Recent studies have shown that the visual component plays an important role in various aspects of communication associated with verbal prosody. For example, several studies have analyzed the audiovisual expressions of affective meanings such as uncertainty (Krahmer & Swerts 2005) and frustration (Barkhuysen, Krahmer & Swerts, 2005). Other studies have analyzed the crucial role of facial gestures in linguistic interpretation (e.g., Srinivasan & Massaro 2003 for question intonation in English, Borrás-Comes & Prieto (2010) for contrastive focus statements and echo questions in Catalan, a.o.). One of the questions that arises is whether the role played by visual cues depends on the saliency of the prosodic cues available to the listeners. The goal of this study is to analyze 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 languages (namely Catalan and Dutch). In Catalan, information-seeking and incredulity yes/no questions have the same phonological contour, L+H* L* HH%, and what distinguishes these two types of yes-no questions is the pitch range expansion of the contour in incredulity questions (see Crespo-Sendra et al. 2010), (see Fig. 1, left graphs). By contrast, in Dutch both question types are expressed through a different pitch contour, namely H* L* HH% for neutral questions, and H* L+H* LH% for incredulity questions (see Fig. 1, right graphs). Our goal is to investigate the interaction between facial expression and intonation in the perception of these two types of questions. Our hypothesis is that Catalan listeners will rely more on facial cues than Dutch listeners, since the latter can make use of a completely different contour type.

To this end, a congruity task and a semantically motivated identification task were carried out for the two languages (20 listeners per language). The materials for the congruity task consisted of 2 pictures (incredulous face, neutral face) which were combined with the two target questions (incredulity and neutral; see Fig. 1). In this task, listeners classified the stimuli as “congruent” or “incongruent” with the facial expression of the picture (see Fig. 1). Results demonstrated that both Catalan and Dutch listeners were able to identify which intonation contour goes with which facial expression (see Fig. 2). The materials for the identification task consisted of 2 audio files x 2 video files x 2 audiovisual files x 2 question types (neutral/incredulity) x 5 speakers (per language). The experiment was run in three conditions: audio-only (AO), video-only (VO) and audiovisual (AV). In this task, participants had to classify the stimuli as either information-seeking or incredulity questions. The results demonstrate that the access to the visual cues (that is, in VO or AV modalities) is more important for the correct identification (see Fig. 3 (a)). As for the AV condition, results with congruent and incongruent materials (see Fig. 3 (b)) demonstrate the following: (a) the congruent stimuli get the clearest identification scores (see “N-Neutral” and “I-Incredulity”); and (b) in Dutch, the perceptual ratings of incongruent stimuli appear to be strongly affected by the gestural and prosodic cues to incredulity, and, in Catalan, the perceptual ratings of incongruent stimuli appear to be strongly affected by the facial expression. That is, the combination of incredulity face/neutral contour in Dutch leads to a 75% of incredulity responses, and the combination of neutral face/incredulity contour to a 60% of incredulity responses. In Catalan, both combinations are dominated by the face; in both cases we can observe almost a 100% of responses (see I-Neutral, and N-Incredulity). To conclude, for the perception of incredulity Catalan listeners give more weight to the facial cues than Dutch listeners. A possible explanation could be that the acoustic information in Catalan is more subtle than in Dutch. That is, in Catalan the two intonation contours identifying both meanings are expressed by the same pitch contour, which only differs in the scaling of the boundary tone. In Dutch, however, both meanings are expressed by different pitch contours.
Figure 1. The graphs on the left show the information-seeking (upper panel) and the incredulity (bottom panel) pitch contour in Catalan and the graphs on the right, the same contrast in Dutch. The pictures on the right show the neutral face (upper panel) and the incredulous face (bottom panel) for the congruity task.

Figure 2. The graph on the left shows proportion of congruent responses in the congruity test in Catalan, and the graph on the right shows the same in Dutch.

Figure 3. The graphs in (a) show proportion of correct identification in the three modalities of the identification task (AO, VO and AV) (left, in Catalan; right, in Dutch). The graphs in (b) show proportion of identification as incredulity question in the AV congruent and incongruent stimuli (left, in Catalan; right, in Dutch).

References
Borràs-Comes, Joan; Prieto, Pilar (2010). ‘Seeing tunes’. The role of visual gestures in tune interpretation. *12th Conference on Laboratory Phonology (LabPhon 12)*. University of New Mexico.