

# The role of alignment and height in the perception of LH contours

Luciana Lucente and Plinio A. Barbosa

Department of Linguistics, Speech Prosody Studies Group,  
State University of Campinas, Brazil

luciana.unicamp@gmail.com, pabarbosa.unicampbr@gmail.com

## ABSTRACT

The DaTo system intends to describe the Brazilian Portuguese intonation using a set of labels to represent dynamic intonational contours. To evaluate what kind of physical parameters are involved in the production of emphasis, a set of perception tests were designed to compare the dynamic contours LH (rising) and >LH (late rising). These contours present the same melodic shape, but differ on alignment and pitch height. The results of three perception tests show that alignment and height cooperate as a dynamical system to produce the function of emphasis.

## 1. Introduction

This work is part of a set of perception tests aiming at evaluating different degrees of emphasis in Brazilian Portuguese (henceforth BP) intonation. BP intonation was annotated according to the DaTo system - *Dynamic Tones of Brazilian Portuguese Intonation* [1] [2]. The DaTo system describes BP intonation by a set of dynamical contours, defined as events that occur along the melodic curve which carry pragmatic and communicative information [2]. These dynamical contours describe six melodic shapes for narrow focus, each one differing in (i) alignment, (ii) type of  $f_0$  movement and (iii) degree of emphasis. Focus in DaTo system is defined as an emphasis in some part of a sentence motivated by a particular discourse situation [3].

The DaTo annotation system uses a set of six dynamical contours (LH, >LH, HLH, HL, >HL and LHL) used for prominence function; and two level contours (L and H), that can be used both for prominence or boundary function.

A previous study evaluated differences between LHL (rising-falling) and HL (falling) contours, which are similar in their shape, but differ in range and alignment [2]. In this paper, a perception test was designed in such a way as to evaluate the degree of emphasis between LH (rising) and >LH (late rising).

In the test reported in [2] the LHL contour in sentence final position was first converted into a HL contour just by changing its original alignment, then converted into a HL contour by changing its alignment and height. The main interest in this test was to analyze the perception of emphasis comparing contours in three situations: LHL (original) and HL from LHL, by changing alignment (HL; alignment); LHL (original) and HL from LHL, by changing alignment and height (HL; alignment + height); HL (alignment) and HL (alignment + height).

The test results showed that the subjects were sensitive to changes in  $f_0$  alignment, but were not so sensitive to changes in  $f_0$  height, which confirms the

importance of the alignment to the perception of emphasis reported in [4], [5] and [6]. The experiment also showed that the lexicon influences the perception of degree of emphasis [2].

Because alignment seems to be a crucial parameter for emphasis assessment, a set of perception tests are proposed here to verify possible differences on emphasis between the LH and >LH contours ( $f_0$  alignment), adding as an experimental condition the judgment of differences on height.

## 2. Alignment, height and degree of emphasis of LH contours

The LH and >LH contours are the most common contours used to signal narrow focus in BP spontaneous speech. It is possible to observe these contours in initial and medial position of declarative utterances and in final position of interrogative utterances. The rising contour pattern (LH) consists of an  $f_0$  falling in a pre-stressed syllable followed by an  $f_0$  rising whose peak is aligned with the middle of the stressed vowel. The same pattern of movement can be observed on late rising contour (>LH), however the peak is aligned later in the stressed vowel, normally during a consonantal interval onset. Both contours >LH and LH can be seen respectively on Fig.1, aligned with the stressed vowel /e/ of the two repetitions of the word “mesmo” (even in this context).

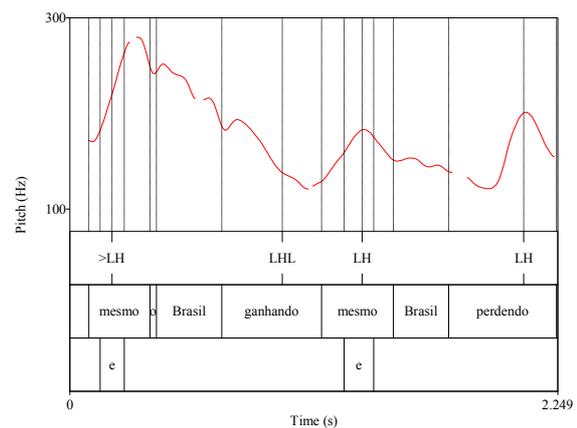


Fig. 1: Examples of >LH and LH contours aligned with two repetitions of the word “mesmo”.

As in the ToBI system framework [7], the DaTo system proposes some annotation sessions to define and discuss the labels attribution. During the sessions on BP intonation annotation the >LH contour was observed as more emphatic than the LH contour.

This observation, together with the observation of differences in form and function of HL and LHL contours, motivated the application of additional perception tests.

Another reason which motivated these tests was the consistent observation of a set of utterances produced by some announcers submitted to speech therapy to improve their fluency [8]. These announcers produced a narrative which was recorded before the therapy, and again the same narrative recorded after the therapy. It is possible to see in these narratives that they changed the LH contour to a >LH one, and increased its height, adding more emphasis in the same words in the narrative after the therapy (see Figs. 2 and 3)

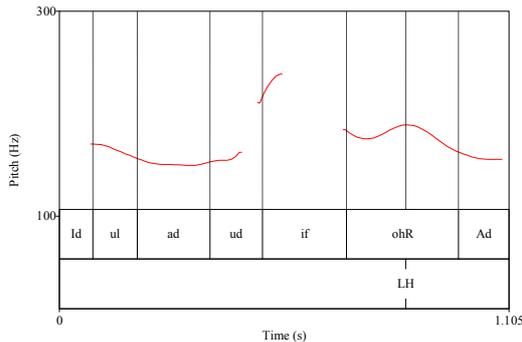


Fig. 2: Example of the narrative before therapy, with the contour LH aligned with the word “fora”(outside).

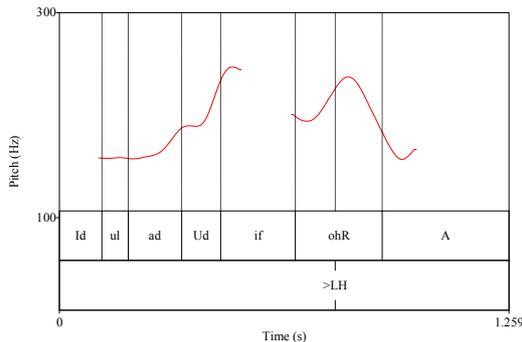


Fig. 3: Example of the narrative after therapy, now with the contour >LH aligned with the same word “fora”(outside).

According to a dynamical system view [9] [10] [11], the cause of the perception of a higher emphasis can lay on alignment changes. This view considers that intonation carries communicative functions which are produced basically by the articulatory system and encoded by f0. According to this paradigm, intonation is conveyed by physical (phonetic) components that signal communicative aspects of speech. For this reason, when intonation phenomena, such as focus, are considered for analysis, the mechanisms involved in producing f0 and the mechanisms involved in articulation of vowels and consonants cannot be analyzed separately.

The working question we want to answer is: is the f0 height and the f0 delay on alignment part of a dynamic system that act together to produce the effect of emphasis, or they are part of distinct systems and each one function independently of the other?

The first experiment with LHL and HL contours showed that height is not so important to the perception of emphasis; however, in this new set of tests the utterances

were manipulated in both their alignments and heights to reevaluate the perception of emphasis.

### 3. Perception experiment

To evaluate the perception of emphasis on rising contours three MFC (Multiple Forced Choice) tests were proposed on Praat software package [12]. The first test evaluated the perception of changes in alignment, the second, changes in height and the third, changes in alignment and height.

For the first test ten utterances containing examples of LH and >LH contours were used. All the utterances were produced by the same female speaker from a spontaneous speech corpus. This corpus was obtained from radio broadcasts and follows an international tendency to use spontaneous speech on speech research, especially in intonation area.

The examples containing >LH contours were modified by the “manipulation” command on Praat and changed into LH contour, and vice-versa.

The manipulation processes consisted in generating two manipulated utterances (M<sup>1</sup> and M<sup>2</sup>) from each one of the ten original (O) records. In the cases of LH contours, the utterances were generated by delaying the f0 peak in 50 ms and 100 ms from the original position, and in cases of >LH the f0 peak was advanced also by 50 ms and 100 ms using the same method, as can be seen in Fig. 4.

These 30 utterances (10 originals + 20 manipulated) were inserted in the MFC experiment for judgment according to the combination below, where “O” is the original sentence, “M<sup>1</sup>” is the first manipulation and “M<sup>2</sup>” is the second manipulation.

- O - M<sup>1</sup>
- O - M<sup>2</sup>
- M<sup>2</sup> - M<sup>2</sup>

This combination applied to each original utterance resulted in 30 stimuli (3 presentations of 10 sentences).

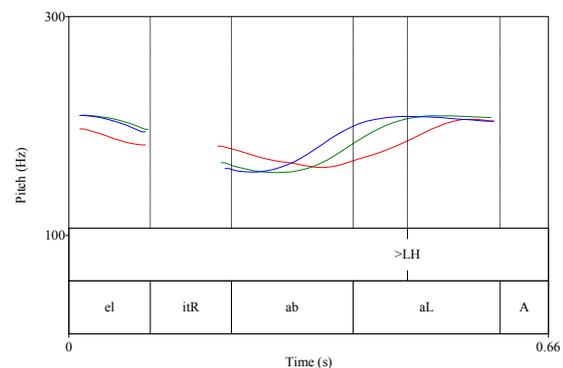


Fig. 4: Original contour (>LH) in red; first modification in green with peak displaced 50 ms from original position (LH); in blue, the second modification with peak displaced 100 ms from original position (LH).

The records of each combination of two utterances were presented to the subjects through headphones. The same utterances were presented simultaneously in written form on a computer screen with the emphatic word in capital letters. The subjects were instructed to judge if the second record was more emphatic, less emphatic or equal

when compared to the first record. For this, they had to click on the respective button at the computer screen, as we can see on Fig. 5 below.



Fig. 5: Example of MFC test on Praat.

Fifteen (five men and ten women) subjects, naïve to the aims of the experiment, participated. All of them are graduate students in their twenties.

The second test presented the same ten original utterances, and consisted in manipulations of f0 height of LH and >LH contours. As in the first test, this test also generated two utterances with manipulation in their heights: the first manipulation (M<sup>1</sup>) with an increase of approximately 1.6 semitones in height, and the second manipulation (M<sup>2</sup>) an increase of approximately 3.5 semitones from the original peak height (see Fig. 6). These semitones scale were defined as such in order to avoid distortion in the manipulation process.

In this second test fifteen (seven men and eight women) different subjects naïve to the aims of the experiment participated, also graduate students in their twenties.

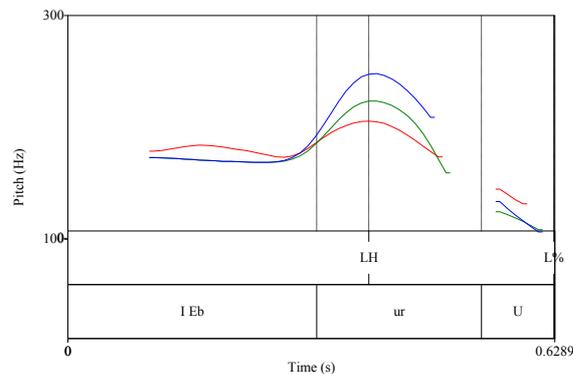


Fig. 7: in red the original contour (LH); in green the first modification with peak increase in approximately 1,6st from original position (LH); in blue the second modification with peak increase in approximately 3.5 st from original position (LH).

The third test was realized following the same protocol of the two preceding tests: a set of manipulations was carried out from the original utterance to evaluate the perception of emphasis. However, in this test, four manipulations were carried out: (i) the first manipulation (M<sup>1</sup>) consisted in changing just the alignment of the original contour peak (O). This change was exclusively a delay of 50 ms from the original alignment between f0

peak and stressed vowel; (ii) the second manipulation (M<sup>2</sup>) consisted in an increase of approximately 2 semitones in the height of the first manipulated contour (M<sup>1</sup>); (iii) the third manipulation (M<sup>3</sup>) consisted in a delay of 100 ms from the original peak (O) without a change in height, in the same way that was made with the first manipulation, (iv) and finally, the fourth manipulation (M<sup>4</sup>), consisted in an increase of approximately 2 semitones from the third manipulation (M<sup>3</sup>). The utterances were inserted in the MFC test according to this combination:

- O - M<sup>1</sup>
- O - M<sup>2</sup>
- O - M<sup>3</sup>
- O - M<sup>4</sup>

This third test was conceived to evaluate the alignment and height of the dynamic contours LH and >LH at the same time. The four manipulated contours can be seen in Fig. 8 below.

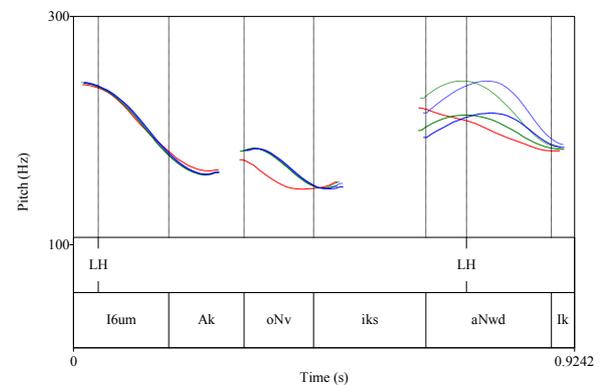


Fig. 8: in red the original utterance (O); in green the first manipulation (M<sup>1</sup>) with a peak delay of 50ms; in dotted green the second manipulation (M<sup>2</sup>) with a increase of approximately 2st from the first manipulation; in blue the third manipulation (M<sup>3</sup>) with a peak delay of 100ms from the original; and in dotted blue the fourth manipulation (M<sup>4</sup>) with a height increase of approximately 2st from the third manipulation.

## 4. Results

The reported experiments were proposed to determine the extent of agreement which can be expected on the basis of chance, and the kappa test was used as an index to measure the level of agreement among the subjects.

The first test results showed a tendency to judge the sentences as being functionally equivalent, showing the following values:  $k = 0.165$ ;  $z = 12.73$ . These values show a significant, but slight agreement for the Kappa test [13]. The coefficient  $k$  also gets very small values when the marginal frequencies differ considerably. In these cases we cannot distinguish between “agreement on the frequency of category use” from “inter-rater agreement” [13].

However, the direct observation of the results shows some combination of utterances O – M<sup>1</sup>, and O – M<sup>2</sup> judged as more emphatic. This seems to indicate that the subjects are not certain about the judgment, which suggests that alignment is not a strong parameter for this judgment.

The results for the second test also showed a tendency by the subjects to judge the sentences as being

functionally equivalent, with values  $k= 0.04$  and  $z= 2.69$ . Although the results point to a judgment that considers the utterances as similar, this slight agreement can be explained due to a tendency by the subjects to judge some utterances as more emphatic than the original. As for alignment criteria, the height alone seems not to be a strong parameter to judge prominence.

The third test shows a kappa agreement coefficient 0.544, with  $z= 11.77$ . This number shows a better agreement between the subjects whose judgment was equal for the combination between O – M<sup>1</sup> and O – M<sup>3</sup>, which changes on alignment alone, and more emphatic for the combination between O – M<sup>2</sup> and O – M<sup>4</sup>, which presents both changes on alignment and height. A better number of agreements were found between O – M<sup>4</sup> combination than in O – M<sup>2</sup> combination. This difference can occur due to the higher delay on alignment (100 ms) from the original peak combined with an increase in height.

## 5. Discussion

For a discussion it is interesting to remind the result of the test with LHL and HL contours [3]. When the LHL contour was converted to a HL contour, the perception of change in the nature of focus is related to alignment. In the three current tests with LH and >LH contours there is no perception of change based on differences in alignment or height.

These differences in the results could be explained if we consider that in the preceding experiment the LHL contour was converted in a different contour: HL. In this case, for a perception of change, just the alignment parameter was sufficient, because the experiment was not testing the degree of emphasis on the same contour as in the present case. For the rising contours LH and >LH the three experiments evaluated differences on emphasis with a same shape of f0 movement. For this reason the emphasis was tested on three different situations.

As asked at the beginning of this paper, the perception of emphasis in some part of an utterance is possible due to an intrinsic cooperation of two different actions: (i) alignment between intonational contour and stressed syllable and (ii) pitch range adjustments.

The observation of the results reported points to the hypothesis that the perception of emphasis based just in changes of alignment or height is weak. Otherwise, the perception of emphasis is clear in situations in which there is cooperation between alignment and height.

Another evidence for this hypothesis is that the utterances where just alignment or height was manipulated were very difficult to judge. One possible reason for this phenomenon can be the absence of a one-dimensional change of contour in natural speech. The manipulation that changes alignment and height at the same time seems more natural for the subjects (as the examples in Figs. 1 and 2) which corroborate with the result of the third test.

## 6. Acknowledgments

This work was supported by State of São Paulo Research Foundation (FAPESP), process number 08/50146-3 The second author acknowledges a grant from National Council of Technological and Scientific Development (CNPq), 300371/2008-0.

## 7. References

- [1]Lucente, L., Barbosa, P. A. “Narrow focus in Brazilian Portuguese: spatial and temporal constraints”, *Proceedings of Fourth Conference on Speech Prosody*. Campinas, pp. 281-284, May 2008.
- [2]Lucente, L., Barbosa, P. A. “Perception Experiment of Intonation in Brazilian Portuguese”, *Abstracts of Fourth PaPI Conference*. Las Palmas, June 2009.
- [3]X Y., Xu, C. X. “Phonetic realization of focus in English declarative intonation”, *Journal of Phonetics* 33, pp. 59-197, 2005.
- [4]Kohler, K. J. “Timing and Communicative Functions of Pitch Contours”, *Phonetica* 62, pp. 88-105, 2005.
- [5]House, D. “Perception of Tone with particular reference to temporal alignment”, *Proceedings of the Symposium Cross-Linguistic Studies of Tonal Phenomena*. Tokio, pp. 203-219, 2003
- [6]Pierrehumbert, J. “Tonal elements and their alignment”, In M. Horne (ed) *Prosody: Theory and Experiment*. Studies Presented to Gosta Bruce, Kluwer, Dordrecht, pp. 11-26, 2000.
- [7] Beckman, M. E., Hirschberg, J., Pitrelli, John F., (1994). *Evaluation of Prosodic Transcription Labeling Reliability in the ToBI Framework*, from [http://www.ling.ohio-state.edu/~tobi/ame\\_tobi](http://www.ling.ohio-state.edu/~tobi/ame_tobi)
- [8] Constantini, A. C; Barbosa, P. A. *Intonation Features in Journalistic Broadcasting*, 2009 (not published)
- [9]Kelso, S. *Dynamic Patterns*. Cambridge: MIT Press, 1995
- [10] Barbosa, P. A. *Incursões em torno do ritmo da fala*. Campinas: Pontes Editores, São Paulo:Fapesp, 2006.
- [11] Xu, Y. “Speech melody as articulatorily implemented communicative functions”. *Speech Communication* 46, pp. 220-251, 2005
- [12] Boersma, P., Weenink, D. “Praat: doing phonetics by computer” (Version 5.1.05) Retrieved May 1, 2009, from <http://www.praat.org/>
- [13] Rietveld, T., van Hout, R. *Statistical Techniques for the Study of Language and Language Behaviour*. Berlin-New York, Mouton de Gruyter, 1993.