vowels pooled together) for males. In final stress words spectral emphasis mean is higher in stressed vowels for both genders. The picture that emerges is the same as acknowledged above: females have the most regular use of spectral emphasis for signalling stress and this is done by dropping the mean value from stressed to post-stressed position.

4. Discussion and conclusion

The results show that only duration distinguishes stressed vowels from unstressed vowels. This is achieved with an effect size (50 %) that is much higher than the effect size of the other two parameters (below 15 % taking the two genders in consideration). Stress explains a higher percentage of variance than speaking style, and differences in how to signal stress are found when examining speaking style. The main difference regards the WR-style, for which the phenomena of final lengthening and hyperarticulation can explain the higher values. Also in this tyle stressed vowels are longer than unstressed vowels. The PR- and SI-styles behave exactly in the same way as regards the duration patterns and statistics. Stressed vowels are more than twice longer than unstressed vowels. As for unstressed vowels, both pre-stressed and post-stressed vowels have the same mean, which is contrary to what was reported by the previous studies with isolated utterances mentioned here. This is probably due to the use of informal interview and the similarity of this style with the reading of the same material by the subjects. There is no special status for final-stress words in terms of stress marking by means of duration: the mean values are exact the same as in the other lexical stress patterns in the PR- and SI-styles.

Post-stressed and stressed vowels exhibit the same mean in terms of the F0 standard-deviation parameters, all styles pooled together. This suggests that stressed and post-stressed positions form a single domain where F0 values are distinguished for linguistic-phonetic and communicative purposes in BP, a proposal already pointed out in a previous study [15]. In these positions pooled together, the mean F0 standard deviation is 0.8 st (females) and 1.0 st (males) in comparison with pre-stressed position (0.6 st). Thus, there is no separate status for stressed position as revealed by this parameter. There is no significant difference between the PR- and SI-styles as for this parameter, with the WR-style exhibiting the highest values. This fact points to more variation in F0 in read, isolated words. The highest mean values in stressed vowels for F0 standard deviation in the WR-style are explained by the joined effect of stress and final boundary position.

Spectral emphasis values have the same mean in pre-stressed and stressed vowels. Post-stressed vowels have signficantly lower values. Since the majority of words are penultimate stress words (which contain one post-stressed vowel), a decrease in spectral emphasis is an indication of stress for the preceding position. This results is in the line of previous studies of BP that used total intensity instead ([1, 2, 3]). Only for spectral emphasis the SI-style differs from the PR-style: it has higher values than in the other two styles, probably due to an increase of vocal effort as a consequence of long high-pitched narrative stances. In fact the overall F0 median in Hz (and st re 1 Hz) is 145 Hz (86 st) for males in the SI-style against 136 Hz (85 st) in the other two styles. As for females, the overall F0 median is 217 Hz (93 st) in the SI-style against 207 Hz (92 st) in the other two styles. Furthermore, females seem to privilege this parameter for signalling stress just after duration, as revealed by the corrected effect size, whereas males use both spectral emphasis and F0 standard deviation in similar ways for that. Although this difference is low, it could be a matter for further investigation of inter-gender differences in the use of the acoustic correlates of lexical stress. There is no difference in mean value across lexical stress patterns for the WR-style for this parameter.

Overall the results suggest that the three parameters examined here signal in different ways lexical stress across the three speaking styles, at least when the word is prominent in the utterance, since circa 80 % of the words are prominent in the PR- and SI-styles and all of them in the WR-style. Furthermore, they point to a similar effectiveness of phrase reading and spontaneous styles in uncovering the word stress acoustic correlates in BP, at least for duration and F0 standard deviation.

5. Acknowledgements

The authors acknowledge a grant 421-2007-2301 from the Swedish Research Council. The first author thanks grant 301387/2011-7 from CNPq.

6. References

[1] Fernandes, N. H. "Contribuição para uma análise instrumental da acentuação e intonação do português", Master's dissertation. University of São Paulo. 1976.

[2] Moraes, J. A. "Corrélats acoustiques de l'accent de mot en portugais brésilien", in Proc. XIth International Congress of Phonetic Sciences, v. 3, pp. 313-316. Tallinn: Estonia, 1987.

[3] Massini, G. "A Duração no estudo do acento e do ritmo em português", Master's dissertation. University of Campinas. 1991.

[4] Barbosa, P. A. "At least two macrorhythmic units are necessary for modeling Brazilian Portuguese duration", in Proc. First ESCA TRW on Speech Production Modeling: From Control Strategies to Acoustics; 4th Speech Production Seminar: Models and Data (pp. 85-88). Autrans, France, 1996.

[5] Barbosa, P. A. "Prominence- and boundary-related acoustic correlations in Brazilian Portuguese read and spontaneous speech", in P. A. Barbosa, S. Madureira, and C. Reis (Eds.), Proceedings of the Speech Prosody 2008 Conference (pp. 257-260). Campinas, 2008.

[6] Eriksson, A., Barbosa, P. A., Åkesson, J., "The acoustics of word stress in Swedish: A function of stress level, speaking style and word accent", These proceedings.

[7] Cintra, G., "Distribuição de padrões acentuais no vocábulo em português", Confluência, 5(3):82–93, 1997.

[8] Boersma,, P. and Weenink, D., "Praat: doing phonetics by computer" (Version 5.1.37) [Computer program], Online: http://www.praat.org.

[9] Traunmüller, H. and Eriksson, A., "The frequency range of the voice fundamental in the speech of male and female adults", Unpublished Manuscript. Online: http://www.ling.su.se/staff/hartmut/aktupub.htm.

[10] Eriksson, A., Thunberg, G. C. and Traunmüller, H. "Syllable prominence: A matter of vocal effort, phonetic distinctness and topdown processing", in P. Dalsgaard, B. Lindberg, H. Benner, and T. Zheng-Hua [Eds.], Proceedings of the 7th European Conference on Speech Communication and Technology (Eurospeech '01) (pp. 399-402). Aalborg, Denmark.

[11] Dytham, C. "Choosing and Using Statistics: A Biologist's Guide", Third ed. Chichester: John Wiley & Sons, 2011.

[12] Siegel, S., Castellan, N. J., "Non parametric statistics for the behavioural sciences", New York: MacGraw Hill Int., 1988.

[13] The R foundation for statistical computing, "The R project for statistical computing", Online: http://www.r-project.org/, 2008.

[14] Poisot, T., "R : test de Scheirer-Ray-Hare", Online: http://www.cafe-sciences.org/billets/r-test-de-scheirer-ray-hare/, 2008.

[15] Madureira, S., Barbosa, P.A., Fontes, M., Spina, D., Crispim, K., "Post-stressed syllables in Brazilian Portuguese as markers", in Proceedings of the XIVth International Congress of Phonetic Sciences.San Francisco,v. 2, 917-920, 1999.