Prominence- and boundary-related acoustic correlations in Brazilian Portuguese read and spontaneous speech

Plínio A. Barbosa

Speech Prosody Studies Group/Dep. of Linguistics
Univ. of Campinas, Brazil
pabarbosa.unicampbr@gmail.com

Abstract

This work presents a function-oriented analysis of the inter-relations between f_0 and syllable-sized duration contours in Brazilian Portuguese read paragraphs and spontaneous speech aiming at: presenting measures of the frequency of prominence-marking and breaking-marking in BP, showing the inter-dependence between f_0 and syllable-sized duration contours, showing the separate contributions of duration and f_0 at minor prosodic boundaries, presenting a semi-automatic method for analysing the correlation between f_0 and normalised syllable-sized duration contours. Contrary to the observations in lab speech for isolated utterances, pitch accents are relatively frequent in BP (from 54 to 73 % of the phonological words). Peaks of syllable-sized duration signal both minor and major prosodic boundaries, having both a phrasal and a sentencial breaking-marking function, as well as a prominent-bearing function.

1. Introduction

As regards the prosodic functions of prominence and phrasing, it is often assumed that in a Romance language such as Brazilian Portuguese (henceforth BP) the use of pitch accents in read speech are not so frequent as in Germanic languages. In [7] it is pointed out that languages such as English and Swedish favour prosodic prominence marking over prosodic phrasing marking, whereas the so-called syllable-timed languages systematically signal both prosodic functions, *not often by the use of pitch accents*. Research on BP prosody seems to point to this very fact, at least as far as isolated read sentences are concerned [3, 15, 18]. The reason for that general pattern is twofold: (1) pitch accents are associated with major prosodic boundaries (whose number is severely limited in isolated sentences), and (2) duration of syllable-sized units (both the phonological syllable and the V-to-V unit) is a crucial parameter for signalling both phrase and lexical stress [1, 16]. Some years ago, however, Moraes [17] claimed that fundamental frequency (f_0) would have a main role for signalling phrase and sentence accent in BP, whereas duration (and intensity) would be mainly used for signalling lexical stress. The analyses of BP read paragraphs in two speakers and of the spontaneous speech of two others reveal that, contrary to the observations in lab speech for isolated utterances, pitch accents are relatively frequent in BP (from 54 to 73 % of the phonological words). Furthermore, peaks of syllable-sized duration signal both minor and major prosodic boundaries, having both a phrasal and a sentencial breaking-marking function, as well as a prominent-bearing function.

This work presents a function-oriented analysis of the inter-relations between f_0 and syllable-sized duration contours in BP read paragraphs and spontaneous speech aiming at: (1) presenting measures of the frequency of prominence-bearing and breaking-marking in BP, (2) showing the inter-dependence between f_0 and syllable-sized duration contours, (3) showing the separate contributions of duration and f_0 at minor prosodic boundaries, (4) presenting a semi-automatic method for analysing the correlation between f_0 and normalised syllable-sized duration contours.

2. Corpora and method of analysis

2.1. Corpora

The Speech Prosody Studies Group is building a database of spontaneous speech formed by utterances spoken by interviewees in radio programs. For this study, a set of utterances containing 176 phonological words from a male and a female subject, both in their thirties and from São Paulo State, was taken for analysis. The female speaker is a psychologist that is often interviewed by the Radio Você in Americana, Brazil. The male speaker is an economist interviewed by the Radio CBN.

Read speech comprises the reading of an excerpt of a well-known Monteiro Lobato's book for children, *Reinações de Narizinho*, by two speakers of São Paulo State, AP and AC (Lobato is an author of the beginning of the XXth century) . The excerpt has two paragraphs containing 66 phonological words. The reason for the choice of this text is the familiarity of the great majority of Brazilian adult speakers (including AC and AP) with the author's stories and characters. For decades TV shows presenting Lobato's characters have been broadcasted in Brazil. It is important to note that the readers' performance cannot be considered child-direct speech. However, the reading was very expressive, including the realisation of frequent emphasised lexical items. Emphasised words were those considered prominent by ten out of ten judges. The use of naturally-sounding utterances in experimental phonetics, including a natural context of realisation, is crucial for obtaining natural, function-oriented patterns of prominence and boundary [13].

2.2. Prosodic patterns of duration and f_0 in BP

Both corpora were manually segmented into V-to-V units in Praat [8], by placing boundaries at vowel onsets (cf [2], where consecutive vowel onsets are used as anchor points delimiting a unit whose boundaries are different from the phonological syllable. Note that the use of this unit in speech research can be traced back at least as far as in [10]). Each interval obtained in this way was labelled by using a broad phonetic transcription system. Traces of f_0 were obtained by using the method of autocorrelation in Praat, followed by a 10-Hz cut-off frequency smoothing procedure.

BP is a right-headed language at the domain of the so-called stress group (a domain formed by a prominent V-to-V unit and the associated non-prominent ones in the same domain). Thus, a phrase stress at the right end of the corresponding stress group delimits this unit. Earlier investigations revealed that duration is the main correlate of phrase stress in BP [1, 2]. That is the reason why stress groups were delimited by automatically detecting phrase stress positions within and across BP connected utterances. This was done by serially applying two techniques for normalising duration, explained earlier [2]: a $z - score$ transform procedure, and a 5-point moving average filtering procedure. The normalisation technique and the detection of phrase stress positions throughout the utterances of both corpora were implemented as a Praat script using the three following sequential steps:

- normalisation of V-to-V durations by computing a $z - score$ by the formula: $z = \frac{dur - \sum_i \mu_i}{\sqrt{\sum_i var_i}}$, where dur is V-to-V duration in ms, the pair (μ_i, var_i), the mean and variance of all phones i within the corresponding V-to-V unit. These descriptors were previously computed from an *ad hoc* reference corpus, where BP phones appear in all possible phonotatically allowed positions in BP [4];

- smoothing of $z - scores$ by using a 5-point moving average filtering procedure (eq. 1), where z^i is the current value of the $z - scores$ computed in step 1;

$$z_{smoothed}^i = \frac{5.z^i + 3.z^{i-1} + 3.z^{i+1} + 1.z^{i-2} + 1.z^{i+2}}{13}$$

(1)

- detection of smoothed $z - score$ maxima at positive to negative derivative transitions.

The correspondance of V-to-V normalised duration peaks with perceived boundary and prominence is very striking in BP (see additional comments in [2]). Peaks of f_0 in BP seem to be more easily perceived in the case of emphasised lexical items, as will be seen in the next section. The Praat script also plots the normalised V-to-V durations in the y-axis, against non-normalised time in the x-axis. Non-normalised time is the right instant of time of the corresponding V-to-V unit in miliseconds. Fig. 1 illustrates the procedure, which allows to compare the moments of occurrence of duration-related peaks, with f_0-related peaks, since f_0 traces, occurring on non-normalised time, can be superimposed onto the durations contours. The ordinate values of normalised duration are indicated with a $+$ sign. Phrase stress positions for this example in "julgar", second "foi", and second "fez". Emphasised items in capitals.

The semi-automatic detection of peaks of normalised duration defines mostly a fine-grained marking of the phrasing function, indicating the boundary strength by means of the normalised value of V-to-V duration immediately before the boundary. Emphasised lexical items usually have a V-to-V unit whose normalised duration is also a local maximum. The underlying function in this case is prominence, though. This is not a problem, since the procedure tracks the "form" of the durational contour, not its function.

Pitch-related prominence was annotated by criteria set in a series of intonation workshops by a group of researchers from the Univ. of Campinas and the Catholic University of São Paulo. During these workshops, where an easy consensus on prominent items was achieved (much easier for spontaneous speech), the function of prominence only was indicated

Figure 1: Example of phrasing (as solid vertical bars) superimposed onto a f_0 trace for the spontaneous utterance "Não vamos nem JULGAR se foi ou não foi, se fez ou não fez.", from the male speaker.
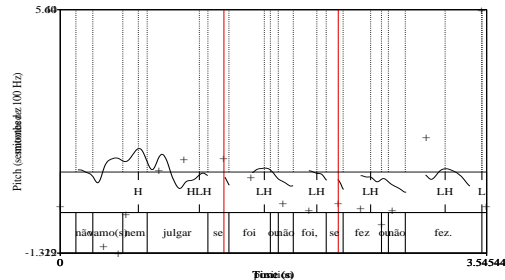
for prominent and emphatically prominent lexical words. After that, the form of the realised pitch accents was classified by using eight labels, rising (LH), late rising (>LH), falling (HL), late falling (>HL), high (H), low (L), falling-rising (HLH), and rising-falling (LHL).

The high peak for rising (LH) is attained by the end of the stressed vowel, whereas it is attained after the stressed syllable for late rising (>LH). The pattern is similar for fallings (HL, and >HL), replacing the low tone as the reference for alignment. The difference between a high tone (H) and a rising contour (LH) is related to the previous low value of f_0: the onset of rising in the LH contour is preceded by a short falling or a value of f_0 at the baseline, functionally associated with the rising (see Figs. 1 and 2). The same holds for low tones (L) vs falling contours (HL), where the falling part is preceded by short risings (see Fig. 2). The falling-rising contour (HLH) has the two high tones at adjacent syllables, where the second tone is lower than the first one (the high tones can occur even on consecutive syllables, which is a counter-example to metrical-based approaches to the Rhythm rule. See [5] for a criticism on that). HLH contours are employed by the four speakers and undergo the perception of a very emphatic item, where secondary stress is marked by the first high tone, and primary stress by the second, together as an increased duration in primary stress position (see Figs. 1, 2, and 5, respectively on the words "julgar", "carecas", and "ouvindo"). A similar account for emphatic items is given by Bolinger in English, with the example of Chínése [9, p. 58]. The reason for considering this contour as a single entity is related to the single function of emphasis. Rising-fallings (LHL) are similar to falling contours (HL), but the falling part in the former is much less sharp than for the latter. Examples of all these contours and tones appear in the figures of this paper.

Low and high tones are also used to respectively signal levelling into the baseline, and rising at the end of major prosodic boundaries, after a pitch accent (there is no phrase accent in this system. See also this decision for European Portuguese in [11]). The diacritics ! and | are used for signalling the onset of periods of downstep and upstep, respectively. The close relation of pitch accents to peaks of duration, and the specificities of BP intonation invited us to coin this system as ToBiPI, Transcription of Brazilian Portuguese Intonation (see the work on narrow focus using this system [14]). As for ToBI workshops [19], four tiers in Praat were used for annotation, but with two differences. The tonal tier does not indicate alignment with a star diacritic, since this alignment is retrieved from the seg-

mentation of the utterance into V-to-V units, in the second tier. From this tier the computation of normalised duration contours is done, serving to indicate phrasing (see Fig. 1), instead of a perceptually-evaluated break indices tier. The third tier contains the orthographic representation of the words, and the fourth tier, pragmatic comments on the context of the utterance realisation. The V-to-V tier and the pragmatic tier are not presented in the figures for the sake of visibility.

The combined action of the annotation for pitch accents, and the phrase stress position detection gives four levels of prominence, as suggested some decades ago by Hirst [12]: unstressed, stressed (non-peaks and peaks of normalised duration, respectively), and prominent, emphatic (perceived prominence, and perceived as emphatically prominent, the latter relatively frequent in spontaneous speech). These levels may be combined: an unstressed V-to-V unit, that is, not exhibiting a peak of normalised duration, can have a pitch accent (see next section). Duration peaks without pitch accents are also possible, as well as the combination of pitch accent and duration peak, the latter a typical indication of a major prosodic boundary (see next section). In the following, an account of the frequency of pitch

Figure 2: Example of spontaneous utterance from the fem. sp.: "Fazer aquilo que eles estão CARECAS de saber fazer."
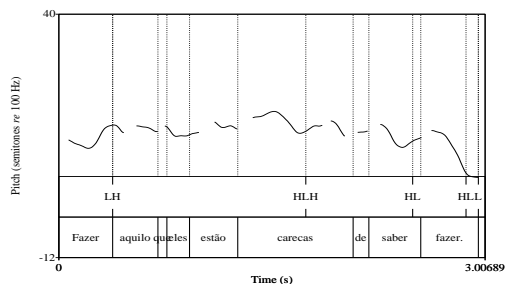
accents and duration peaks, as well as their interdependence is presented in read and spontaneous speech. The relation to major and minor prosodic boundaries, and the implementation of emphasis and secondary stress in BP is also highlighted. The pattern that emerges is very different from the early accounts of intonation of isolated utterances in BP.

3. Frequence of pitch accents in BP and its relation to duration peaks

The general melodic pattern for isolated declarative utterances in both BP [17, 15, 6] and European Portuguese [11] is a prominence (usually a high tone) in the first stressed syllable of the utterance, and a falling contour at the end of the utterance. This gives a fairly low frequency of pitch accents per phonological word in long utterances. This is not the case both for read paragraphs (at least of the type used here), and spontaneous speech. If boundary-marking and prominence-marking are considered interchangeably, the inter-relation between peaks of duration and pitch accents is given in table 1, as probabilities. A priori probabilities are the proportion of pitch-accents (or duration peaks) in the total of phonological words. Conditional probabilities consider the co-occurrence between a duration peak with a pitch accent over the total number of duration peaks ($p(f_0/dur)$) or pitch accents ($p(dur/f_0)$). A significant difference (test of proportions) between a priori and conditional probability signals a

dependence between pitch accent and duration peak.

Table 1: *Probab. of pitch accent (f_0) and duration peak (dur) in % of phonological words. Speech rate (# V-to-V/s), speaker and speaking style are indicated. Stars signal significant differences between a priori and conditional probabilities ($\alpha = 0.02$).*

sp (sp.sty)	$p(f_0)$	$p(f_0/dur)$	$p(dur)$	$p(dur/f_0)$	rate
F (spont.)	63 *	79 *	49 *	63 *	5.8
M (spont.)	73	80	48	56	5.4
AC (read)	54	66	56 *	76 *	4.6
AP (read)	70	83	65	74	4.3

Note that the proportion of pitch accents is over 50 % for both speaking styles, that is, there is a pitch accent for every one or two phonological words. The frequence seems to depend more on speaker (AP and the male speaker of the spontaneous speech) than on style. The frequency of duration peaks (more frequently related to phrasing) is more stable across speakers, but slightly more frequent in read speech. There is a dependence between duration peak and pitch accent for the female speaker in spontaneous speech, as well as for speaker AC in read speech: for the latter, a pitch accent implies 76 % of chance of a duration peak. For the female speaker both are inter-related. This inter-relation is confirmed when the analysis is restricted to major prosodic boundaries in read speech (utterance boundaries, clause and subject-predicate boundaries): 98 % (speaker AP), and 100 % (AC) of the time, both pitch accent and duration peak occur in the same lexical item, usually in the stressed vowel for pitch accents, and in the stressed or pre-pausal V-to-V unit for duration peaks. Fig. 3 illustrates, however, a case where the major boundary (subject-predicate) is signalled only by a duration peak in the word *palmas* in the sentence *Uma chuva de palmas acolheu suas palavras.* (A roar of claps welcomed his words), just before the VP.

All emphasised words receiving pitch accent also have a duration peak located to the right of the stressed V-to-V units. Here, two functions realised by distinct resources could be distinguished: one of prominence, and the other of boundary marking aligned to primary stress. Additionally, part of the emphasised items, both in spontaneous and read speech, also signals a secondary accent with the leading tone of a HLH contour.

Figure 3: F_0 trace, word segmentation and duration peaks (as solid vertical bars) for the read utterance "Uma CHUVA de palmas acolheu suas palavras.", speaker AP.
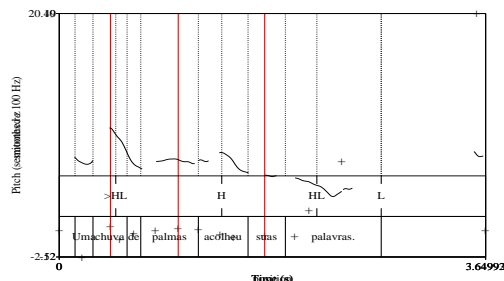
As regards minor boundaries in read speech, of all the lexical items preceding these boundaries in speaker AC, 35 % are realised by duration peaks and 38 % of pitch accents only.

The proportion is respectively 21 % and 27 % for speaker AP. Speaker AP prefers to signal minor boundaries by associating both duration and f_0 parameters (52 % of the cases), while speaker AC does the same only 27 % of the time.

Figures 4 and 5, from the female speaker in spontaneous speech, illustrate other types of contour. The final LHL contour at the end of the utterance in Fig. 4 gives a very emphatic rendition of the verb "é", as well as the HLH contour of the word "ouvindo", Fig. 5. An emphatic item can be realised for almost all types of contours, cited here, since it depends on function, rather than form. Peaks of V-to-V duration sig-

Figure 4: Utterance "É MUITA colocação de FOFOCA em cima de um acontecimento que ninguém sabe realmente o que É.", fem. speaker, sp. speech.
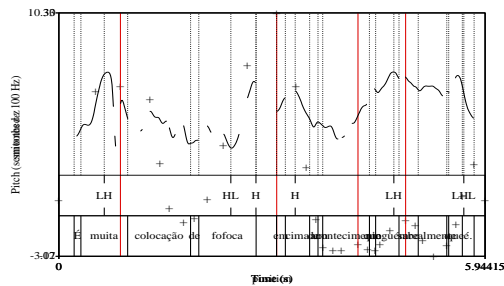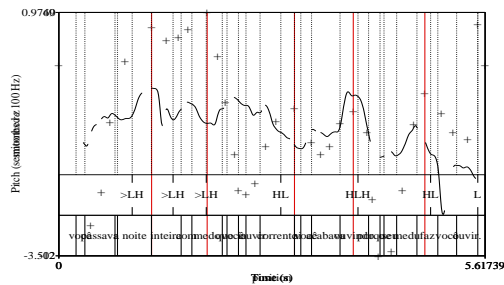

Figure 5: Utterance "Você passa a NOITE INTEIRA com MEDO de ouvir corrente, e você acaba OUVINDO, porque o medo faz você ouvir.",fem. speaker, sp. speech.

nal both minor and major prosodic boundaries, having both a phrasal and a sentencial breaking-marking function, as well as a prominent-bearing function. In signalling prominence, duration peaks were often associated with pitch accents, particularly at major boundaries, but not only. There are inter-individual differences in this kind of association, mainly at minor boundaries.

I thank Prof. S. Madureira for her suggestions.

4. References

[1] Barbosa, P.A., 1996. At least two macrorhythmic units are necessary for modeling Brazilian Portuguese duration. *Proc. of the 1st ETRW on Speech Production Modeling and 4th Speech Production Seminar*, Autrans, France. May 20-24th, 85-88.

[2] Barbosa, P.A., 2007. From syntax to acoustic duration: a dynamical model of speech rhythm production. *Speech Communication*, 49, 725-742.

[3] Barbosa, P.A., Arantes, P., 2003. Investigation of non-pitch-accented phrases in Brazilian Portuguese: no evidence favoring stress shift. *Proc. of the XVth ICPhS*. Barcelona, Spain, 135-143.

[4] Barbosa, P.A. et al., 1999. Aiuruetê: a high quality concatenative text-to-speech system for Brazilian Portuguese. *Proc. of Eurospeech'99*. Budapest, Hungary. 5, 2059-2062.

[5] Barbosa, P.A., Arantes, P., Silveira, L.S., 2004. Unifying stress shift and secondary stress phenomena with a dynamical systems rhythm rule. *Proc. of the Second Int. Conf. on Speech Prosody*. Nara, Japan, 49-52.

[6] Barbosa, P. A., Lucente, L., Silveira, L. S., Madureira, S., 2007. Intonational Patterning in the Paulista Brazilian Portuguese Variety: Two Case Studies in the Framework of AMPER. *Actas do III Cong. Int. de Fon. Exp.*, Santiago de Compostela, Spain, 2005, 117-129.

[7] Beckman, M.E., 1992. Evidence for speech rhythms across languages, In *Speech, perception, production and linguistic structure*, Y.Tohkura, E. Vatikiotis-Bateson, and Y. Sagisaka (eds.). Amsterdam: IOS Press, 457-463.

[8] Boersma, P., Weenink, D., 2005. Praat: doing phonetics by computer. Accessible in <http://www.praat.org/>, version 4.6.21.

[9] Bolinger, D., 1986. *Intonation and its parts: melody in spoken English*. London: Edward Arnold.

[10] Classe, A., 1939. *The Rhythm of English Prose*. Oxford: Blackwell.

[11] Frota, S., 2002. Nuclear falls and rises in European Portuguese: a phonological analysis of declarative and question intonation. *Probus*, 14, 113-146.

[12] Hirst, D.J., 1977. Intonative features. A syntactic approach to English intonation. *Janua Linguarum Series Minor* 139. Mouton: The Hague.

[13] Kohler, K.J., 2007. Beyond Laboratory Phonology. The phonetics of speech communication, In *Experimental Approaches to Phonology*, M.-J. Solé, P.S. Beddor, M. Ohala (eds.) Oxford: OUP, 41-53.

[14] Lucente, L, Barbosa, P.A., 2008. Narrow focus in Brazilian Portuguese: spatial and temporal constraints. This vol.

[15] Madureira, S., 1994. Pitch patterns in Brazilian Portuguese: an acoustic phonetic analysis. *Vth Australian Int. Conf. on Speech Science and Technology*, Perth, Australia, 156-161.

[16] Massini, G., 1991. A duração no estudo do acento e do ritmo em português. Master's thesis. Univ. of Campinas.

[17] Moraes, J.A., 1998. Intonation in Brazilian Portuguese. In *Intonation Systems. A Survey of Twenty Languages*, D. Hirst, and A. Di Cristo (eds.). Cambridge: CUP, 179-194.

[18] Moraes, J.A., 1999. F0 declination in Brazilian Portuguese in read and spontaneous speech. *Proceedings of the XIVth ICPhS*. San Francisco, USA, 2323-2326.

[19] Silverman, K. et al., 1992. ToBI: a Standard for Labeling English Prosody. *Proceedings of the 2nd International Conference on Spoken Language Processing*. Banff, Canada, v. 2, 867-870.