VISUAL AND AUDITORY CUES OF ASSERTIONS AND QUESTIONS IN BRAZILIAN PORTUGUESE AND MEXICAN SPANISH: A COMPARATIVE STUDY

Luma da Silva Miranda, Carolina Gomes da Silva, João Antônio Moraes, Albert Rilliard

Journal of Speech Sciences (9): 73-92. 2020.
Available at: http://revistas.iel.unicamp.br/joss

VISUAL AND AUDITORY CUES OF ASSERTIONS AND QUESTIONS IN BRAZILIAN PORTUGUESE AND MEXICAN SPANISH: A COMPARATIVE STUDY

MIRANDA, Luma da Silva[1]*
GOMES DA SILVA, Carolina[2]
MORAES, João Antônio[3]
RILLIARD, Albert[4,3]

[1]Eötvös Loránd University
[2]Federal University of Paraíba
[3]Federal University of Rio de Janeiro, CNPq
[4]Université Paris-Saclay, CNRS, LIMSI

Abstract: *The aim of this paper is to compare the multimodal production of assertions and questions in two different languages: Brazilian Portuguese and Mexican Spanish. Descriptions of the auditory and visual cues of these speech acts are presented based on Brazilian and Mexican corpora. The sentence "Como você sabe" was produced as an assertion and an echo question by ten speakers (five male) from Rio de Janeiro and the sentence "Apaga la tele" was produced as an assertion and a yes-no question by five speakers (three male) from Mexico City. The speech acts intonational patterns were described in terms of F0 movements and annotated in the nuclear region of the contours with ToBI system. Momentary facial muscular changes (namely Action Units) located in the upper and lower part of the face as well as head movements were used to analyze the facial expressions. The acoustic description showed that Brazilian Portuguese assertions are produced with a falling F0 nuclear configuration (H+L*L%) and echo questions with a rising F0 nuclear configuration (L+<H*L%). Mexican Spanish assertions present two types of F0 nuclear configurations, either a low flat nuclear F0 (L*L%) or a falling-rising (L+H*L%) nuclear F0, whereas Mexican Spanish yes-no questions are produced with a low nuclear F0 followed by a rising boundary tone (L*LH%). The outcome of the visual analysis indicates that, whereas Brazilian Portuguese assertions are visually produced with blink and right head tilt and Mexican Spanish assertions with lip stretcher, lowering the eyebrows, tightening the eye-lid and wrinkling the nose can be considered question markers in both language varieties.*

Keywords: Question; Assertion; Audiovisual prosody; Brazilian Portuguese; Mexican Spanish.

---

*Corresponding author: miranda.luma@btk.elte.hu

1 Introduction

A large number of studies have shown that the multimodality of speech is part of human communication. According to Kendon (2004), in addition to spoken expressions, a great amount of information regarding speakers' goals, attention, ideas, feelings and attitudes is transmitted by visible body actions. Those actions, which are called *gestures* when they constitute part of an utterance, can often be employed to achieve either the same or a similar communicative purpose as that which is verified in spoken expressions. Kendon (2004: 5) argues that "when speakers use gesture they do so as an integral part of the act of producing an utterance". Additionally, the way in which gestures are organized with speech is varied. For instance, gestures can be used either in conjunction with spoken language (e.g., question markers) or as complements to spoken messages (e.g., pointing gestures).

Facial expressions are recognized as an important signal of others' intentions in a spoken interaction (Schmidt and Cohn, 2008). The study of facial gestures is traditionally related to the analysis of emotional speech thanks to Charles Darwin's (1872) seminal work "The expressions of the emotions in man and animals" which pointed out, from an evolutionary point of view, the similarities of facial expressions between humans and other species (e.g. mammals) triggered by emotions. However, facial movements also change due to communicative interactions, even in other mammals, such as primates (Waller *et al.*, 2015; Scheider *et al.*, 2016) and domestic dogs (Kaminski *et al.*, 2017).

In human communication, facial information is processed at an early age. Language acquisition studies have shown that the integration of auditory and visual modalities in speech perception (Lewkowicz, 2003; Lalonde and Holt, 2014) is relevant even for infants. For instance, according to the literature (Lewkowicz and Hansen-Tift, 2012), children's attention shifts from the gaze area of the face to the speaker's mouth between 8 to 12-months of age. Moreover, research findings with adult participants provide evidence that the auditory percept is affected by visual information (McGurk and MacDonald, 1976).

Visual speech has been documented as a source of information for a variety of speech functions. As for expressive functions, several studies confirmed that visual cues helped listeners to decode emotions (Barkhuysen *et al.*, 2010) and attitudes (Moraes *et al.*, 2010; Crespo Sendra *et al.*, 2013; Moraes and Rilliard, 2014). In addition, multimodal investigations on linguistic meanings conveyed by prosody at the sentence level, such as the intonation of sentence types (House, 2002; Srinivasan and Massaro, 2003; Borràs-Comes and Prieto, 2011; Cruz *et al.*, 2017; Miranda *et al.*, 2019; Miranda *et al.*, 2020b), prominence (Hadar *et al.*, 1983; Swerts and Krahmer, 2004), focus (Dohen and Loevenbruck, 2009) and speech segmentation (De la Cruz-Pavía *et al.*, 2019), also provided evidence that the visual channel benefits the comprehension of the interlocutor's communicative intentions.

According to Gili Fivela (2018), although prosody has been unimodally explored, the integration of auditory and visual information is considered crucial in both speech comprehension and production. It is worth noting that the multimodal production of speech is not only related to the mechanics of speech production, such as articulatory gestures of speech sounds (e.g., bilabials x non-bilabials; rounded vowels), but also to the facial gestures used to convey the same message as the verbal signal (Gili Fivela, 2018).

In this paper, the multimodal production of Brazilian Portuguese (BP, henceforth) assertions and echo questions as well as Mexican Spanish (MS, henceforth) assertions and yes-no questions is analyzed. Yes-no questions are produced when the speaker is seeking for a new information and the expected answer is yes/no. In BP, echo questions were chosen because they are an underexplored type of question. According to Frota *et al.* (2015a), the echo question is a

type of yes-no question used to express a lack of understanding of a preceding utterance in the communicative interaction, which also demands a yes-no response, similar to those of neutral yes-no questions. In addition, neutral yes-no questions and neutral echo-questions are expressed in Portuguese with a similar nuclear intonational contour in terms of pitch accents.

On the one hand, many studies have already showed that assertions and questions are produced with specific auditory cues in BP (Moraes 1998, 2008; Frota *et al.*, 2015a; Miranda, 2019) and MS (Sosa, 1999; Gomes da Silva, 2019). While the assertion contour in BP presents a falling F0 in the last nuclear stressed syllable, BP echo questions are produced with a rising F0 in the nuclear region of the contour. Frota *et al.* (2015a) verified this intonational pattern of echo questions in four Brazilian regional varieties: São Paulo, Minas Gerais, Bahia and Rio Grande do Sul. Similarly, the MS intonation studies described differences between assertions and yes-no questions in the nuclear position of the melodic contour. While the yes-no questions are produced with a high F0 at the end of the utterance, the assertions present two types of movements: either a falling F0 in the last nuclear stressed syllable or a circumflex configuration (Sosa, 1999; De-la-Mota *et al.*, 2010). In another study of the Mexican variety of Puebla, Willis (2005) noted that both the downward pattern and the circumflex pattern were used in all contexts by the speakers, although there were individual preferences. In addition, Sosa (1999) considers that the peak associated with the nuclear accented syllable of the declarative statement is possibly related to differences in the expressive value of the statement. The author states that there is no focus in this production, despite the particular circumflex F0 configuration.

On the other hand, few studies investigated the role of the visual channel in the production of the assertion and question intonation either in BP (Peres *et al.*, 2011; Miranda, 2019; Miranda *et al.*, 2019; Miranda *et al.*, 2020b) or in MS (Gomes da Silva, 2019). As far as we know, multimodal analysis can be found only in another Spanish variety, such as in the study of Gónzalez-Fuentes (2015), which described the audiovisual characteristics of ironic utterances produced by a Spanish professional humorist. According to the author's investigation, there are gesture marks aligned with prosodic marks to distinguish the irony of a statement from another statement produced without that attitude. In other words, from an audiovisual perspective, the author showed that both verbal (prosody) and non-verbal variation (gestures) are important for the production and perception of irony.

The question that then naturally arises is whether assertions and questions can be distinguished by means of visual cues in conjunction with auditory cues in BP and MS, which are two typologically similar languages that employ intonation patterns as a resource to mark the difference between assertions and questions without any morphosyntactic strategies. The novelty of this paper lies in the investigation of the auditory and visual cues in the production of assertions and questions cross-linguistically in order to provide an explanation to how the facial gesture contributes to the utterance meaning of which it is part, as pointed out by Kendon (2004). This way, cross-linguistic cues for pragmatic meanings such as interrogativity can be found. To our knowledge, multimodal analysis comparing assertions and questions in these two languages has not been made previously.

In the audiovisual prosody literature, the interrogative meaning is conveyed by specific gestures. For instance, Kendon (2004), when analyzing the pragmatic uses of the Italian hand gesture called *grappolo*, a gesture in which "the hand is held with palm upwards with all digits drawn together so that they are in contact with one another at their tips" (Kendon, 2004: 228), described different movement patterns of this gesture and some of them were related to the action of asking a question. For example, when the speaker asks a question, just requesting information, "the hands close to the *grappolo* from a partially open pose" (Kendon, 2004: 230) and "the *grappolo* is oscillated forward and back" (Kendon, 2004: 231). On the other hand,

when the speaker produces a more expressive type of question, such as a surprised question, "the *grappolo* is held on a supine forearm and is moved upwards and sometimes somewhat toward the speaker several times" (Kendon, 2004: 231).

In this paper, the visual analysis of assertions and questions focuses on gestures produced by the speakers' face while uttering the speech acts. It is worth noting that most of the works that compared the visual production of assertions and questions analyzed different types of questions and pragmatic meanings, such as yes-no/polar questions, echo questions and wh-questions.

House (2002) proposed visual cues for the interrogative and declarative modes in Swedish to be manipulated in a talking head. The eyebrow lowering and vertical head tilting were cues that conveyed yes-no questions and, for statements, a smile plus a short up-down head nod and eye narrowing. Srinivasan and Massaro (2003) demonstrated that echo questions were produced in American English with a significant eyebrow raising and head tilt, whereas the statement presented little or no eyebrow raise plus an insignificant head movement.

In Catalan, Borràs-Comes and Prieto (2011) demonstrated that the characteristics of a facial gesture expressing an echo question are the combination of brow lowerer and head backward. When comparing the visual cues of Catalan and Dutch polar questions, Borràs-Comes *et al.* (2013) showed that, in question production, there were similar distributions of gaze and eyebrow raisings in both languages. Torreira and Valtersson (2015) also analyzed polar questions and statements in a conversation corpus with French speakers. Based on the description of the data, the authors found out that, although the majority of the questions were not accompanied by body movements, when either the raised eyebrow or the head nod were present, both movements occurred in questions. The authors showed that not only polar questions but also statements are produced in most of the utterances with the gaze towards the addressee.

Cruz *et al.* (2015) investigated the production of yes-no question in European Portuguese and found out that this pragmatic usage is accompanied by head movements (either up-down or back-forward) along with eyebrow raising, while statements are produced with up and down head movement. Recently, Cruz *et al.* (2019) described the visual production of sentence types in Portuguese Sign Language (LGP): whereas the statement can be produced with an up-down head nod, yes-no questions were produced with eyebrow lowering along with head nods and wh-questions with eyebrow lowering plus a head up movement. In addition, Miranda *et al.* (2019) showed that BP wh-questions are also produced lowering the eyebrow plus turning the head right.

Overall, the aforementioned works point out that the production and perception of speech is multimodal. Furthermore, Cruz *et al.* (2017, 2015) shows that facial gestures are produced according to the sentence type (e.g., assertions and yes-no questions) and the pragmatic meaning of a specific sentence (e.g., broad focus statements and narrow focus statements). As this study is also built upon this assumption, it remains to be seen which types of similarities and differences in the production of assertions and questions can be found in BP and MS. Hence, the aim of this work is to compare the auditory and visual cues in the production of assertions and questions in two different languages: Brazilian Portuguese and Mexican Spanish, which are the languages spoken in the two most populous countries in Latin America.

Based on the BP and MS intonation literature as well as the variation found in the general outcome of the production analysis of visual cues, such as eyebrows and head movements, displayed in interrogative meanings (Cavé *et al.*, 1996; House, 2002; Srinivasan and Massaro, 2003; Borràs-Comes and Prieto, 2011; Cruz *et al.*, 2015; Torreira and Valtersson, 2015; Miranda *et al.*, 2019; Miranda *et al.*, 2020b), the hypotheses of this study are: (1) both

languages present different auditory cues for assertions and questions and (2) speakers of both languages produce different facial expressions for each speech act. After this introduction, this paper is divided into the following sections: the description of the method; the results of the BP and MS auditory and visual production analysis of the two speech acts; the general discussion of findings and the conclusion of the study.

2 Method

In this section, the details of the design of Brazilian Portuguese and Mexican Spanish corpora and the method of description of the auditory and visual cues of this study are presented.

2.1 Brazilian Portuguese and Mexican Spanish corpora

2.1.1 Recording procedure and structure of the corpora

Prior to the recording sessions at the Acoustic Phonetics Laboratory at UFRJ, a consent form regarding the scientific use of the recorded data was signed by the ten BP speakers. A SONY NEX-F3 video camera was used to record the upper body and face of the speakers who were seated against a dark background in a sound-attenuated room. Whereas the camera was positioned 90 cm from the speakers, a Zoom H4 recorder was positioned 20 cm from the speaker's mouth, outside the view of the camera. The synchronization of the audio waves and the recorded video was made in the software Vegas Pro 14 (Magix, 2016).

In the BP corpus, the sentence "*Como você sabe*" was produced as an assertion ("As you know.") and as an echo question ("As you know?"). The pragmatic context of each speech act was explained before the beginning of the recording session. For instance, in the assertion context, the participant was asked to imagine that two people are enrolled in the same course and both have a scheduled exam. Then, one of them asks: "How do you know that the test is tomorrow?" and the other answers: "As you know.", which can be paraphrased as "In the same way that you know." In the production context of the echo question, the speaker had to imagine that someone who was talking to him said: "As I know.", but he is not sure he heard it correctly and asks: "As you know?" just to check if he understood the previous sentence correctly. In the recording sessions, the type of utterance to be produced ten times by each BP speaker was described by the experimenter, alternating assertions and echo questions, until ten repetitions of each was reached. A total of two hundred utterances were collected. For this article, only the eighth and ninth repetitions of each sentence type produced by all speakers were chosen, in order to avoid a qualitative selection of the visual material. Hence, forty sentences were selected (2 sentence types x 2 repetitions x 10 speakers) to the auditory and visual analyses.

As for the MS data, the audio and video recordings were made in two different stages. The first male informant was filmed and recorded at the Acoustic Phonetics Laboratory at UFRJ, while the others were recorded in a laboratory of the "*Escuela Nacional de Antropología e Historia*", in Mexico City. It should be noted that, whereas the audio of the five informants was recorded with a portable recorder in .wav format, the video was filmed with a portable digital camera camcorder, SONY HX400V.

The five MS speakers produced two different speech acts with the sentence "*Apaga la tele*" ("Turn off the TV"). Participants were asked "*Qué hace Manuel antes de acostarse?*" ("What does Manuel do before going to bed?") in two different pragmatic contexts. In the assertive one, the speaker knows what Manuel does and answers the question ("*Qué hace*

Manuel antes de acostarse?"/"What does Manuel do before going to bed?"), producing an assertion ("*Apaga la tele*."/ "He turns off the TV."). In the interrogative context, the speaker does not know the answer for this question ("*Qué hace Manuel antes de acostarse?*"/"What does Manuel do before going to bed?") and guess what Manuel does by asking a yes-no question ("*Apaga la tele?*"/ "Does he turn off the TV?"). The investigator instructed the participants about the utterance types to be produced, alternating eight speech acts (command, request, supplication, suggestion, advice, challenge, yes-no question and assertion), which were repeated three times by each speaker. At the end of this process, a hundred and twenty utterances were collected. In this article, only assertions and yes-no questions are investigated; thus, thirty sentences were collected (2 sentence types x 3 repetitions x 5 speakers) for the auditory analysis. For the visual analysis, only twenty stimuli (2 sentence types x 2 repetitions x 5 speakers) were used.

2.1.2 Participants

Ten native speakers of BP from Rio de Janeiro recorded the corpus and they were either undergraduate or graduate students from Federal University of Rio de Janeiro. The mean age of the BP participants was 28.5 years. Five Mexican Spanish native speakers from Mexico City recorded the corpus and they were either undergraduate or graduate students and their mean age was 30.0 years.

2.2 Acoustic description and phonological notation

In this study, forty BP and thirty MS sentences were analyzed. The phonological notation of the nuclear region of the assertive and interrogative intonational contours was made following the Autossegmental-Metric model (Pierrehumbert, 1980). Sp_ToBI (Prieto and Roseano, 2018) was used for analyzing MS data and Portuguese_ToBI (Frota *et al.*, 2015b) for BP as well as descriptions provided by Frota *et al.* (2015a) and Moraes (1998, 2008). Results of the acoustic description are presented in section 3.1 of this article.

2.3 Visual description

The FACS manual developed by Ekman *et al.* (2002) was used to describe the facial movements performed by the speakers while uttering the stimuli for both corpora. The muscular activities that make momentary changes in the face are called Action Units (AU, henceforth). This manual is widely known in the academic research. Recently, the FACS manual was adapted to also taxonomize the facial expressions of other species, such as DogFACS (*The Dog Facial Action System*) developed by Waller *et al.* (2013), allowing cross-species comparisons of facial movements.

Based on previous audiovisual studies (Moraes *et al.*, 2012; Cruz *et al.*, 2017), the AUs located in the upper and lower part of the face as well as the head movements were selected, due to the recurrent results of visual description, showing that eyebrow and head movements are employed in the visual production of questions. The twenty-two AUs chosen to describe the facial movements in the BP and MS corpus are: inner brow raiser (AU 1), outer brow raiser (AU 2), brow lowerer (AU 4), upper lid raiser (AU 5), cheek raiser and lid compressor (AU 6), lid tightener (AU 7), nose wrinkler (AU 9), lip corner puller (AU 12), lip corner depressor (AU 15), chin raiser (AU 17), lip stretcher (AU 20), lips part (AU 25), blink (AU 45), head turn left (AU 51), head turn right (AU 52), head up (AU 53), head down (AU 54), head tilt left (AU 55), head tilt right (AU 56), head forward (AU 57), head back (AU 58) and, finally, up and down

head movement (AU 85). In addition, only the presence of AUs was recorded in this study, not the intensity of each facial movement.

The annotation of the two sets of data, including forty BP stimuli (2 sentence types x 2 repetitions x 10 speakers) and twenty MS stimuli (2 sentence types x 2 repetitions x 5 speakers), was made independently by two of the authors. The two annotators watched each video clip, using the 22 AUs to describe the facial gestures. The presence of each feature was marked based on the subjective impression of the annotators. The number of observations of each AU for each of the 40 BP and 20 MS stimuli and for each annotator constitutes the dataset with 60 stimuli. However, in two stimuli (two assertions produced by male speakers, one from each language group), the speakers did not show any AUs. They were removed from the dataset; this limits the total number of stimuli to 58. In order to evaluate the interrater agreement between the two judges, Cohen's Kappa was calculated (Cohen, 1960). It measures the interrater accuracy, corrected for agreement by chance.

The labels given by the two annotators were pooled in a contingency matrix having as lines each stimulus (58 lines) and as columns the AUs (22 columns). Each cell contains the number of times one AU was observed for a given stimuli by the two annotators. So, to study the variation of AU across illocutions, languages and speakers, this contingency table was submitted to a correspondence analysis (CA, henceforth; for details, see Husson *et al.* 2017) that extracts the main dimensions that explain the variation of annotations.

The distribution of the stimuli coordinates along the main axes of the CA are indices of the differences marked in the speakers' performances. The magnitude of these differences, according to the four levels of Modes and Languages (ML: assertion-MS, assertion-BP, question-MS, question-BP), are evaluated using a one-way multivariate analysis of variance (MANOVA; see Field *et al.*, 2012), using Pillai's trace.

The dependent variables of the MANOVA were the factor scores of stimuli along the main dimensions of the CA. Due to the small size of the data set, it is not possible to evaluate a potential effect of the random selection of speakers, whose specificities were polled together for the MANOVA. After the MANOVA, follow-up Kruskal-Wallis tests were run to evaluate the significance of each dependent variable; a Bonferroni adjusted alpha level of 0.0125 was used. Then, pairwise comparison between the four levels of the independent variable ML were run to check which levels significantly differ from the other; a Games-Howell post-hoc test was used, with Tukey adjustment of the *p* values. The results of the visual description are exposed in section 3.2 of this article.

3 Results

The outcome of the production analysis, including the auditory as well as the visual data of BP and MS, is presented in this section.

3.1 Auditory analysis

Specific intonational contours mark the difference between BP assertions and echo questions. Both BP F0 contours can be seen in Figure 1.

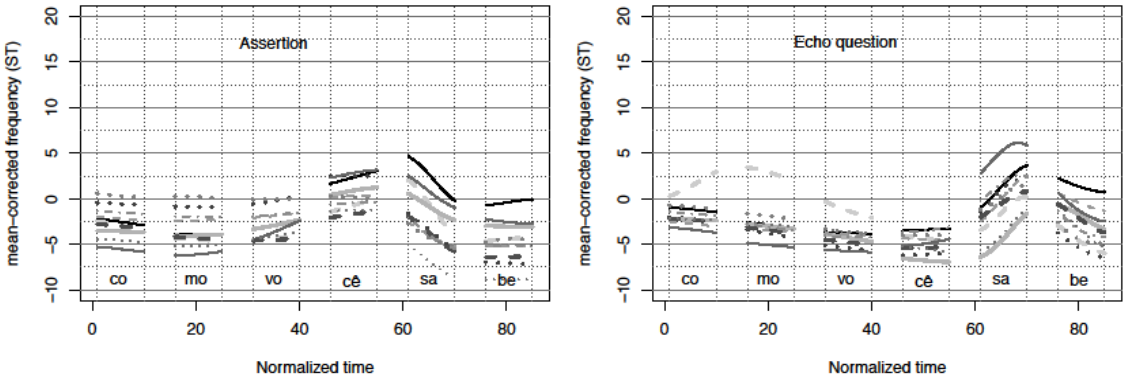

Figure 1: Example of the F0 contours of the sentence "*Como você sabe*" produced as an assertion (to the left) and as an echo-question (to the right) by each one of the ten speakers (mean of ten repetitions), retrieved from Miranda (2019).

Regarding the BP assertions, the results showed that there is a flat F0 movement on the prenuclear region of the intonational contour. In the nuclear region of the assertions, there is a rising F0 movement on the pre-stressed syllable "cê", which is followed by a F0 falling movement on the nuclear stressed syllable "sa". The BP echo-question contour starts, overall, with a slight falling F0 movement on the prenuclear region of the contour until the nuclear pre-stressed syllable "cê". After that, a rising F0 movement takes place at the nuclear stressed syllable "sa" and the F0 alignment is at the right edge of the nuclear stressed syllable. This rising movement is followed by a final F0 fall in the post-stressed syllable "be".

Note that in Fig. 1 one BP speaker produced the echo question with a rising-falling F0 movement in the prenuclear region of the contour. This type of F0 configuration is also predicted for yes-no questions in BP intonation (Moraes 1998, 2008; Miranda, 2015). Miranda *et al.* (2020b) applied perceptual identification tests with the same production data set of the present study as auditory stimuli and the production of echo questions of all speakers was recognized as echo questions, which shows that both F0 configurations represent the same illocution (i.e., the functional value of the utterance is the same). Furthermore, Miranda (2015) showed in a perceptual analysis of BP yes-no questions that the absence of the prenuclear rising-falling movement of the neutral yes-no question does not compromise the perceptual identification of the illocution. That may explain the variation of the F0 configuration found in the prenuclear region of this type of interrogative intonational contour.

Based on the BP intonation literature (Moraes 1998, 2008), we propose the phonological notation of the BP assertive intonational contours as H+L*L% in the nucleus. The echo questions receive the pitch accent L+<H*L% in the nuclear region, with the diacritic (<) to indicate a late nuclear F0 peak. The results from the Rio de Janeiro variety are in line with Frota *et al.*'s (2015a) description of BP echo questions, since the nuclear rising F0 plus the low boundary was also verified in the São Paulo and Minas Gerais varieties. Although echo questions in the Bahia and Rio Grande do Sul varieties also present a rising nuclear F0 movement, they exhibit a high boundary tone. Our findings also support Frota *et al.*'s (2015a) argument that the neutral echo questions and neutral yes-no questions in Portuguese are remarkably similar in terms of intonational movements. For a detailed analysis of phonetic information of the assertions and echo question speech acts in BP regarding F0, intensity and duration measures, see Miranda *et al.* (2020a).

Concerning the MS description of the intonational patterns, we can observe different F0 patterns for assertions and yes-no questions (cf. Figure 2).

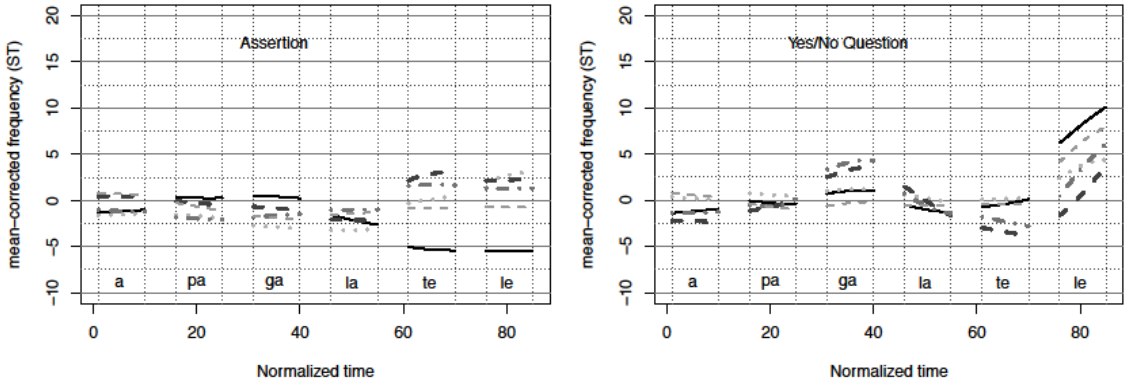

Figure 2: Example of the F0 contours of the sentence "*Apaga la tele*" produced as an assertion (to the left) and as a yes-no question (to the right) by each one of the five speakers (mean of three repetitions), adapted from Gomes da Silva (2019).

In the MS assertions, there is a slightly falling F0 movement in the prenucleus, and the nucleus presents a low pre-stressed syllable "la" followed by a high stressed syllable "te" that is maintained as a plateau in the post-stressed syllable "le". The occurrence of this pattern shares similarities with the MS circumflex F0 contour, regarding the nuclear low pre-stressed syllable and high stressed syllable, as previously described by Sosa (1999), De-la-Mota et al. (2010) and Prieto and Roseano (2018). We also verified the occurrence of another F0 contour in the assertion production characterized by a low initial peak followed by a falling F0 in the nuclear stressed syllable "te" that is extended to the final post-stressed syllable "le" (cf. Figure 2). The downward pattern for MS assertions is described in De-la-Mota *et al.* (2010) as well.

Thus, we can propose by the Sp_ToBI notation two types of F0 nuclear configuration in the assertive contour: L+H*L% and L*L%. It is worth noting that while the former F0 configuration is regularly recognized by MS listeners, the latter presents low recognition rates (Gomes da Silva, 2019). In addition, productions of the assertive speech act with a rising nuclear F0 configuration were also verified, although they were not recognized in perceptual tests as assertions (Gomes da Silva, 2019).

As for MS yes-no questions, the MS melodic contour presents a rising F0 movement with a displaced peak at the beginning of the pre-stressed nuclear syllable "la", low nuclear stressed syllable "te" and rising boundary tone (cf. Figure 2), as described in the literature (cf. Sosa, 1999; De-la-Mota *et al.*, 2010, Guimarães, 2018). According to the prosodic notation of Sp_ToBI (Prieto and Roseano, 2018), the yes-no question receives the notation L*LH% for the nucleus.

To sum up, both languages present specific auditory cues that differentiate the two speech acts. The main difference between the assertive and interrogative speech acts in BP and MS relies on the F0 configuration in the nuclear region of the contour.

3.2 Visual analysis

After the presentation of the description method of the AUs made in section 2.3 of this article, in this section the results of the statistical analysis are presented. The agreement between the two annotators of the BP and MS visual stimuli was measured by Cohen's Kappa (1960). A Kappa of 0.404, significantly superior to chance ($z = 14.7$, $p < 0.05$), was observed. This supports the grouping of both raters' scorings.

The five first dimensions of the CA explain more than half of the variance and were kept for further analysis (considering the remaining ones as noise). Results of the CA output are summarized in Table 1 for the columns of the contingency matrix (Action Units). It shows the main links between the first five dimensions (a selection criterion based on the contribution of the column to the axis that shall be above the mean contribution, i.e. 5, and the axis that shall have a squared cosine above 0.2 for this column. See Abdi and Williams, 2010).

Table 1: Factor scores (fs), contribution (ct) and cos^2 (squared cosines, multiplied by 100 and rounded for convenience) of each columns of the contingency matrix (i.e. the AUs) for the first five dimensions of the CA. Negatively scored AU are shown in italics. AUs that have the main link with a given axis are shown in bold face.

AU	Dim 1			Dim 2			Dim 3			Dim 4			Dim 5		
	fs	ct	cos^2	fs	ct	cos^2	fs	ct	cos^2	fs	ct	cos^2	fs	ct	cos^2
01	*-0.5*	*3*	*11*	0.9	9	31	*-0.1*	*0*	*0*	*-0.1*	*0*	*1*	0.3	1	3
02	*-0.6*	*5*	*15*	**1.1**	**22**	**55**	*-0.1*	*0*	*0*	*-0.4*	*3*	*7*	0	0	0
04	**1**	**17**	**49**	*-0.4*	*3*	*8*	0.1	0	1	*-0.1*	*0*	*0*	***-0.6***	***13***	***21***
05	*-0.1*	*0*	*1*	**1.1**	**16**	**55**	0.2	0	1	*-0.4*	*3*	*8*	*-0.1*	*0*	*0*
06	1.3	3	16	1	2	9	*-0.1*	*0*	*0*	**1.6**	**7**	**27**	0.7	1	5
07	**1.1**	**16**	**50**	*-0.3*	*1*	*3*	0.2	1	2	*-0.1*	*0*	*0*	*-0.6*	*7*	*12*
09	**1.3**	**9**	**26**	0.9	5	13	0.1	0	0	**1.4**	**18**	**33**	0.1	0	0
12	*-1.4*	*5*	*17*	*-0.7*	*2*	*5*	*-0.5*	*1*	*2*	1.1	4	11	*-0.4*	*1*	*1*
15	1.5	1	7	0.1	0	0	1.1	1	4	0.9	1	2	0.7	1	2
17	1.2	3	12	1.3	5	15	*-0.5*	*1*	*2*	**2**	**14**	**34**	0.4	1	1
20	*-0.6*	*2*	*7*	***-1.1***	***10***	***26***	*-0.5*	*3*	*7*	0.2	0	1	**1.1**	**15**	**26**
25	*-0.7*	*3*	*13*	0.1	0	0	*-0.2*	*0*	*1*	*-0.2*	*1*	*1*	*-0.6*	*5*	*12*
45	***-0.5***	***7***	***20***	*-0.3*	*3*	*7*	0	0	0	*-0.2*	*1*	*3*	*-0.3*	*5*	*9*
51	0.3	0	1	*-0.4*	*1*	*2*	**1.7**	**13**	**24**	0.9	4	8	0.5	1	2
52	*-0.5*	*2*	*7*	0.1	0	0	*-0.5*	*2*	*7*	0.8	7	18	*-0.1*	*0*	*0*
53	0.4	0	1	*-0.4*	*0*	*1*	*-0.1*	*0*	*0*	*-0.5*	*1*	*2*	*-1.7*	*9*	*17*
54	0.5	4	8	*-0.7*	*9*	*18*	*-0.5*	*6*	*9*	***-0.7***	***13***	***20***	0.5	6	8
55	0	0	0	0	0	0	**2.4**	**46**	**64**	*-0.5*	*2*	*3*	0.8	7	7
56	***-1.1***	***14***	***32***	*-0.8*	*8*	*15*	0.2	0	1	**1**	**17**	**27**	*-0.2*	*1*	*1*
57	0.7	4	9	0.2	0	0	***-1.2***	***18***	***28***	*-0.1*	*0*	*0*	0.9	14	16
58	*-0.9*	*0*	*2*	2.3	4	12	0.1	0	0	*-1.3*	*1*	*4*	*-0.3*	*0*	*0*
85	0.1	0	0	*-0.4*	*1*	*2*	1.4	7	17	*-0.3*	*0*	*1*	1.6	13	22

The performed Action Units can be described as follows (AUs are given in decreasing order of association with the axis):

- Dimension 1 is built positively on AU 4 (eyebrow lowerer), AU 7 (lid tightener), and AU 9 (nose wrinkler) – and negatively on AU 56 (head tilt right) and AU 45 (blink)
- Dimension 2 is built positively on AU 2 (outer brow raiser) and AU 5 (upper lid raiser) – and negatively on AU 20 (lip stretcher)
- Dimension 3 is built positively on AU 51 (head turn left)
- Dimension 4 is built positively on AU 17 (chin raiser), AU 9 (nose wrinkler) and AU 6 (cheek raiser and lid compressor)
- Dimension 5 is built positively on AU 20 (lip stretcher) – and negatively on AU 4 (eyebrow lowerer)

Table 2: Factor scores (fs), and cos^2 (squared cosines, multiplied by 100 and rounded for convenience) of each category of the supplementary variable Mode and Language (that groups rows with these characteristics) for the first five dimensions of the CA. Negatively scored AUs are shown in italics. Categories that are well represented by a given axis are shown in bold face.

Supl. var.	Dim 1		Dim 2		Dim 3		Dim 4		Dim 5	
	fs	cos^2	fs	cos^2	fs	cos^2	fs	cos^2	fs	cos^2
Ass MS	*-0.4*	*18*	***-0.7***	***48***	*-0.4*	*15*	0.1	2	0.2	6
Ass BP	***-0.5***	***58***	0.0	0	0.3	17	*-0.2*	*5*	0.2	6
QE MS	**0.5**	**53**	0.0	0	*-0.2*	*5*	0.0	0	*-0.1*	*2*
QE BP	**0.3**	**37**	**0.3**	**26**	0.0	0	0.1	2	*-0.2*	*10*

Table 2 shows the factor score and squared cosines for the supplementary (i.e. it did not take part in the analysis) variable ML, which groups set of rows (stimuli) of the contingency matrix that share the same characteristics (same mode and language). It shows which of the first five dimensions have a good representation of the levels of ML (the axis that shall have a squared cosine above 0.2. See Abdi and Williams 2010). Levels of ML can be described as follow:

- MS assertions are linked negatively with the second axis
- BP assertions are linked negatively with the first axis
- MS and BP questions are linked positively with the first axis

The MANOVA run on factor score for the first five dimensions showed that there was a statistically significant difference between the four types of ML ($V = 0.80$, $F(15, 156) = 3.78$, $p < 0.05$), on the combined dependent variables. Follow-up Kruskal-Wallis tests showed that only the spread of stimuli on the first two dimensions was done with a statistically significant difference between ML levels.

Pairwise comparisons between groups applied to these two dimensions (cf. Figure 3 and Table 3) showed that dimension 1 exhibits significant differences in the position of BP assertions with respect to the position of both BP and MS questions (the distance being greater for BP speakers). Dimension 2 showed significant differences in the position of MS assertions from the positions of the three other categories (BP assertion, MS and BP questions, in increasing order).

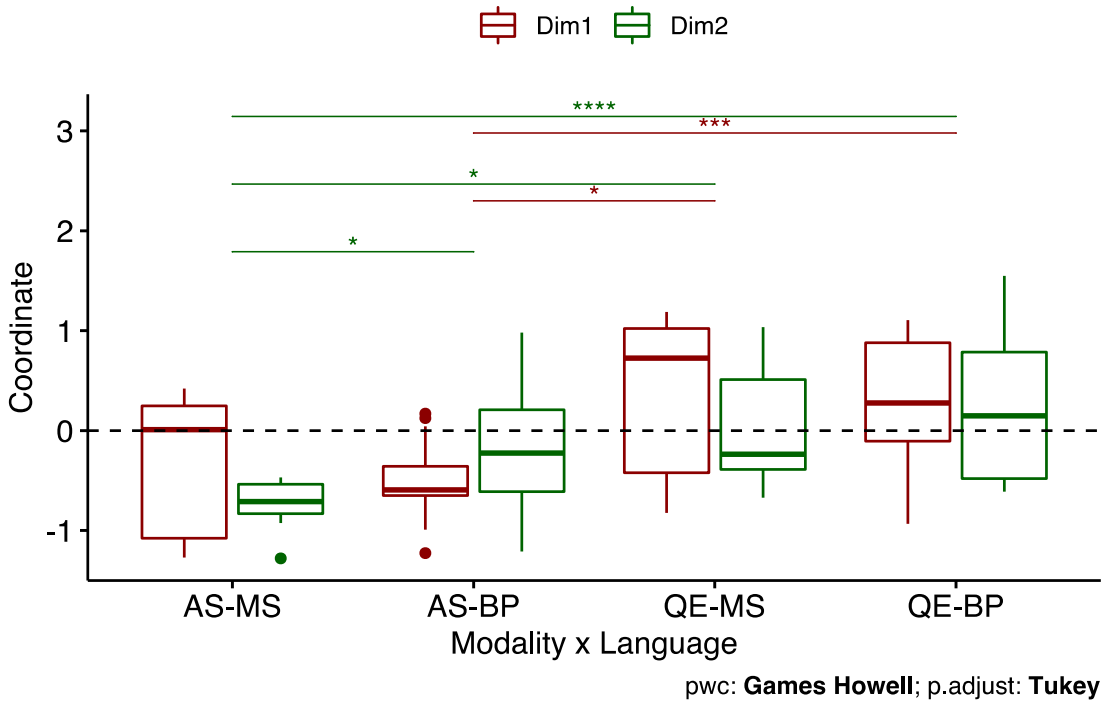

Figure 3: Boxplots showing the distribution of the stimuli's coordinates over the first (red or left boxes) and second (green or right boxes) dimensions of the correspondence analysis, grouped by modes (assertion AS, question QE) and languages (MS, PB); significant differences between groups are indicated by the lines above the plot (see text). The dashed horizontal line indicates the center of the coordinate system.

Table 3: Adjusted p values are reported for the two dependent variables (DV) that showed variations, output of pairwise comparison (between two levels of the ML independent variable), based on Games Howell post-hoc tests. Significant comparisons are shown in bold face.

DV	Level 1	Level 2	Adjusted p
Dim1	AS-MS	AS-PB	0.890
Dim1	AS-MS	QE-MS	0.202
Dim1	AS-MS	QE-PB	0.185
Dim1	**AS-PB**	**QE-MS**	**0.030**
Dim1	**AS-PB**	**QE-PB**	**0.001**
Dim1	QE-MS	QE-PB	0.973
Dim2	**AS-MS**	**AS-PB**	**0.013**
Dim2	**AS-MS**	**QE-MS**	**0.014**
Dim2	**AS-MS**	**QE-PB**	**0.000**
Dim2	AS-PB	QE-MS	0.822
Dim2	AS-PB	QE-PB	0.284
Dim2	QE-MS	QE-PB	0.892

The result of the MANOVA first shows that assertions and questions differ in terms of AU used to perform them. As described earlier, the first CA dimension opposes BP assertions (on the negative part of the axis) to the two types of questions (on the positive part of the axis). Thus, AU 56 (head tilt right) and AU 45 (blink) are the most related with BP assertions, while

AU 4 (eyebrow lowerer), AU 7 (lid tightener) and AU 9 (nose wrinkler) are the most related to questions for both language groups.

The results for the second dimension show that MS assertions differ from the other expression along this axis, these expressions tending to be located on the negative part of the axis–thus MS assertions are mostly performed with AU 20 (lip stretcher).

Examples of the prototypical facial gestures for each speech act in both languages can be seen in Figures 4 to 7.

Figure 4: BP Female speaker producing an assertion (above) with AU 55 (head tilt left) and an echo question (below) with AU 4 (brow lowerer) and AU 7 (lid tightener).

Figure 5: BP male speaker producing an assertion (above) with AU 53 (head up) and an echo question with the AU 4 (eyebrow lowerer), AU 7 (lid tightener), AU 9 (nose wrinkler), AU 57 (head forward) and AU 17 (chin raiser).

Figure 6: MS male speaker producing an assertion (above) with AU 20 (lip stretcher) and a yes-no question (below) with AU 4 (brow lowerer) and AU 9 (nose wrinkler).

Figure 7: MS male speaker producing an assertion (above) with head forward (AU 57) and AU 55 (head tilt left) and a yes-no question (below) with AU 4 (brow lowerer), AU 9 (nose wrinkler) and AU 57 (head forward).

In summary, the visual analysis of BP and MS data set showed that the stimuli can be sorted according to their intended speech acts on the basis of visual patterns produced by the speaker: thus, there is a coherence in the way speaker uses visual cues while performing assertions and questions. The analysis revealed that questions are produced in both languages with the same facial gestures, including eyebrow lowering, lid tightening and nose wrinkle. For assertions, there are different facial expressions: in BP, the head tilt movement plus blink are characteristic visual actions and, in MS, this speech act is linked to lip stretching as a visual cue.

4 General discussion

In this paper, the multimodal production of questions and assertions in Brazilian Portuguese and Mexican Spanish were described and compared. Our hypotheses, stated in the introduction of this study, was that both languages would present specific auditory and visual cues for characterizing each speech act.

The acoustic analysis showed that both languages encode the speech acts with specific prosodic cues. Whereas, in BP, a rising F0 in the nuclear stressed syllable is typical of echo questions, and a falling F0 in the nuclear stressed syllable is typical of assertions, in MS yes-no questions are characterized by a low stressed syllable followed by a rising boundary tone, and the assertions are characterized by either a falling-rising or a flat low F0 movement in the nucleus.

The visual analysis in the Brazilian Portuguese and Mexican Spanish data disclosed different facial expressions produced by speakers when uttering assertions. Whereas BP assertions were produced with head tilt right plus blink, the MS results showed that assertions are produced with lip stretcher. The facial expressions for both assertions showed a relatively reduced expressivity, which can be explained by the neutral character of the speech act (i.e., without expressive marks). On the other hand, both BP and MS questions are produced with brow lowerer, lid tightener and nose wrinkler, which may be related to the search for unknown information from the speaker conveyed by the speech act. The conclusion is that lowering the brow, tightening the eyes, and wrinkling the nose are question markers for the two types of questions in both languages.

Additionally, in Portuguese Sign Language (Cruz *et al.*, 2019), eyebrow lowering was also identified as a question marker for both wh-questions and yes-no questions; the difference is related to the up head movement in the former and down head movement in the latter. BP wh-questions were produced with eyebrow lowering as well, plus head turning right (Miranda *et al.*, 2019). In Catalan (Borràs-Comes and Prieto, 2011), echo questions were produced with eyebrow lowerer and head backward. In Swedish (House, 2002), yes-no questions were produced by a talking head with eyebrow lowering and vertical head tilting as visual cues, although the author concluded that these visual cues had little effect in the perceptual identification. Therefore, Brazilian Portuguese, Mexican Spanish, Portuguese Sign Language, Catalan and Swedish languages seem to share the same visual cue as a question marker.

It is also worth mentioning that there are other visual cues considered question markers in previous bimodal studies. Eyebrow raising and head tilt were the visual cues found in English echo questions (Srinivasan and Massaro, 2003). In European Portuguese (Cruz *et al.*, 2015), yes-no questions were also produced with eyebrow raising and either up-down or back-forward head movements. Furthermore, in Catalan and Dutch polar questions, eyebrow raising was also detected as a visual cue. In French (Torreira and Valtersson, 2015), polar questions were more often produced in conversational interactions with eye gaze towards the addressee. However, while the gaze also occurred with statements, eyebrow raising and head movements, despite the low frequency, always occurred with questions rather than statements. Hence, eyebrow raising is considered a question marker in American English, European Portuguese, French, Catalan and Dutch languages.

Up until now, different types of questions and pragmatic meanings have been analyzed in previous works; furthermore, general observations are found in the literature. For instance, Cruz *et al.* (2019), analyzing facial expressions in Portuguese Sign Language, proposes that the eyebrow movement pattern is related to the interrogativity meaning (i.e., the pragmatic value of the utterance), whereas the head movement pattern seems to be related to the type of interrogative sentence (e.g., yes-no question and wh-questions). Nevertheless, the author also stated that, in other sign languages studies, eyebrow patterns seem to distinguish interrogative

sentence types: furrowed eyebrows for wh-questions and raised eyebrows for yes-no questions, which is also mentioned by Borràs-Comes and Prieto (2011). Additionally, in line with these authors, a similar result was found in Brazilian Sign Language (Libras), since furrowed brow plus head up/elevation were described for wh-questions and raising eyebrows along with head lowering for yes-no questions (Paiva *et al.*, 2016).

We conclude that there are similar visual cues conveying question meaning (yes-no questions, echo questions and wh-questions) in different languages, especially regarding the eyebrow movements, which can be often accompanied by head movements. In the literature, several works indicated that eyebrow movements combined with head movements are also used to signal other speech functions, such as prosodic focus (Borràs-Comes and Prieto, 2011). However, De la Cruz-Pavía *et al.* (2019) have pointed out that the head movements seem to be more relevant than eyebrow movements in signaling prominence and focus, as well as prosodic phrase boundaries, due to its stronger visual saliency (i.e., the head occupies a larger surface than the eyebrows). Studies measuring the saliency of each visual cue described in the literature for the expression of question meaning are still needed.

As for the facial gestures of BP echo questions, the visual analysis revealed that one speaker presented a specific facial muscular activity (cf. Fig. 5): the chin raiser (AU 17), which can be considered a "shrug" gesture investigated by Debras (2017). According to this author, the "shrug" gesture is a kinetic ensemble that can be expressed by different body movements, such as hand gesture, movement of the shoulder, a particular facial expression and head movement. A "*prototypical*" shrug includes a combination of some of these movements, but it may occur with just one of these gestures. For instance, the shoulder shrug is a commonly employed form and the mouth shrug can be expressed by the face only. Debras (2017) showed that, in previous studies on the literature, shrugs express (inter)personal attitudes related to some functions. For instance, the "mouth shrug", which can be performed by the face only, is used to convey the pragmatic meaning of "disclaimer", expressing messages such as 'I don't know', 'It's nothing to do with me' or 'I don't understand'.

In her study, in which the expression of shrug gestures, including the "mouth shrug", was analyzed in co-occurrence with speech, Debras (2017) argued that the shrug is a more complex and dynamic network of related forms and functions, rather than just an emblem. Some of these functions identified in the corpus analysis can be divided into three groups: epistemic meaning (epistemic indetermination and common ground), attitudes (incapacity, inaction and submissiveness) and affect (indifference and rejection). Taking this distinction into account, in the case of the facial expression composed of raised chin gesture (AU 17) plus the lip corner depressor (AU 15) and nose wrinkler (AU 9) employed by the Brazilian speaker in the production of echo questions, the "mouth shrug" expresses in this context an epistemic indetermination, that is, a sign of doubt linked to the pragmatic meaning of the speech act of echo question. Further investigation on shrug gestures in co-occurrence with speech in other types of sentences in BP is needed to determine which other functions can be expressed by this gesture.

In sum, our findings showed that the acoustic component together with the gesture component are produced to reach the speakers' communicative aim in both languages. The analysis also revealed that questions are produced with more facial movements than are assertions. In addition, BP and MS data presented the same visual cues as questions markers in our visual analysis. The results of this paper support previous investigations on the multimodality of speech production (Gili Fivela, 2018), by showing that there are facial gestures which are pragmatically used to mark the illocutionary force of an utterance, as stated by Kendon (2004: 5).

5 Conclusion

This paper presented an analysis of the production of assertions and questions in two different languages: Brazilian Portuguese and Mexican Spanish. Our first hypothesis was confirmed. Questions and assertions in both languages showed specific prosodic strategies. In BP the F0 rising in the nuclear stressed syllable (L+<H*L%) distinguishes echo questions from assertions, which present a falling nuclear F0 configuration (H+L*L%). In MS, assertions present either a falling-rising F0 nuclear configuration (L+H*L%) or a flat low nuclear F0 (L*L%), whereas yes-no questions show a low nuclear F0 in the stressed syllable followed by a rising boundary tone configuration (L*LH%).

The second hypothesis was partially confirmed since only BP and MS assertions presented different visual cues, whereas questions in both languages were produced with similar facial gestures. The visual description revealed that questions in both languages are produced with eyebrow lowering, which is a recurrent question marker found in other languages investigated in the audiovisual prosody literature, plus lid tightening and nose wrinkle. The facial expression for assertions is different in both languages: while in BP this speech act is produced with blink and head tilt right, in MS, the visual cue is the lip stretcher.

The next step, in furthering this study, will be to compare the perceptual salience of these acoustic and visual cues in both languages with identification tests. In addition, based on the positive results of this study, further investigation of the multimodal production of other sentence types in different languages are encouraged. For example, the multimodal analysis of other directive speech acts in Brazilian Portuguese and Mexican Spanish, including commands and supplications, is left for future research.

Acknowledgments

Part of this research has been funded by the scholarship 88882.331896/2015-01 from the Brazilian Federal Agency for Support and Evaluation of Graduate Education–CAPES awarded by the first author during her PhD course at Federal University of Rio de Janeiro, Brazil. We also thank the two anonymous reviewers for their careful reading of our manuscript and their insightful comments and suggestions.

REFERENCES

1. Abdi H, Williams, LJ. *Correspondence Analysis. In: Salkind NJ* (Ed.), Encyclopedia of Research Design, Thousand Oaks, CA: Sage, 2010.
2. Barkhuysen P, Krahmer E, Swerts M. *Cross-modal and incremental perception of audiovisual cues to emotional speech.* Language and speech, 53 (1), pp. 3-30, 2010.
3. Borràs-Comes J, Kaland C, Prieto P, Swerts M. *Audiovisual Correlates of Interrogativity: A comparative analysis of Catalan and Dutch.* Journal of Nonverbal Behavior, 38, pp. 53-66, 2013. DOI: https://doi.org/10.1007/s10919-013-0162-0
4. Borràs-Comes J, Prieto P. 'Seeing tunes.' *The role of visual gestures in tune interpretation.* Laboratory Phonology, 2 (2), pp. 355–380, 2011.
5. Cavé C, Guaitella I, Bertrand R, Santi S, Harlay F, Espesser R. *About the relationship between eyebrow movements and F0 variations.* Proceedings of the 4th International Conference on Spoken Language Processing, Philadelphia, EUA, ICSLP, pp. 2175–2179, 1996.
6. Cohen J. *A coefficient of agreement for nominal scales.* Educational and psychological measurement, 20(1), 37-46, 1960.
7. Crespo Sendra V, Kaland C, Swerts M, Prieto P. *Perceiving incredulity: the role of intonation and facial gestures*. Journal of Pragmatics, 47, pp. 1-13, 2013.
8. Cruz M, Swerts M, Frota S. *Do visual cues to interrogativity vary between language modalities? Evidence from spoken Portuguese and Portuguese Sign Language*. Proceedings of the 15th International Conference on Auditory-Visual Speech Processing 10-11 August 2019, Melbourne, Australia, pp. 1–5, 2019.
9. Cruz M, Swerts M, Frota S. *The role of intonation and visual cues in the perception of sentence types: Evidence from European Portuguese varieties*. Laboratory Phonology, 8 (1), 23, 2017.
10. Cruz M, Swerts M, Frota S. *Variation in tone and gesture within language*. Proceedings of the 18th International Congress of Phonetic Sciences, Glasgow, UK: The University of Glasgow, paper number 452, 2015.
11. Darwin C. *The Expression of the Emotions in Man and Animals*. London: J. Murray, 1872.
12. Debras C. *The shrug: Forms and meanings of a compound enactment*. Gesture, *16* (1), pp. 1–34, 2017. DOI: https://doi.org/10.1075/gest.16.1.01deb
13. De la Cruz-Pavía I, Werker JF, Vatikiotis-Bateson E, Gervain J. *Finding phrases: The interplay of word frequency, phrasal prosody and co-speech visual information in chunking speech by monolingual and bilingual adults*. Language and Speech, 2019. DOI: https://doi.org/10.1177/0023830919842353
14. De-la-Mota C, Butragueño PM, Orozco L, Prieto P. *Mexican Spanish Intonation*. In: Pietro P, Roseano P. (org.). Transcription of Intonation of the Spanish Language. München: Lincom Europa, pp. 319-350, 2010.
15. Dohen M, Loevenbruck H. *Interaction of audition and vision for the perception of prosodic contrastive focus.* Language and Speech, 52 (2/3), pp. 177–206, 2009.
16. Ekman P, Friesen WV, Hager JC. *The Facial Action Coding System*. Salt Lake City: Research Nexus, 2002.
17. Field AP, Miles J, Field Z. *Discovering Statistics Using R*. London: SAGE Publications Ltd, 2012.
18. Frota S, Cruz M, Svartman FRF, Collischonn G, Fonseca A, Serra CR, Oliveira P, Vigario M. *Intonational variation in Portuguese: European and Brazilian varieties*. In: Frota S, Prieto P. (Org.). Intonation in Romance. 1ed. Oxford: Oxford University Press, v. 1, pp. 235-283, 2015a.
19. Frota S, Oliveira P, Cruz M, Vigário M. *P-ToBI: tools for the transcription of Portuguese prosody*. Lisboa: Laboratório de Fonética, CLUL/FLUL, 2015b. ISBN: 978-989-95713-9-6. [http://labfon.letras.ulisboa.pt/InAPoP/P-ToBI/]
20. Gili Fivela B. *Multimodal analyses of audio-visual information: Some methods and issues in prosody research*. In: Feldhausen I, Fliessbach J, Vanrell MM (Eds.). Methods in prosody: A Romance

language perspective (Studies in Laboratory Phonology 4). Berlin: Language Science Press, pp. 83-122, 2018.

21. Gomes da Silva C. *A prosódia de atos de fala no espanhol da Cidade do México*. Rio de Janeiro, 2019. Tese de Doutorado em Língua Espanhola – Faculdade de Letras, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2019.
22. González-Fuente S. *La prosodia audiovisual de la ironía verbal: un estudio de caso*. Revista Española de Lingüística 45/2, pp. 73-103, 2015.
23. Guimarães DP. *Análise prosódica de enunciados interrogativos totais de conversas coloquiais de fala espontânea na variedade mexicana*. Dissertação de Mestrado em Língua Espanhola. Faculdade de Letras, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2018.
24. Hadar U, Steiner, TJ, Grant EC, Clifford Rose F. *Head movement correlates of juncture and stress at sentence level.* Language and Speech, 26, pp. 117–129, 1983.
25. House D. *Intonational and visual cues in the perception of interrogative mode in Swedish.* Proceedings of the 7th International Conference on Spoken Language Processing, Denver, Colorado, pp. 1957-1960, 2002.
26. Husson F, Lê S, Pagès J. *Exploratory Multivariate Analysis by Example Using R*. 2nd edition. Chapman & Hall/CRC, 2017.
27. Kaminski J, Hynds J, Morris P, Waller B. *Human attention affects facial expressions in domestic dogs.* Sci Rep 7, 12914, 2017. https://doi.org/10.1038/s41598-017-12781-x
28. Kendon A. *Gesture: Visible Action as Utterance*. Cambridge: Cambridge University Press, 2004.
29. Lalonde K, Holt RF. *Preschoolers Benefit From Visually Salient Speech Cues*. J Speech Lang Hear Res., 58 (1), pp. 135–150, 2014. Doi: 10.1044/2014_JSLHR-H-13-0343
30. Landis JR, Koch CG. *The measurement of observer agreement for categorical data*. Biometrics, *33* (1), pp. 159–174, 1977.
31. Lewkowicz DJ. *Infants' Perception of the Audible, Visible, and Bimodal Attributes of Multimodal Syllables*. Child Development, 71, pp. 1241-1257, 2003. doi:10.1111/1467-8624.00226
32. Lewkowiz DJ, Hansen-Tift, AM. Infants deploy selective attention to the mouth of a talking face when learning speech. *Proceedings of the National Academy of Sciences*, Jan 2012, 109 (5), pp. 1431-1436, 2012. DOI: 10.1073/pnas.1114783109
33. McGurk H, MacDonald J. *Hearing lips and seeing voices.* Nature, 264, pp. 746–748, 1976.
34. Miranda LS. *Estudo fonético-perceptivo da entoação de enunciados assertivos, interrogativos e exclamativos do português do Brasil: uma análise multimodal*. 243 f. Tese (Doutorado em Letras Vernáculas – Língua Portuguesa) – Faculdade de Letras, Universidade Federal do Rio de Janeiro, 2019.
35. Miranda LS. *Análise da entoação do português do Brasil segundo o modelo IPO*. 161f. Dissertação de mestrado (Letras Vernáculas – Língua Portuguesa) – Faculdade de Letras, Universidade Federal do Rio de Janeiro, Brazil.
36. Miranda L, Moraes J, Rilliard A. *Statistical modeling of prosodic contours of four speech acts in Brazilian Portuguese.* Proceedings of the 10th International Conference on Speech Prosody, May 25-28, Tokyo, Japan, p. 404-408, 2020a. DOI: 10.21437/SpeechProsody.2020-83.
37. Miranda L, Swerts M, Moraes J, Rilliard A. *The role of the auditory and visual modalities in the perceptual identification of Brazilian Portuguese statements and echo questions*. Language and Speech, 2020b. DOI: https://doi.org/10.1177/0023830919898886
38. Miranda LS, Moraes J, Rilliard A. *Audiovisual perception of wh-questions and wh-exclamations in Brazilian Portuguese*. Proceedings of the 19th International Congress of Phonetic Sciences, August 5-9, Melbourne, Australia, pp. 2941–2945, 2019.
39. Moraes J. The pitch accents in Brazilian Portuguese: analysis by synthesis. *Proceedings of the 4th International Conference on Speech Prosody*, Campinas, Brazil, pp. 389-397, 2008.
40. Moraes J. *Intonation in Brazilian Portuguese*. *In*: Hirst D, Di Cristo A. (eds.). Intonational Systems: a survey of twenty languages. Cambridge. MIT Press, 1998.

41. Moraes J, Rilliard A. *Illocution, attitudes and prosody: a multimodal analysis*. In: Raso T, Mello H. (Eds.). Spoken Corpora and Linguistic Studies. Amsterdam: John Benjamins, 2014.

42. Moraes J, Miranda LS, Rilliard A. *Facial gestures in the expression of prosodic attitudes in Brazilian Portuguese.* Proceedings of Seventh GSCP International Conference Speech and Corpora, Belo Horizonte, Brazil, pp. 157–161, 2012.

43. Moraes J, Rilliard A, Mota B, Shochi T. Multimodal Perception and production of attitudinal meaning in Brazilian Portuguese. *Proceedings of the International Conference on Speech Prosody*, 5, Chicago, USA, paper 340, 2010.

44. Paiva FAS, Martino JM, Barbosa PA, Benetti A, Silva IR. *Um sistema de transcrição para língua de sinais brasileiras: o caso de um avatar.* Revista do Gel, 13 (3), pp. 12-48, 2016.

45. Peres DO, Raposo de Medeiros B, Ferreira Netto W, Baia MFA. The role of the visual stimuli in the perception of prosody in Brazilian Portuguese. *Proceedings of Fifth Conference on Laboratory Approaches to Romance Phonology*, Somerville, MA, USA, pp. 136–141, 2011.

46. Pierrehumbert J. *The phonology and phonetics of English intonation*. Bloomington: Indiana University Linguistics Club. PhD thesis, MIT, 1980. [Published 1987 by IULC edition, Bloomington, IN.].

47. Prieto P, Roseano P. Prosody: Stress, Rhythm, and Intonation. In: Geeslin KL. (ed.) *The Cambridge Handbook of Spanish Linguistics*. Cambridge: Cambridge University Press, pp. 211-236, 2018.

48. Scheider L, Waller BM, Oña L, Burrows AM, Liebal K. *Social use of facial expressions in hylobatids*. PloS One 11, e0151733, 2016.

49. Schmidt KL, Cohn JF. 2018. *Human Facial Expressions as Adaptations: Evolutionary Questions in Facial Expression Research*. Am J Phys Anthropol. Suppl 33, pp. 3–24, 2001. Doi: 10.1002/ajpa.2001

50. Sosa JM. *La entonación del español. Su estructura fónica, variabilidad y dialectología*. Madrid: Cátedra, 1999.

51. Srinivasan RJ, Massaro DW. P*erceiving prosody from the face and voice: Distinguishing statements from echoic questions in English.* Language and Speech, 46 (1), pp. 1–22, 2003.

52. Swerts M, Krahmer E. *Congruent and incongruent audiovisual cues to prominence. In*: Bel B, Marlin I. (eds.), Proceedings of 2nd International Conference on Speech Prosody, Nara, Japan, pp. 69-72, 2004.

53. Torreira F, Valtersson E. *Phonetic and visual cues to questionhood in French*. *Phonetica*, 72, pp. 20-42, 2015.

54. Vegas Pro software, Version 14 of Vegas Pro. Copyright © [2016] *MAGIX*. Software available at: https://www.vegascreativesoftware.com/.

55. Waller BM, Caeiro CC, Davila-Ross M. *Orangutans modify facial displays depending on recipient attention*. *PeerJ 3*, e827, 2015.

56. Waller BM, Peirce K, Caeiro CC, Scheider L, Burrows AM, McCune S, Kaminski J. *Paedomorphic facial expressions give dogs a selective advantage*. *PloS one*, 8 (12), e82686, 2013.

57. Willis, EW. *Tonal levels in Puebla Mexico Spanish declaratives and absolute interrogatives*. In: Gess R, Rubin EJ (eds.), Theorical and experimental approaches to Romance languages, pp. 351–363, 2005.