Prosodic analysis of Brazilian Portuguese attitudes

Albert Rilliard[1], João Antônio de Moraes[2], Donna Erickson[3], Takaaki Shochi[4]

[1]LIMSI-CNRS, France; [2]Laboratório de Fonética Acústica, FL/UFRJ/CNPq, Brazil
[3] Showa Music University, Japan; [4] University Bordeaux 3, France
rilliard@limsi.fr, jamoraes3@gmail.com, ericksondonna2000@gmail.com, shochi38@gmail.com

Abstract

This paper presents the prosodic analysis of a corpus of Brazilian Portuguese attitudes. Attitudes are separated between the social and propositional categories, and performed either with an assertive or an interrogative modality. Previous studies show the particular relevance of prosodic cues for propositional attitudes, while visual cues are more relevant for social ones. This paper shows that this greater relevance of prosody for propositional attitudes is also observed on the prosodic parameters' variations – and enhance particularly the clearly different and prototypical F_0 contours that distinguished such expressions. The importance of lexical stress on the contour is emphasized, strengthening the global coherence of the speaker's prosodic strategies.

Index Terms: prosodic attitudes, Brazilian Portuguese

1. Introduction

The expression of a speaker's opinion, belief and knowledge to his interlocutor is partly performed through the use of prosodic attitudinal expressions [1]. The use of such prosodic strategies constitutes an important part of the speaker's engagement in his speech [2] and may contribute for an important part of the semantic content of utterances. For example, a sentence produced with an ironic tone of voice will certainly not carry the same meaning than a more its more neutrally performed counterpart. Such prosodic attitudes differ from emotional expressions in that they are voluntarily produced during the interaction, in a given social setting where the attitudes are conventionally encoded for a language and a culture, and may vary with them [3].

Typologies of attitudinal expressions vary with authors and their points of interest (e.g. [4], [5]). The present study is based on a separation between two categories of attitudes (already used by [4] and [6]): propositional and social attitudes. The propositional ones address the propositional content of the sentence (e.g. doubt, obviousness, irony), while social ones refer to the interpersonal relationship between the speaker and the receiver (e.g. politeness, irritation, arrogance). [1] proposes a similar distinction between what she calls propositional and behavioral categories of attitudes.

This study describes the prosodic analysis of a corpus of such attitudes in Brazilian Portuguese (BP). The attitudes have been perceptually validated in previous studies [7,8], and the present paper will focus on the prosodic parameters relevant to such a perception. After describing the corpus of BP attitudes, the process of prosodic analysis is detailed, and the main results observed on the corpus are given.

2. Attitudinal corpus

The set of attitudes used in this study is based on the distinction between propositional and social attitudes introduced above, with a supplementary distinction between the assertive or interrogative modes of the carrier sentences.

These attitudes described here are the following:
Assertive mode:

- *Social*: arrogance (ARR), authority (AUT), contempt (CONT), irritation (IRR), politeness (POL) and seduction (SED);
- *Propositional*: doubt (DOU), irony (IRO), incredulity (INC), obviousness (OBV) and surprise (SUR).

Interrogative mode:

- *Social*: arrogance (ARR), authority (AUT), contempt (CONT), irritation (IRR), politeness (POL) and seduction (SED);
- *Propositional*: confirmation (CONF), incredulity (INC), rhetoricity (RET) and surprise (SUR).

The labels and the number of attitudes vary according to the sentence mode (11 attitudes for assertion, 10 for interrogation), as some attitudes are incompatible with some modes (e.g. obviousness with interrogation). All attitudes were performed by two native speakers of BP (a female and a male), on a set of five sentences from 1- to 6-syllable long and with varying lexical stress position (cf. table 1). The sentences don't have any particular meaning in relation to the attitudes and the modes. Their performances were audio-visually recorded using high quality equipment.

Table 1. *Sentences used for the attitudes, with their length (L., in syllables), the position of the lexical stress on their last word and an English translation.*

BP sentence	L.	Stress	Translation
Tá	1	oxytone	OK
Vaidançar	3	oxytone	she / he is going to dance
Dançava	3	paroxytone	she/he danced
Roberta vaidançar	6	oxytone	Roberta is going to dance
Roberta dançava	6	paroxytone	Roberta danced

These attitudes (including the neutral assertive and interrogative sentences), performed on the 5 sentences, were recorded (in the audio and video modality) in three repetitions by the two speakers, resulting in 690 stimuli. The recordings were phonetically aligned by hand, using Praat [9].

3. Perceptual validation

In order to assess the pertinence of the speakers' performances, one repetition of each attitude from the last sentence of table 1 were chosen in order to perform perception tests, separately for the assertive and interrogative modes. These attitudes have been presented in three modalities (audio-only, visual-only, audio-visual) to native BP listeners who had to recognize the performed attitudes, among the possible attitudes in a given mode and category (propositional or social). The perception results are fully described in [7,8], and provide a validation of the pertinence of the above

described prosodic parameters. Figure 1 presents the mean recognition scores obtained by each attitude, in each three modality, for the two modes and for propositional or social attitudes.
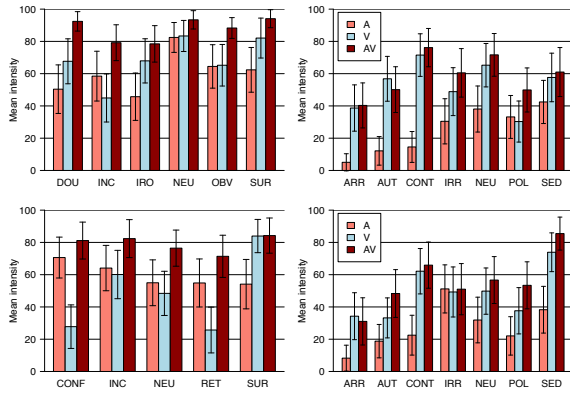

Figure 1: *Identification of propositional assertive (top left), social assertive (top right), propositional interrogative (bottom left) and social interrogative attitudes (bottom right). First column (pink), only audio stimulus, second column (blue), visual, third column (brown), audio-visual.*
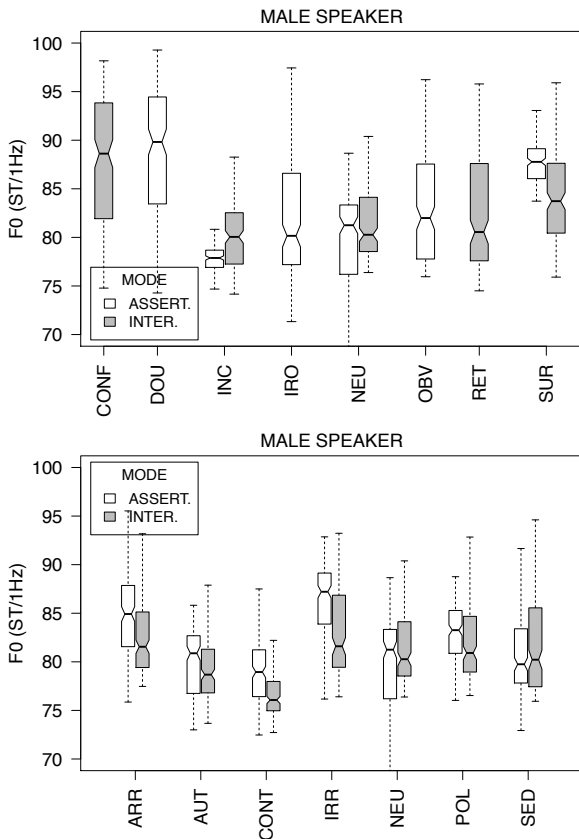

Figure 2: *Dispersion of the F0 values measured for the male speaker on the propositional (top) and the social (bottom) attitudes, for assertive and interrogative modes.*

The most important result that was learned from these perception tests concerns the relative importance of visual and audio modality to the recognition of the two categories. While the visual cues clearly outperformed the audio cues for the recognition of social attitudes, it seems that audio cues are generally more important than the visual ones for the propositional ones (mostly for propositional interrogatives).

This primary use of audio cues for signaling information relating to the propositional content of utterances rather than information relating to the interpersonal relationship during a face-to-face interaction is interesting and led us to a complete analysis of the prosodic variation of this attitudinal corpus.

4. Prosodic analysis

From each of the 690 stimuli, the following prosodic parameters were extracted: the fundamental frequency (F_0, expressed in semitones), the intensity (in dB), and the phonemic duration (expressed in z-score, following [10] method). Both F_0 and intensity where measured on each vowel, at three points (at 10, 50, 90% of the vowel's length).
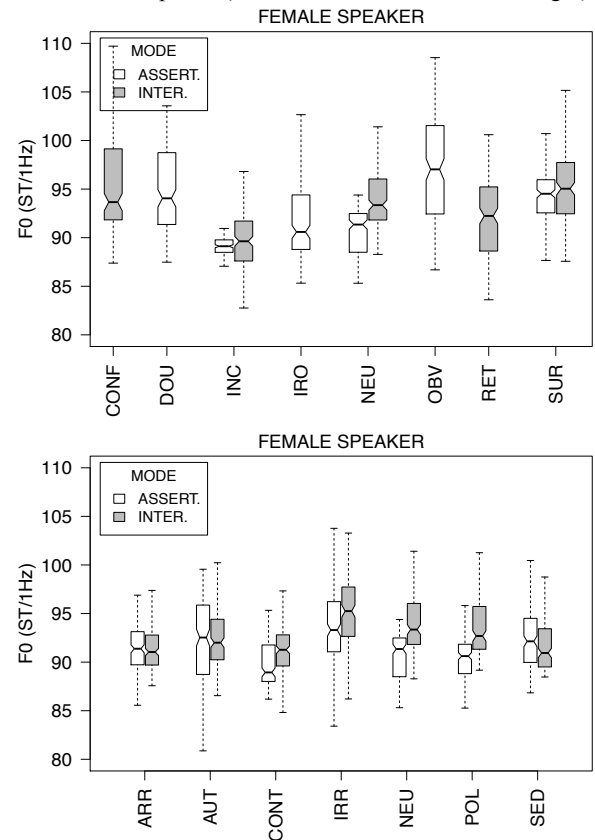

Figure 3: *Dispersion of the F0 values measured for the female speaker on the propositional (top) and the social (bottom) attitudes, for both assertive and interrogative modes.*

4.1. Means values over sentences

As it has been claimed by e.g. [5], the mean distribution of pitch over sentences already gives indication on the type of attitude: a high or low pitch − regarding to the speaker's mean laryngeal frequency, constitute a first kind of indices.

Figures 2 and 3 present the distributions, for F0 values, for each attitude over all sentences. An inspection of these distribution show that, in each category of attitude, some

different patterns of distributions are observed: attitudes with high mean pitch and wide distribution (e.g. CONF and DOU), attitudes a low and flat pitch (e.g. INC), etc. – supporting the above hypothesis.

A comparison of propositional and social attitudes shows a tendency to a wider distribution of the measured parameters in the case of the former, supporting the analysis of perceptual results of a higher importance of prosodic cue for these attitudes (cf. table 2). This is mainly marked for F_0 and Z-duration parameters.

Table 2. *Mean (standard deviation) of F_0, Z-duration and intensity observed for each category of attitude, and for each speaker (Female & Male).*

Category	Spk.	F_0 (st)	Z-duration	Intensity
Propositional	F	93.4 (5.1)	0.39 (1.31)	68.9 (6.5)
Social	F	92.1 (3.4)	-0.25 (0.61)	68.8 (6.2)
Propositional	M	83.8 (6.3)	0.38 (1.26)	68.5 (4.9)
Social	M	81.6 (4.7)	-0.24 (0.68)	66.8 (5.9)

4.2. Prototypical contours

Meanwhile, the means and distribution of prosodic parameters can hardly allow distinguishing between such a complex set of attitudes. The evolution of these parameters across time and with respect to the carrying sentence's morphosyntactic structure shall also play a role. To assess such an importance of prosodic contours, they have also been inspected for all three prosodic parameters (cf. figure 4 and 5 for the F_0 contours in the interrogative mode). Interestingly, the shapes of contours for propositional attitudes (fig. 4) are characteristically different for each one, while the shapes of

social attitudes (fig. 5) tend to be more similar. For example, for the 1-syllable long sentence (first columns in figure 4) propositional attitudes show a large diversity of contours (rising, falling, flat-rising…), while the contours observed for social attitude are all rising – with small differences of pitch mean. The increase of sentences length shows the evolution of the global contours' shapes that tend to conserve a similar shape, whatever the length (under some constraints of minimal length). Such an observation is in line with [11] principle of "prosodic movement expansion" shown on French prosodic attitudes.

Visual inspection also shows the influence of the linguistic constraints of prosody on the global contours of attitudes: a main difference between [11] description and BP attitude is linked with the importance and varying position of lexical stress position in BP. Whereas lexical stress in French always occurs at the final syllable, the described corpus proposes a systematic variation of oxytone and paroxytone words at the end of sentences. So, the two 3- and 6-syllable long sentences (respectively oxytone at the 2nd and 4th columns, and paroxytone at the 3rd and 5th columns) have a different morphosyntactic constraint that imposes a varying position of the main F0 peak and lengthening. This is especially clear for the CONF attitude (2nd line), where the final slope occurs on stressed syllable – for both speakers. Such a phenomenon can also be seen for other attitudes. The other parts of the contours remain similar across sentences. For the segmental duration similar phenomenon are observed: large lengthening of the stressed syllables are observed for some propositional attitudes (e.g. for IRON), while social attitudes' duration patterns tend to be more comparable across attitudes.
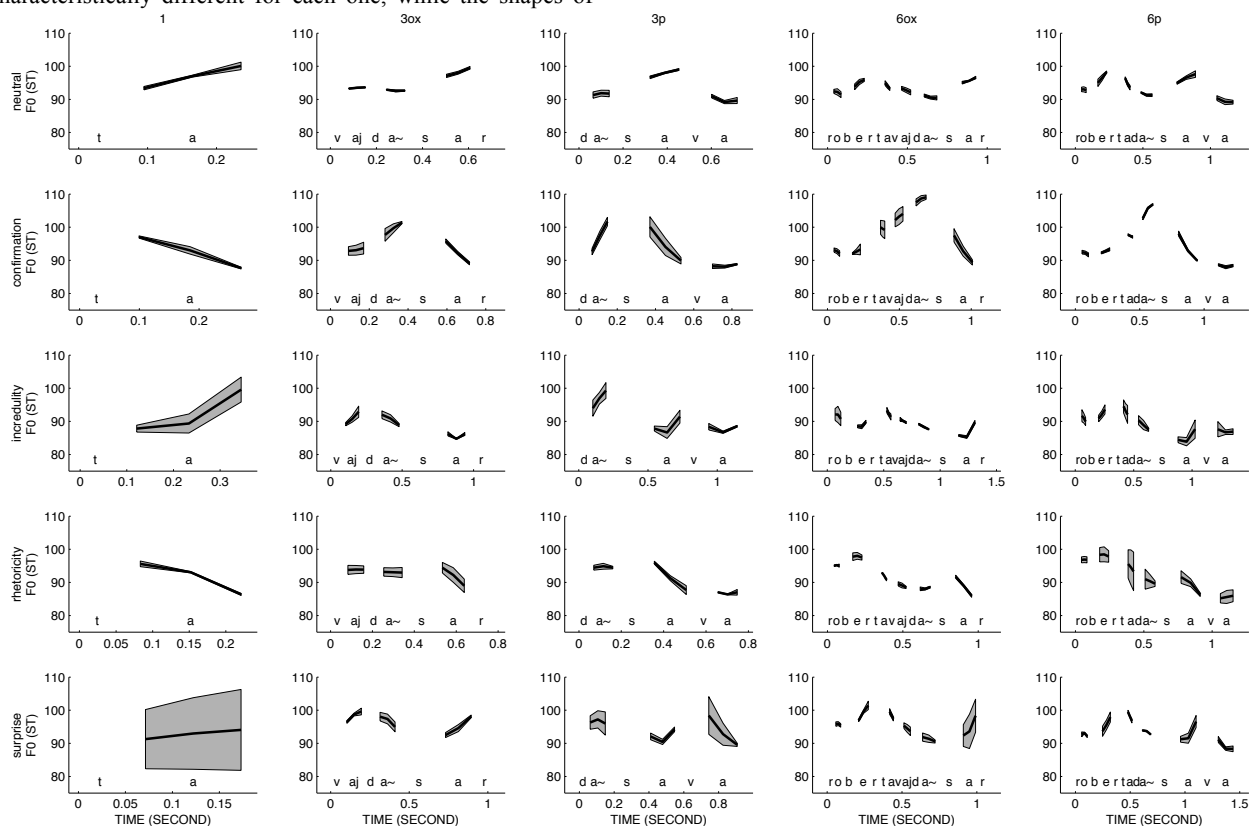

Figure 4: *F0 contours (mean of 3 repetitions in black, standard deviation in gray) for the 5 interrogative sentences (in columns) with the 4 propositional attitudes plus the neutral interrogation (first row), as performed by the female speaker.*
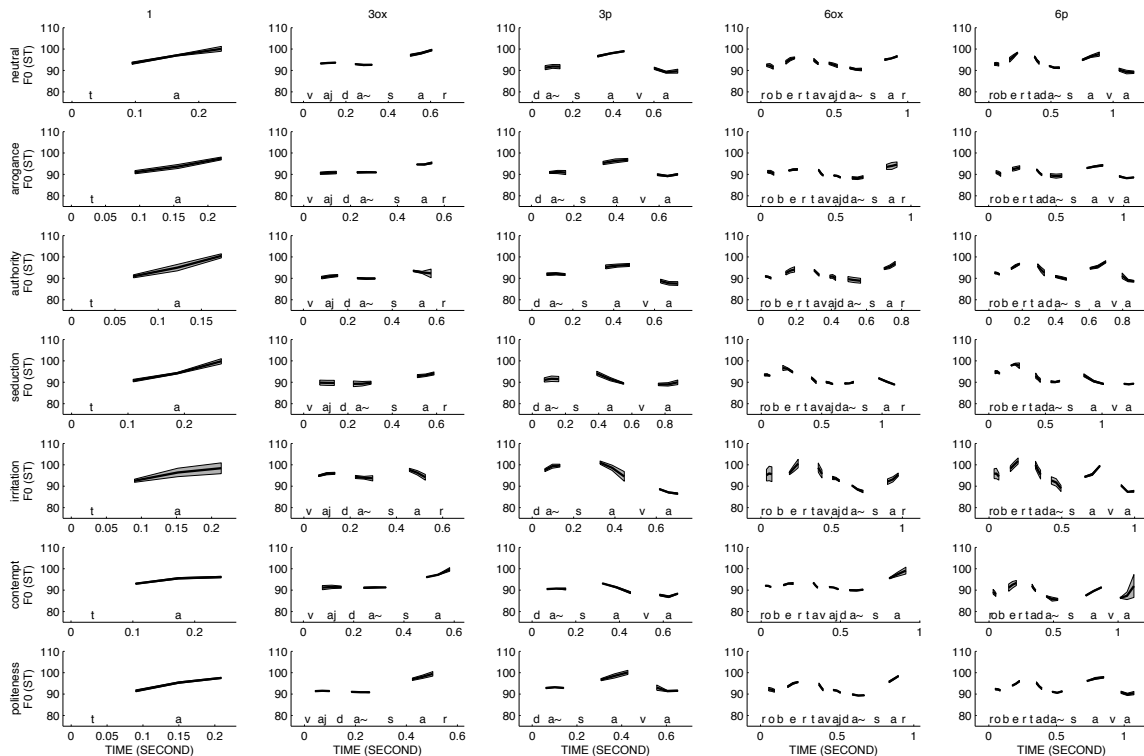
Figure 5: *F0 contours (mean of 3 repetitions in black, standard deviation in gray) for the 5 interrogative sentences (in columns) with the 6 social attitudes, plus the neutral interrogation (first row), as performed by the female speaker.*

5. Discussion& conclusion

This paper has presented a prosodic analysis of the variation induced by attitudinal expressions into the prosodic parameters of a set of BP sentences. These modifications affect the speaker's mean register, pitch range and rhythm. To rate the efficiency of mean prosodic patterns to convey attitudinal expressions would require perceptual tests based on a gating paradigm to check whether e.g. a high start followed by a slope at the beginning of a sentence will be systematically perceived as an expression of rhetoric question (cf. [12] for such an experiment on Japanese attitudes).

The modifications also affect the sentences' prosodic contour. Prototypical strategies have been observed for each propositional attitude, and are reproduced in a similar fashion over speakers for several attitudes – but not for all. The CONF attitude show a rise until the last stressed syllable for the female speaker, while the male speaker tend to produce a high plateau, but both make a steep slope on the stressed syllable. This shows that several communication strategies may coexist in a same language, with common grounds. This variation may be accounted for by gender differences, but more investigation is required to confirm this hypothesis. Preceding perception results have also shown such inter-speakers differences, with higher performances obtained by either the female or the male speaker on several attitudes. To describe the possibility of strategic variations inside a given attitudinal expression would require a larger set of speakers to be recorded and analyzed.

6. Acknowledgements

7. References

[1] Wichmann A., "The attitudinal effects of prosody, and how they relate to emotion", Proceedings of the ISCA Workshop on Speech and Emotion, Newcastle, North Ireland, 143-148, 2000.

[2] Daneš F., "Involvement with language and in language", Journal of Pragmatics, 22, 251–264, 1994.

[3] Rilliard, A., Shochi, T., Martin J.C., Erickson D. & Aubergé, V., "Multimodal indices to Japanese and French prosodically expressed social affects", Language & Speech 52(2/3), 223-243, 2009.

[4] Martins-Baltar M., "De l'énoncé à l'énonciation : une approche des fonctions intonatives", Paris: Didier, 1977.

[5] Gu W., Zhang T., and Fujisaki H., "Prosodic Analysis and Perception of Mandarin Utterances Conveying Attitudes", Proc. of Interspeech 2011, Firenze, Italy, 1069-1072, 2011.

[6] Fónagy I., Bérard E. & Fónagy J.,"Clichés mélodiques". Folia Linguistica, 17, 153-185, 1984.

[7] de Moraes J. A., Rilliard A., Mota B. & Shochi T., "Multimodal perception and production of attitudinal meaning in Brazilian Portuguese", Proc. of Speech Prosody 2010, Chicago, 2010.

[8] de Moraes J.A., Rilliard A., Erickson D. & Shochi T., "Perception of attitudinal meaning in interrogative sentences of Brazilian Portuguese", Proc. of ICPhS 2011, Hong-Kong, 2011.

[9] Boersma P. & Weenink D., "Praat: doing phonetics by computer" (Version 5.1.05) [Computer program], 2011.

[10] Campbell, N., "Automatic detection of prosodic boundaries in speech", Speech Communication, 13, 343-354, 1993.

[11] Morlec Y., Bailly G. & Aubergé V., "Generating prosodic attitudes in French: Data, model and evaluation", Speech Communication, 33, 357-371, 2001.

[12] Shochi, T., Aubergé, V., Rilliard, A. "Nihongobogowashawaha-tsuwanoichibuwokiitetaidowochikakudekiruka ?", Bulletin de l'enseignement du japonais en France, 4, 65-75, 2009.