

PERCEPTION OF ATTITUDINAL MEANING IN INTERROGATIVE SENTENCES OF BRAZILIAN PORTUGUESE

João Antônio de Moraes^a, Albert Rilliard^b, Donna Erickson^c & Takaaki Shochi^d

^aLaboratório de Fônica Acústica, FL/UFRJ/CNPq, Brazil; ^bLIMSI-CNRS, France;

^cShowa Music University, Japan; ^dUniversity Bordeaux 3, France

jamoraes3@gmail.com; rilliard@limsi.fr; ericksondonna2000@gmail.com; shochi38@gmail.com

ABSTRACT

We present here a perceptual analysis of Brazilian Portuguese prosody expressing eleven different attitudes using an interrogative mode. The attitudes are separated between 6 social, and 4 propositional expressions plus a neutral expression, and are performed by two native speakers (a female and a male). An earlier work studied the declarative mode. The aim of the current study is to perceptually evaluate whether the mode (e.g., declarative or interrogative) influences the perception of attitudes. The multimodal perception of the attitudes is compared across modalities. Our results show higher scores for the propositional attitudes, particularly in the audio modality, while visual information is more important for social expressions, regardless of whether the attitudes are performed in the interrogative or declarative mode.

Keywords: prosodic attitudes, audio-visual prosody, yes-no question, Brazilian Portuguese

1. INTRODUCTION

‘Prosodic attitude’ generally refers to the expression of social affects, voluntarily controlled by the speaker [2], linked to the culture of the speaker. They differ from basic emotional expressions, which may be more spontaneous and universal in nature [6, 9]. Attitudinal meaning may either address the propositional content of the sentence (e.g. incredulity, obviousness, irony), or be directed towards the interlocutor (e.g. politeness, irritation, arrogance) [2, 3]. As prosody is expressed multimodally [8], the importance of both audio and visual information is studied.

It has been shown that audio cues might have a more important role for propositional attitudes [5], than for socially motivated attitudes; for the latter, the visual cues played a more crucial role. This paper aims at reproducing these perception results, obtained with a declarative sentence mode, on interrogative sentences. Audio-visual realizations of 11 prosodic attitudes of Brazilian Portuguese

(BP), produced on an interrogative sentence were recorded, perceptually evaluated and analyzed. The perceptual proximities between attitudes are described, allowing a comparison with their counterpart attitudes performed on the same sentence with a declarative mode [5].

2. CORPUS

Extending [4, 5], 11 attitudes were selected, all expressed with an interrogative mode. Sentences were recorded by two native BP speakers (1 female and 1 male). They consist of 6 social attitudes, 4 propositional ones plus a neutral question, namely:

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

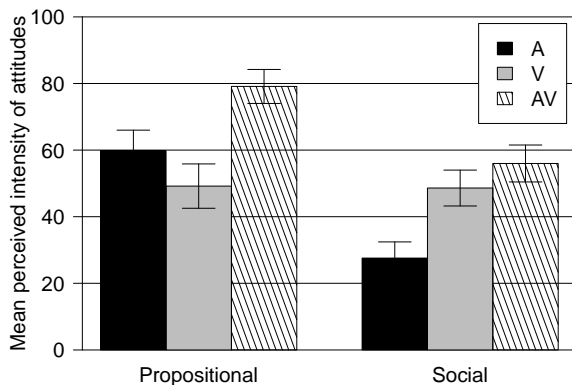
Each attitude was performed on the same semantically neutral 6-syllable long Portuguese interrogative sentence (yes/no question) “*Roberta dançava?*” (“*Was Roberta dancing?*”). The labels (in Portuguese) were followed by a description of the context in which these attitudes commonly appears. For instance, the description for *surprise* was “*Surprise is a yes-no question in which the speaker conveys surprise by the fact that Roberta was dancing*” Speakers stood in a sound-proof room, facing a video camera and a microphone. The audio was digitalized at 22.5kHz, and video was encoded with a 784*576 pixels resolution.

3. PERCEPTION TEST

The social and propositional attitudes (with the neutral expression added to both sets), were separately presented to listeners in their audio-only, visual-only and audio-visual modalities. Subjects had to recognize the presented attitude during a forced-choice paradigm, amongst a list of

5 or 7 possible answers, which include all attitudes of the given group plus the neutral expression. The presentation order of attitudes and of the modalities was balanced across subjects: half were presented with audio-only, visual-only, audio-visual, while the other half were presented with visual-only, audio-only, audio-visual modalities. In each group, half the subjects were presented with social attitudes first, and the other half, with propositional attitudes first. Each stimulus was played twice, in order to ease the listeners' task. Subjects had to give their answers by selecting on a slider the relative intensity of the perceived attitude (one slider per possible attitude was provided). The scale ranged from "barely marked attitude" to "very marked attitude". The performance from each speaker and each attitude was randomized inside each group and modality. Each subject has to rate 72 stimuli. 30 native speakers of BP (23 women, 7 men, with a mean age of 29 years) participated in the experiment. None reported any perception problem.

Figure 1: Mean intensity of correct answers for propositional (left) and social (right) interrogative attitudes, for both speakers in audio (A) visual (V) and audiovisual (AV).



4. RESULTS

4.1. Analysis of variance

Results were analyzed using two ANOVAs (cf. table 1), run separately for propositional and social attitudes. The GLM repeated-measure procedure of SPSS was used. There was one between-subject factor: the order of presentation of modalities (OM, fixed), and three within-subject factors: the modality (M) of presentation (3 levels: audio, visual, audiovisual), the speaker (S, 2 levels: female, male) and the attitudes (At, 5 or 7 levels according to the attitudes' group). The intensity scores were used as the dependent variable.

Table 1: Repeated measures ANOVAs on intensity scores. Insignificant interactions of within-subjects and between-subjects factors are omitted.

<i>Propositional attitudes</i>					
	<i>df</i>	<i>df error</i>	<i>F</i>	<i>p</i>	<i>Partial η²</i>
<i>OM</i>	1	28	2.8	0.10	0.09
<i>Modality</i>	2	56	36.2	0.00	0.56
<i>Speaker</i>	1	28	0.5	0.51	0.02
<i>Attitude</i>	5	140	8.6	0.00	0.23
<i>M * S</i>	2	56	0.9	0.42	0.03
<i>M * At</i>	10	280	11.8	0.00	0.30
<i>S * At</i>	5	140	1.7	0.16	0.06
<i>M * S * At</i>	10	280	1.4	0.18	0.05
<i>Social attitudes</i>					
<i>OM</i>	1	28	1.5	0.22	0.05
<i>Modality</i>	2	56	64.7	0.00	0.70
<i>Speaker</i>	1	28	0.5	0.50	0.02
<i>Attitude</i>	6	168	14.8	0.00	0.35
<i>M * S</i>	2	56	0.9	0.39	0.03
<i>M * At</i>	12	336	4.0	0.00	0.12
<i>S * At</i>	6	168	13.0	0.00	0.32
<i>M * S * At</i>	12	336	2.0	0.02	0.07

For both the propositional and social attitudes, the modality (Fig. 1), the attitude (Fig. 2) and the interaction between them have a significant effect. A t-test shows that propositional attitudes received significantly higher scores than social ones. Note that the greater number of social attitudes (6) compared to that of the propositional attitudes (4), may contribute to the lower scores for the social attitudes; although the mean scores obtained by the neutral expression inside each group do not differ significantly.

Inside each group of attitudes, a post-hoc Tukey's HSD test showed that for both groups the audio-visual modality received significantly higher scores than other modalities. For propositional attitudes, the audio information received higher scores than visual, while for social expressions the visual modality brings more efficient cues than the audio ones. The modality for both sets of attitudes has the stronger effect on listeners' responses (cf. η^2 in table 1). Attitude (and the interaction between attitude and modality) has the second more important effect for both groups of attitudes.

The interaction between the speaker and the attitude is also significant (and has an important effect) in the case of social attitudes. This difference between attitudes performed by the two speakers is shown on Fig. 3: the female-performed *irritation* received significantly higher scores than the male one, while a reverse pattern is found for

arrogance. This underlines the importance of individual strategy and individual performances for the expression of social affects, as well as possible differences between genders.

Figure 2: Mean intensity rating for each attitude, all speakers and modalities averaged.

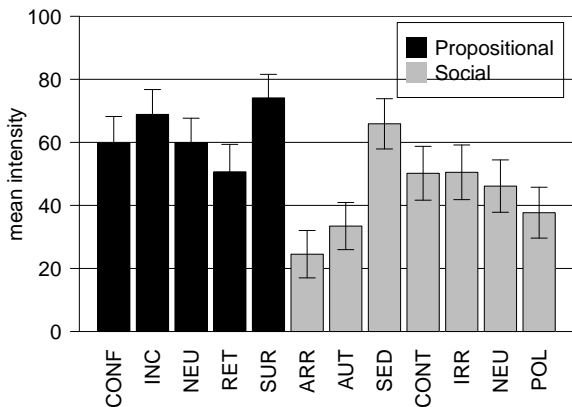
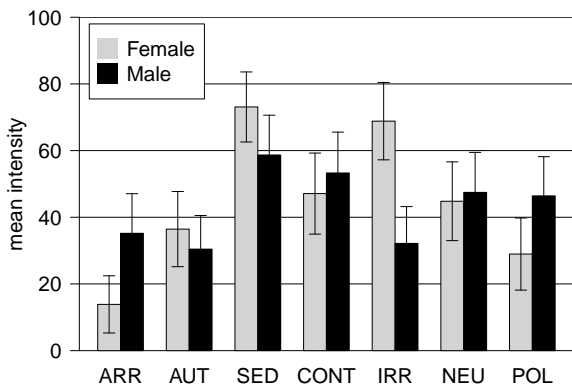


Figure 3: Mean intensity rating for social attitudes performed by either the female or male speaker, all modalities averaged.



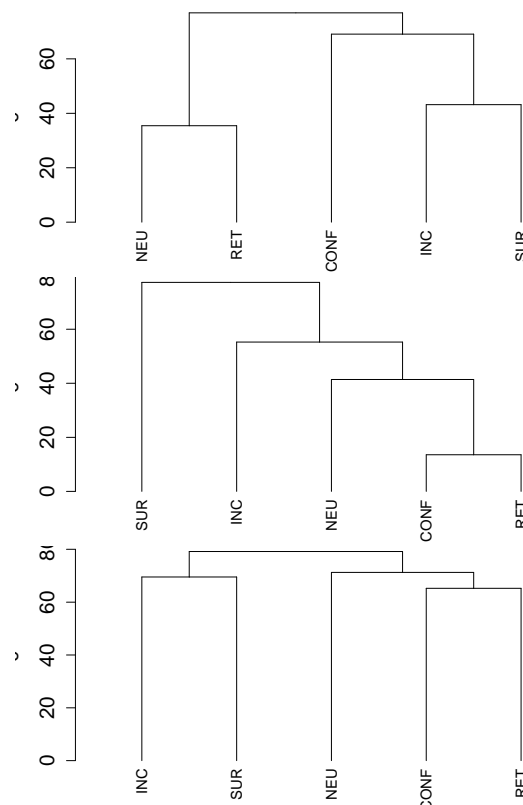
4.2. Hierarchical clustering

Confusion matrices, summing up the distribution of subjects' answers for each attitude over all the possible answers were calculated, separately for each of the three modalities and for each group of propositional or social attitudes (averaged over the two speakers). The Euclidean distances between all pairs of lines of these matrices were used as the perceived distance between corresponding attitudes. These distances were then used as the input of an agglomerative hierarchical clustering, resulting in dendrograms of the perceived distances between attitudes (cf. Fig. 4 and 5, y-axis is the perceived distance).

Amongst propositional attitudes in the audio-only modality, confusions are observed between the neutral and the rhetoric questions on one hand, and the incredulous and surprise questions on the

other. Confirmation questions are clearly distinguished from the others. Within the visual modality, listeners show confusions between rhetoric and confirmation questions, and to a lesser extent, neutral questions; surprise and incredulous questions are distinguished. Listeners are able to efficiently combine the information from both modalities, allowing a clearer dissociation of propositional attitudes in the audio-visual modality.

Figure 4: Dendrograms for propositional attitudes, for each modality (from the top: audio, visual, audio-visual). Results for both speakers are averaged.

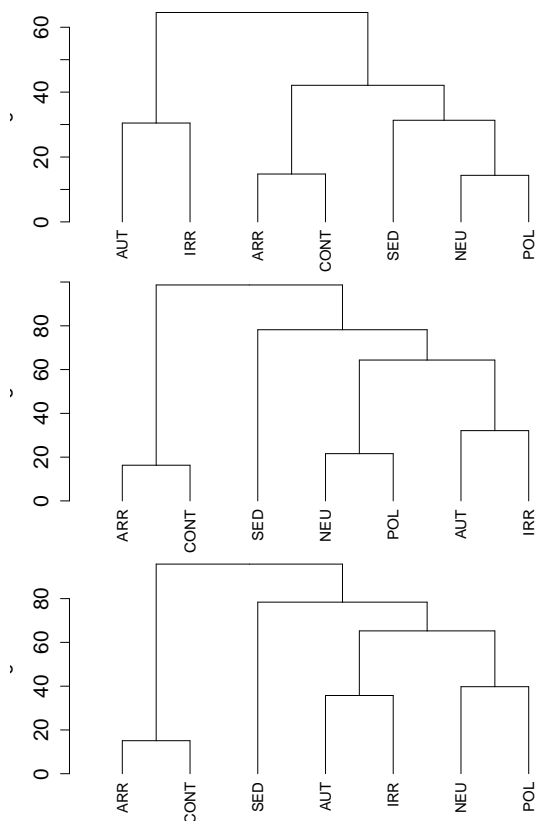


More confusion is observed amongst the social attitudes. With the help of the audio modality, listeners distinguished three main classes of expressions to regroup the following attitudes: (1) polite, neutral and – to a lesser extent – seductive questions; (2) arrogant and contemptuous questions; and (3) authoritative and irritated questions.

The visual modality allows subjects to distinguish the seductive expression from neutral and polite ones, but shows confusions similar to the above for the other social expressions. Thus listeners still show three groups of meta-expressions in the audio-visual modality, regrouping politeness with neutral, arrogance with

contempt, authority with irritation, and seduction stands alone. These clusterings remain with a higher discrimination rate.

Figure 5: Hierarchical clustering for social attitudes, in the three modalities (from the top: audio, visual, audio-visual). Results for both speakers are averaged.



5. DISCUSSION & CONCLUSION

As supposed in the introduction, the propositional and social attitudes are not perceived the same way. The propositional attitudes, which directly address the linguistic content of sentences, were expected to be more strongly related to speech – and therefore to acoustic variations. Conversely, social attitudes are addressed to the receiver in addition to the sentence's meaning; they are therefore more linked to the management of the face-to-face interaction between the speaker and its receiver – and may be expressed with a higher degree of multimodality.

The results clearly support this description: the audio-only modality brings most of the information for the propositional attitudes, while the visual modality is mainly used for the social ones. Meanwhile, both modalities play an important role for perception and allow to even disambiguate the other: the multimodal condition receives the best discrimination scores. These

results corroborate the findings of preceding studies on BP social affects [4, 5]. Specifically, they suggest that the utterance mode (interrogative or declarative) has no effect on the perception of propositional or social attitudes.

Finally, the repartition of the social attitudes in 4 main expressive groups is interesting as it can be to some extent linked to the notions of imposition of the speaker towards the receiver and of the relative power of the speaker with regard to the receiver, described by [1]. This is also in line with the main expressive dimensions found in works on social affect throughout other languages [7].

Acoustic parameters for part of this data are currently analyzed. First results indicate that fundamental frequency shows a wider range of variation and more different contour shapes across propositional than social attitudes. Future work will extend the prosodic and acoustic analyses on the perception corpus to the complete corpus recorded from these two speakers in order to see more clearly what acoustic cues characterize propositional vs. social attitudes in Brazilian Portuguese. Work is currently underway to compare acoustic and perceptual characteristics of propositional and social attitudes in a variety of cultures, such as French, American English, and Japanese, etc.

6. REFERENCES

- [1] Brown, P., Levinson, S. 1978. Universals in language usage: politeness phenomena. In Goody E. N. (ed.), *Questions and Politeness. Strategies in Social Interaction*. Cambridge: Cambridge University Press, 56-324.
- [2] Fónagy, I., Bérard, E., Fónagy, J. 1984. Clichés méthodiques. *Folia Linguistica* 17, 153-185.
- [3] Martins-Baltar, M. 1977. *De L'énoncé à L'énonciation: Une Approche Des Fonctions Intonatives*. Paris: Didier.
- [4] Moraes, J.A. 2008. The pitch accents in Brazilian Portuguese: analysis by synthesis. *Proc. Speech Prosody 2008*, Campinas, 389-397.
- [5] Moraes, J.A., Rilliard, A., Mota, B., Shochi, T. 2010. Multimodal perception and production of attitudinal meaning in Brazilian Portuguese. *Proc. Speech Prosody 2010*, Chicago.
- [6] Scherer, K.R., Wallbott, H.G. 1994. Evidence for universality and cultural variation of differential emotion response patterning. *J. Person. Soc. Psy.*, 66(2), 310-328.
- [7] Shochi, T., Rilliard, A., Aubergé V., Erickson, D. 2009. Intercultural perception of English, French and Japanese social affective prosody. In Hancil, S. (ed.), *The Role of Prosody in Affective Speech*. Bern: Peter Lang, 31-59.
- [8] Swerts, M., Kraemer, E. 2005. Audiovisual prosody and feeling of knowing. *J. Memory Language* 53(1), 81-94.
- [9] Zinck, A., Newen, A. 2008. Classifying emotion: a developmental account. *Synthese*, 161, 1-25.