ATTITUDINAL PATTERNS IN BRAZILIAN PORTUGUESE INTONATION: ANALYSIS AND SYNTHESIS

João Antônio de Moraes & Cirineu Cecote Stein

Laboratório de Fonética Acústica
Universidade Federal do Rio de Janeiro, Brazil
jamoraes@gbl.com.br; cirineustein@uol.com.br

Abstract

The main goal of this paper is to investigate the prosodic manifestation of the following attitudinal states: consideration, despair, disappointment, irony, justification, obviousness, and uncertainty. The sentence *O Carlos Alberto já sabe*. [Carlos Alberto already knows it.] was pronounced by a subject, who tried to convey each of these attitudes. Afterwards, it was presented to 20 panelists, which were asked to identify the original intention of each enunciation. The test results showed that the attitudes were, in general, correctly identified. The acoustic analysis revealed that the attitudinal patterns make use of distinct prosodic parameters in their manifestation: some are linked to segmental duration, be it global or localized; in other cases, the decisive prosodic component is the fundamental frequency. Auditory tests using speech resynthesis turned it possible to evaluate the relative weight of the prosodic characteristics identified in the analysis.
KEYWORDS: prosody, intonation, attitudinal meaning

1. Introduction

The study of the expression of emotions and attitudes in speech, started in the first decades of the 20th century [4], still concerns linguists and psychologists directly [1, 3, 5, 6, 7, 8]. Since the 1960's on, two groups of affective states are normally distinguished: the raw emotions, such as fear and joy; and the stylized emotions, socially "filtered", such as irony and justification, which have been sometimes called attitudes [2, 4]. In this paper, we will deal with the expression of some attitudes of this second group. In addition to the proper description of the patterns, we are interested in investigating the relative weight of the prosodic parameters involved in the expression of the different attitudes, as well as observing if its manifestation is global, that is, if it extends through the sentence as a whole; or otherwise punctual, located in specific privileged portions of the enunciation, such as the nuclear stress. This would bring these attitudes closer to the sentence modalities, and would make their patterns more easily liable to receive a phonological description.

2. Method

The sentence *O Carlos Alberto já sabe*. [Carlos Alberto already knows it.] was pronounced by an adult female speaker, born in Rio de Janeiro, experienced in phonetics, with 13 distinct intonational patterns, in a way to express the following attitudinal states: consideration, critics, despair, disappointment, disbelief, disdain, drunkenness, indifference, irony, justification, meditation, obviousness, and uncertainty, besides the "neuter" version of the sentence.

The 13 versions of the sentence were recorded and submitted to the analysis of a panel of 20 volunteer undergraduates, all of them native speakers of Brazilian Portuguese, for Test 1. After the listening of the sentences, the panelists should indicate, for each of them, one of 13 options, corresponding to the 13 attitudinal states.

Based on the observation of the contours relative to the seven patterns better recognized by the panelists, 39 versions of the sentence were produced, using speech resynthesis. Some characteristics of fundamental frequency and duration were modified in a controlled way, with the intention of gradually changing the "neuter" enunciation into each of the attitudinal versions. The computer program WinPitch v.1.92 was used.

A second auditory test – Test 2 – was applied to a panel of another 20 undergraduates, which allowed an evaluation of the effects of each modification implemented with resynthesis.

3. Results

3.1. Analysis of Test 1

The results of the auditory test showed that the rate of recognition of seven attitudes – consideration, despair, disappointment, irony, justification, obviousness, and uncertainty – was equal or superior to 45%, about six times superior to the expected index in a chance analysis, which would be just 7.69%. Although "drunkenness" has reached a recognition index of 100%, we decided not to consider it among the results presented here, because it cannot be considered a proper "attitude". The pitch curves related to the "neuter" assertion and those seven better recognized attitudes can be seen in figures 1 to 8. Their main prosodic characteristics are summarized in Table 1.

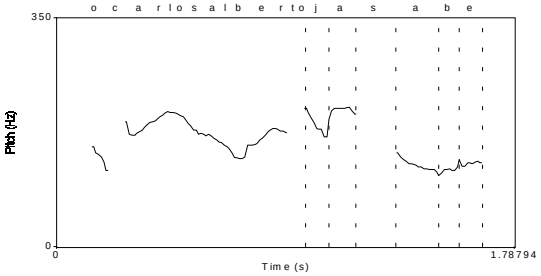


Figure 1: *Melodic contour of the "neuter" assertion*

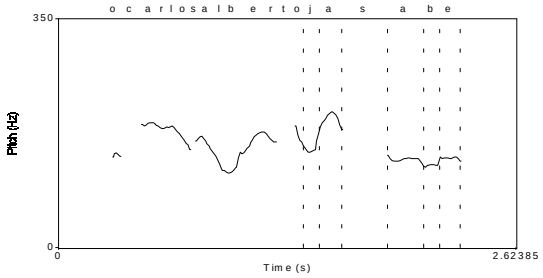


Figure 2: *Melodic contour of the "consideration" assertion*

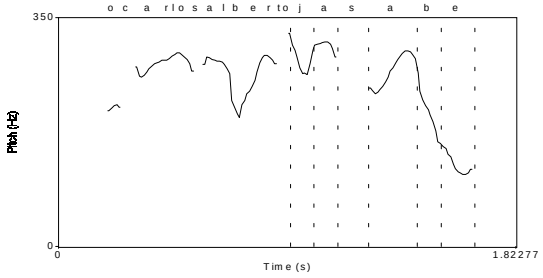


Figure 3: *Melodic contour of the "despair" assertion*

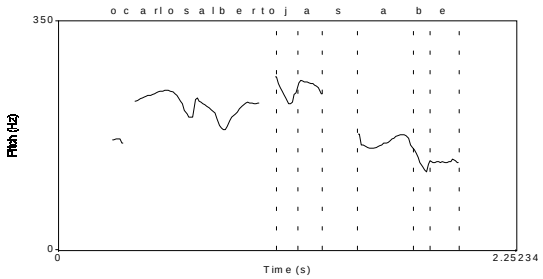


Figure 4: *Melodic contour of the "disappointment" assertion*

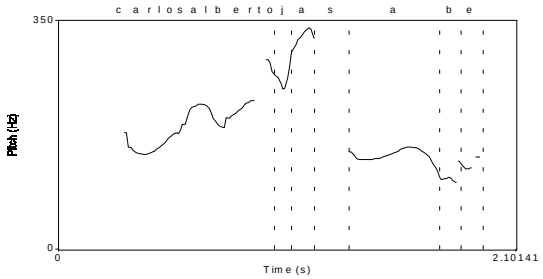


Figure 5: *Melodic contour of the irony" assertion*

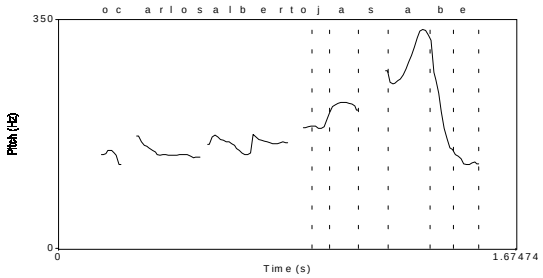


Figure 6: *Melodic contour of the "justification" assertion*

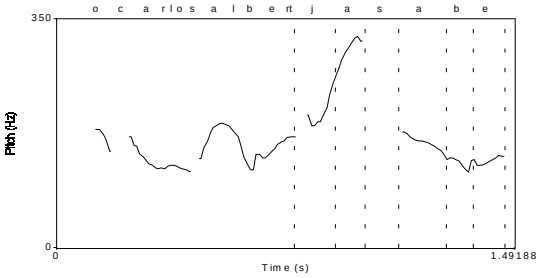


Figure 7: *Melodic contour of the "obviousness" assertion*

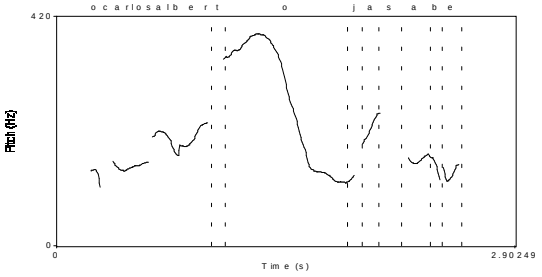


Figure 8: *Melodic contour of the "uncertainty" assertion*

The patterns of despair, uncertainty, irony, justification, and obviousness, present, at the global level, an important FO variability. The first two also show an average melodic level far superior to the other patterns. Except for the melodic patterns of disappointment and consideration, the other attitudes are distinguished by a suppression of the declination line, typical of the "neuter" enunciation. At the local level, two major patterns are observed: one with the final stressed syllable in a high level and (predominantly) with a rising configuration (despair, justification); the other presenting the final stressed syllable in a low level, with a falling configuration (neuter, disappointment, consideration, obviousness), or a slightly rising one (irony). "Obviousness" and "irony" are also characterized by presenting a pre-stressed syllable in a high level. Concerning the behavior of duration, the patterns of "uncertainty" and "consideration" present a global duration clearly superior to the other ones. At the local level there is a final lengthening, particularly observed in the last stressed vowel, in "disappointment" and overall, in "irony".

3.2. Analysis of Resynthesis Processes and Test 2

Based on the observation of the prosodic behavior of the attitudes with a good index of recognition (equal or superior to 45%, in Test 1), we established the prosodic changes to be implemented during the resynthesis of the "neuter" sentence. The parameters of FO, duration, or both, were manipulated, thus creating, in this way, 39 new variants. These changes could be established globally, occurring in the whole sentence (or in most of it), or locally, changing prosody in specific points of the sentence, especially the final stressed and pre-stressed syllables.

With the 39 resynthesized variants, a second recognition test was performed, involving, as in the first test, another group of 20 volunteer undergraduates. In the second test, the aim was to measure the groups of stimuli created by successive modifications of the "neuter" assertion to simulate each of the

attitudinal patterns that were identified as related to those attitudes. The stimuli of each group, for example the eleven stimuli simulating the attitude of irony, were presented, and the panelists were asked to judge their quality concerning the focused attitude, attributing one of 5 possible degrees, from "0" (no similarity with the attitude) to "4" (maximum similarity). The average of the degrees attributed to each stimulus by the panelists was considered as an index of judgment to the proposed pattern, i.e., as the index approaches to "4" the bigger the success of the resynthesis.
Following, we present the stimuli based on the "neuter" pattern, and the evaluation performed by the panelists. The indexes from "0" to "4", between square brackets, following each stimulus, correspond to the average degree obtained by it. The FO and duration values were chosen considering the prosodic behavior observed for the natural assertions analyzed.

3.2.1. Consideration

(i) rising-falling FO modulation in the final stressed syllable "sa", starting at 125Hz, rising to 135Hz in its middle, and falling to 119Hz at its end [1.3];
(ii) duration increase in the syllable "sa", in 31% [1.95];
(iii) total sentence duration increase, in 31% [2.6];
(iv) association of (i) and (ii) [2.3];
(v) association of (i) and (iii) [2.5].

3.2.2. Despair

(i) FO transposition to a higher register (increase of 100Hz) from the beginning up to the final pre-stressed syllable [0.8];
(ii) rising-falling FO modulation in the final stressed and post-stressed syllables, starting at 231Hz in the stressed vowel, peak at 294Hz, and falling to 173Hz in the post-stressed vowel [1.0];
(iii) F0 transposition of the whole sentence to a sharper register (increase of 100Hz) [0.8];
(iv) association of (i) and (ii) [2.2].

3.2.3. Disappointment

(i) FO transposition to a higher register (increase of 80Hz) from the beginning up to the final pre-stressed syllable [0.55];
(ii) rising-falling FO modulation in the final stressed and post-stressed syllables, starting at 148Hz in the stressed vowel, peak at 161Hz, and falling to 114Hz in the post-stressed vowel [1.8];
(iii) FO transposition of the whole sentence to a higher register (increase of 80Hz) [1.45];
(iv) association of (i) and (ii) [1.0];
(v) duration increase of 31% in the word "sabe" [2.05];
(vi) association of (i) and (v) [1.5];
(vii) association of (ii) and (v) [1.9];
(viii) association of (iii) and (v) [2.3];
(ix) association of (i), (ii) and (v) [3.0].

3.2.4. Irony

(i) FO transpostion to a higher register (increase of 50Hz) in the final stressed syllable [0.8];
(ii) rising-falling Fo modulation in the final pre-stressed syllable, starting at 207Hz, peak at 323Hz, and falling to 142Hz at its end [1.3];
(iii) FO transposition to a higher register (increase of 50Hz) in the word "alberto" [0.45];
(iv) association of (i) and (ii) [1.05];
(v) association of (i), (ii) and (iii) [0.85];
(vi) increase of the final stressed syllable duration in 81% [2.35];
(vii) association of (i) and (vi) [2.9];
(viii) association of (ii) and (vi) [2.9];
(ix) association of (iii) and (vi) [2.5];
(x) association of (i), (ii) and (vi) [2.9];
(xi) association of (i), (ii), (iii) and (vi), with FO modulation in the syllable "al", in an ascending-descending format, attacking at 215Hz, with peak at 219Hz, and minimum point at 202Hz [3.2].

3.2.5. Justification

(i) rising-falling FO modulation in the word "sabe", starting at 224Hz, peak at 326Hz at the end of the stressed vowel, falling to 129Hz, at the post-stressed vowel [3.65];
(ii) association of (i) and FO transposition to a lower register (decrease of 20Hz) in "carlos" [3.4].

3.2.6. Obviousness

(i) rising-falling FO modulation in the pre-stressed syllable "já", starting at 192Hz, peak at 313Hz, ending at 247Hz [1.85];
(ii) FO transposition to a slightly higher register (increase of 40Hz) in "sa" [1.7];
(iii) association of (i) and (iii) [3.0];
(iv) association of (iii) and FO modulation in the word "carlos", starting at 161Hz, ending at 138Hz; and a rising-falling F0 modulation in the syllable "s al", starting at 139Hz, peak at 200Hz, ending at 165Hz [3.6].

3.2.7. Uncertainty

Because of the restrictions imposed by the computer program version used for the production of the resyntheses, it was not possible to octuplicate the duration of the syllable "to", to simulate the pattern of uncertainty originally recognized as such (Test 1). Nonetheless, we developed the four following resyntheses, and submitted them to analysis:
(i) Fo modulation in the word "sabe", in a falling general configuration, with peak in 312Hz at the beginning of the syllable, falling to 190Hz in the middle, rising to 208Hz at its end, and a final fall to 165Hz in the post-stressed syllable [0.15];
(ii) association of (i) with FO modulation in "alberto ja", starting at 226Hz, falling to 179Hz at the end of the syllable "al", rising to 312Hz at the end of the syllable "ja" [0.5];
(iii) association of (ii) with FO modulation in "to", in a falling configuration, starting at 265Hz and ending at 222Hz [0.75];
(iv) association of (iii) with FO peak elevation in the syllable "to" to 386Hz [1.4].

4. Conclusions

Based on the observation of the prosodic patterns investigated, we concluded that the expression of the attitudes is revealed in different ways in Brazilian Portuguese, both in terms of its domain, which can include the whole sentence or just one of its parts, and in terms of its prosodic parameters. Thus, attitudes such as despair and consideration reveal global manifestations, while justification and obviousness are located in portions of the enunciation, more specifically in the nuclear stress. Additionally, while "despair", "justification", and

"obviousness" are expressed mainly by pitch, "consideration" is expressed by the duration pattern, while "irony", and "disappointment", by a combination of pitch and duration.

5. References

[1] Banse, R.; Scherer, K., 1996. Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology*, vol. 70 (3): 614-636.

[2] Fónagy, I., 2000. *Languages within Language: an evolutive approach.* Amsterdam: John Benjamins.

[3] Ladd, D. R.; Scherer, K.; Silverman, K., 1986. An integrated approach to studying intonation and attitude. In *Intonation in Discourse*, Johns-Lewis, C. (ed.). London: Croom Helm, 125-138.

[4] Léon, P., 1976. De l'analyse psychologique à la catégorisation auditive et acoustique des émotions dans la parole. *Journal de Psychologie*, vol. 3-4: 305-323.

[5] Léon, P., 1993. *Précis de phonostylistique*. Paris: Nathan.

[6] Scherer, K.; Ladd, D. R.; Silverman, K., 1984. Vocal cues to speaker affect: testing two models. *Journal of the Acoustical Society of America*, 76 (5): 1346-1356.

[7] Scherer, K.; Bänziger, T., 2004. Emotional expression in prosody: a review and an agenda for future research. *Speech prosody 2004*, Nara, Japan.

[8] Schröder, M., 2003. Experimental study of affect bursts. *Speech Communication*, 40: 99-116.

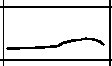
attitude / parameter	N	C	Dp	Di	Ir	J	O	Un
F0 medium/s.d.	168/31	159/24	**252/48**	206/40	200/65	197/61	182/66	**241/98**
F0 maximum	212	205	313	257	335	332	318	387
F0 minimum	112	114	111	127	107	128	118	116
melodic variability	100	91	**202**	130	**228**	**204**	**200**	**271**
DL	+	+	-	+	-	-	-	-
F0 final stressed syllable								
total duration	1528	**2008**	1475	1728	1670	1401	1363	**2360**
medium syllable duration	170	223	164	192	186	156	151	262
final word duration	493	685	550	680	773	442	462	523
final stressed syllable duration	332	464	320	425	572	260	270	313
final stressed vowel duration	181	204	188	**245**	**413**	152	154	178

Table 1 – Behavior of the prosodic parameters: N: neuter; C: consideration; Dp: despair; Di: disappointment; Ir: irony; J: justification; O: obviousness; Un: uncertainty; s.d.: standard deviation; DL: declination line. FO values in Hz; duration values in msec.