Article

The Interplay between Syllabic Duration and Melody to Indicate Prosodic Functions in Brazilian Portuguese Story Retelling

Plinio A. Barbosa * and Luís H. G. Alvarenga

Department of Linguistics, Universidade Estadual de Campinas (UNICAMP), Campinas 13083-970, SP, Brazil; l255365@dac.unicamp.br
* Correspondence: pbarbosa@unicamp.br

Abstract: This paper investigates the relationship between syllabic duration and F0 contours for implementing three prosodic functions. Work on rhythm usually describes the evolution of syllable-sized durations throughout utterances, rarely making reference to melodic events. On the other hand, work on intonation usually describes linear sequences of melodic events with indirect references to duration. Although some scholars have explored the relationship between these two parameters for particular functions, to our knowledge, there has been no investigation on the systematic correlation between syllabic duration and F0 values throughout narrative sequences. Based on a corpus of story retelling with nine speakers of Brazilian Portuguese from two regions, our work investigated the interplay between syllabic duration and melody to signal three prosodic functions: terminal and non-terminal boundary marking and prominence. The examination of local syllabic duration maxima and four F0 descriptors revealed that these maxima act as landmarks for particular F0 shapes: for non-terminal boundaries, the great majority of shapes were increasing and increasing–decreasing patterns; for terminal boundaries, almost all shapes were decreasing F0 patterns; and for prominence marking, the great majority of shapes were high tones across the stressed syllable. Time series analyses revealed significant correlations between duration and specific F0 descriptors, pointing to a ruled interplay between F0 and syllabic duration patterns in Brazilian Portuguese story retelling.

Keywords: prosodic function; prosodic annotation; story retelling

1. Introduction

Intonation and speech rhythm research is practically exclusively associated with the respective studies of fundamental frequency (F0) patterns and syllabic duration throughout utterances, with a few exceptions reviewed below. Those who work with the description and modeling of a melodic contour focus almost exclusively on the description of the form and progression of F0 contours and levels in time, with indirect or secondary references to syllabic duration. This is the case of the American ToBI annotation system (Silverman et al. 1992), which is largely used in phonology-based intonation research in several languages to which the ToBI system was adapted (see, for instance, G-ToBI for German and SP-ToBI for Spanish). As for prosodic breaks, ToBI only marks the strength of a break in essentially two levels (3 or 4) to direct the analyses on matters of the sequence and organization of mono- and bitonal events linked to the realization of pitch accents and boundary tones. Some criticism appeared after ten years of use of the ToBI system. The main drawbacks concern the low inter-annotator reliability for the choice of pitch accent tones (Wightman 2002), as well as the possible circularity in not dissociating form from function (Hirst 2005). These two drawbacks are avoided in the work reported here by a procedure we used some time ago (Barbosa 2010) and which is similar to the rapid prosodic transcription used by Cole and collaborators (Cole et al. 2010; Cole and Shattuck-Hufnagel 2016). The method they proposed consists of asking listeners to evaluate how the speakers split their production into smaller parts by signaling breaks and how they highlighted words by listening to the

corresponding audio files. To accomplish both tasks, the listeners were invited to mark the breaks with bars (/) and highlighted words with circles. The strength of these two functions (boundary and prominence) is considered as directly related to the proportion of the listeners' choices.

From the phonetic point of view, models of intonation have relied on the realization of focus and boundary marking functions from simple phonological descriptions of the sentence with codes which refer to F0 levels and contours. Examples of such models were proposed for languages such as Swedish (Bruce 1977, 1982; Gårding and Bruce 1981), Dutch (t'Hart 1984), English (t'Hart 1984; van Santen and Möbius 2000), and Japanese (Fujisaki and Hirose 1984), among others. Botinis et al. (2001) revised some of these models, pointing out that these phonetic models combine global trends and local changes in the F0 contours for implementing focus. In particular, the model by van Santen and Möbius (2000) combined the duration of accented and the following non-accented syllables to generate the F0 contours associated with pitch accents.

Recent work on Brazilian Portuguese (BP) analyzed the F0 contour and the role of the visual modality in the realization of both wide and narrow focus in declarative and interrogative sentences. The authors evaluated the roles of F0, syllabic duration, and intensity in the signaling of the focal function (Carnaval et al. 2022; Miranda et al. 2021, 2022). Also, in BP, Teixeira et al. (2018) studied the relevance of more than 100 acoustic prosodic descriptors for signaling terminal and non-terminal boundaries, with the aim of producing an algorithm for the automatic detection and classification of prosodic boundaries based on data of the C-Oral-Brazil corpus. Their work revealed the need to combine duration and F0 to achieve a better performance in predicting prosodic boundaries in spontaneous speech.

As for work on other languages, a comparative study of Mandarin and English investigated the joint manifestation of F0 contours and syllabic duration changes associated with tone, intonation, and duration in English (Xu 2009). A more recent work modeled the joint manifestation of F0 contours and syllabic duration for focus implementation in Emirati Arabic (Alzaidi et al. 2023). Work by Christodoulides (2018), which investigates the relative importance of prosodic–acoustic parameters to signal different kinds of boundaries in French spontaneous speech, points to the higher relevance of silent pause duration. As for prominence marking in English, the work by Herment-Dujardin and Hirst (2002) pointed out the relative relevance of F0 and duration parameters, of their combination as well as of semantic content. Their work also took into consideration a corpus of spontaneous speech. With controlled experiments, work on the prosodic cues for signaling prominence in German, both in adults (Holzgrefe et al. 2012) and in 8-month infants (Wellmann et al. 2012), revealed that the combination of pitch change and preboundary lengthening is a reliable cue for perceiving a boundary in speech.

Recently, for Greek, Arvaniti et al. (2024) used functional principal component analysis (FPCA) to evaluate the trade-off between F0 shape and duration in a corpus of 13 speakers arranged in different pairs to read dialogs. The first two components of the FPCA revealed that "F0 curves with a less pronounced dip and an earlier and lower peak were associated with longer accented vowel duration" and that "lower F0 curves, particularly those low in the preaccentual region, were associated with longer duration of that region." (pp. 292 and 293). None of these works, however, analyzed the systematic correspondence of F0 descriptors extracted from the F0 trace and syllabic duration for realizing focus and for signaling prosodic boundary throughout long stretches of unscripted speech.

In the same direction, it is important to highlight that long-standing research on speech rhythm rarely makes reference to F0 contour events. This line of research has been pointing to the primacy of syllable duration for marking prosodic boundaries and prominence in different languages (see Leemann et al. 2016 in eight languages/varieties; Streefkerk 1997 in Dutch; Gussenhoven and Rietveld 1992 in English; and Barbosa 2007 in BP).

Recent work on the notion of macro-rhythm considered the timing of melodic events such as pitch accents throughout utterances but did not take into account the analysis of

the duration of entire syllabic sequences regardless of their accentedness (Jun 2014; Wehrle et al. 2020).

It is the interplay between F0 and syllabic duration patterns that we propose to investigate here by examining the cross-correlation between the corresponding time series in places where terminal and non-terminal boundaries and prominence are realized in storytelling. Speech productions in two dialects (São Paulo and Rio de Janeiro dialects) are analyzed. These two dialects were chosen because they have been the object of the majority of BP intonation studies.

Time series correlations capture the systematic correspondence between two variables. If both series are identical, the correlation is 1, whereas if they are completely unrelated on statistical grounds, the correlation is 0. In between, the technique is able to capture systematic positions where both series have local maxima or minima, which is the aspect we want to investigate here. To the best of our knowledge, the previous literature did not explore systematic correlations of these paired series. This also applies for work on BP, for which very limited corpora can be found, usually investigating the realization of local functions in read utterances.

As for work on narratives and storytelling, no systematic exploration of the relationship between melodic and duration levels has been carried out. The work by Oliveira (2012) investigated narratives in BP on the role of prosodic parameters such as F0, speech rate, and pause for the segmentation of narrative sections. He pointed to the cyclical character of speech rate which accompanies the change across these sections.

Other studies on storytelling proposed prosodic modification rules from neutral speech to endow speech synthesis systems with this speaking style. They have been developed for languages such as Malay (Ramli et al. 2016), Hindi (Verma et al. 2015), and French (Doukhan et al. 2011). The first one points to the prosodic differences between storytelling and the neutral speeches of two professional storytellers, while the second one evaluates the differences regarding distinct emotions conveyed during the narratives of five laypeople telling stories to children. The work on French, on the other hand, uses a single professional storyteller and assesses differences in melodic, durational, and intensive prosodic parameters between other speaking styles, as well as across narrative sections. None of these studies explore the retelling of a story, which has a strong component of cognitive load (see Dixon and Gould 1996 for the difference between storytelling and story retelling; see Pratt et al. 1989; Skehan and Foster 1999 for the study of processing load in story retelling). The number of speakers is also lower than the number of participants in our study.

In the present study, we fill these two gaps by describing and quantifying the convergence between F0 descriptors and syllabic duration in story retelling with a corpus that includes more speakers than usually found for this kind of study as well as analyzing running speech, and not isolated utterances, as is the case in the great majority of studies investigating BP. Recently, we investigated the interplay between F0 descriptors and normalized duration maxima in three speakers from São Paulo state, Brazil, for both reading and story retelling recorded in 2009 (Barbosa 2024). The results of cross-correlations between the positions of prominence and each series of four F0 descriptors (F0 median, F0 range, F0 rise and fall rates) revealed that the two speaking styles differ in the sense that the F0 range and F0 rise/fall rate are more correlated with peaks of duration in story retelling than F0 median is in reading. In this style, F0 median and F0 rise rate peaks are mostly aligned with duration maxima, which is also due to rising contours being the most frequent shape for realizing this function in reading. A rising contour for marking prominence is the most frequent contour in story retelling, but it is preceded by F0 fall rate minima and with a large F0 range. This is related to the fact that an F0 fall with higher rates precedes the typical rising of the prominent F0 contour either inside the longer V-to-V interval or inside the immediately preceding unit, which signals that the alignment of an F0 fall is more relevant for preparing a variable following an F0 rise.

As for the cross-correlation between the non-terminal boundary positions with the F0 descriptor series, with the exception of one speaker, the F0 range and F0 rise rate maxima correlate more with duration peaks, with higher values for story retelling, which can be taken as a characteristic of this style: the peaks of risings and extended ranges may indicate in BP that the speaker has more to say. Finally, the same work pointed to the correlation between terminal boundary positions with the F0 descriptor series, showing that F0 median minima and F0 range maxima are the descriptors with higher values of correlation in reading. This can be seen as a characteristic of this style when BP speakers signal terminal boundaries.

Based on previous work, our hypotheses concerning the correlation of one or more F0 descriptor with duration maxima are as follows: (1) F0 range and F0 rise/fall rate maxima are the most relevant correlated descriptors when signaling prominence; (2) the F0 range and F0 rise rate maxima are the most relevant descriptors at non-terminal boundaries; and (3) F0 median minima and F0 range maxima are the most relevant descriptors at terminal boundaries, which are less frequent in the case of story retelling due to the nature of the task.

2. Methodology

2.1. Corpus

The BELÉM corpus is formed by readings and retellings of the story about the origin of the Portuguese Belém pastries by Brazilian and Portuguese male and female speakers with a range of 30 to 45 years of age from the state of São Paulo. For the building of this corpus, the first recording was made in 2009. In 2022, the recordings were resumed to add to the BELÉM corpus Brazilian speakers from different dialectal regions of Brazil, starting with speakers from Rio de Janeiro, as well as extending the number of speakers from São Paulo. Furthermore, a shortened version of the original text (753 words instead of 1568), which can be found in Appendix A, was used as the text for the new retellings. These new recordings include six subjects from São Paulo (3 females and 3 males) and three subjects from Rio de Janeiro (2 females and 1 male). As in 2009, this new recruitment was based in the traditional friend-of-a-friend approach in sociolinguistics started by the friends of the second author in Brazil. Their narratives had between 220 and 330 words (2 to 4 min). The additional participants were between 20 and 30 years old at the time of recording and were college students of different majors. These recordings are available in the Figshare platform [https://doi.org/10.6084/m9.figshare.25383190.v1 (accessed on 5 July 2024)] Only data from the new recordings, based on the shortened version of the Belém pastries story, are used for the analyses presented here.

Due to restrictions related to the COVID-19 pandemic, the participants themselves used the Easy Voice app on their own cell phones to make all recordings. The nine subjects read the text and soon after retold the story in their own words. Only the retellings are considered for analysis here. Because this app allows choosing among different codifications, instructions were given to record all audio files in PCM format (WAV) at a sampling rate of 48 kHz. The first author, who is a trained phonetician, further evaluated all audio files for the presence of noise that would impact computing the acoustic parameters. No recordings were discarded. The recordings were then resampled at 16 kHz and leveled to the same maximum intensity level at 65 dB.

2.2. Acoustic–Prosodic Parameters for Analysis

A Praat script, Prosody Descriptor Extractor (Barbosa 2020), henceforth PDE, was used to extract statistical descriptors of prosodic–acoustic parameters based on syllabic duration and fundamental frequency (F0). The script is accompanied by a manual, and examples of input and output data are also given.

A syllabic duration series was obtained by normalizing and smoothing the duration of the sequence of V-to-V intervals, that is, of all the segments delimited by two immediately consecutive vowel onsets.

The reason for using V-to-V intervals instead of phonological syllables is threefold: (1) Dogil and Braun (1988) conducted psycholinguistic experiments that showed that vowel onset tracking is a fundamental property of speech signal processing in our brain. (2) This property of the brain activity was also pointed out by Chistovich and Ogorodnikova (1982) by examining post-stimulus temporal neuronal responses to speech. They reported amplified neuronal responses to portions of energy increase typical of C-V transitions, accompanied by response suppression in regions where energy decreased (typically around V-C transitions). (3) Furthermore, a segmentation based on vowel onsets has the advantage of being detectable under moderately noisy conditions (Barbosa 2010).

The perceptual impression of speech rate is also primarily associated with the tracking of vowel onsets and not syllable onsets, as was experimentally tested by Pompino-Marschall (1991) with German subjects. This reveals that the perception of the syllable sequence relies on the detection of the nucleus of the syllables, more often occupied by vowels. Because V-to-V intervals are syllable-sized units, which mark the flow of syllables throughout utterances, we refer to the V-to-V interval duration as the "syllabic duration".

Figure 1 illustrates, in the second tier, the segmentation and labeling of the V-to-V intervals for the excerpt "e aí no fim ele dormiu..." (and then he eventually slept) of a story retelling carried out by a female speaker from Rio de Janeiro. The content of the other tiers is explained later on in this text.
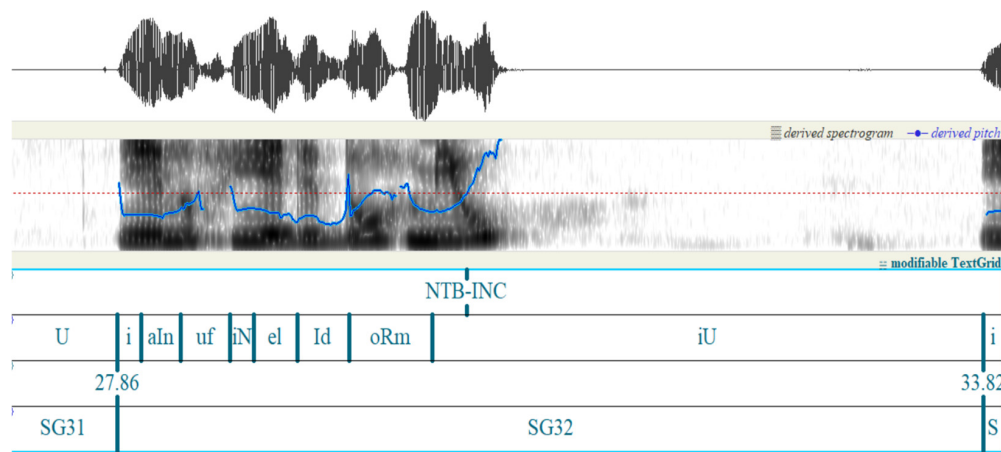

Figure 1. Waveform (**top**), F0 trace (blue) superimposed to a broadband spectrogram (**middle**) and annotation tiers of the excerpt "e aí no fim ele dormiu..." (and then he eventually slept) of a story retelling from a female speaker from Rio de Janeiro. From top to bottom, tier 1 codes the prosodic function and F0 shape, tier 2 the V-to-V intervals with their labels, and tier 3 the value of the normalized interval duration at the end of the stress group delimited in tier 4.

For normalizing the duration of the V-to-V intervals, the PDE script uses the z-score transformation given in Equation (1), where dur is the V-to-V duration in ms and the pair (μ_i and var_i) denotes the reference mean and variance in ms of the phones within the corresponding interval. These reference descriptors are found in the file TableOfReal included with the script.

$$z = \frac{dur - \sum_i \mu_i}{\sqrt{\sum_i var_i}} \tag{1}$$

A smoothing technique is then used which consists of serially applying a smoothing technique carried by a 5-point moving average filter given by Equation (2) to the sequence of z-scores obtained from the previous stage.

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

This technique minimizes the effects of intrinsic duration and number of segments in the V-to-V unit, as well as attenuates the minor effects of duration variation related to the realization of lexical stress in the speech chain. Local peaks of smoothed z-scores are then detected by tracking the position for which their discrete first derivative changes from a positive to a negative value.

Previous research demonstrated good correspondence between these local peaks of syllabic duration and the perception of both prominent and pre-boundary syllables with correlations between 69 and 82% for reading (Barbosa 2008), provided that silent pauses, when applicable, were included in the corresponding V-to-V interval, as can be seen in Figure 1 for "iU". A local peak of smoothed z-scores is considered here an index of prosodic strength and the duration of the silent pause, when present, is an integral part of the signaling of this strength. It is related to the fact that the longer a silent pause is, the stronger the perception of a boundary (Sanderman and Collier 1995). The same applies when a local peak of smoothed z-scores signals the prominence of a syllable in a word. This is why these maxima were taken as indicators of the right edges of stress groups in BP, a language for which syllabic duration is the main parameter for signaling lexical and phrase stress (Fernandes 1976; Massini 1991; Barbosa 1996). When a silent pause is part of the V-to-V interval and it has a duration peak, this does not mean that it is part of the stress group but only that the stretch of sound at the left of this pause ends the stress group. For the purpose of this study, the interval between two immediately consecutive smoothed z-score maxima is called the "stress group". The stress group delimited in such a way is taken as a prosodic constituent that ends in a prominent unit (often an informational focus) or a prosodic boundary, either terminal or non-terminal.

The PDE script also delivers 12 F0 descriptors for the intervals of the tier specified by the user, here, the V-to-V intervals tier. From these 12 parameters, we selected four F0 descriptors: F0 median, F0 range (F0 maximum minus F0 minimum), and F0 rise and F0 fall mean rates. The latter two descriptors were computed as the first derivatives of smoothed and interpolated F0 contour. Smoothing and interpolation employed embedded Praat functions with 5 Hz as the cut frequency for smoothing and a quadratic function for interpolation, avoiding small oscillations and gaps of the F0 contour before computing the derivative. These four F0 descriptors in particular were chosen due to their relevance for signaling prominence and boundary in the literature (see Mittmann and Barbosa 2016 for a review). This relevance is illustrated in Figure 1, where the values of the V-to-V smoothed z-scores at the right ends of stress groups 31 and 32 (27.86 and 33.82) correspond to V-to-V units followed by long pauses. Furthermore, by the end of stress group 32, an F0 rise (INC) signals a non-terminal boundary (NTB). In those cases, the F0 range and F0 rise rate are better descriptors of F0 shape in signaling this type of boundary.

It is exactly the relation between the four F0 descriptors and duration maximum positions that realize each one of three prosodic functions, which are further discussed in the next section. We illustrate this relation in Figure 2 before presenting the prosodic functions in the next section. By examining the very first terminal boundary position at the extreme left of Figure 2, one can observe a syllabic duration peak coinciding with a local minimum of the F0 contour. The "meeting" of these two landmarks, a local maximum of syllabic duration and a local minimum of F0, unequivocally marks a terminal boundary of declarative utterances for a Brazilian listener (Moraes 1998). What is more, the fact that the relevance of both F0 minima and lower rates of F0 decrease, as we will see later on, allows the listener to distinguish the terminal boundary from the non-terminal boundary, such as the ones shown in the antepenultimate and penultimate positions in this figure. In the penultimate position, the non-terminal boundary is marked by both a local duration peak and an F0 peak, soon followed by an F0 fall. The two dashed positions, on the other hand, indicate prominences by F0 rising (LH) and falling (HL) contours, respectively. In both positions, the duration peaks are less salient in contrast to the salient local peaks of the boundary positions, as revealed by their heights.

Figure 2. Original (above, red line) and smoothed F0 contour (above, superimposed, dark blue being the smoothed one), syllabic duration contour (light blue, middle) and positions of V-to-V duration maxima where the following functions are realized: terminal (solid line), non-terminal (dotted line), and prominence (dashed line).

Analyses of the parameters investigated here were carried out for two regions (São Paulo and Rio de Janeiro) and two genders (males and females). The reason to split between the two genders and two regions is to suggest future work on the contribution of these factors, as well as to be able to engage with the previous literature, as presented in the Discussion section.

2.3. Prosodic Function Annotation

For each stress group automatically generated by the PDE script, the two authors independently annotated, based on their respective perception, which one of the three functions studied here, i.e., the terminal boundary (TLB), non-terminal boundary (NTB), or prominence (PRO), was realized at its right edge. After that stage, both authors discussed cases of initial disagreement until they achieved consensus. This was carried out for less than 1% of the labels given. Previous research revealed that even lay listeners perceive these functions with quite a fair to high degree of inter-rater reliability (Raso and Mello 2012; Cole and Shattuck-Hufnagel 2016).

The detection of the positions where one of the three functions investigated here was realized and signaled by a peak of normalized/smoothed syllabic duration had performance compatible with previous tests (e.g., success rates between 69 and 82% for reading in Barbosa 2008). As regards missing errors, 17% of the prosodic events were not detected. In addition, the procedure produced circa 6% of false alarms. A false alarm (FA) stands for a duration peak that does not correspond to any of the functions studied here. Thus, the procedure has a success rate of 77%, which means that the use of syllable-size duration for detecting both prominence and boundary is quite successful. The relative frequency of hesitations was 7% for São Paulo and 5% for Rio de Janeiro speakers. Gender-wise, the hesitations were 10% for males and 11% for females.

Hesitations were not analyzed here. If one of these three functions was realized at places not at the right edge of the stress groups, the label NDT (not detected) before the label for the performed function was used. For instance, NDT_NTB signals that the specific non-terminal boundary was not realized at the right edge of the automatically delimited stress group. The total number of labels, excluding the FA labels, was 500 for the speakers of São Paulo and 272 for the speakers of Rio de Janeiro distributed in the following way, according to the region: 34 TLB, 296 NTB, and 170 PRO for São Paulo and 39 TLB, 153 NTB, and 80 PRO for Rio de Janeiro.

For each boundary function, both terminal and non-terminal, the shape of the melodic contour that ended the stress group was labeled as follows: increasing (INC), decreasing (DEC), increasing–decreasing (INCDEC), leveled (LEV), and decreasing with a slow F0 rate

(SLDEC). All decreasing contours with a falling rate between 2.5 and 15 Hz every 50 ms were classified as a case of slow decreasing (SLDEC); those with a falling rate below 2.5 Hz every 50 ms were classified as LEV and the others as DEC.

As for the prominence function, we marked the F0 shapes by conventional symbols from intonation research (monotonal and bitonal units): leveled high F0 peak (H), falling contour (HL), and rising contour (LH), where the right tone of the bitonal contours was aligned with the stressed syllable. This choice of annotation allows for a more direct comparison with the previous literature, which uses such a convention. The H, HL, and LH labels were chosen according to criteria found in Lucente ([2012](#)) referring to the F0 shape during the stressed vowel in this way: the LH label refers to a rising F0 preceded by an F0 fall; the H label refers to a less sharp increase just before and leveled during the stressed vowel and not preceded by an F0 fall; the HL label refers to an F0 fall during the stressed vowel preceded by an F0 rise.

2.4. Statistical Analyses

To quantify the systematicity of the relationships between each F0 descriptor and the positions in which a F0 shape signals a function among the three studied here, we calculated the cross-correlation of these two time series for each function: (1) a series of values of 0 and 1 along the succession of V-to-V units in which 0 meant that the particular function was not realized in the current unit, and 1 if it was realized there, and (2) a series of each one of the four F0 descriptor values computed in each V-to-V unit in Hertz. This was conducted for each speaker. To do this, we used the *ccf()* function for cross-correlation between two time series available in the *tseries* package on the R statistical platform (R Project n.d., version 4.3.1). For a similar analysis of dialogs analyzing the correspondence between F0 and intensity contours, see Buder and Eriksson ([1999](#)).

These cross-correlations were computed around a window of five V-to-V units before and after the duration peak to investigate for which lag these landmarks have the higher correlation with a particular F0 contour descriptor. A lag of 0 (zero) means that the maximum duration position and the F0 descriptor series are compared directly, without moving one series in relation to the other, that is, the F0 descriptor maximum or minimum corresponds exactly with the duration maximum position. A lag of 1 (one) means that the maximum duration position series is compared with the F0 descriptor series moved one V-to-V unit to the right, allowing us to investigate whether the series of F0 descriptor minima or maxima one V-to-V unit to the left of the duration maximum is better correlated with the latter parameter. A lag of -1 (minus one) means that the maximum duration position series is compared with the F0 descriptor series moved one V-to-V unit to the left, allowing us, on the other hand, to investigate whether the series of F0 descriptor minima or maxima one V-to-V unit to the right of the duration maximum achieves higher correlations. The same reasoning applies for higher lag values. The significance level for the cross-correlations was set to 0.05.

Proportion tests were used for comparing the significance of proportions of functions between the two regions studied here and between speakers, both at the 0.05 significance level. For this purpose, the *prop.test()* function in R was used.

3. Results

The great majority of studies on BP intonation, both from phonological and phonetic perspectives, have investigated either the dialect of São Paulo (Madureira 1994, 2016; Madureira and Fontes 1997) or of Rio de Janeiro (Moraes 1998, 2008; Carnaval et al. 2022). For the sake of comparison, in the following, we refer to significant differences and commonalities found between the two dialects. Furthermore, new findings beyond the analysis of controlled speech with isolated utterances usually found in those previous studies were obtained in the present one.

Figure [3](#) shows the relative frequency of the prosodic functions under scrutiny here for the two regions. Non-terminal boundaries are the most common function realized in

story retelling, with proportions higher than 70%. According to a proportion test, terminal boundaries are significantly more frequently used by participants from Rio de Janeiro (14%, ranging from 11 to 16%) than participants from São Paulo (7%, ranging from 3 to 12%). The amount of data for this comparison (a total of 39 TLB in São Paulo and 34 TLB in Rio) is higher than the usual minimum number of 30 data points suggested for paired comparisons including proportion tests (Dowdy and Wearden 1991). The differences in proportion between the two regions for the other two functions are not significantly different.
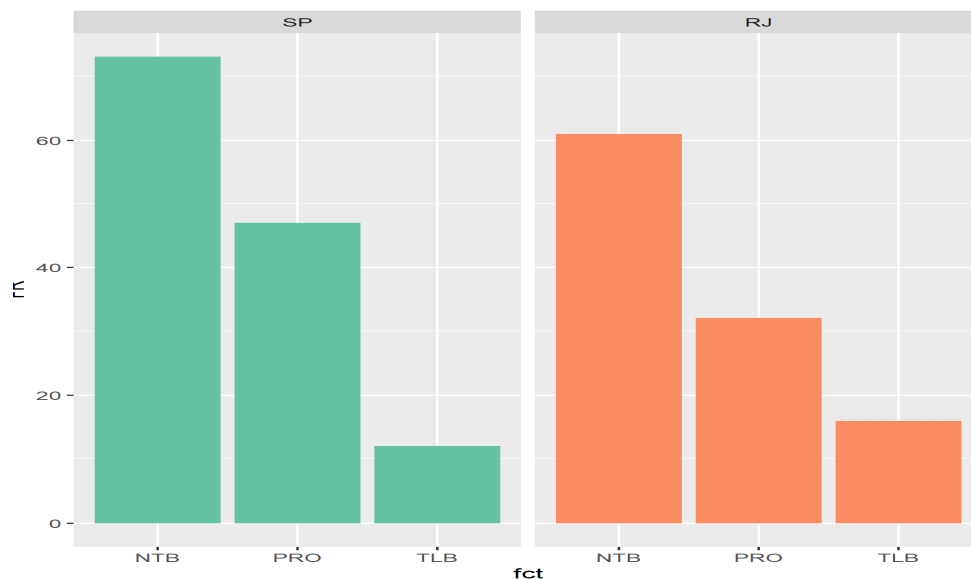
Figure 3. Relative frequencies of the three functions in story retelling according to dialect (**left**, SP = São Paulo, and **right**, RJ = Rio de Janeiro).

Considering the limited number of speakers, it is important to check if the results are relatively homogeneous across speakers or are an artifact of computing the average. This does not seem to be the case, as can be seen in Figure 4. There are, however, some exceptions. Figure 4 shows that male speaker AM from São Paulo has proportions that differ from the general tendency either for males or for São Paulo speakers: he uses the prominence function more than signals non-terminal boundaries. Furthermore, speakers MV (male) and SP (female) from São Paulo use a very low frequency of terminal boundaries.
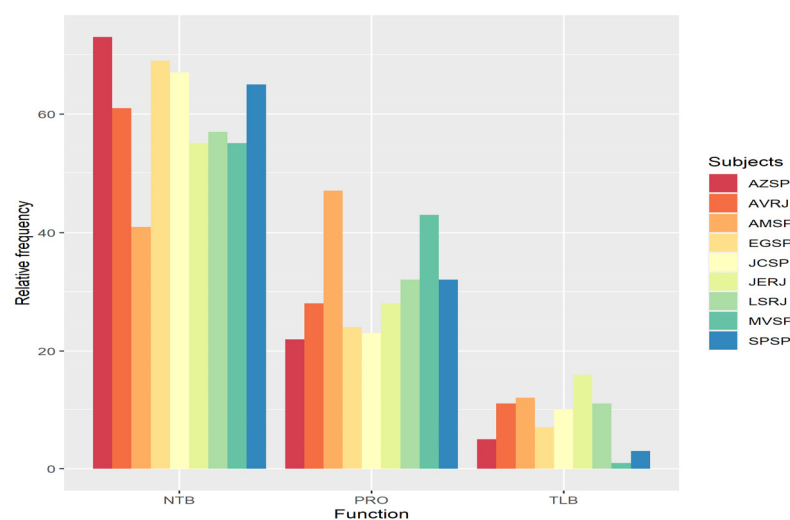
Figure 4. Proportion of the three functions in story retelling according to subject. The first two letters stand for the subject and the two last ones, the dialectal origin (SP or RJ).

Table 1 indicates the relative frequencies of the F0 shapes associated with each of the functions in story retelling according to dialect. The results are pooled for all speakers from each dialect. The two most frequent shapes for non-terminal boundaries are level and increasing, making up circa 86% in RJ and 80% in SP of all shapes, with a significant preference for increasing contours in São Paulo against level in Rio, as confirmed by a proportion test (X^2 = 4.0, *p* = 0.04 in São Paulo, X^2 = 8.3, *p* = 0.004 in Rio). For prominence, the most frequent shape is the high tone (H) followed by the LH contour in both dialects. Together, they account for 80% of all shapes in both RJ and SP. As for the terminal boundaries, the decreasing shape is by far the most relevant melodic form for realizing this function.

Table 1. Proportions (%) of the melodic forms associated with each of the functions in story retelling according to dialect. Significantly different proportions for the same function and shape in each dialect is indicated with an asterisk (*).

Dialect	Function	Shape	Shape Proportion per Function
RJ	NTB	LEV	51.91 *
		INC	35.24 *
		SLDEC	7.04
		INCDEC	5.76
	PRO	H	58.75
		LH	21.25
		HL	13.75
		HLH	6.25
	TLB	DEC	74.99
		INCDEC	19.43
		SLDEC	5.54
SP	NTB	INC	44.9 *
		LEV	36.46 *
		INCDEC	11.13
		SLDEC	7.4
	PRO	H	43.5
		LH	24.08
		HL	18.79
		HLH	13.5
	TLB	DEC	85.27
		INCDEC	8.82
		SLDEC	5.88

The only two significant differences across gender were that (1) male speakers use the slow decreasing shape more than female speakers, a property already found in a previous study in BP reading (Barbosa and Mareüil 2016), and that (2) there is a significant preference for the high tone in female speakers in the prominence position (circa 61% against circa 42% in males).

With regard to the cross-correlations between the time series whose contours were illustrated in Figure 2, only significant results are shown in Table 2, both by dialect and by gender. Only correlations within the window of lags −2 and 2 are taken into account due to the extension of the phonological words which generally include one to two prestressed and one to two post-stressed syllables. As can be seen in Table 2, higher correlations were found for lag 0, with only a single exception (TLB for females for the F0 rise, lag −2). Only

the two highest significant correlations are shown because the others are lower than 5%, not taken into consideration here.

Table 2. The two highest significant correlations between position of syllabic duration peak and F0 descriptor value according to function, gender and dialect. When not mentioned otherwise, the correlations shown here are for lag 0. The abbreviation "ns" stands for non-significant for any of the four F0 descriptors.

PRO	NTB	TLB
SP		
F0range (0.06)	F0range (0.36)	F0fall (0.07)
F0rise (0.06)	F0rise (0.17)	F0range (0.04)
RJ		
F0fall (0.12)	F0range (0.15)	ns
F0range (0.09)	F0rise (0.06)	
Male		
F0range (0.06)	F0range (0.27)	ns
F0fall (0.06)	F0rise (0.17)	
Female		
F0fall (0.14)	F0range (0.32)	F0range (0.06)
F0rise (0.12)	F0rise (0.12)	F0rise (0.05, lag −2)

F0 range and F0 rise are the most relevant descriptors for non-terminal boundaries when correlations with duration maxima are concerned. With the exception of speakers from Rio de Janeiro, correlations between the F0 range and duration maxima for São Paulo and for both genders are between circa than 30% and 40%. This means that there is a tendency for the F0 range to be higher just before non-terminal boundaries, where syllabic duration maxima are positioned.

The F0 range is a relevant descriptor for the correlation of F0 and duration for the prominence function for males but not females. Looking at the data according to regions, the F0 range is the most relevant descriptor as far as correlations with duration are concerned in both regions. The higher correlation of F0 fall with duration in Rio and in both genders suggest that an F0 fall in the syllabic unit where there is a maximum of duration is a relevant cue for marking prominence in BP.

It is likely that the non-significant correlations for terminal boundaries for the less represented dialect (RJ) and for male speakers is due to the small amount of data for these cases. For São Paulo, however, the correlations of F0 range and F0 fall with duration maxima are the most relevant descriptors. This is compatible with the decreasing shape. Female speakers, on the other hand, seem to privilege the convergence of duration maxima with an F0 rise two syllabic units before the position of the local maximum, which suggests their falls are much more variable and converge less with duration maxima.

Figure 5 shows individual values of cross-correlations between the position of syllabic duration peak and the four F0 descriptor values for lags −5 to 5. The importance of the splitting of the data according to speaker is to show that the correlations shown in Table 2 are not a result of averaging but an effect found in the great majority of the speakers. The results shown here are split according to lag, function and subject. If we consider lags between −2 and 2, which would imply examining up to two syllables around the stressed syllable, where usually duration maxima are found, a lower spread of the correlation values across the individuals for a particular combination of function and F0 descriptor signals a higher consistency of the interplay between syllabic duration and F0 for that particular function. Following this rationale, for non-terminal boundaries, F0 range and F0 rise are, in that order, the most relevant descriptors, which confirms the pooled values of Table 2.
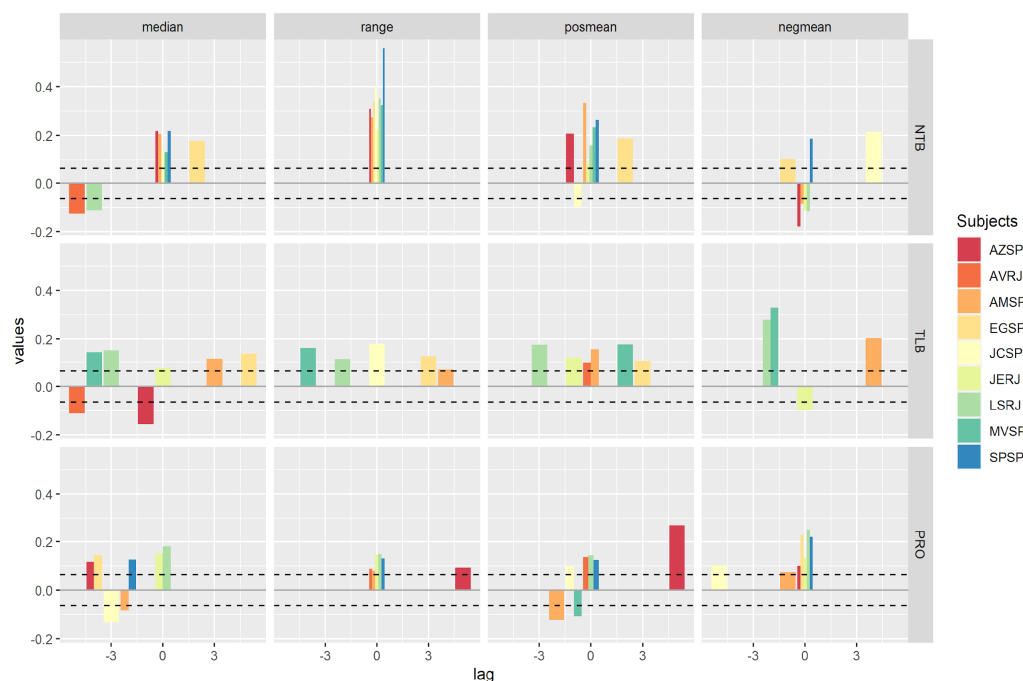
Figure 5. Cross-correlations between position of syllabic duration peak and each one of the four F0 descriptor values (columns) according to function (rows, labels in the right), lag (*x*-axis) and subject (color). Non-significant correlations are not presented. Median stands for F0 median, range stands for F0 range, posmean stands for mean F0 rise rate and negmean stands for mean F0 fall rate.

For prominence, the bars are more concentrated in the windows [−2, 2] for the F0 range, F0 fall, and F0 rise in that order. For terminal boundaries, the situation is more complex, presenting much more inter-individual variation with a higher concentration in the same lag window for F0 falls, as expected, with cross-correlations around 30% for the subjects with the two shades of green (LSRJ and MVSP). This value is similar to the correlations for non-terminal boundaries with the F0 range as a descriptor.

An important aspect of F0 dynamics concerns the rates of rises and falls, as well as the amount of the F0 range where significant correlations apply. For São Paulo speakers, terminal boundaries are signaled by decreasing F0 contours, with a median rate of 2.4 Hz/50 ms against 3.5 Hz/50 ms in other positions (corresponding to the points in the data series which includes positions for the two other functions and a position where none of the three functions studied here are realized), that is, F0 falls are slower when realizing terminal boundaries (see Barbosa 2024 for similar results for the same dialect).

As for F0 rises before non-terminal boundaries, the rate values for São Paulo speakers vary from 8.6 Hz/50 ms in the window where non-terminal boundaries are realized against 5.0 Hz/50 ms elsewhere, and for Rio speakers, they vary from 6.2 Hz/50 ms in the window where non-terminal boundaries are realized against 5.0 Hz/50 ms elsewhere. This means that F0 rises are faster when realizing non-terminal boundaries in both varieties but with a larger difference in São Paulo (see Barbosa 2024 for similar results for another corpus of São Paulo speakers).

A similar pattern applies for prominences for São Paulo speakers in terms of F0 rises: 7.1 Hz/50 ms in the window where prominences are realized against 5.4 Hz/50 ms elsewhere; for Rio speakers, the figures are 7.2 Hz/50 ms in the window where prominences are realized against 5 Hz/50 ms elsewhere, a very similar result in comparison with São Paulo. As for the F0 range, the patterns are as follows: 28.8 Hz in the window where prominences are realized against 20.0 Hz elsewhere for São Paulo; for Rio, the values are 31.6 Hz in the window where prominences are realized against 17.2 Hz elsewhere. As it can be seen, the figures for rates are quite close in both varieties.

In the prominent position, female speakers have F0 falls of 8.4 Hz/50 ms in the window where prominences are realized against 4.1 Hz/50 ms elsewhere and 3.8 Hz/50 ms in the window where prominences are realized against 2.7 Hz/50 ms elsewhere for male speakers.

4. Discussion

With respect to the previous literature on the prosody of BP, the results of this study add to the current knowledge on the matter, first by analyzing a higher number of speakers and by exploring spontaneous speech in the story retelling style. In fact, with just a few exceptions, such as the study of focus by Carnaval et al. (2022) with four speakers from Rio de Janeiro, the majority of past and recent studies on BP relies on results obtained from isolated utterances (see, for instance, Moraes 1998, 2008; Madureira 1994; Miranda et al. 2021, 2022). Moreover, none of the studies on spontaneous speech investigated the frequency of F0 shapes and the relation of F0 descriptors to syllabic duration maxima in a systematic way. The work by Lucente (2012) with speakers from São Paulo, for instance, occasionally pointed to some aspects of the shapes of F0 at boundary position, while the work by Teixeira et al. (2018), with speakers from Minas Gerais, studied the relative importance of duration and F0 descriptors for the signaling of boundaries but not the amount of their correlation. Although in their material, instances of narrative excerpts can be found, their analysis did not consider these instances as an independent factor of investigation.

Our results point to the fact that in story retelling, the most realized function is non-terminal boundary marking, followed by the signaling of prominence. The preference for non-terminal boundaries is compatible with the need to chain stretches of speech to tell a story and to call attention to the most relevant remembered facts. Non-terminal boundaries and prominences represent almost 90% of the instances of the three functions studied here. Non-terminality is signaled by increasing and level F0 shapes where the first one has fast F0 rises which produces expanded F0 ranges partially synchronized with syllabic duration maxima in the unit just before the boundary. To the best of our knowledge, a new finding of this study is to report that faster F0 rises are associated with duration maxima in realizing non-terminality. The relevance of F0 rise peaks for non-terminality is a consequence of the use of increasing contours for realizing non-terminal boundaries.

An important aspect of the F0 dynamics for terminal boundaries which is significant, despite the limited number of data (about 30 data points per dialect), is the use of slower falls for signaling terminality in comparison to the rate of falls elsewhere, a finding not pointed out by previous studies on BP intonation.

As for prominence, the F0 range is expanded when realizing this function accompanied by earlier synchronization between the F0 fall and duration maxima within the syllabic unit where prominence is realized mainly by an H tone. This fall allows F0 to reach low values which are associated with a lengthened syllabic unit. This behavior is similar to the findings by Arvaniti et al. (2024) on the trade between duration and F0 around accented units, where F0 valleys accompany longer stressed syllables when realizing pitch accent.

A certain amount of interspeaker variability is part of prosodic studies on speaking style (see Perkell et al. 2002; Yoon 2014; Barbosa 2022, inter alia) and this is not different for story retelling, as shown in Figures 4 and 5, already commented upon here. A further investigation of prosodic differences across speakers, including the ones studied here, associated with the study of the effect of different retellings to the listeners could contribute to coaching on storytelling, making listening to stories a more pleasant experience to a target audience. Studies on the relation of poetry declamation and pleasantness (Wagner and Betz 2023 for German; Barbosa 2022 for Brazilian and European Portuguese) have results in this direction.

The rising shape (LH), the second most frequent shape for signaling prominence in the present study, is described by Moraes (2008) as a default realization of narrow focus for the dialect of Rio de Janeiro and later on by Lucente (2012) for the São Paulo dialect, the former for read speech and the latter for spontaneous speech. Nevertheless, both authors only

described the LH shape as being the most common in the dialect they studied, without computing its frequency, as we showed in Table 1: circa 21% in Rio and 24% in São Paulo. In our study, it is the second most frequent and not the most frequent, as in Moraes' and Lucente' studies. As for the terminal boundaries, the decreasing shape is the most relevant melodic form for realizing this function. It is proposed as the prototypical realization at the right end of neutral declaratives in BP by Moraes (1998, 2008) in read speech. Based on the present study, terminal boundaries are realized mostly with the same deceasing shape in story retelling.

Some findings from the current study deserve further investigation. One of them is the significantly distinct proportions of instances of terminal boundaries between the two regions, with a higher proportion for Rio de Janeiro (14% against 7% in São Paulo, see Figure 3). Having additional data from the two regions could reveal, if this difference in proportion is confirmed, that speakers from Rio de Janeiro complete thematic excerpts of the story being told more often than speakers of São Paulo. Another finding is related to interspeaker differences specially referring to the frequency of terminal and non-terminal boundaries. This could benefit studies of the effect of different storytelling and story retelling on listeners in terms of different degrees of pleasantness depending on the uses of the types of boundaries (see Figure 4 for some differences across speakers). The differences across gender, like the finding that female speakers are faster in signaling prominences by a previous sharp fall also deserve a more extended investigation with more speakers and a balanced corpus in this respect. As the study by Barbosa and Mareüil (2016) showed, this could contribute to a perception of more musical prosody in female speakers in BP.

5. Conclusions

The results of this study derive from the use of a new methodology for investigating the relation between the two main prosodic parameters (syllable duration and F0 contours) for signaling prominence and boundaries. The results presented here stress the importance of investigating the correspondence between F0 contours and syllabic duration contours to further understanding how prosodic functions are realized.

The examination of local syllabic duration maxima and the four F0 descriptors revealed that these maxima act as landmarks for particular F0 shapes: for non-terminal boundaries, the great majority of shapes were increasing and increasing–decreasing patterns; for terminal boundaries, almost all shapes were decreasing F0 patterns; and for prominence marking, the vast majority of shapes were high tones across the stressed syllable.

Time series analyses revealed significant correlations between duration and specific F0 descriptors pointing to a ruled interplay between F0 and syllabic duration patterns in Brazilian Portuguese story retelling. The cross-correlation values obtained in our analysis of the data indicate that the right edge of stress groups in BP, primarily marked by peaks of normalized duration of syllable-size units (the V-to-V unit) is the characteristic place where duration and F0 landmarks meet.

What is more, expanded F0 ranges and faster or slower rates of F0 contours are significant aspects of the dynamics of boundary marking in BP story retelling, findings that could stimulate cross-linguistic work on the prosody of storytelling and story retelling.

Appendix A

Manuel desceu a escada, maldizendo a sua vida. Por mais que se esforçasse, estava sempre se metendo em problemas. Tristonho, foi ver o irmão Bernardo de Santa Maria. E este deu um sorriso, com uma ternura bondosa. Muito baixo, gordo, de careca brilhante e duas bochechas encarnadas como maçãs maduras, ele poderia ser uma figura rústica se não fossem os olhos cor de bronze, mais brilhantes do que estrelas. Olhos de um homem sereno, inteligente e feliz.

Bernardo, em vez de fazer perguntas, o mandou trabalhar ao lado do forno onde os irmãos faziam grandes pães redondos. Um calorzinho bom devolveu-lhe as forças. E colocou as mãos à obra sem saber o porquê sentia a cabeça girar e um vazio no estômago. Mas o motivo era bem simples. O pão quente e o leite morno exalavam um cheiro de dar água na boca. Não se atreveu, no entanto, a pedir nada. Afinal de contas, estava de castigo.

Para enganar a fome, enterrou as mãos na massa até ao cotovelo e fez uma grande bola. Depois olhou em volta de soslaio. Talvez não reparassem se ele comesse um bocado de massa crua. Ninguém reparou. Então, em vez de fazer os movimentos certos, começou a amassar ao acaso dobrando e redobrando aquela mistura de farinha e água que iria se transformar em pão. Uma ideia genial passou pela sua cabeça. Se fosse buscar outros mantimentos; Em vez de pão poderia fazer uma bola para rechear de carne. Ou um doce. Assim poderia ir comendo bocadinhos disto e daquilo sem que ninguém percebesse.

Foi dito e feito! Guloso como era, começou pela manteiga, depois foi ao leite, raspas de limão, ovos... O frei Bernardo o olhava discretamente sem dizer nada. Manuel tinha recuperado as cores e girava pela cozinha na maior correria.

Em cima da mesa, em vez de um pão redondo, tinha várias tigelinhas de massa cheias de creme. Meteu tudo no forno e foi se sentar. Estava cansadíssimo! Alguns minutos de descanso não seriam notados. Ele se encostou na parede, fechou os olhos e deixou que uma dormência suave fosse tomando conta do seu corpo e espírito. Ele estava tão bem ali! Flutuando dentro de si mesmo que não ouviu os sinos anunciarem que amanhecia. Nem foi à igreja para rezar a Prima. Escorregou para debaixo da mesa e dormia profundamente deitado no chão.

Quando, algumas horas depois, a voz do prior se fez ouvir na cozinha, Manuel quase desmaiou de susto. Não pensava em nenhuma desculpa possível para se justificar. Levantou-se passado de vergonha, com a roupa cheia de manchas, o cabelo bagunçado e os olhos inchados. E foi naquela triste figura que o prior entrou no refeitório. Ali estava o frei Diogo, os outros monges e irmãos, todos com ar muito solene. Com certeza iam expulsá-lo do convento. Abaixou a cabeça, esperando ouvir palavras terríveis, e só então reparou que, na mesa estava um tabuleiro cheio de pastéis que alguém polvilhou de canela.

– Este rapaz é um exemplo para todos nós – exclamou o prior, apontando para o tabuleiro com um gesto breve. – Ele fez uma bela ação, merece uma recompensa.

Manuel ainda olhou à procura do tal rapaz que seria recompensado, e ficou perplexo quando percebeu que era ele próprio, porque não se lembrava de ter feito nada excepcional.

– Ele passou a noite trabalhando sem descanso – continuou o prior no mesmo tom de aprovação.

– Com certeza não poupou esforços para inventar os melhores pastéis que já comi na minha vida. Podem até ser vendidos para fora, o que ajudará muito as finanças do mosteiro.

Até frei Diogo acenou que sim. E não é que ele parecia contente? O prior se virou com um sorriso aberto:

– A partir de hoje você será o nosso doceiro chefe! O pobre rapaz ficou sem saber se ria ou se chorava. Nunca lhe tinham feito um elogio! As palavras do prior encheram seu peito com uma alegria nova, desconhecida. Mas o pior é que não se lembrava da receita inventada por acaso na noite anterior. Era preciso confessar a verdade. E coragem?

Foi o irmão Bernardo quem o socorreu. Primeiro, fez sinal para que ficasse quieto. Depois, a sós, explicou que tinha reparado em todas as voltas que ele deu, em tudo o que usou e que juntos fariam uma nova dose nesse mesmo dia.

– Fui eu quem colocou a canela – confessou. – E ficaram tão deliciosos que só por minha conta já comi seis! Você será famoso, Manuel. Você e os seus pastéis de Belém!

References

Alzaidi, Muhammad S. A., Yi Xu, Anqi Xu, and Marta Szreder. 2023. Analysis and computational modelling of Emirati Arabic intonation: A preliminary study. *Journal of Phonetics* 98: 101236. [CrossRef]

Arvaniti, Amalia, Argyro Katsika, and Na Hu. 2024. Variability, overlap, and cue trading in intonation. *Language* 100: 265–307. [CrossRef]

Barbosa, Plinio A. 1996. At least two macrorhythmic units are necessary for modeling Brazilian Portuguese duration. Paper presented at 1st ESCA Tutorial and Research Workshop on Speech Production Modeling: From Control Strategies to Acoustics & 4th Speech Production Seminar: Models and Data, Autrans, France, May 20–24; pp. 85–88.

Barbosa, Plinio A. 2007. From syntax to acoustic duration: A dynamical model of speech rhythm production. *Speech Communication* 49: 725–42. [CrossRef]

Barbosa, Plinio A. 2008. Prominence- and boundary-related acoustic correlations in Brazilian Portuguese read and spontaneous speech. Paper presented at Speech Prosody 2008 Conference, Campinas, Brazil, May 6–9; pp. 257–60.

Barbosa, Plinio A. 2010. Automatic Duration-Related Salience Detection in Brazilian Portuguese Read and Spontaneous Speech. Paper presented at Speech Prosody, Chicago, IL, USA, May 10–14; p. 100067.

Barbosa, Plinio A. 2020. Prosody Descriptor Extractor [Software Program]. Available online: https://github.com/pabarbosa/prosody-scripts/tree/master/ProsodyDescriptorExtractor (accessed on 5 July 2024).

Barbosa, Plinio A. 2022. Pleasantness and wellbeing in poem declamation in European and Brazilian Portuguese depends mostly on pausing and voice quality. *Frontiers in Communication* 7: 855177. [CrossRef]

Barbosa, Plinio A. 2024. The interplay between syllabic duration and melody to signal prosodic functions in reading and story retelling in Brazilian Portuguese. Paper presented at Speech Prosody 2024 Conference, Leiden, The Netherlands, July 2–5; pp. 1075–79. [CrossRef]

Barbosa, Plinio A., and Philippe B. Mareüil. 2016. Pics mélodiques prétoniques en portugais brésilien: Une étude quantitative. Paper presented at Actes de la Conférence Conjointe JEP-TALN-RECITAL 2016, Paris, France, July 4–8; pp. 527–53.

Botinis, Antonis, Bjorn Granström, and Bernd Möbius. 2001. Developments and paradigms in intonation research. *Speech Communication* 33: 263–96. [CrossRef]

Bruce, Gösta. 1977. *Swedish Word Accents in Sentence Perspective*. Lund: Gleerup.

Bruce, Gösta. 1982. *Developing the Swedish Intonation Model*. Working Papers. Lund: Department of Linguistics and Phonetics, Lund University, p. 22.

Buder, Eugene H., and Anders Eriksson. 1999. Time-series analysis of conversational prosody for the identification of rhythmic units. Paper presented at 14th International Congress of Phonetic Sciences, San Francisco, CA, USA, August 1–7, vol. 2, pp. 1071–74.

Carnaval, Manuela, João A. Moraes, and Albert Rilliard. 2022. Focus types in Brazilian Portuguese: Multimodal production and perception. *DELTA. Documentação de Estudos em Linguística Teórica e Aplicada* 38: 1–34. [CrossRef]

Chistovich, Ludmilla A., and Elena A. Ogorodnikova. 1982. Temporal processing of spectral data in vowel perception. *Speech Communication* 1: 45–54. [CrossRef]

Christodoulides, George. 2018. Acoustic Correlates of Prosodic Boundaries in French. A Review of Corpus Data. *Revista de Estudos da Linguagem* 26: 1531–49. [CrossRef]

Cole, Jennifer, and Stephanie Shattuck-Hufnagel. 2016. New methods for prosodic transcription: Capturing variability as a source of information. *Laboratory Phonology* 7: 8. [CrossRef]

Cole, Jennifer, Yoonsook Mo, and Soondo Baek. 2010. The role of syntactic structure in guiding prosody perception with ordinary listeners and everyday speech. *Language and Cognitive Processes* 25: 1141. [CrossRef]

Dixon, Roger A., and Odette N. Gould. 1996. Adults telling and retelling stories collaboratively. In *Interactive Minds: Life-Span Perspectives on the Social Foundation of Cognition*. Edited by Paul B. Baltes and Ursula M. Staudinger. Cambridge: Cambridge University Press, pp. 221–41.

Dogil, Grzegorz, and Gunter Braun. 1988. *The PIVOT Model of Speech Parsing*. Wien: Verlag.

Doukhan, David, Albert Rilliard, Sophie Rosset, Martine Adda-Decker, and Christophe d'Alessandro. 2011. Prosodic analysis of a corpus of tales. Paper presented at Twelfth Annual Conference of the International Speech Communication Association, Florence, Italy, August 27–31; pp. 3129–32.

Dowdy, Shirley, and Stanley Wearden. 1991. *Statistics for Research*, 2nd ed. New York: John Wiley & Sons.

Fernandes, Norma H. 1976. Contribuição para uma análise instrumental da acentuação e intonação do português. Master's thesis, State University of Sao Paulo, São Paulo, Brazil.

Fujisaki, Hiroya, and Keikichi Hirose. 1984. Analysis of voice fundamental frequency contours for declarative sentences of Japanese. *Journal of the Acoustical Society of Japan* 5: 233–42. [CrossRef]

Gårding, Eva, and Gösta Bruce. 1981. A presentation of the Lund model for-Swedish intonation. In *Nordic Prosody II*. Edited by T. Fretheim. Trondheim: Tapir, pp. 33–39.

Gussenhoven, Carlos, and Antonius C. Rietveld. 1992. Intonation contours, prosodic structure and pre boundary lengthening. *Journal of Phonetics* 20: 283–303. [CrossRef]

Herment-Dujardin, Sophie, and Daniel J. Hirst. 2002. Emphasis in English: A perceptual study based on manipulated synthetic speech. Paper presented at Speech Prosody 2002, Aix-en-Provence, France, April 11–13; pp. 379–82.

Hirst, Daniel J. 2005. Form and function in the representation of speech prosody. *Speech Communication* 46: 334–47. [CrossRef]

Holzgrefe, Julia, Caroline Schröder, Barbara Höhle, and Isabell Wartenburger. 2012. Processing of prosodic boundary cues as revealed by event-related brain potentials. Paper presented at the 13th Conference on Laboratory Phonology (LabPhon 13), Stuttgart, Germany, July 27–29.

Jun, Sun-Ah. 2014. Prosodic typology: By prominence type, word prosody, and macro-rhythm. In *Prosodic Typology II*. Oxford: Oxford University Press, pp. 520–39.

Leemann, Adrian, Marie-José Kolly, Yang Li, Ricky Chan, Geraldine Kwek, and Anna Jespersen. 2016. Towards a typology of prominence perception: The role of duration. Paper presented at 8th International Conference on Speech Prosody (SP2016), Boston, MA, USA, May 31–June 3; pp. 445–49.

Lucente, Luciana. 2012. Aspectos dinâmicos da fala e da entoação do português brasileiro. Ph.D. thesis, Universidade Estadual de Campinas, Campinas, Brazil.

Madureira, Sandra. 1994. Pitch Patterns in Brazilian Portuguese: An Acoustic-Phonetic Analysis. Paper presented at Fifth Australian International Conference on Speech Science and Technology, Perth, Australia, December 6–8; pp. 156–59.

Madureira, Sandra. 2016. Intonation and variation: The multiplicity of forms and senses. *Dialectologia: Revista Electrònica* Special Issue VI: 57–74.

Madureira, Sandra, and Mário A. Fontes. 1997. Fundamental contours in Brazilian Portuguese words. Paper presented at Intonation: Theories, Models and Applications, Athens, Greece, September 18–20; pp. 211–14.

Massini, Gladis. 1991. A duração no estudo do acento e do ritmo em português. Master's thesis, State University of Campinas, Campinas, Brazil.

Miranda, Luma S., João A. Moraes, and Albert Rilliard. 2022. Effects of F0 movements, intensity, and duration in the perceptual identification of Brazilian Portuguese wh-questions and wh-exclamations. *DELTA. Documentação de Estudos em Linguística Teórica e Aplicada* 38: 31–29. [CrossRef]

Miranda, Luma S., Marc Swerts, João A. Moraes, and Albert Rilliard. 2021. The Role of the Auditory and Visual Modalities in the Perceptual Identification of Brazilian Portuguese Statements and Echo Questions. *Language and Speech* 64: 3–23. [CrossRef]

Mittmann, Maryualê M., and Plinio A. Barbosa. 2016. An automatic speech segmentation tool based on multiple acoustic parameters. *CHIMERA. Romance Corpora and Linguistic Studies* 3: 133–44. [CrossRef]

Moraes, João A. 1998. Intonation in Brazilian Portuguese. In *Intonational Systems: A Survey of Twenty Languages*. Edited by Daniel Hirst and Albert Di Cristo. Cambridge: MIT Press.

Moraes, João A. 2008. The pitch accents in Brazilian Portuguese: Analysis by synthesis. Paper presented at Speech Prosody 2008, Campinas, Brazil, May 6–9; pp. 389–97.

Oliveira, Miguel, Jr. 2012. A study on speech rate as a prosodic feature in spontaneous narrative. *Alfa: Revista de Linguística* 56: 623–51. [CrossRef]

Perkell, Joseph S., Majid Zandipour, Melanie L. Matthies, and Harlan Lane. 2002. Economy of effort in different speaking conditions. I. A preliminary study of intersubject differences and modeling issues. *The Journal of the Acoustical Society of America* 112: 1627–41. [CrossRef]

Pompino-Marschall, Bernd. 1991. The syllable as a prosodic unit and the so-called P-centre effect. *Forschungsberichte Institut für Phonetik und Sprachliche Kommunikation der Universität München* 29: 66–124.

Pratt, Michael W., Cheryl Boyes, Susan Robins, and Judy Manchester. 1989. Telling tales: Aging, working memory, and the narrative cohesion of story retellings. *Developmental Psychology* 25: 628. [CrossRef]

R Project. n.d. The R Foundation for Statistical Computing. Available online: http://www.r-project.org/ (accessed on 5 July 2024).

Ramli, Izzad, Noirani Seman, Norizah Ardi, and Nursuriati Jamil. 2016. Rule-based storytelling text-to-speech (TTS) synthesis. *MATEC Web of Conferences, EDP Sciences* 77: 04003. [CrossRef]

Raso, Tommaso, and Heliana Mello. 2012. The C-ORAL-BRASIL I: Reference corpus for informal Brazilian Portuguese. In *Lecture Notes in Computer Science*. Berlin and Heidelberg: Springer, pp. 362–67.

Sanderman, Angelien A., and René Collier. 1995. Prosodic phrasing at the sentence level. In *Producing Speech: Contemporary Issues. For Katherine Safford Harris*. Melville: American Institute of Physic, pp. 321–32.

Silverman, Kim, Mary Beckman, John Pitrelli, Mari Ostendorf, Colin Wightman, Patti Price, Janet Pierrehumbert, and Julia Hirschberg. 1992. ToBI: A standard for labelling English prosody. Paper presented at International Congress of Spoken Language Processing, Banff, AB, Canada, October 13–16; pp. 867–70.

Skehan, Peter, and Pauline Foster. 1999. The influence of task structure and processing conditions on narrative retellings. *Language Learning* 49: 93–120. [CrossRef]

Streefkerk, Barbertje M. 1997. Acoustical correlates of prominence: A design for research. *Proceedings of the Institute of Phonetic Sciences of the University of Amsterdam* 21: 131–42.

t'Hart, Johan. 1984. A Phonetic Approach to Intonation: From Pitch Contours to Intonation Patterns. In *Intonation, Accent and Rhythm: Studies in Discourse Phonology*. Berlin and New York: De Gruyter, pp. 193–202.

Teixeira, Barbara, Plinio A. Barbosa, and Tommaso Raso. 2018. Automatic detection of prosodic boundaries in Brazilian Portuguese spontaneous speech. In *International Conference on Computational Processing of the Portuguese Language*. Cham: Springer International Publishing, pp. 429–37.

van Santen, Jan. P., and Bernd Möbius. 2000. A quantitative model of F0 generation and alignment. In *Intonation: Analysis, Modelling and Technology*. Dordrecht: Springer Netherlands, pp. 269–88.

Verma, Rashmi, Parakrant Sarkar, and Krothapalli S. Rao. 2015. Conversion of neutral speech to storytelling style speech. Paper presented at 2015 Eighth IEEE International Conference on Advances in Pattern Recognition, Kolkata, India, January 4–7; pp. 1–6.

Wagner, Petra, and Simon Betz. 2023. Effects of Meter, Genre and Experience on Pausing, Lengthening and Prosodic Phrasing in German Poetry Reading. Paper presented at Interspeech 2023, Dublin, Ireland, August 20–24; pp. 2538–42.

Wehrle, Simon, Francesco Cangemi, Harriet Hanekamp, Kai Vogeley, and Martine Grice. 2020. Assessing the intonation style of speakers with autism spectrum disorder. Paper presented at 10th International Conference on Speech Prosody, Tokyo, Japan, May 25–28; pp. 809–13.

Wellmann, Caroline, Julia Holzgrefe, Hubert Truckenbrodt, Isabell Wartenburger, and Barbara Höhle. 2012. How each prosodic boundary cue matters: Evidence from German infants. *Frontiers in Psychology* 3: 33536. [CrossRef] [PubMed]

Wightman, Colin. W. 2002. ToBI or not ToBI? Paper presented at Speech Prosody 2002, Aix-en-Provence, France, April 11–13; pp. 25–29.

Xu, Yi. 2009. Timing and coordination in tone and intonation: An articulatory-functional perspective. *Lingua* 19: 906–27. [CrossRef]

Yoon, Tae-Jin. 2014. Speaker variation in English prosodic boundary. *Linguistic Research* 31: 1–23.