Title page

Title: **Perceiving incredulity: the role of intonation and facial gestures**

Authors: Verònica Crespo Sendra
Grup d'Estudis de Prosòdia / Dept. de Traducció i Ciències del Llenguatge
Universitat Pompeu Fabra
carrer de Roc Boronat, 138
08018-Barcelona
Spain
veronica.crespo@upf.edu
Tel. (34) 628 120 126
Fax. (34) 93 542 1617

Constantijn Kaland
Faculty of Arts
Communication & Cognition
NL - 5000 LE Tilburg,
The Netherlands
c.c.l.kaland@uvt.nl

Marc Swerts
Faculty of Arts
Communication & Cognition
NL - 5000 LE Tilburg,
The Netherlands
m.g.j.swerts@uvt.nl

Pilar Prieto
Grup d'Estudis de Prosòdia / Dept. de Traducció i Ciències del Llenguatge
Universitat Pompeu Fabra
carrer de Roc Boronat, 138
08018-Barcelona
Spain
pilar.prieto@upf.edu

Highlights

- We analyze the interaction between facial gestures and intonation in the distinction between information-seeking and incredulity yes/no questions in two languages
- We examine changes in intonation in both languages and the interaction with gestures.
- The importance of visual cues in the identification of tune meaning is more relevant when both question types have the same contour.

Manuscript

Perceiving incredulity: the role of intonation and facial gestures

Abstract

Recently, some studies have revealed that facial gestures can play an important role in teasing out the meaning of interrogative sentence types in a particular language (Srinivasan & Massaro 2003; Borràs-Comes & Prieto 2011; among others). However, less is known about potential cross-linguistic differences. This paper investigates the interaction between facial gestures and intonation in the distinction between information-seeking and incredulity yes/no questions in two languages (i.e., Catalan and Dutch) which use different prosodic strategies to express the distinction between these two types of interrogatives. While Dutch uses two phonologically distinct intonational contours, Catalan uses the same pitch contour with a distinction in pitch range. Twenty listeners of Catalan and twenty listeners of Dutch performed a perception experiment with audio-only, video-only, and audiovisual stimuli in congruent and incongruent intonation and gestural combinations. The results reveal that there is a contrast between Dutch and Catalan listeners in the perceptual processing of these sentences. While Dutch participants rely more on intonational differences, Catalan participants use the facial expression cues to a greater extent. All in all, the results show that both languages express pragmatic contrasts both at the intonation and facial expression levels, and native speakers are highly sensitive to the relative weight of these cues at the perceptual level.

Keywords: gestures, intonation, information-seeking questions, incredulity questions intonational contrasts, Catalan, Dutch.

1. Introduction

Recent studies have demonstrated that communicative functions signaled by auditory prosody tend to be supported by visual information as well in the form of specific facial expressions (Krahmer & Swerts 2005, Dijkstra, Krahmer & Swerts 2006, Borràs-Comes & Prieto 2011, and others). Some studies have shown that visual information partially duplicates the information of auditory cues and facilitates the decoding process (Srinivasan & Massaro 2003, House 2002, Dohen & Lœvenbruck 2009). Yet, little is known about the relative contributions of gestures and intonation information in the decoding of linguistic meaning.

Interestingly, many studies have suggested that, in sign languages, facial expression may play a role which is similar to that served by intonation in speech (see, e.g., Nespor & Sandler 1999, Reilly McIntire & Bellugi 1990, Sandler 2005, Wilbur 2000, Dachkovsky & Sandler 2009). For example, Dachkovsky & Sandler (2009) concluded that like auditorily perceived intonational melodies, the visual component in sign language provide a meaningful accompaniment to the words and sentences of language. For instance, their results points to similarities between spoken and signed languages, such as the range of analogous functions attributed to H tones and Brow Raise, since they both signal continuation or forward reference and are also characteristic of polar questions, the if-clause of conditionals and other structures. This relationship suggests that it would be well worth investigating further the relative weight of auditory and gestural modalities within non-sign languages.

To gain more insight into possible relationships between auditory and visual cues, the current study will look into the interaction between visual and acoustic cues in the conveyance of pragmatic information in questions. To date only a handful of studies have analyzed the role of facial gestures in the linguistic interpretation of questions (see, e.g., Srinivasan & Massaro 2003, House 2002, Granström & House 2005 for question and statement intonation in English; Borràs-Comes & Prieto 2011 for contrastive focus statements and echo questions in Catalan and others.). For example, Srinivasan & Massaro (2003) performed five perception experiments to investigate the role of facial expression and intonation in distinguishing statements from questions in English. They concluded that statements and questions are distinguished both intonationally and visually but that intonational cues have greater influence as compared to the visual cues. Their data were consistent with the Fuzzy Logical Model of Perception (FLMP), which predicts that both auditory and visual modalities will influence the perception of prosody, and that the influence of one modality will be greater when the information provided by the other modality is ambiguous.

Along the same lines, House (2002), carried out two perception experiments to test whether similar visual cues could influence the perception of question and statement intonation in Swedish. In the first experiment, only the intonational cues were manipulated. The visual component of the stimuli was held constant and contained no

head, eye or eyebrow movements. Subjects were asked to listen to each stimulus while looking at the face and then decide whether the speaker intended to make a statement or ask a question. Results of the first experiment indicated that both a widened F0 (Fundamental Frequency) range on the focal accent and time alignment properties of the rise and peak make important contributions to the perception that an utterance is interrogative. In the second experiment, the audio stimuli from the first experiment were presented with two different visual cue movement configurations. Results of this experiment were similar to the results of the first one and thus House concluded that the influence of visual cues on auditory cues is marginal.

The results reported above are at variance with recent ones found for Catalan. For example, in two perception experiments Borràs-Comes & Prieto (2011) investigated the relative contribution of visual and pitch accent cues in conveying the prosodic perception of statements vs. echo questions in Catalan. Experiment 1 used a pitch range auditory continuum combined with two congruent and incongruent video clips showing the facial gestures that are characteristic of the two pragmatic meanings. Experiment 2 used the same auditory continuum in combination with another continuum for facial gestures produced using a digital image-morphing technique. Results of both experiments revealed a consistent and strong effect of visual cues in the listener's decisions, but also a consistent effect of the auditory stimulation.

Borràs-Comes, Pugliesi & Prieto (2011) carried out a follow-up of the abovementioned study and investigated the same contrast in Catalan by means of a computer-generated 3D talking face. They investigated the AV interaction between intonational and gestural information in the detection of incredulity questions (which are used to express the denial of a discourse-activated assumption) compared to narrow focus statements. Results agreed with the FLMP model of perception in that the influence of one source of information is greater when the other is ambiguous, i.e., in this case the effect of intonation decreases as the incredulity gesture becomes more salient.

Further research is needed to explain why differences have been found between the results of previous studies. For example, it may be that such differences are due to differences between the respective languages studied. These contradictory results in the literature and a lack of cross-linguistic studies investigating the differences in the weight of gestures and acoustic cues motivate this paper. This study will concentrate on the interaction between facial gestures and intonation in question interpretation in two languages, Catalan and Dutch. The main goal of this study is to contribute to the investigation of the role facial gestures cues in question interpretation. To this end, we have chosen these two languages because we hypothesize that they display an asymmetry in the main tonal correlates that cue the difference between information-seeking and incredulity questions (also known as counter-expectational questions). That is, the two languages are expected to mark the difference between the two question types through different intonational strategies—Catalan differentiates between the two meanings by means of a graded intonation contrast whereas Dutch uses a categorical

intonation contrast; but the two languages seem to use similar visual cues. This contrast constitutes a unique opportunity to test whether listeners from these two languages can use different processing strategies in the course of semantic interpretation. According to the FLMP theory, we hypothesize that Catalan listeners will rely more heavily on visual information since acoustic information is weaker than gestural patterns; while Dutch speakers will rely more on acoustic signals.

The research question described above fits into a tradition of research that explores the extent which cues signaled by one modality relate to those in another modality. Evidence suggests that languages may differ in the weight they assign to specific linguistic cues. In recent years, the phenomenon of linguistic *mirativity* has generated increasing interest among researchers in the field of semantics and pragmatics, as well as in the field of linguistic typology (DeLacey 2001). The definition of linguistic *mirativity* is the linguistic marking of new or unexpected information on the part of speakers. The existing literature on this phenomenon claims that such marking can be conveyed through the intonation system in some languages and through the morphosyntactic system in others (DeLancey 2001). In languages like Tibetan, mirativity is conveyed by means of its morphosyntactic system. For example, a sentence like *nga-'i nang-la shmi yod* 'There is a cat in my house' expresses the normal fact that the household pet is at home, while a sentence *isand nga-'i nang-la shmi 'dug* 'There is a cat in my house' would be appropriate when speaking about a unknown or unexpected cat. But DeLancey (2001) points out that other languages that do not mark this category in the morphosyntactic system can mark it by means of intonation; for example, "the mirative intonation contour (in English) is an exaggerated version of the declarative intonation, with the tonic rise considerably higher" (DeLancey 2001, p. 377).

As noted above, the perception of incredulity questions in audiovisual stimulus materials has been investigated before in Borràs-Comes & Prieto (2011) and Borràs-Comes, Pugliesi & Prieto 2011. However, their study was limited to one language. Our goal is to compare the interaction between facial gestures and intonation in the expression of two different question types (i.e., information-seeking and incredulity yes/no questions) in two different languages, namely Catalan and Dutch. From a pragmatic point of view, information-seeking questions are used when speakers seek unknown information, without previous expectations or knowledge about the answer. On the other hand, incredulity questions are questions with a strong presuppositional component, that is, questions that are used to express surprise and incredulity about the contextual facts (e.g., *You're going by plane?!*, asked when the speaker cannot believe that their interlocutor is flying to a nearby city that can be easily accessed by train or car). In Catalan, like in other Romance languages, information-seeking questions and incredulity questions have an identical syntactic surface structure, since they both are intended to elicit an answer of either "yes" or "no". From a prosodic point of view, these types of questions have been studied by Crespo-Sendra et al. (2010), who concluded that both information-seeking and incredulity yes/no questions in Catalan are prototypically characterized by the same intonational nuclear configuration, namely L*

; yet they exhibit a difference in the nuclear pitch range, which is wider in the case of incredulity questions (see Crespo-Sendra et al. 2010). Figure 1 below shows schematic diagrams of the nuclear configuration of the information-seeking yes/no question *Mandarines?* 'Tangerines?' (solid lines) and the incredulity yes/no question *Mandarines?* 'Tangerines?!' (dotted lines) in Catalan.

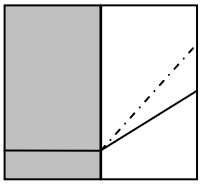

Figure 1. Schematic diagram of an information-seeking yes/no question (solid lines) and an incredulity yes/no question (dotted lines), in Valencian Catalan. Shaded rectangle indicates the accented syllable.

By contrast, through observation of how the two question types are produced in Dutch (see section 5.2.2.), we hypothesize that they tend to be expressed through different nuclear configuration contours, namely L* HH% for information-seeking yes/no questions, and L+H* LH% for incredulity questions. Haan (2001) found four frequent question contours in Dutch labeled in ToDI system (Gussenhoven 2005); L+H* LH% (H*L H%, in ToDI transcription) is the most frequent for yes/no questions in Dutch with L+H* HH% (H* H%) and L* HH% (L*H H%). Figure 2 shows schematic diagrams of the nuclear configuration of the information-seeking yes/no question *Mandarijnen?* 'Tangerines?' (left panel) and the incredulity yes/no question *Mandarijnen?* 'Tangerines?!' (right panel) in Dutch.

L* HH% L+H* LH%

Figure 2. Schematic diagrams of an information-seeking yes/no question (left panel) and an incredulity yes/no question (right panel) in Dutch. Shaded rectangles indicate accented syllables.

As Figures 1 and 2 show, both Catalan and Dutch express the difference between information-seeking questions and incredulity questions through different intonation strategies. In Catalan, the main difference between the two question types lies in the pitch range difference in the whole contour; while in Dutch, the main difference lies in the nuclear configuration (L* HH% vs. L+H* LH%). In other words, while Dutch uses a specific very specialized contour to express incredulity, Catalan uses the same basic contour as that seen in information-seeking questions.

This article addresses two related questions regarding the perceptual processing of the audiovisual markers of information-seeking question vs. incredulity question meanings in Catalan and in Dutch. First, in languages that express mirativity by prosodic means,

how important are facial gestural correlates to this distinction with respect to pitch accent cues? Second, are there differences in the relative weight of the acoustic and facial cues when there is an intonational contrast between the two question types? In order to answer these questions, we will, like Borràs-Comes & Prieto (2011), examine the bimodal perception of a prosodic contrast by using by means of an identification task involving congruent and incongruent pitch accent and facial cue information. To this end the experiment was divided into three modalities: audio only, video only, and audiovisual. The stimuli were the one-word utterances *Mandarines*? (for Catalan) and *Mandarijnen*? (for Dutch), both meaning 'Tangerines?'. The task of the participants was to identify the intended meaning (i.e., whether the speaker was asking an information-seeking yes/no question or an incredulity yes/no question) for each part of the experiment, that is, when exposed to an exclusively audio stimulus, when exposed to an exclusively visual stimulus, and when exposed to a combined audio-visual stimulus in which intonational and visual information was congruent in some instances and incongruent in others.

The article is organized as follows. Section 2 shows the methodology of the experiment conducted. Section 3 describes the results of the experiment. Finally, concluding remarks are offered in section 4, together with the implications of these results for understanding the relative role of visual information and intonation in the perception of incredulity in questions.

2. Methodology

Identification tasks have been used to investigate the role of acoustic and visual cues that allow listeners to recognize a particular pragmatic meaning. The identification experiment presented here was divided into three tasks: audio only (AO), video only (VO) and combined audiovisual (AV). During each task, participants were asked to respond after exposure to a stimulus according to how they would interpret the question in a real situation.

2.1. Participants

In total, 20 native speakers of Catalan (12 females and 8 males) and 20 native speakers of Dutch (15 females and 5 males) between 20 and 35 years old took part in this experiment. The Catalan participants participated on a voluntary basis while the Dutch participants took part in the experiment as partial fulfillment of course credits. All the participants had normal or corrected to normal vision and good hearing.

2.2. Materials

As has been noted, there has been no previous research that compares the facial gestures that characterize incredulity questions in Catalan and Dutch. In order to obtain the stimuli for the tasks we recorded 5 native speakers of each language. To make sure that

participants in the experiments focused on the audiovisual competition between the two target pragmatic meanings, we selected a very short utterance that would contain the target intonational cues and facial gestures in a synchronous way. In order to obtain the AV stimuli for the experiment, subjects were asked to read in an expressive fashion the two responses to the prompt questions shown in (1), with dialogue (1a) involving an information-seeking question and dialogue (1b) exemplifying an incredulity question. Subjects were given no prior instructions of any sort as to how to express these pragmatic meanings, to make sure that their visual gestures would be representative of what they would display spontaneously. The audiovisual recordings of all ten speakers were carried out in quiet research rooms at the Universitat Pompeu Fabra, for Catalan speakers, and at Tilburg University, for Dutch speakers. Speakers were seated on a chair in front of a digital camera that recorded their upper body and face at 25 frames per second.

(1) a. **Information-seeking question**

—You are talking with your partner about the party that you are organizing for tonight. He is talking about the dessert, and after saying that you already have apples, bananas and pears, he asks you what else you could buy. You make a tentative proposal in the form of a question about tangerines. (Recorded response: *Tangerines?*)

b. **Incredulity question**

—You enter your home and see that your mother is eating tangerines. You know that your mother doesn't like tangerines. Unable to believe your eyes, you ask her about the tangerines. (Recorded response: *Tangerines!?*)

Analysis of the recordings

After recording the audiovisual stimuli, we assessed qualitatively the facial gesture cues that seemed to be most effective and representative for each pragmatic meaning. One of the facial expressions that correlates most clearly with the perception of an information-seeking question is the raised eyebrow movement and downward head movement followed by an upward head movement. It is worth noting that in sign languages, eyebrow movements are raised in yes/no questions in American Sign Language (Baker-Shenk 1983, Grossman 2001, Grossman & Kegl 2006), Swedish Sign Language (Bergman 1984), British Sign Language (Kyle & Woll1985) and Sign Language of the Netherlands (De Vos, van der Kooij & Crasborn 2009). As for a question conveying incredulity, the characteristic facial expression involves a furrowing of the brows and a squinting of the eyes, often accompanied by a downward head movement followed by a backward head movement (also noted in Borràs-Comés & Prieto 2011). The data were coded using the Facial Action Coding System (FACS: Ekman, Friesen & Hager 2002) for type of brow movement as well as head movement. FACS measurement units are Action Units —AUs— and head movements are labeled as Ms. As argued by De Vos, van der Kooij & Crasborn (2009) and Borràs & Prieto (2011), three AUs are relevant in

the production of eyebrow movements: AU 1, the Inner Brow Raiser; AU 2, the Outer Brow Raiser; and AU 4, the Brow Lowerer. Analysis of the visual component of our recordings showed that for information-seeking interpretations the most common facial expression consisted of a combination of action units AU1+2 (Inner and Outer Brow Raisers) and head movement M59 (Head Down and Head Up). For incredulity question interpretations, the most common pattern was a combination of AU4 (Brow Lowered), M59+58 (Head Down and Head Back) and squinting of the eyes.

Following a full analysis of the facial expressions of all five speakers for each language, we concluded that one of the most effective gestural cues for the distinction between information-seeking questions and incredulity questions was the pattern of eyebrow movements. Figure 3 shows two representative stills of the peak gesture from the video clips as the subjects utter an information-seeking question (left panels) and then an incredulity question (right panels), with a Catalan speaker shown in the upper row and a Dutch speaker shown in the lower. Several studies agree that significant gestures are aligned with pitch accents, and both give the sense of a particular pragmatic meaning; this is reported in studies such as Cassell et al. (1994) and Loehr (2004). The two images in each set correspond to the peak of the facial gesture, which approximately coincides with the middle of the stressed nuclear syllable.

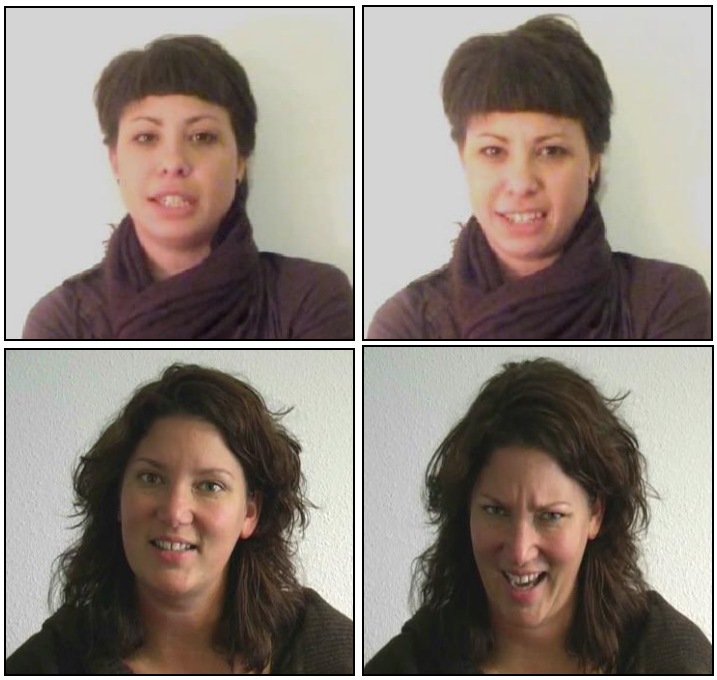

Figure 3. Photos on the left show the neutral expression characteristic of an information-seeking yes/no question while photos on the right show the expression that typifies an incredulity yes/no question. Upper panels are of a Catalan speaker, lower ones of a Dutch speaker.

The prosodic information obtained in this set of audiovisual recordings served as a basis for the preparation of the audio, visual and audiovisual stimuli to be used in this experiment. In Catalan, information-seeking and incredulity yes/no questions had the

same intonation nuclear configuration contour, L* HH%, with the main difference manifesting itself in terms of the total pitch range, which is higher in incredulity questions (see Crespo-Sendra et al. 2011). Figure 4 shows the waveforms and F0 contours of the sentence *Mandarines*? 'Tangerines?' produced with an information-seeking question (upper panel) and incredulity question meaning (lower panel) by a Catalan speaker.

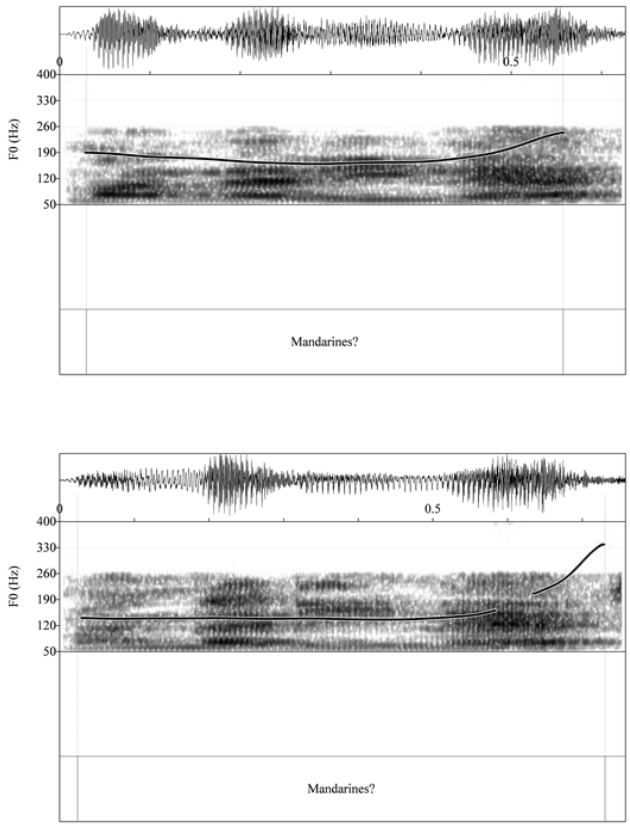

Figure 4. Waveforms, F0 contours, and orthographic transcription of the utterance *Mandarines?* produced with an information-seeking meaning (upper panel) and an incredulity meaning (lower panel) in Catalan.

In Dutch, the two question types are expressed through different nuclear configuration contours, namely L* HH% for information-seeking questions and L*+H LH% for incredulity questions. Figure 5 shows the waveforms and F0 contours of the sentence *Mandarijnen*? 'Tangerines?' produced with an information-seeking question (upper panel) and incredulity question meaning (lower panel) by a Dutch speaker.

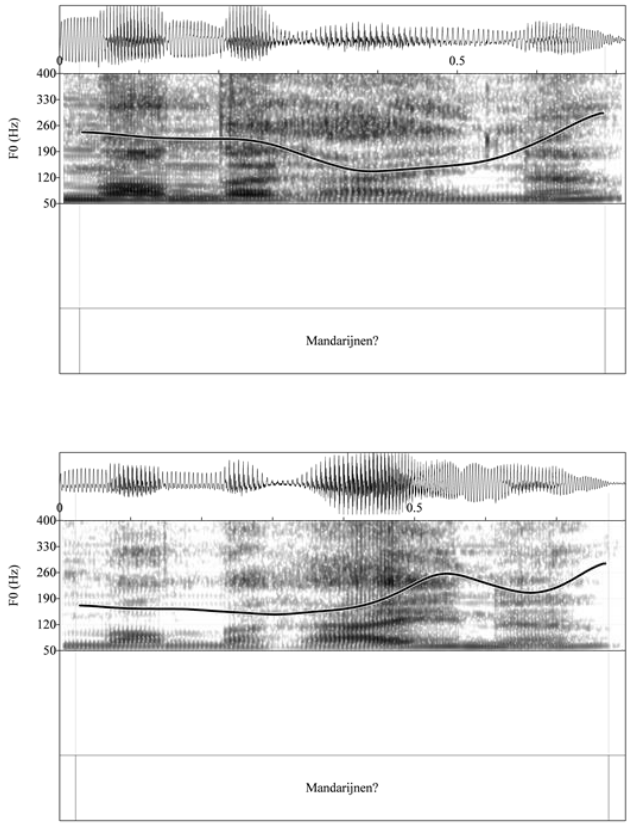

Figure 5. Waveforms, F0 contours and orthographic transcription of the utterance *Mandarijnen*? produced with an information-seeking meaning (upper panel) and an incredulity meaning (lower panel) in Dutch.

2.3. Procedure

For the identification task explained below, three kinds of stimuli were prepared and the experiment was divided accordingly into these three different parts: audio stimuli only (AO), video stimuli only (VO) and combined audiovisual stimuli (AV). The AO and VO stimuli were extracted from the original audiovisual recordings using Adobe Premiere (Adobe Systems Incorporated 2002), which allows the audio and video tracks to be saved separately.

The AO task consisted of listening to audio content. We have seen in Figures 4 and 5 (Section 5.2.2.) the prosodic characteristics of the two pairs of interrogative utterances in Catalan (Figure 4) and Dutch (Figure 5). In this task participants had to decide whether they regarded the intonation of each audio stimulus they heard as "neutral", i.e., information-seeking, (which they indicated by pressing the "N" key) or "incredulous" (for which they pressed the "I" key). We obtained a total of 600 responses per language in this task (2 audio files (informative, incredulity) x 5 speakers x 3 blocks x 20 participants).

The VO task consisted of watching the video stimuli extracted from the audiovisual recordings but without the sound track. Again, Figure 3 (Section 2.2.) illustrates the facial characteristics of the two meanings in both languages. Note that participants were asked to make judgements about film clips, i.e., moving images, whereas the pictures in figure 3 are merely stills intended to give the reader an idea of the stimuli. As in the first task, participants had to decide whether they regarded the intonation of the speaker shown in each video clip they watched as "neutral" (which they again indicated by pressing the "N" key) or "incredulous" (the "I" key). We obtained a total of 600 responses per language in this task (2 video files (informative, incredulity) x 5 speakers x 3 blocks x 20 participants).

The AV task consisted of watching and listening to combined audiovisual stimuli based on the original audiovisual recordings. In this task there were congruent and incongruent videos. Congruent stimuli consisted of the original audiovisual recordings, untouched. The incongruent stimuli were prepared by manipulating the original recordings using Adobe Premiere (Adobe Systems Incorporated 2002). These incongruent variants consisted of various combinations of information-seeking and incredulity cues for each speaker. We thus obtained altogether a total of 4 stimuli per speaker: 2 congruent (neutral and incredulous) and 2 incongruent. (neutral face-incredulous intonation; and, incredulous face-neutral intonation). The final 20 stimuli (4 variants × 5 speakers) were pretested for naturalness to see whether any manipulation artefacts could be detected, which turned out not to be the case. It was felt that it would be impossible for experiment participants to detect the manipulations, and auditory and visual cues were very well aligned. Indeed, when asked at the end of the full experiment, none of the participants indicated that they had noticed manipulation effects in the stimuli. In this task once again participants had to decide whether they perceived the audiovisual stimuli as "neutral" or "incredulous". We obtained a total of 1.200 responses per language in this task (2 audiovisual files (informative, incredulity) x 2 situations (congruous, incongruous) x 5 speakers x 3 blocks x 20 participants).
This kind of task with incongruent audiovisual files has been used in many studies to investigate the contribution of various factors (see, for example, Massaro et al. 1996, Gelder & Vroomen 2000, Dijkstra, Krahmer & Swerts 2006~~,~~ and Swerts & Krahmer 2008).

Prior to each of the three parts of the experiment, a brief training session was conducted in which participants were shown or played two repetitions of each of the two pragmatic meanings. These training recordings feature a sixth speaker of each language, that is, a speaker who did not feature in the experimental stimuli. When the experimental stimuli were then presented to the participants, they were presented in a randomized order in order to prevent learning effects. Likewise, after the second part of the experiment (the audio only task) participants were asked to watch a 5-minute documentary to prevent their transferring learning from this task to the third part of the experiment, the combined audiovisual task.

Finally, following the three experimental tasks, participants carried out a minor task which was intended as a kind of control to confirm that the unmanipulated original recordings would be unambiguously interpreted by all participants. In this control task, participants were shown ten audio-only and ten video-only recordings of the five speakers of their own language each producing one token of an information-seeking yes/no question and one token of an incredulity yes/no question. As they listened or watched, participants were told to evaluate the degree of incredulity that they perceived on a five point scale (1 = "speaker is very neutral"; 5 = "speaker is very incredulous"). We obtained a total of 400 responses per language in this task (2 visual files (informative, incredulity) x 2 audio files (informative, incredulity) x 5 speakers x 20 participants). The factor speaker did not have a significant effect in the participants' results.

All the participants performed the full sequence of four tasks (three experimental plus one control) but did the VO and AO experimental tasks in different orders to control for the effect of task order on the results. Thus, for each language group, while 10 did the tasks in the order AO, VO, short documentary, AV and control, the other 10 did the tasks in the order VO, AO, short documentary, AV and control. No differences were found with regards to the order.

The experiment was set up in a quiet research room at the Universitat Pompeu Fabra (for Catalan participants) and Tilburg University (for Dutch participants). The presentation of the stimuli was prepared using E-Prime (Psychology Software Tools Inc., 2009). Participants performed the tasks individually seated in front of a laptop in a quiet room and listening to the stimuli through headphones. The entire experiment lasted about 30-35 minutes.

3. Results

This section describes the results for the three congruent modalities (AO, VO and AV), the results for the audiovisual modality with congruent and incongruent stimuli compared.

3.1. Audio Only, Video Only, Audiovisual (congruent stimuli)

The two graphs in Figure 6 show the mean correct response rate (y axis) (whereby "1" indicates that 100% of participant responses were consistent with the question type intended by the recorded speaker and "0" indicates that none of the participant responses matched the question type as expressed by the speaker) for information-seeking question stimuli (black bars) and incredulity question stimuli (grey bars) as a function of the three experimental modalities (x axis) for the two languages (Catalan, left panel, and Dutch, right panel).

In Catalan, the average rate of correct responses for the AO modality is 0.73 for information-seeking questions and 0.74 for incredulity questions. For the VO modality, the average rate of correct responses is 0.96 for information-seeking questions and 0.98 for incredulity questions. And for the AV modality, the average rate of correctness is 0.96 for information-seeking questions and 0.98 for incredulity questions. This demonstrates that the AO stimuli were not very strong. By contrast, results for the VO and AV stimuli demonstrate that visual cues were basic in determining participants' decisions.

In Dutch, the corresponding average rates of correct responses are, for the AO modality 0.69 for information-seeking questions and 0.88 for incredulity questions; for VO, 0.98 for information-seeking questions and 0.96 for incredulity questions; and, for AV, 0.85 for information-seeking questions and 0.96 for incredulity questions. This means that the proportion of matching responses to AO incredulity stimuli is higher for Dutch than for Catalan, a fact that might be explained by the specific and marked contour type of incredulity questions in Dutch. Results for the VO modality are quite similar for both pragmatic meanings and both languages. Looking at the results for the AV modality, however, we see that while Dutch participants are able to correctly identify incredulity cues, there is a lower proportion of matching responses for information-seeking questions —this despite the fact that audiovisual stimuli they were exposed to were congruent, that is, participants were watching authentic unmanipulated video clips and still found them ambiguous to a small extent. This fact could be because the L* HH% of information-seeking questions can also be used in Dutch—albeit less frequently—to convey incredulity in a yes/no question with an expanded pitch range.

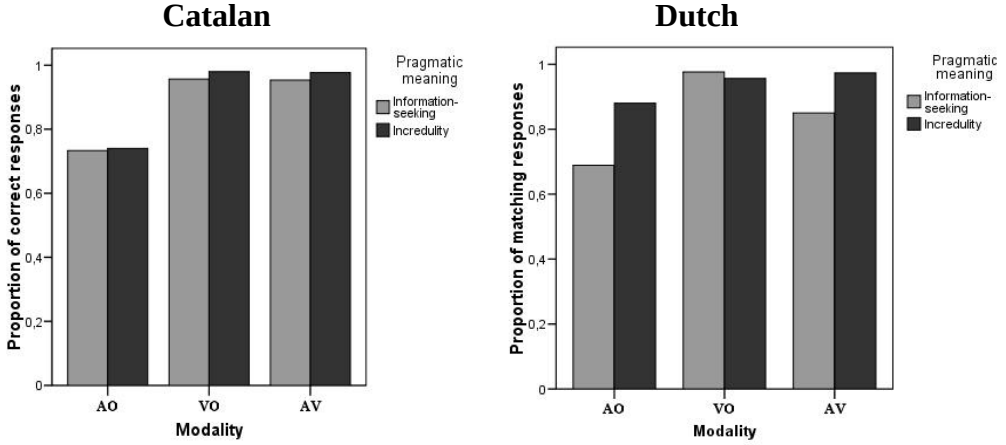

Figure 6. Proportion of correct identification in the three modalities of the identification task (AO, VO and AV) for the two languages (left graph, Catalan; right graph, Dutch).

These immediate observations about the results were supported by a statistical test. An ANOVA repeated measures design was carried out with two within-subjects independent factors: experimental modality (three levels: auditory only, visual only, audiovisual)~~,~~ and pragmatic meaning (two levels: information-seeking, incredulous).

The between-subjects factor was language (two levels: Catalan, Dutch). The dependent variable was the proportion of correct responses. The analysis revealed a significant main effect of modality (F (2, 76) = 30.342, p < .001, η_p^2= .444). The interaction between language and modality was not significant (F (2, 76) = 1.186, p = .311, η_p^2= .030). We can thus claim that there were no significant differences between the three experimental modalities in either language. However, the statistical test revealed a significant effect for pragmatic meaning (F (1, 38) = 74.297, p < .001, η_p^2= .662). The analysis also showed a significant two-way interaction between pragmatic meaning and language (F (1, 38) = 6.154, p = .018, η_p^2<.139). Finally, the test revealed a significant three-way interaction between experimental modality, pragmatic meaning and language (F (2, 76) = 9.021, p < .001, η_p^2= .192).

3.2. Audiovisual (congruent and incongruent stimuli)

The two graphs in Figure 7 show the incredulity responses rate (y axis) (whereby "1" indicates 100% correct labelling of a stimulus as an incredulity answer and "0"indicates 100% correct labelling of a stimulus as an neutral answer) for facial expression information-seeking stimuli (black bars) and incredulity facial stimuli (grey bars) as a function of the Contour Types (x axis) for the two languages (Catalan, left panel, and Dutch, right panel).

For Catalan, the results revealed an average rate of incredulity responses of 0.04 for the congruous information-seeking meaning (neutral facial expression + neutral intonation) and an average rate of incredulity responses of 0.98 for the congruous incredulity meaning (incredulous facial expression + incredulous intonation). By contrast, the average rate of incredulity responses in the incongruous contexts was 0.20 for incredulity intonation juxtaposed with the neutral facial expression, and 0.96 for the incredulous facial expression with the information-seeking question intonation. This means that given an incongruent situation, Catalan listeners rely heavily on the facial expression in distinguishing between the two question types.

By contrast, in Dutch, the results revealed an average rate of incredulity responses of 0.15 for the congruous information-seeking meaning (neutral facial expression + neutral intonation) and an average rate of incredulity responses of 0.97 for the congruous incredulity meaning (incredulous facial expression + incredulous intonation). In contrast with congruent situations and in contrast with Catalan, the average rate of incredulity responses in the incongruous contexts was 0.55 for information-seeking intonation juxtaposed with the incredulity facial expression, and 0.76 for the incredulous facial expression with the information-seeking question intonation. Thus, though Dutch speakers seem to rely strongly on facial cues, there is evidence that these cues are not enough to obtain a 100% correct response~~s~~ rate, and they show a clear influence of intonation contours.

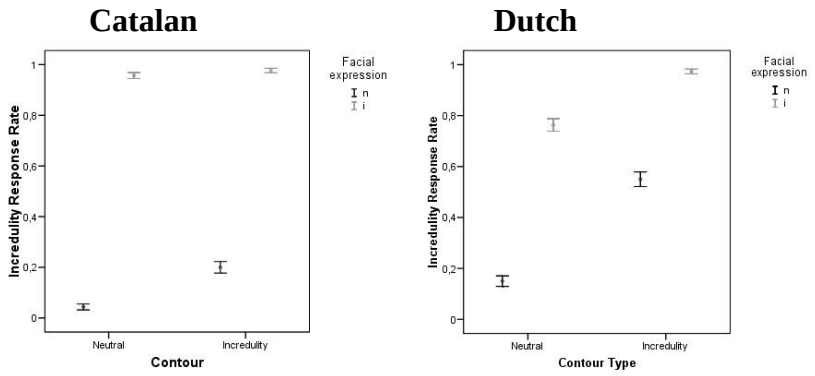

Figure 7. Proportion of identification of stimulus as an incredulity question for AV congruent and incongruent stimuli (upper panel, Catalan; lower, Dutch).

An ANOVA repeated measures test was carried out with two within-subjects independent factors: facial expression (two levels: neutral, incredulous) and auditory stimulus (two levels: information-seeking question, incredulity question). The between-subjects factor was language (two levels: Catalan, Dutch). The dependent variable was the proportion of incredulity responses. The analysis revealed a significant main effect of facial expression stimulus (F (1, 38) = 453. 562, $p < .001$, $\eta_p^2 = .923$). The interaction between language and facial expression was also significant (F (1, 38) = 26.040, $p < .001$, $\eta_p^2 = .407$), so we see that facial expression has a similar effect in each language. For auditory stimulus, the statistical test also revealed a significant effect (F (1, 38) = 48.140, $p < .001$, $\eta_p^2 = .559$). The analysis also showed a significant two-way interaction between auditory stimulus and language effect (F (1, 38) = 14.607, $p < .001$, $\eta_p^2 = .278$). Finally, the three-way interaction between facial expression, auditory stimuli and language turned out not to be significant (F (1, 38) = .952, $p = .335$, $\eta_p^2 = .024$).

4. Conclusions

In this paper, we have investigated the relative importance of pitch variation and facial gestures in the perception of the contrast between information-seeking yes/no questions and incredulity yes/no questions in Catalan and in Dutch by using congruent and incongruent multimodal stimuli. Our general goal was to assess the role of audio and visual cues in the linguistic detection of incredulity in two languages that are different in their intonational encoding of incredulity questions.

Several conclusions can be drawn from the results of these experiments with regard to the asymmetric perception of information-seeking question and incredulity question prosody across the two languages. Results from identification task in three sensory modalities (AO, VO, AV) confirm that there are indeed differences between Catalan and Dutch listeners in the perceptual processing involved in distinguishing between the

two pragmatic meanings. First, our results demonstrate that visual cues (that is, in VO or AV modalities) play a stronger role than auditory cues (AO) in both languages and induce higher degrees of correct identification. The results from the unimodal tasks (AO, VO) show that the role of facial expression is more important than intonation for detecting the target meanings, a conclusion that holds true for both Catalan and Dutch. In the AO task in Catalan, listeners produced the same correct identification responses rates for both pragmatic meanings. By contrast, in Dutch, the specific incredulity contour induced a 30% increase in correct identification responses.

Finally, the results from the AV condition with congruent and incongruent materials demonstrate that the congruent stimuli achieve higher identification scores. Yet for the identification of incongruent stimuli, while Dutch listeners' perceptual ratings are strongly affected by both gestural and prosodic cues to incredulity, Catalan listeners' perceptual ratings are more strongly affected by facial expression. In sum, for the perception of incredulity questions, Catalan listeners give more weight to facial cues than Dutch listeners do. Our explanation is that the acoustic information in Catalan is more subtle than in Dutch, as is confirmed by the lower AO results in Catalan. That is, in Catalan the respective intonation contours identifying the two meanings are expressed by the same pitch contour, which differs only in the scaling of the boundary tone. In Dutch, however, the two meanings are expressed by different pitch contours.

The results presented here agree with the FLMP model of speech perception, namely that an ambiguous or weaker cue in one modality enhances the role of the other modality. This is most clearly seen in the results of the AV experiment, where in incongruent situations the stronger cues gain a enhanced role in participants' decisions. Thus, in Catalan, decisions are strongly affected by facial expression because the difference in intonation between the two question types is less salient (i.e., they have the same nuclear configuration By contrast, in Dutch, the strongest cues are both the incredulity facial expression and the incredulity pitch contour; in other words, listeners identify a neutral facial expression with an incredulity intonation contour as incredulous, and they also perceive an incredulous meaning when the stimuli are an incredulous facial expression combined with a neutral intonation contour.

This paper represents an attempt to address the current paucity of research on the relevance of facial gestures to intonational meaning. In general, the results of this set of experiments reveal the importance of visual cues in the identification of tune meaning, thus confirming the results of recent studies (Swerts & Krahmer 2005, Borràs & Prieto 2011, and others.). Moreover, the results of the present study show that two different languages that express mirativity in a prosodic fashion assign different weights (in both perception and production) to visual and intonational cues. Finally, we believe that the results presented here support the idea that gestures and intonation must be investigated together when the perception of a given meaning is analyzed.

Acknowledgments

*Parts of this paper were presented at the *Phonetics and Phonology in Iberian Conference*, 21-23 June 2011, at Universitat Rovira I Virgili in Tarragona (Spain). We are grateful to the audience at this conference for their helpful comments and discussion, especially Carlos Gussenhoven, and to the GrEP group for their help and comments. This research has been funded by projects FFI2009-07648/FILO and by the CONSOLIDER-INGENIO 2010 Programme CSD2007-00012 (both awarded by the Ministerio de Ciencia e Innovación), by project 2009 SGR 701 (awarded by the Generalitat de Catalunya) and by the *Atles interactiu de l'entonació del valencià* (awarded by the Acadèmia Valenciana de la Llengua).

References

Adobe Systems Incorporated (2002). Adobe Premiere. Version 6.5.

Baker-Shenk, C. (1983). *A microanalysis of the nonmanual components of questions in American Sign Language*. Berkeley: University of California.

Bergman, B. (1984). Non-manual components of signed language: Some sentence types in Swedish Sign Language. In F. Loncke, P. Boyes Braem, and Y. Lebrun (Eds.), *Recent research on European sign languages (Proceedings of the European Meeting of Sign Language Research, Brussels)*. Lisse: Swets & Zeitlinger, 49-59.

Borràs-Comes, J., & Prieto, P. (2011). "Seeing tunes". The role of visual gestures in tune interpretation. *Laboratory Phonology* 2.2, 355-380

Borràs-Comes, J., Pugliesi, C., & Prieto, P. (2011). "Audiovisual competition in the perception of counter-expectational questions". In Giampiero Salvi, Jonas Beskow, Olov Engwall and Samer Al Moubayed (Eds.): *Proceedings of the 11th International Conference on Auditory-Visual Speech Processing* (AVSP2011). (Volterra, Aug 31 - Sep 3, 2011), pp. 43-46. Stockholm: KTH Computer Science and Communication, Department of Speech Music and Hearing. [ISBN: 978-91-7501-079-3]

Cassell, J.; Pelachaud,C.; Badler, N.; Steedman, M.; Achorn, B.; Becket, T.; Douville, B.; Prevost, S., & Stone, M. (1994) Modeling the interaction between speech and gesture. Proceedings of the 16th Annual Conference of the Cognitive Science Society, Georgia Institute of Technology, Atlanta, USA.

Crespo-Sendra, V., Vanrell, M.M., & Prieto, P. (2010) "Infomation seeking questions and incredulity questions: gradient or categorical contrast?". *Proceedings of Speech Prosody*, (11-14 May, Chicago), 100164: 1-4. ISBN: 978-0-557-51931-6.

de Gelder, B., & Vroomen, J. (2002). The perception of emotions by ear and by eye. *Cognition and Emotion*, 14(3), 289-311.

de Vos, C., van der Kooij, E., & Crasborn, O. (2009). Signals: Combining Linguistic and Affective Functions of Eyebrows in Questions in Sign Language of the Netherlands. Language and Speech June/September 2009 vol. 52.

Dachkovsky, S., & Sandler, W. (2009).Visual intonation in the prosody of a sign language. *Language and Speech*, 52 (2/3), 287-314.

DeLancey, S. (2001) The mirative and evidentiality. Journal of Pragmatics, 33, pp. 369-382.

Dijkstra, C., Krahmer, E., & Swerts, M. (2006). Manipulating Uncertainty: The contribution of different audiovisual prosodic cues to the perception of confidence. *Proceedings of the Third International Conference on Speech Prosody*, Dresden.

Dohen, M., & Lœvenbruck, H. (2009). Interaction of audition and vision for the perception of prosodic contrastive focus. *Language and Speech*, 52(2/3), 177-206.

Ekman, P., Firesen, W.V., & Hager, J. C. (2002). *The Facial Action Coding System CD-ROM*. Salt Lake City, UT: Research Nexus.

Granström, B., & House, D. (2005). Audiovisual representation of prosody in expressive speech communication. *Speech Communication*, 46, 473–484

Grossman, R. (2001). *Dynamic facial expressions in American Sign Language: Behavioral, neuroimaging, and facial coding analyses for deaf and hearing participants*. Boston, MA: Unpublished doctoral dissertation, Boston University.

Grossman, R., & Kegl, J. (2006). To capture a face: A novel technique for the analysis and quantification of facial expressions in American Sign Language. *Sign Language Studies*, 6 (3), 273-305.

Gussenhoven, C. (2005). Transcription of Dutch Intonation (ToDI). In Sun-Ah Jun (Ed.), *Prosodic Typology and Transcription: A Unified Approach* (pp. 1-33). Oxford University Press.

Haan, J. (2001). Speaking of questions: an exploration of Dutch question intonation. PhD dissertation. Radboud University of Nijmegen. Utrecht: LOT

House, D. (2002). Intonation and visual cues in the perception of interrogative mode in Swedish. *Proceedings of ICSLP* 2002, 1957–1960.

Krahmer, E., & Swerts, M. (2005). How children and adults produce and perceive uncertainty in audiovisual speech. *Language and Speech*, 48 (1), 29-54.

Kyle, J., & Woll, B. (1985). *Sign language: The study of deaf people and their language*. Cambridge: Cambridge University Press.

Loehr, D. P. (2004). Gesture and Intonation. Doctoral Dissertation, Georgetown University, Washington, DC.

Massaro, D., Cohen, M., & Smeele, P. (1996). Perception of asynchronous and conflicting visual and auditory speech. *Journal of the Acoustical Society of América* 100(3), 1777-1786.

Nespor, M., & Sandler, W. (1999). Prosody in Israeli Sign Language. *Language and Speech*,42, 143-176.

Psychology Software Tools Inc. (2009). E-Prime (version 2.0). Computer Program. On line < http://www.pst-net.com/ >

Prieto, P., Torres-Tamarit, F., & Vanrell, M. M. (2008). Categorical perception of mid boundary tones in Catalan. Paper presented at the Third TIE Conference on Tone and Intonation 2008 (TIE), Lisboa (Portugal), September, 15-17. URL: http://optimitza.cat/mvanrell/presentacions.html

Rathcke, T., & Harrington, J. (2010). The variability of early accent peaks in Standard German. In C. Fougeron, B. Kühnert, M. D'Imperio, & N. Vallée (Eds.), *Laboratory Phonology 10* (pp. 533-555). Berlin / New York : Mouton de Gruyter.

Reilly, J. S., McIntire, M., & Bellugi, U. (1990). The acquisition of conditionals in American Sign Language: Grammaticized facial expressions. *Applied Psycholinguistics*, 11, 369-392.

Sandler, W. (2005). Prosodic constituency and intonation in a sign language. *Linguistische Berichte*, 13, 59-86.

Swerts, M., & Krahmer, E. (2005). Audiovisual prosody and feeling of knowing. *Journal of Memory and Language*, 53(1), 81-94.

Srinivasan, R. J., & Massaro, D. W. (2003). Perceiving from the face and voice: Distinguishing statements from echoic questions in English. *Language and Speech*, 46(1), 1-22.

Vanrell, M.M., Mascaró, I., Torres-Tamarit, F., & Prieto, P. (2010). When intonation plays the main character: information- vs. confirmation-seeking questions in Majorcan Catalan. *Proceedings of Speech Prosody 2010* (11-14 May, Chicago), 100168: 1-4. ISBN: 978-0-557-51931-6.

Wilbur, R. B. (2000). Phonological and prosodic layering of non-manuals in American Sign Language. In K. Emmorey & H. Lane (Eds.), *The signs of language revisited* Mahwah, NJ: Lawrence Erlbaum Associates, 215–247.