Emotion expression in speech acts in Brazilian Portuguese: production and perception

Manuela Colamarco and João Antônio de Moraes

Universidade Federal do Rio de Janeiro, Brazil
m.colamarco@ufrj.br; jamoraes2@br.inter.net

Abstract

This work investigates the relation between linguistic and paralinguistic categories in terms of intonation, in order to observe how the manifestation of three primary emotions affect pitch contours which are typical of four speech acts in Brazilian Portuguese (BP).

1. Introduction

The prosodic expression of linguistic categories such as lexical stress, focus or speech acts (statements, questions, etc.) is generally seen as being all or none, and placed in specific parts of the segmental string, while paralinguistic intonational features are considered as having a gradient nature, and affect globally the acoustic signal [4, 8, 10]. Most works that investigates the effects of emotion in the acoustic signal are based in only a single type of speech act, usually statements, and occasionally questions. Few works look into the relation between both categories (linguistic *vs*. paralinguistic). Uldall's classic study [12] constitutes an exception. Through speech synthesis she analyzes how 10 pairs of attitudes affect the intonational contours of statement, yes/no question, wh-question and order in English. In the present study, we cross four speech acts (statement, yes/no question, request and order) with four emotional patterns (three primary or prototypical emotions [2], sadness, joy and anger plus the neutral form), in order to contribute for the description of the emotive speech and try to answer two more specific questions about the speech act *vs*. emotion interaction, namely (i) are they truly orthogonal and independent prosodic dimensions, as generally considered (from both production and perception perspectives), or does this interaction bring forth substantial changes in the generally described patterns, as to modify their phonologic representation; (ii) in case that the first hypothesis is true for production, do the melodic contours group basically by emotions or by speech acts?

2. Method

2.1. *Corpus*

A native female speaker of the Rio de Janeiro dialect uttered the sentence "Prepara a panqueca" [pɾeˈpaɾa pɐ̃ˈkɛkɐ] [Make the pancake] in 16 different ways, statements, yes/no questions, requests and orders, combined with four emotional states, namely joy, sadness, anger and neutral. The utterances had been inserted in specific contexts as they were presented to the speaker, in order to offer a better guidance for her performance. Moreover, she was free to repeat each utterance as many times as she wanted until a satisfactory result was achieved both by her and by the authors of this research judgments.

2.2. Acoustic Analysis

The utterances were acoustically analyzed with *Praat* software. Six prosodic parameters were established for analysis and description: the global average pitch level, the global pitch span (here represented by the F0 standard deviation [11]), the pitch level and span in the last stressed syllable, which plays a crucial role for the phonological representation of the pattern and, finally, the global intensity and the global duration [2, 5, 9].

2.3. Perceptual Analysis

To validate the quality of the *corpus* and, basically to observe which categories are confused the most, the 16 stimuli were submitted to a perception test, in which the subjects were asked to identify the speech acts and the emotional states in a forced choice test.

3. Results

3.1. Acoustic Analysis

3.1.1. Statement

The melodic contour of the neutral statement has been represented as L + H*, for the prenuclear accent, and H + L* L%, for the nuclear accent [6] (Fig. 1).

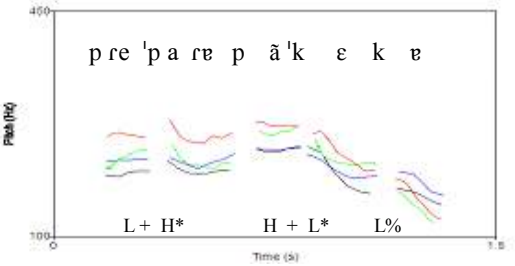


Figure 1: *Melodic contours of the statement "Prepara a panqueca", expressing sadness (blue), joy (green), anger (red) and neutral (black); its duration has been normalized.*

As one can observe in Fig. 1, no emotional state disfigures the melodic pattern of statement; all modify it in a secondary way, though the slight raise expected between the pre-stressed and the stressed syllables, in the prenuclear accent, does not appear clearly in the emotive utterances. So, according to the Chart 1 below, both, the neutral and sad statements present a lower pitch level – global and local –, as well as a lower global pitch span and average intensity when compared to joy and anger statements. Furthermore, the utterance expressing sadness is slower than the other ones. The final stressed syllable behavior

has shown a melodic variation that is greater in the neutral utterance, smaller in the one expressing sadness and even smaller in the one expressing joy.

- ◄——————————► +				
GPL	**NEU** (199 Hz)	**SAD** (211 Hz)	**JOY** (212 Hz)	**ANG** (230 Hz)
GS	**SAD** (19.2 Hz)	**NEU** (25.4 Hz)	**JOY** (37.1 Hz)	**ANG** (45.8 Hz)
LPL	**NEU** (197 Hz)	**SAD** (203 Hz)	**JOY** (219 Hz)	**ANG** (230 Hz)
LS	**JOY** (12.5 Hz)	**SAD** (13.9 Hz)	**ANG** (21.8 Hz)	**NEU** (25.3 Hz)
IN	**SAD** (65 dB)	**NEU** (69 dB)	**JOY** (71 dB)	**ANG** (75 dB)
GD	**NEU** (1.12 s)	**ANG** (1.18 s)	**JOY** (1.2 s)	**SAD** (1.24 s)

Chart 1: *Acoustic measures of the statements pronounced according to the following four emotional states: neutral (NEU), sadness (SAD), joy (JOY) and anger (ANG), disposed, from the lowest to the highest, from the left to the right, according to the following six parameters: global average pitch level (GPL); global span (GS); local average pitch level (LPL); local span (LS); intensity (IN) and global duration (GD).*

3.1.2. Yes/no question

The yes/no question melodic pattern is described as being L + H*, for the prenuclear accent, and L + <H* L%, for the nuclear accent (Fig. 2); the diacritic "<" represents the late alignment of tone H to the final stressed syllable [6,7].

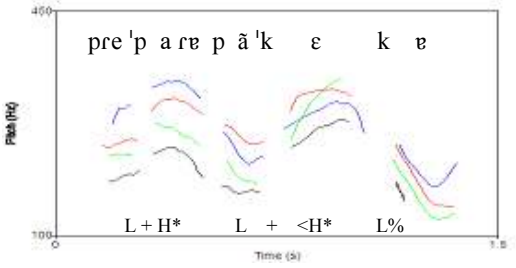


Figure 2: *Melodic contours of the yes/no question "Prepara a panqueca?"; see caption under Fig. 1.*

None of the emotional patterns affected the yes/no question melodic contour enough to disfigure it either. It is important to remark that the F0 rising rate on the final stressed syllable – which takes a H* tone – varies much, having its maximum in joy and its minimum in anger, and that this also reflects in the final stressed syllable span, and, specially in the case of anger, in which it affects the manifestation of the F0 peak alignment to the right, since it turns out less distinct as, instead of a rising, there is almost a plateau.

In Chart 2, one observes that the neutral utterance and the one expressing sadness present the lowest values of local pitch level and global pitch span, when compared to those expressing joy and anger. Surprisingly, sadness has the highest pitch level of all and anger has a relatively low intensity.

- ◄——————————► +				
GPL	**NEU** (220 Hz)	**JOY** (229 Hz)	**ANG** (264 Hz)	**SAD** (268 Hz)
GS	**NEU** (37.5 Hz)	**SAD** (52 Hz)	**ANG** (58.3 Hz)	**JOY** (65.4 Hz)
LPL	**NEU** (264 Hz)	**SAD** (294 Hz)	**JOY** (314 Hz)	**ANG** (324 Hz)
LS	**ANG** (8.91 Hz)	**SAD** (12.9 Hz)	**NEU** (14.4 Hz)	**JOY** (25.9 Hz)
IN	**NEU** (45 dB)	**ANG** (64 dB)	**SAD** (68 dB)	**JOY** (69 dB)
GD	**ANG** (1.21 s)	**NEU** (1.26 s)	**SAD** (1.38 s)	**JOY** (1.44 s)

Chart 2: *Acoustic measures of the interrogative utterances; see captions under Chart1.*

3.1.3. Request

The typical melodic contour for request is represented by H + H*, as for its prenuclear accent and L + >H* L%, as for its nuclear accent; here, the diacritic ">"signals the early alignment of tone H to the final stressed syllable [6, 7] (Fig. 3).

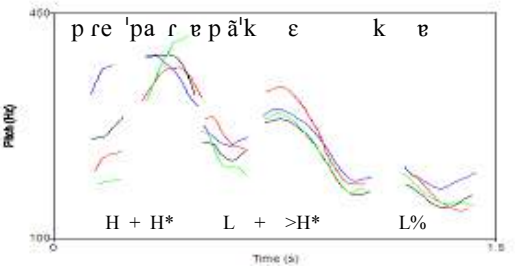


Figure 3: *Melodic contours of the request "Prepara a panqueca"; see captions under Fig. 1.*

Again, no emotional state disfigures the melodic pattern of requests (Fig. 3). It should be observed, though, that the melodic contour of the first stressed syllable has a rising contour in joy, rising-falling in anger and falling in sad and neutral utterances. In Chart 3, one can see that the local pitch span is smaller in neutral and in sad utterances and larger in those that express joy and anger. Also, the total duration of the utterance is longer in the neutral and in the sad states and smaller in the joy and anger ones. Uncommonly, sadness has relatively high values for both global and local pitch levels and average intensity.

- ◄——————————► +				
GPL	**JOY** (232 Hz)	**NEU** (243 Hz)	**ANG** (257 Hz)	**SAD** (262 Hz)
GS	**SAD** (65.7 Hz)	**ANG** (70.1 Hz)	**NEU** (72.3 Hz)	**JOY** (74 Hz)
LPL	**NEU** (230 Hz)	**JOY** (242 Hz)	**SAD** (251 Hz)	**ANG** (267 Hz)
LS	**SAD** (42.8 Hz)	**NEU** (46.1 Hz)	**JOY** (48.6 Hz)	**ANG** (59.6 Hz)
IN	**NEU** (55 dB)	**JOY** (58.0 dB)	**SAD** (58.4 dB)	**ANG** (64 dB)
GD	**NEU** (1.04 s)	**ANG** (1.10 s)	**JOY** (1.15 s)	**SAD** (1.20 s)

Chart 3: *Acoustic measures of the requests; see captions under Chart 1.*

3.1.4. Order

The melodic pattern of order is represented as H + H*, for the prenuclear accent and as H + L* L% for the nuclear accent [6] (Fig. 4).

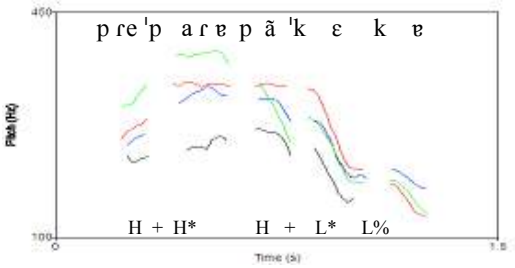
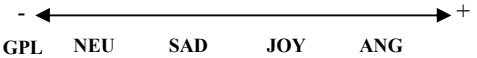


Figure 4: *Melodic contours of the order "Prepara a panqueca!"; see captions under Fig. 1.*

The melodic pattern of order was not disfigured either by any of the emotional patterns (Fig. 4). As shown in Chart 4, the general tendency of the parameters is confirmed by (i) global and local pitch level and global and local pitch span, presenting lower values in neutral and sadness and higher ones in joy and anger; and (ii) total duration of the utterance, longer in neutral and sad utterances and shorter in joy and anger ones.



GPL	**NEU** (227 Hz)	**SAD** (250 Hz)	**JOY** (270 Hz)	**ANG** (274 Hz)
GS	**NEU** (31.2 Hz)	**SAD** (53.6 Hz)	**ANG** (84.6 Hz)	**JOY** (70 Hz)
LPL	**NEU** (193 Hz)	**SAD** (229 Hz)	**JOY** (232 Hz)	**ANG** (269 Hz)
LS	**SAD** (32.4Hz)	**NEU** (35 Hz)	**JOY** (38.4 Hz)	**ANG** (47.2 Hz)
IN	**NEU** (62 dB)	**JOY** (69 dB)	**SAD** (71 dB)	**ANG** (74 dB)
GD	**ANG** (0.96 s)	**NEU** (1.01 s)	**JOY** (1.06 s)	**SAD** (1.20 s)

Chart 4: *Acoustic measures of the orders; see captions under Chart1.*

3.2 Perception test

For the development and application of the perception tests, the computer program PsyScope [3] was used. Sixteen subjects, all native speakers of Brazilian Portuguese and either undergraduate or graduate students of Language at the Federal University of Rio de Janeiro participated in the test.

To evaluate if the subjects recognized the different speech acts and expressive patterns, we applied a first perception test in which they listened only to the statement utterances and should recognize the emotion expressed in them. Following, they listened only to neutral utterances and had to recognizable the speech act in question. All the speech acts had a statistically significant recognizable rate, that applies also to the expressive patterns, except joy that was mistaken for anger and the neutral utterance. In a second test, that purposed to better identify the most frequent types of confusion, each one of the 16 utterances were presented in a computer screen two times in a row. After the first hearing, the subject was questioned about the modality of the sentence, having to choose between the four options: statement, yes/no question, request or order. Given the answer, he listened for the second time the same stimuli and, then, acknowledged the emotion expressed in the utterance, again deciding between four options, this time: neutral, sadness, joy and anger. The results of this test may be seen in the confusion matrix (table 1).

One observes, in the highlighted cells of the confusion matrix, diagonally, the number of votes in which the subjects' interpretation coincides with the speaker's intention when she pronounced the utterances.

The statistical analysis of the data was made with the computer program *R* [13]. Two proportion tests were applied to the results of the second perception test. In the first, we verified which of the utterances had a statistically relevant number of votes for a particular speech act/emotion association. The proportion test showed that the associations that had four or more votes (p-value < 0.05) obtained a significant index of identification. These are signaled in green diagonally in the confusion matrix. In the second test, we examined if the number of votes attributed to the speech act/emotion association that had had the best recognition by the subjects was different, from a statistic standpoint, from the number of votes attributed by them to the second most voted association for the same utterance. The cases that obtained p-values under 0.05 in both tests are signaled in blue in the confusion matrix. The cases in which the number of "wrong" votes for a particular pattern was significantly greater than the number of votes for the expected pattern are signaled in red.

Table 1: *Confusion Matrix drawn up from the results in the perception test submitted to 16 subjects. Vertically, the speaker's intention as she pronounced the utterances; horizontally, the subjects' interpretation; diagonally, the votes that correctly identified with the speaker's intention.*

		Statement				Y/N Question				Request				Order			
		NEU	SAD	JOY	ANG	NEU	SAD	JOY	ANG	NEU	SAD	JOY	ANG	NEU	SAD	JOY	ANG
Statement	NEU	7	1											3			5
	SAD		14												2		
	JOY	9	1	2										2			2
	ANG	1			1												14
Question	NEU					15			1								
	SAD		1			1	9			1	4						
	JOY					1	1	13	1								
	ANG					2		1	11								2
Request	NEU						1			8	6	1					
	SAD									3	13						
	JOY					2				4	4	6					
	ANG					4	1		1	3	1		4				2
Order	NEU	1	1											3	3		8
	SAD	3			1									7	2	1	2
	JOY	3		1						1		3	2	3			3
	ANG													1		1	14

4. Discussion and conclusions

From the production point of view speech acts and expressive patterns are independent categories, once the emotions do not disfigure the melodic contours which are typical of the various speech acts. This is confirmed by the fact that the generally proposed phonological representation for a neutral utterance is also applied to expressive utterances.

Emotional patterns do not always affect the different speech acts in the same way in BP. We verified that the relation between the emotional patterns (the one which presents the highest melodic level or the largest span, for instance) is different in every speech act, just as observed by Uldall [12] for her data. One can only deduce that there seems to be a general tendency, which is not always complied: the neutral utterances and those that express sadness present lower values for pitch level, pitch span and average intensity and a higher value for duration; in the utterances expressing joy and anger, pitch level and pitch span have the higher values, and duration has a lower value.

Anger and joy present a similar behavior to that generally described in other languages: an increase in the pitch span, pitch level (both local and global) and average intensity of the melodic contours which are typical for statements, yes/no questions and orders. Although these emotional states affect the F0 curves of the speech acts in BP very similarly, they were not confounded in the perception test, and this gives us evidence that the vocal quality would have a relevant role in distinguishing joy and anger utterances.

The typical melodic contours for utterances which express sadness have behaved in a slightly different way according to that generally observed. When compared to the neutral utterances, they present, in many instances, not a decrease but an increase in the pitch span, in the pitch level and even in the average intensity. Bezooyen [2] described a fact that is similar to this one, occurring in Dutch. We assume that not as an indication that BP has a different sadness pattern, but that what we hold under the "sadness" label would more properly correspond to the desperation pattern. This, even though semantically very similar to sadness, in the intonation point of view differs from it precisely because presents a considerable increase in the pitch span, pitch level and average intensity [1]. One should also mention that interestingly, in the request, a speech act that when combined with sadness, conveys supplication, average pitch level of both the whole utterance and of the final stressed syllable was very high.

Concerning the perception, however, we observed that the speech acts and the expressive patterns intertwine. The main confusions observed in the perception test seem to have explanations in different levels.

On one hand, the melodic patterns that have a closer relation, like statement and order, or question and request are naturally the most misunderstood by the subjects, once the utterances are presented to them out of any context.

On the other hand, the structure of the sentence with no subject, that would suit the manifestation of the four mentioned speech acts, favors the interpretation of the utterances as being directive – request and order – instead of statement and yes/no question.

Finally, there is a relation between certain speech acts and emotions from a pragmatic point of view. Thus, sadness is naturally associated to request and anger is associated to order, i.e., when recognizing a request, the subjects tended to consider it sad and, when identifying an order, there was an inclination to chose anger as being the emotion expressed by the particular utterance. Conversely, some pragmatic conflicts were also observed between the speech acts and the emotion states: a joyful order does not seem to be natural, such as an angry request. The lack of regularity among the votes of the subjects for these two cases shows that they considered these utterances to be "strange", and could not identify what they had listened to: there seems to be pragmatic limitations to the orthogonality ideally proposed. To sum up, the independence observed in production is not kept in the perceptual domain.

5. REFERENCES

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

[2] Bezooyen, R., 1984. *Characteristics and Recognizability of vocal expression of emotion.* Dordrecht: Foris.

[3] Cohen, J., MacWhinney, B., Flatt, M. and Provost, J., 1993. PsyScope: An interactive graphical system for designing and controlling experiments in psychology laboratory using Macintosh computers. *Behavior Methods, Research, Instruments, and Computers,* 21, 257-271.

[4] Ladd, D. R., 1996. *Intonational Phonology*. Cambridge: Cambridge University press.

[5] Léon, P., 1993. *Précis de Phonostylistique parole et expressivité.* Paris: Natan.

[6] Moraes, J., 2007. Intonational phonology of Brazilian Portuguese, paper presented at the Workshop on Intonational Phonology: understudied or fieldwork languages, ICPhS 2007 Satellite Meeting, Saarbrücken, 5/08/2007, extended abstract available at http://www.linguistics.ucla.edu/ people /jun/ Workshop 2007 CPhS/Moraes-BR.pdf.

[7] Moraes, J. and Colamarco, M., 2007. Você está pedindo ou perguntando? To be published in *Revista de Estudos da Linguagem, FALE-UFMG*.

[8] Scherer, K., 2000. A Cross-Cultural Investigation of Emotion Inferences from Voice and Speech: Implications for Speech Technology. *ICSLP 2000.* Beijing, vol.2, 379-382.

[9] Scherer, K.R.; Bänziger, T., 2004. Emotional expression in prosody: a review and an agenda for future research. *Proceedings of Speech Prosody*. Nara, 359-366.

[10] Scherer, K.R.; Johnstone, T.; Sangsue, J., 1998. L'état émotionnel du locuteur : facteur négligé mais non négligeable pour la technologie de la parole. *Proceedings of the XXIIes Journées d'Etudes sur la Parole*. Martigny.

[11] Scherer, K. R.; Ladd, D. R.; Silverman, K. A., 1984. Vocal cues to speaker affect: Testing two models. *Journal of the Acoustical Society of America,* 76*,* 1346-1356.

[12] Uldall, E., 1973. Attitudinal Meanings conveyed by intonation contours. In: Jones, W. E. and Laver, J. (eds.) *Phonetics in Linguistics A Book of Readings.* London: Logman, 324-335.

[13] R Development Core Team, 2004. "R: A language and environment for statistical computing."R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-00-3, http://www.R-project.org.