Iris Hübscher*, Laura Wagner and Pilar Prieto

Three-year-olds infer polite stance from intonation and facial cues

https://doi.org/10.1515/pr-2017-0047
Submitted October 31, 2017; Accepted May 11, 2018

Abstract: Despite the evidence that infants are sensitive to facial cues and prosody for the detection of emotion, we have contradictory evidence regarding the use of these cues by older preschool and school children when inferring both emotional and politeness stance. This study assessed preschool aged children's sensitivity to intonational and facial cues signalling a speaker's polite stance in requestive speech acts with controlled lexical and contextual materials. Thirty-six 3-year-old American English-speaking children performed a forced-choice decision task which investigated whether children at this age use pitch and/or facial cues to infer a speaker's affective stance in either audio-only, visual-only or audio-visual presentation modalities, when lexical cues are controlled for. Results showed that (a) children at three years can infer a speaker's polite stance equally well in all three conditions (audio-only, visual-only and audio-visual) and thereby (b) unlike previous research, in the present task both intonation and facial cues are equally strong cues in children's understanding of a speaker's polite stance in requestive speech acts. The authors discuss especially the implications of this early use of intonation to detect politeness, relating it to other previous research on children's ability to infer meaning from pitch.

Keywords: politeness development; affective stance; intonation and facial cues; multimodal request comprehension; emotion recognition

***Corresponding author: Iris Hübscher**, URPP Language and Space, University of Zurich, Zurich, Switzerland; Universitat Pompeu Fabra, Translation and Language Sciences, ORCID-ID: 0000-0002-3215-7294, Barcelona, Catalonia, E-Mail: iriscarmela.huebscher@uzh.ch
Laura Wagner, The Ohio State University, Department of Psychology, Columbus, OH, United States, E-Mail: wagner.602@osu.edu
Pilar Prieto, Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Catalonia; Universitat Pompeu Fabra, Translation and Language Sciences, Barcelona, Catalonia, E-Mail: pilar.prieto@upf.edu

1 Introduction

The ability to infer another person's interpersonal stance is crucial and central to our communicative interactions. In day-to-day interactions listeners are confronted with a complex set of cues happening simultaneously. Consider, for example, the utterance *Pass me the bread*. The sociopragmatic meaning of this utterance can vary greatly depending on whether the speaker is being polite, friendly or impolite, that is, the stance that the speaker is taking. In order to infer a speaker's polite stance not only what is said but also how it is said needs to be decoded. Recent studies have highlighted the importance of the multimodal encoding of (im)politeness and the relevance of investigating prosodic, facial and body cues alongside the traditionally analyzed lexical structure of communicative acts (Culpeper, Bousfield, and Wichmann 2003; Culpeper 2011; Brown and Prieto 2017; Langlotz and Locher 2017). In an investigation of a sketch staged in Monty Python's Flying Circus: Live from the Grill-O-Mat, Langlotz and Locher (2017) demonstrated how an English-speaking shop assistant switches from taking a positive and polite stance to taking an aggressive and impolite one by changing intonation and facial expression: "While negative stance is expressed through more stress and vocal force, positive stance is mediated through the higher pitched voice and rising intonation. This is usually accompanied by a smiling vs. a grim facial expression" (Langlotz & Locher, 2017: p. 313). Even though the importance of emotional cues in politeness research was noted already in Brown and Levinson's (1987) seminal work on politeness, it is only recently that this relationship has been given serious attention within interpersonal pragmatics (for an overview see Locher and Langlotz 2008; Culpeper 2011; Spencer-Oatey 2011; Langlotz and Locher 2013; Locher and Koenig 2014; Langlotz and Locher 2017). More specifically, as Langlotz and Locher (2017: 315) observe, "emotions thus play an important part in arriving at emic judgments on relational work so that they have a clear place in the theoretical arguments of interpersonal pragmatics". Interpersonal relationships constitute the most important source for emotions. Andersen and Guerrero (1998: 64) argue that "the primary elicitor of emotions is interpersonal interaction". From an interactional perspective, the focus lies on external emotional "representations" in communication, compared to the "internal" psychological research perspective on people's emotions (Bänziger et al. 2010). People often strategically induce emotional states in others as a way of achieving interpersonal goals (Andersen and Guerrero 1998). This implies that when analyzing communicative acts of relational work, as Langlotz and Locher (2017) point out, emotional signals accompanying the messages should be considered. In the current study we set out to investigate whether and how preschool aged children infer a speaker's polite affective stance (from now on referred to as

polite stance) by considering the role played by intonational and facial cues. We will start out by reviewing previous research that focusses on children's ability to infer a speaker's emotional state through prosody and gesture, as well as the few existing studies that look at how children perceive politeness through intonation and facial cues.

1.1 Early sensitivity to prosody and facial cues in the detection of emotion

While the present study focusses on how preschool children (aged 3) interpret affective meanings conveyed through prosody (specifically intonation) and facial expression, there is clear evidence that infants gain access to emotional meaning through various types of prosodic and gestural encoding at very young ages. Previous studies have shown that 3.5-month-old infants have a sophisticated ability to recognize emotion in facial expressions (Kahana-Kalman and Walker-Andrews 2001; Farroni et al. 2005), yet while infants in their first year of life seem to be able to distinguish emotional expressions, the ability to explicitly name such emotions only appears much later. There is evidence that by two years of age children can already appropriately match the name of at least some of these emotions with the corresponding facial expressions (Izard 1971; Markham and Adams 1992; Widen and Russell 2003, 2008; Nelson and Russell 2011). Furthermore, preschool children can recognize happiness, sadness, anger, fear and disgust on the basis of facial expressions (e. g. Harrigan 1984; Denham et al. 1990; McClure 2000; Tracy et al. 2005; Gagnon et al. 2010; Székely et al. 2011; Gagnon et al. 2014; Kayyal and Widen 2015; Nelson and Russell 2015). However, other research suggests that learning to recognize expressions is a protracted process which lasts into the school years (see amongst others, Herba et al. 2006; Gao and Maurer 2010; Roberson et al. 2010; Nelson and Russell 2012; Widen 2013).

Though research on infants' and young children's sensitivity to emotions has largely concentrated on the identification of facial expressions, several studies have also shown infants' early sensitivity to (a) the emotional valences of speech as expressed through prosodic cues, and (b) the correspondences between prosodic and gestural cues in the expression of emotions. Mastropieri and Turkewitz (1999) showed that newborns react differently when presented with vocal cues expressing different emotions in their first language. Furthermore, 5-month-old infants have been found to react differently depending on the accompanying affective facial expression such as smiling in response to approval vocalizations (Fernald 1993). They are also able to distinguish between sad and angry vocalizations when they are accompanied by matching facial expressions (Vaillant-Mo-

lina et al. 2013) and can match both positive and negative affective vocalizations with the corresponding face (Walker-Andrews and Grolnick 1983). Results obtained by both electrophysiological and brain imaging studies have confirmed these results (e. g. Grossmann et al. 2010; Zhang et al. 2014). Nonetheless, the above-mentioned studies have either used prosody discrimination tasks or facial expression-prosody matching tasks, neither of which really show that infants are attributing affective stance to the speaker. More helpful in this respect are studies that include behavior regulation tasks with toddlers. For example, the results of Mumme et al.'s (1996) study using a novel toy paradigm with 12-month-old children showed that fearful emotional prosody conveyed by meaningful utterances was sufficient to elicit an appropriate behaviorial reaction in children. Similarly, Vaish and Striano (2004) demonstrated that 12-month-old infants could be persuaded to cross a visual cliff purely on the basis of positive vocalizations. However, the target utterances used also contained semantic meaning, and in fact there are few studies that allow us to conclude that prosody was the only factor at work in emotion recognition tasks. One example is Hoicka and Wang (2011), who showed that 15-month-olds were able to make prosody-meaning associations independently of situational and lexical context. In this study, the experimenter produced either humorous or sweet vocal cues, the only difference between the two utterances being the mean values for pitch, with higher values in the humorous utterances. After infants were exposed to the vocal cues, they observed the experimenter performing either matching or mismatching behavior (i. e., either a humorous action or a "sweet" action), with a neutral facial expression in all conditions. The infants looked longer at the mismatched behavior than at the matching behavior. These results showed that 15-month-olds are able to develop expectations about an adult's behavior exclusively on the basis of prosody. Finally, some studies have also assessed infants' early ability to express their own emotions through differentiated prosodic patterns and facial expressions (Papaeliou et al. 2002; Scheiner et al. 2003; Oller et al. 2013; Jhang and Kimbrough 2017). Thus, infants are not only able to perceive emotions early, they also learn to express affective cues themselves through protophones and facial affect valences within the first 3 months of life (Jhang and Oller 2017).

1.2 Nonlinear development of emotional and pragmatic comprehension in preschool years

Though the above-mentioned results point to an early (and parallel) development of infants' understanding of emotional meaning encoded through both prosody and facial cues, this seems to change as children get older and start acquiring the

lexical and grammatical features of language. There is some evidence that infants' and toddlers' use of prosodic and facial cues begins to change once they start to develop a lexicon. Friend (2001) investigated 15- and 16-month-olds' sensitivity to prosody as well as facial cues vs. lexical content. Before the children were given the opportunity to play with a novel object, they saw videos of a speaker who conveyed either an approving or disapproving message. The behavior of those children who understood the lexical meaning of the message was better regulated through lexical content than through prosody or facial cues, making receptive vocabulary a significant predictor of the children's behavior. This led Friend to conclude that there is a transition stage from affective to linguistic meaning around 15 months of age. A similar tendency was found by Lawrence and Fernald (1993). In their study, while 9-month-olds were better regulated through tone of voice than through lexical content, 18-month-olds were better regulated through lexical content.

Furthermore, it has been found that when presented with multiple cues to emotion, young children tend to rely on the lexical cues when prosodic cues and the lexical message conflict (Friend and Bryant 2000; Friend 2001; Morton et al. 2003). For example, Morton et al. (2003) carried out three experiments with 4- to 10-year-old children and adults. In experiment 1, children were exposed to 20 sentences with matching or mismatching propositional information (such as *My mommy gave me a treat* vs. *I lost my sticker collection*) and paralinguistic cues (happy vs. sad prosody). 'Happy' prosody sentences were produced with higher pitch level, greater pitch and loudness variation and a faster speaking rate, while 'sad' prosody was produced with lower average pitch, attenuated pitch and loudness variation and slower speaking rate relative to the happy paralanguage. In experiment 2, in order to remove any possible interference from the propositional content, the sentences were presented in a foreign language (Italian), and in experiment 3, the sentences were low-pass filtered to remove potential distraction from the semantic content. The overall results of these experiments showed that 9- to 10-year-old children and adults judged the speaker's feelings by *how* s/he spoke whereas children aged 8 or younger judged the speaker by *what* s/he said. However, results from experiment 2 and 3 showed that even 4-year-old children were able to attribute an emotion to the affective paralanguage of both the foreign language and the low-passed filtered speech, although accuracy improved with age. Waxer and Morton (2011) confirmed this lexical bias in 6-year-olds, since the children in their study proved to be inflexible in their interpretations of conflicting speech cues when they had to decide between emotions based on words and emotions based on prosody (happy/neutral/sad paralanguage). Aguert et al. (2013) suggested that it is not that lexicon or context invokes a bias, but rather that prosody plays a 'subordinate' role when it is in competition with situational

context, a phenomenon that persists even into children's early teenage years. In their study, Aguert et al. (2013) presented 5- to 13-year-old children with a judgment task in which they were exposed to animal figures located in neutral situations, and then asked them to judge the emotional state (happy or sad) of the animal on the basis of prosody alone since both lexical and contextual cues were devoid of emotional valence. The utterances the children heard while viewing the animals were five syllables long and deliberately made unintelligible, with the syllables being randomly mixed. The prosody employed was described as either positive (happy) or negative (sad). The children were then asked to judge the animal's emotional state by pointing to a drawing with either a happy face or a sad face with which the children had been previously familiarized during a pretest. They were also asked to verbally explain their judgment. The results of the experiment showed that the youngest children in the study (the 5-year-olds) struggled to infer the speaker's emotional state on the basis of prosody alone and only performed at chance level. The authors concluded from this that there are no specific biases involved, but that prosody is simply a difficult cue for preschool children.

These above-mentioned results stand in stark contrast to other developmental findings showing that from very early in development (and also later), children are extremely sensitive to prosodic cues, which in fact act as bootstrapping mechanisms for language development (see Hoehle 2009: for a review). In addition, a number of studies have shown that toddlers can regulate their behavior depending on prosodic cues alone (see Hoicka and Wang 2011: , amongst others). Why do infants clearly detect and respond to emotional cues in infancy and early childhood (and in a variety of tasks), but then struggle to detect prosody in the preschool years? Apart from the fact that at later ages children need to integrate information coming from prosody with other contextual and lexical cues, some authors have suggested that while happy and sad contours may be accessible to babies in infancy, they may lose their iconicity through reinterpretation during the language acquisition process, and the late learning of connections between pitch and emotion could be due to the "complexity of pitch-contour patterning in the language as a whole" (Quam and Swingley 2012). Also, as Aguert et al. (2013) note, the fact that the few studies available reflect a variety of experimental designs is not ideal and has really yielded only fragmentary knowledge about the ability of children to detect emotional stance from prosody.

Regarding the comprehension of meanings through facial expression, there are a number of papers that have examined children's ability to recognize emotion through both facial expressions and affective prosody (amongst others McCluskey and Albas 1981; Nowichi and Duke 1994; Nelson and Russell 2011; Quam and Swingley 2012). In particular, several studies have demonstrated an advantage in the preschool years compared to vocal expression. For example, Nelson and

Russell (2011) tested preschool children's (3–5 years old) ability to label emotions (happiness, sadness, anger and fear) based on video clips which were produced in four different conditions: face-only, body posture-only, voice-only and multi-cue (i. e., face + body + voice). Results showed that preschoolers were equally able to provide labels in the face-only and multi-cue conditions, yet they were significantly less accurate in the body posture condition. However, the children performed worst in the voice-only condition. Furthermore, Quam and Swingley (2012) found that while 4- and 5-year-olds were consistent in detecting happy or sad prosody (created by manipulating the low-pass filter on the audio recording) to decide whether a puppet had succeeded or failed at a task, 2- and 3-year-olds depended more on facial-gestural and body language cues. Clearly there remains considerable room for further exploration of young children's sensitivity to affective stance on the basis of prosody, facial expression and lexical information, and the interaction and relative importance of these factors.

1.3 Children's sensitivity to polite stance

Children's multimodal acquisition of politeness in requests is greatly under-researched and little is known about children's ability to infer a speaker's polite stance on the basis of intonation and/or facial cues only. Also, at this moment, still surprisingly little is known about children's sensitivity to a speaker's polite stance and the role affect plays therein. Focussing on first grade children's and adults' understanding of the sociality of emotions, Camras et al.'s (1985) study showed that children expect there to be a relationship between speaker affect and directive choice. Particularly, angry speakers are expected to be less polite than happy or neutral speakers. The authors concluded from this that adults and children sometimes make inferences about a speaker's emotions based on his or her directive choice. In another study, Shochi et al. (2009) focussed on children's understanding of politeness from a multimodal point of view. Investigating 9- and 10-year-old Japanese children, they found that facial cues were beneficial for the processing of politeness and impoliteness meanings (Shochi et al. 2009). They also found that, in comparison to adults, children relied on facial cues earlier than auditory information to understand politeness meanings. One of the only and most comprehensive studies with a focus on preschool children's developing perception of politeness by including intonation as a cue is Bates' (1976) study. Bates (1976) experimentally tested whether 60 Italian children aged 3–6 could perceive politeness as encoded through lexical cues or through prosody only. The children were asked to judge which frog puppet made the most polite request as the experimenter varied the puppet's use of different lexical cues and tones of

voice. Tone of voice varied from harsh to gentle and the verbal message *dammi un dolce* 'give me a candy' was either accompanied by *per favore* 'please' or not. The results showed that the children had acquired *per favore* 'please' as a politeness marker by age 3 but the use of gentle intonation as a strategy only reached significance after 4 years of age. Nonetheless, Bates argued that the younger children may have judged the harsh tone as 'nicer' because they found it amusing, hinting at an earlier sensitivity to intonation after all.

While at this stage it is not clear whether preschool aged children can infer a speaker's polite stance from prosodic and/or facial cues only, our hypothesis is that by age 3 children will be able to successfully detect a change in polite stance by only accessing changes in these specific cues. Despite contradictory results in the literature on preschool children's sensitivity to emotional states encoded through prosody and facial gesture, we claim that the use of a child-directed pragmatically relevant task will allow us to adequately assess this issue.

2 The current study

The current study intends to determine whether 3-year-old children are able to distinguish a polite stance from a non-polite stance in requestive speech acts exclusively on the basis of prosodic (in particular intonation) cues, solely on facial cues, and on the basis of the two combined, removing any possible lexical or contextual bias. To do so, we compared children's behavior on a between-subjects polite stance comprehension task where a set of requests were presented to the child which always contained the same polite stance lexical cue (the word *please*, e. g., *Can you give me the ball, please*) and the same neutral speech act situation as a situational prompt. We adapted Bates' (1976) experimental procedure whereby the subject is asked to give an object to the person who seems to be asking more nicely. Crucially, in order to investigate children's sensitivity to intonational patterns and facial cues, the experimental materials were presented in three different between-subject modalities: (a) audio-only (AO), with just verbal and prosodic cues available; (b) visual-only (VO), with exclusively non-verbal cues available; and (c) audio-visual (AV), with both verbal and non-verbal cues available. This is an adaptation of the methodology used in Hübscher, Esteve-Gibert, Igualada et al. (2017) and originally created by (Armstrong et al. 2014). Importantly, we chose to use intelligible speech but to keep the lexical cues constant in both the polite affect and the non-polite affect conditions, thus controlling for lexical content. Both intonational and facial cues were varied, with a falling nuclear configuration (L* L%) and frown signaling the non-polite affect condition, and

a rising nuclear configuration (L+H* H%) and smile signaling the polite affect condition.

Based on the previous literature, we had several hypotheses regarding children's ability to access a speaker's polite stance through intonational and gestural information. Despite incongruent findings in the previous literature, we hypothesized that the use of a child-directed and pragmatically relevant task would allow us to adequately assess children's ability to infer meaning based on facial cues and intonation only, and the combination of both. First, we hypothesized that, in line with Quam and Swingley (2012) and Nelson and Russell (2011) children should be able to detect a speaker's polite stance best in the AV condition, where both intonational and facial cues are present. However, contrary to some other previous research, and more consistent with the early exploitation of pitch and facial cues for communicative functions, we hypothesized that both intonation and facial cues are strong cues in the age under investigation, just as in earlier ages, and children should be able to infer to a certain degree another person's polite stance through those cues in the AO and VO conditions.

3 Method

3.1 Participants

Thirty-six (18 female and 18 male) American English-speaking children participated in a between-subjects polite stance comprehension task. Five additional children had to be excluded for various reasons (they were bilingual, had developmental problems, or dropped out of the experiment). Their ages ranged from 2;10,22 to 4;0,16 (mean age: 3 years and 5 months, *SD* = 0.31). The participants were mostly high socioeconomic status white children visiting the Center of Science and Industry, a science education center in Columbus, Ohio (Wagner et al. 2015). Parental consent was obtained before the experiment. Children were given an animal-shaped stamp on the hand as a reward for participation. All research practices and consent forms were reviewed and approved by the Ohio State University Social and Behavioral Sciences Institutional Review Board.

3.2 Preliminary elicitation study with adults

In order to create the stimulus materials to be used in the study, it was first necessary to determine the prosodic and facial cues characteristics that are used in

American English to encode polite and non-polite stances. To this end, a free discourse elicitation task was carried out with ten American English native speakers who were students or researchers at Ohio State University. The participants were asked to imagine themselves in the company of a child and read a set of context prompts in which they had to make a request to that child. To give them a better idea of the target age group, next to the text describing the context the participants were provided with a picture of a 3-year-old child. Two examples of these contextual prompts are given in (1) and (2) (see appendix for all prompts). In (1) the context prompts the speaker to ask for help in a polite way (i. e., a polite affect request) while context (2) prompts the speaker to produce a command rather than a question (i. e., a non-polite affect request).

(1) **Polite stance condition**: You have both your hands full of plates on your way to the kitchen and you have just dropped a fork. Ask the child nicely to give you the fork.
(2) **Non-polite stance condition**: The child is very excited and plays continuously with a noisy toy but you want him/her to be quiet. You're quite annoyed. Tell the child to give you the toy.

Video recordings were made of the participants as they responded to the prompts, yielding recordings of a total of 100 requests (10 requests × 10 speakers). Analysis of the recordings showed that the lexical structures used most often by speakers were the conventional request questions *Can you give me xy?* (with or without *please*) and *Could you give me xy?* (with or without *please*) in the polite affect condition and the imperative *Give me xy* (with or without *please*) in the non-polite affect condition. The recordings were then analyzed for prosodic content using MAE_ToBI (Beckman et al. 2005). MAE_ToBI is an annotation system used for labeling intonation and prosody in databases of spoken Mainstream American English (for a consensus account of English intonation and prosody on which the MAE_ToBI system is based, see Beckman et al. 2005). The frequency distribution of the types of intonation contours displayed in the non-polite vs. polite affect requests can be seen in Table 1.

Table 1: Prosodic cues displayed in the requests

	Non-polite affect	Polite affect
L* L%	49 (98 %)	7 (14 %)
L* H%	0	5 (10 %)
L+H* H%	1 (2 %)	38 (76 %)

The results of the prosodic analysis showed that there was a clear preference for a falling (L* L%) nuclear configuration in the non-polite affect condition and a rising (L+H* H%) nuclear configuration in the polite affect condition. Finally, the recordings were analyzed in terms of the facial cues that speakers had used. It was found that while in the non-polite affect condition participants displayed a stern facial expression, in the positive affect condition they consistently displayed a polite smile.

On the basis of this analysis, it was decided that in the subsequently prepared experimental stimulus materials the non-positive affect condition would be characterized by the L* L% nuclear configuration prosodically, and a stern facial expression, while the polite affect condition would be characterized by the L+H* H% nuclear configuration and a smiling facial expression. Regarding the lexicon, we decided to add *please* at the end of the conventional request structure *'Can you give me xy'* since it is a politeness cue that is usually taught to children early on in order to make a request sound acceptable. In order to control for lexical input, the actual utterance would be the same in both polite and non-polite affect conditions.

3.3 Experimental materials

Six female students or researchers from Ohio State University volunteered to take part in preparing the stimulus materials. They were individually videotaped while producing a set of sentences with the same target structure *'Can you give me the [toy] please'*. The name of one of six toy items, namely *ball*, *shark*, *frog*, *bear*, *horse* and *duck*, was inserted in the sentence with a different toy per speaker. The speakers each produced one request in two conditions, namely non-polite and polite stance, in which they replicated the main intonational and facial features of the non-polite and polite stance conditions described in section 3.2. A total of 12 sentences were obtained (1 target structure × 1 toy × 2 affect conditions × 6 speakers).

The four panels in Figure 1 illustrate the pitch contours and facial expressions used in the two AV conditions.

In order to be sure that our stimuli were clearly distinguished as portraying polite stance vs. non-polite stance, we performed a preliminary control survey with the online survey platform SurveyGismo. Ten adult American English speakers were recruited and were asked to judge each of the six paired video recordings in terms of which one of each pair depicted the more polite request. Their responses were 100 % consistent across raters.

PowerPoint presentations (PPTs) were then created for each of the three between-subject experimental conditions (AO, VO and AV). Each PPT contained

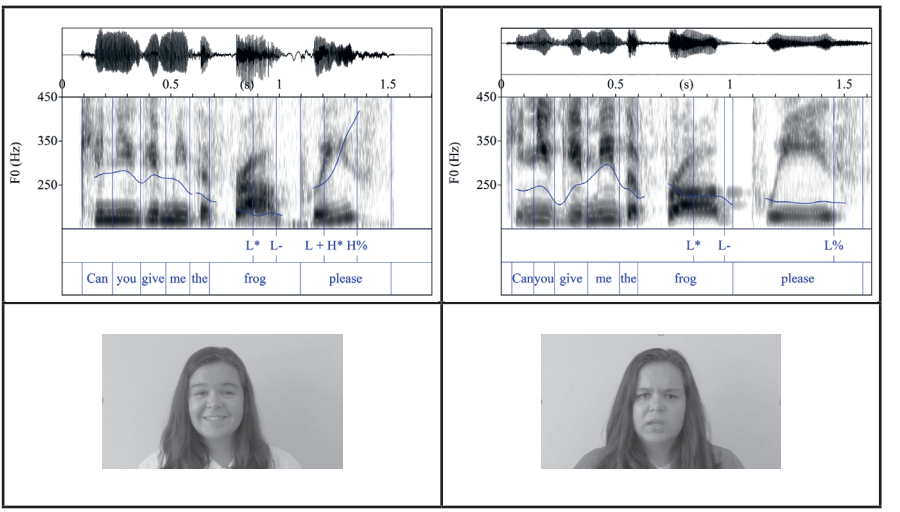

Figure 1: Pitch tracks, spectrograms and waveforms (top panels) and facial-gestural cues (bottom panels) for the polite affect request (left panels) and non-polite affect request (right panels) stimuli used in the AV condition.

eight slides in the following order: a familiarization slide, three test trials, the familiarization slide again, and finally three more test trials. In the AV version of the PPT, the full audiovisual recordings of the speakers were embedded in the trial slides. The PPT presentation for the VO condition was identical except that the audio track was eliminated from each of the six videos. Thus, the children would only be exposed to facial cues without any audio cues. Finally, the PPT presentation for the AO condition included the audio track in the trial slides but instead of video content subjects were shown two grey squares under which they could merely see the outline of a face, this was done to remove any possible information from facial cues but enable participants to associate the sound with a person. In order to counterbalance the presentation order of the stimuli within each trial, we created four different PPTs for each modality of presentation.

3.4 Procedure

First, each child was randomly assigned to one of the three conditions, namely AO (12 children), VO (12) and AV (12). The child was seated next to the experimenter, a native American English speaker, on the floor in a quiet room in the Center of Science and Industry Museum. As a warm up, the child was asked to name each

of the six toy items that would be seen in the experimental trials, namely a ball, shark, frog, bear, horse and duck. For the experiment proper, the child faced a laptop on which the polite stance comprehension task was presented with two empty buckets placed between the child and the computer.

To familiarize him/her with the test procedure, the child was told that s/he was going to play a game. Two animated stars appeared at the left and right of the screen. After the star on the left said "I'm the blue star", the experimenter passed a fuzzy ball to the child and asked *Can you put this fuzzy ball in the bucket right in front of the blue star?* The same procedure was then followed for the red star on the right. Then the child was presented with the test trials. In each trial, the screen showed two embedded videos (in the AV and VO condition PPTs) or still pictures (in the AO condition PPT) of two speakers, one to the left, the other to the right. Because it was the same speaker in both videos or pictures, the experimenter referred to them as "sisters that look very much alike" (see Figure 2). Each trial slide also depicted one of the target toy items in the top right or left corner of the screen (e. g., in Figure 2, a ball at top right).

Figure 2: Example of a test trial PPT slide, showing position of stimulus recordings or photos and one of the toy items. In this case, the 'sisters' both portray a neutral expression because the slide is from the audio-only PPT.

The experimenter then directed the child's attention to the toy item by saying "What do you think they want? Both of them are going to ask you to give them the ball. You have to listen/watch very carefully. Can you give the ball to the person who asks more nicely?". The two embedded videos (or audio tracks in the AO condition) were then played consecutively, one in which the speaker made a request while displaying the polite affect cues and the other in which the speaker made the same request while displaying the non-polite affect cues. The child then

indicated which one of the two speakers s/he thought had asked more nicely by placing the named object (e. g., the ball) in the bucket that was in front of that person. In each trial a new set of "sisters" and a different toy was displayed.

After the first three test trials, the child was again shown the familiarization activity with the stars and then performed the remaining three test trials. In half of the trials, the polite affect stimulus recording was presented first and was located on the left side of the screen; in the other half, the polite affect choice was presented second and was located on the right side of the screen. Thus, as noted, in total each child performed six test trials and two familiarization trials. The whole experiment lasted about 5 minutes and was videotaped.

3.5 Coding

For the coding of responses, a trial was coded as "correct" and scored "1" if the child put the toy into the bucket in front of the speaker who had produced the "polite affect" cues, in other words, the speaker who had seemed nicer in making their request. In their responses, children quite often pointed first towards the person they thought seemed nicer and only afterwards put the object into the bucket in front of that person. However, those children who pointed first always put the object in the corresponding bucket. The combination of "polite affect" cues varied according to the modality, in the AV condition consisting of a verbal request (e. g., *Can you give me the ball please?*) with a rising nuclear configuration (L+H* H%) and a smile; in the AO condition a rising nuclear configuration in the audio playback but a neutral facial expression in a still photo; and in the VO condition a smile in the video playback but no audio track. If the child put the bucket in front of the speaker who displayed "non-polite affect" cues, this was counted as an "incorrect" response and scored as "0". A total of 216 trials were analyzed (6 test trials × 36 participants).

4 Results

In order to test overall performance in each condition, a binomial test was applied to the data. A bar graph illustrating the mean ratios of correct responses for each modality is shown in Figure 3.

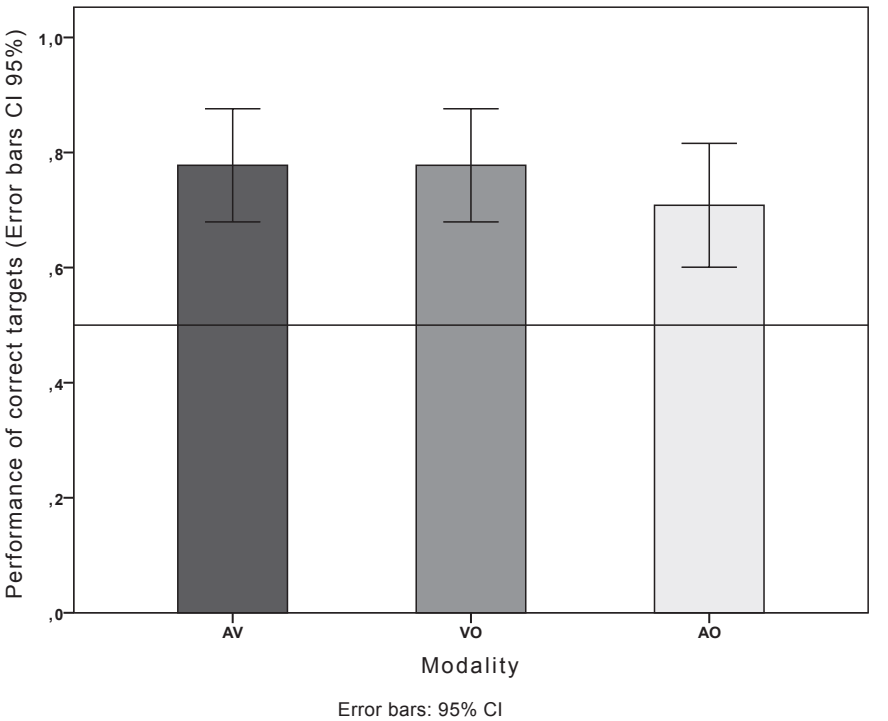

Figure 3: Mean ratio of correct responses by modality (AV = audio-visual, VO = visual-only, AO = audio-only)

The results of the binomial test indicated that the ratio of correct interpretations of polite stance was .78 in conditions AV and VO, thus greater than the chance ratio of .50 ($p < .001$, 1-sided), as predicted. Contrary to our predictions, however, the ratio of correct interpretations of polite stance in the AO condition was .71, also greater than chance ($p < .001$, 1-sided). Thus, the children performed significantly better than chance in all three conditions, indicating that they were sensitive to both prosodic and visual characteristics of polite stance. To assess potential differences between modalities, a second Generalized Linear Mixed Model (GLMM) was performed in which Choice was set as the dependent variable, Modality was set as fixed factor and Subject and intercept were set as random factors. Again, modality in the GLMM analysis was not found to be significant, with $F(2, 213) = 0.315$, $p = .671$. In other words, the children performed as well in the AO condition as in the other two conditions. Odd ratios (Exp(β)) were calculated with the AV condition as baseline in order to analyze effect sizes with logistic regressions.

AV vs. AO: (β =-.485, *SE* = .596, *p* = .416, *Exp(β) = .616)*
AV vs. VO: (β =-.066, *SE* = .609, *p* = .914, *Exp(β) = .936)*

Odd ratios represent the odds that an outcome will occur given a particular exposure and in comparison to the odds of the outcome occurring in the absence of that exposure (Szumilas 2010). Odd ratios superior to 1 are associated with higher odds of outcome compared with the baseline category. In the two comparisons, the odd ratios were inferior to 1. In the first case (AV-VO) the β regression coefficient indicates that the probability of a child in the VO condition detecting a speaker's polite stance is about .066 times higher than for a child in the AV condition. In the second case (AV-AO) the β regression coefficient indicates that the probability of a child in the AO condition detecting a speaker's polite stance is about .485 times higher than for a child in the AV condition. Even though the difference between AV and AO compared to AV and VO is bigger based on the effect sizes, there is no statistically significant difference between the conditions, implying that this tendency might only be found in the present data sample and not in the total American English population in the USA which is represented by the sample.

5 Discussion and conclusions

This study tested 3-year-old children's understanding of a speaker's polite stance in requests by analyzing their ability to exploit audio cues (intonation), visual cues (facial expression) and the combination of both in a situation where lexical cues were controlled for. The results of the study showed that children performed significantly above chance in all three modalities of presentation, namely audio-visual, visual-only and audio-only. In the experimental task, children had to choose from two possible answers and the only cues they could access in order to detect a difference between the two requests were either intonation or facial cues, or both, since lexical content and context were controlled for. The results of the experiment add interesting new insights to research on children's ability to infer politeness from both intonational and facial cues at a very early age.

First, the results confirm that facial expression is a cue that children can broadly use to access polite meanings conveyed by speakers. These results are consistent with other studies suggesting that preschoolers have the ability to access both emotional information and more complex pragmatic meanings which might require them to understand the other person's pragmatic perspective through gestural and facial cues (Butcher and Goldin-Meadow 2000; Kelly

2001; Armstrong et al. 2014; Hübscher et al. 2017; Armstrong and Hübscher 2018). Furthermore, interestingly, children performed similarly well in the visual-only as in the audio-visual condition and there was no additive effect in the audio-visual condition.

Second, the children performed as well in the audio-only condition as in the other two conditions, suggesting that children's sociopragmatic understanding of intonation is as good at this age as is their understanding of facial cues encoding the same meaning. This is the first study showing that 3-year-olds can infer a speaker's polite affective meaning just as well from intonation as they can from facial cues. While most previous research has neglected to investigate children's ability in inferring politeness meaning from intonation and facial/gestural cues, those studies that did, either found later sensitivity to intonational cues (Bates 1976), or found that facial cues play an important scaffolding role in older children's understanding of politeness meanings (Shochi et al., 2009).

Thus the present results differ from the results of other studies that have found a lack of sensitivity in pre-school and primary school children to prosodic cues to a speaker's emotional state relative to lexical or contextual cues (Morton and Trehub 2001; Morton et al. 2003; Waxer and Morton 2011; Quam and Swingley 2012; Aguert et al. 2013). The question that arises is why the present findings are so discrepant with prior studies on children's ability to infer emotional/pragmatic meaning from intonational cues. There are several plausible explanations which might account for this difference. First of all, it might be due to the different nature of the experimental speech stimuli employed. In research on emotional state encoded through intonation, the experimental stimuli have often been pseudo-utterances or low-pass filtered speech, with children being told that the animal characters displayed were speaking another language. While this strategy certainly controls for any possible lexical interference or bias, it may have significantly affected the children's ability to interpret the pragmatic intonation of the message. In the present study, in order to avoid these issues, children were exposed to a pragmatically relevant task involving a request situation whereby they were asked to judge a real person's intelligible polite stance as encoded through either lexical and prosodic cues alone, or facial expressions alone, or all of these things at once. Regarding the materials, they were exposed to actual video recordings of child-directed speech produced by a set of speakers.

Second, in most studies on children's sensitivity to (emotional) meaning encoded through prosody, prosody is actually poorly defined and is often referred to as a "paralinguistic cue". For example, Morton and Trehub (2001) characterize the target prosodic cues used in their experiment as "affective paralanguage", without giving any further information. Aguert et al. (2013) likewise fail to offer information about the prosodic cues that their child participants were exposed

to beyond the observation that the prosody was either happy or sad and thus emotionally salient. Even when such information is provided, it may be that the prosodic features selected to cue emotion were excessively subtle. For instance, though Aguert et al. (2013) define the prosodic cues they employed with sufficient detail, informing us that their positive-affect prosody cues had a high F0 mean accompanied by a large F0 range while their negative-affect cues had a lower, more monotonous F0 pattern, the question remains as to whether these differences were sufficiently salient as cues to emotion. In order to avoid this issue, we defined and controlled for the specific contours of the intonational cues employed in our experiment (basically rising and falling interrogative pitch contours) in accordance with the Autosegmental Model (Pierrehumbert 1980; Jun 2010).

Third, the experimental design could have played a role as well. In the current study children were tested in a forced-choice task in which they were exposed to two video stimuli in sequence. It may be that this experimental set-up actually helped children to perform better compared to other designs where children were exposed to one stimulus at a time and then had to make a decision between two possible answers. Though at this point it is not possible to draw any firm conclusions about the potential effects of this difference in experimental design, it might be worthwhile to explore this issue further. Futhermore, when analyzing the understanding of emotions/polite stance it must be borne in mind that the affects' valence depends on context, which is itself partially generated by societal factors. Thus, in order to understand the intentionality of affect, an appropriate contextualization and social context must be provided (see e. g. Wilutzky 2015; Khu et al. 2017). Finally, there might be a difference in the way preschool children access the various meanings encoded through prosody. It thus remains to be tested whether with the current experimental paradigm children could just as easily access a person's emotional state as they do someone's polite stance.

While our results differ from those found in some prior studies, we would like to highlight the fact that the sensitivity to intonation by 3-year-old children we report here is consistent with reports in the literature. Many studies in fact suggest that prosody helps infants detect emotional information, and there are at least three studies with children aged 12 to 36 months showing that prosodic information in isolation was sufficient to regulate their behavior. For example, Mumme et al. (1996) showed that 12-month-old children regulated their behavior when hearing emotional prosody that made them afraid. Similarly, Hoicka and Wang (2011) showed that 15-month-olds responded appropriately when the experimenter produced either humorous or sweet vocal cues combined with mismatched and matched behavior. These results also seem to be in line with Hübscher et al.'s (2017) study which showed that 3-year-old children are able to attribute mental states (in this case, epistemic stance) to another person on the basis

of intonation only, and crucially that they are better at detecting this epistemic stance through intonation than through lexical marking. Together, these results suggest that between 1 and 3 years of age children are able to make the appropriate semantic links between prosody and pragmatic and emotional information (see also Armstrong and Hübscher 2018; Esteve-Gibert and Prieto 2018).

To sum up, clearly intonation and facial cues have been grossly overlooked in most studies on children's developing understanding of politeness, and the current results provide new evidence that those cues might play a much bigger role in this process. This is the first study of its kind to show that young (3-year-old) children are sensitive to sociopragmatic meaning conveyed through intonation and facial cues and that they are able to metacognitively react to it. This has implications for parents, caregivers and preschool teachers because it suggests a means of scaffolding children's sociopragmatic awareness, which often solely focusses on verbal content. However, judging from our findings, it seems clear that neither prosodic (in particular intonational) nor facial cues should be forgotten. One possible limitation of the present study is that the data set includes a relatively small number of children in each condition and future research should address whether the trends observed in the present study are applicable to a greater number of children. In this context it would also be interesting to compare American English-speaking children with children from other language backgrounds, to see whether the currently obtained results are comparable, or whether there are cross-cultural differences in children's development of politeness understanding. For example, it might well be that languages which do not display an intonational contrast between polite and non-polite requests offer a greater difficulty for children to grasp the prosodic cues of politeness. Research should also consider testing children on perspective-taking tasks, such as for example theory of mind and emotion detection to investigate more closely individual differences. Finally, it should be noted that this study has examined only one speech act, e. g., requests, at a specific age and it would be interesting in future research to widen up the scope and investigate whether children's development of politeness behaves similarly across different speech acts and also across a wider age span. In general, more studies with accurate descriptions of intonational patterns are needed (and which include pragmatically relevant situations) to start assessing how children develop their prosodic awareness in relation to social and pragmatic meanings.

Acknowledgements: We would like to express our gratitude to Faith Stagge, whose assistance in collecting the data at the Center of Science and Industry in Columbus, Ohio, was invaluable. Many thanks also to the students and researchers at Ohio State University who participated in our discourse elicitation task and

the stimulus recordings. We are grateful to the linguistics team at Ohio State University who helped with some native speaker inputs along the way. Thank you to Shari Speer and her team at the Psycholinguistics Lab (Kiwako Ito and Kathryn Kimberley in particular), where some preliminary results were presented, and also to Mary Beckman and Ajou Chen for making valuable comments. Thanks also go to Joan Borràs-Comes for helping with the statistics.

Funding: This research has been funded by a research grant awarded by the Spanish Ministry of Science and Innovation (PGC2018-097007-B-I00 'Multimodal Language Learning (MLL): Prosodic and Gestural Integration in Pragmatic and Phonological Development') and by a grant awarded by the Generalitat de Catalunya (2017 SGR _ 971) to the Prosodic Studies Group. The first author also acknowledges a travel grant received by the Department of Translation and Language Sciences at the Universitat Pompeu Fabra.

References

Aguert, Marc, Virginie Laval, Agnès Lacroix, Sandrine Gil & Ludovic Le Bigot. 2013. Inferring emotions from speech prosody: Not so easy at age five. *PLoS ONE* 8(12), 1–9. doi:10.1371/journal.pone.0083657

Andersen, Peter A. & Laura K. Guerrero. 1998. *Handbook of communication and emotion: Research, theory, applications, and contexts*. San Diego: Academic Press.

Armstrong, Meghan. E., Núria Esteve-Gibert & Pilar Prieto. 2014. *The acquisition of multimodal cues to disbelief.* Paper presented at Speech Prosody 2014, May 20–23, Dublin, Ireland

Armstrong, Meghan. E. & Iris Hübscher. 2018. Children's development of internal state prosody. In Pilar Prieto & Núria Esteve-Gibert (Eds.), *Prosodic Development in First Language Acquisition* 272–293. John Benjamins.

Bänziger, Tanja, Marcello Mortillaro & Klaus Scherer. 2010. Introducing the Geneva Multimodal Emotion Portrayal (GEMEP) corpus. In Klaus Scherer, Tanja Bänziger, & Etienne Roesch (Eds.), *Blueprint for affective computing: A sourcebook* 271–294. Oxford: Oxford University Press.

Bates, Elizabeth. 1976. *Language and context: the acquisition of pragmatics*. New York: Academic Press.

Beckman, Mary E., Julia Hirschberg & Stefanie Shattuck-Hufnagel. 2005. The original ToBi system and the evolution of the ToBi framework. In Sun Ah Jun (Ed.), *Prosodic Typology: The Phonology of Intonation and Phrasing* 9–54. Oxford: Oxford University Press.

Brown, Lucien & Prieto Prieto. 2017. (Im)politeness: prosody and gesture. In Jonathan Culpeper, Michael Haugh, & Dániel Kádár (Eds.), *The Palgrave Handbook of Linguistic (Im)politeness* London: Palgrave Macmillan.

Brown, Penelope & Steven Levinson. 1987. *Politeness: Some universals in language use*. Cambridge, UK: Cambridge University Press.

Butcher, Cynthia & Susan Goldin-Meadow. 2000. Gesture and the transition from one- to two-word speech: When hand and mouth come together. In Dacid McNeill (Ed.), *Language and Gesture* 235–257. New York: Cambridge University Press.

Camras, Linda A., Toni M. Pristo & Mari J. Brown. 1985. Directive choice by children and adults: Affect, situation and linguistic politeness. *Merrill-Palmer Quarterly* 31. 19–31.

Culpeper, Jonathan. 2011. "It's not what you said, it's how you said it!": prosody and impoliteness. In Linguistic Politeness Research Group (Ed.), *Discursive approaches to politeness* Vol. 8, 57–83. Berlin: Mouton de Gruyter.

Culpeper, Jonathan, Derek Bousfield & Anne Wichmann. 2003. Impoliteness revisited: With special reference to dynamic and prosodic aspects. *Journal of Pragmatics* 35(10–11). 1545–1579. doi:10.1016/S0378-2166(02)00118-2

Denham, Susanne A., Marcia Mckinley, Elizabeth. A. Couchoud & Robert Holt. 1990. Emotional and behavioral predictors of peer status in young preschoolers. *Child Development* 61. 1145–1152.

Esteve-Gibert, Núria & Pilar Prieto. 2018. Early development of prosody-meaning interface. In Núria Esteve-Gibert & Pilar Prieto (Eds.) *The Development of Prosody in First Language Acquisition.* 228–246. Amsterdam: John Benjamins.

Farroni, Teresa, Mark H. Johnson, Enrica Menon, Luisa Zulian, Dino Faraguna & Gergely Csibra. 2005. Newborns' preference for face-relevant stimuli: Effects of contrast polarity. *Proceedings of the National Academy of Sciences of the United States of America* 102(47). 17245–17250. doi:10.1073/pnas.0502205102

Fernald, Anne. 1993. Approval and disapproval: infant responsiveness to vocal affect in familiar and unfamiliar languages. *Child Development* 64(3). 657–674. doi:10.1111/j.1467-8624.1993.tb02934.x

Friend, Margaret. 2001. The transition from affective to linguistic meaning. *First Language* 21(63), 219–243. doi:10.1177/014272370102106302

Friend, Margaret & Judith B. Bryant. 2000. A developmental lexical bias in the interpretation of discrepant messages. *Merrill-Palmer Quarterly* 46(2). 342–369. Retrieved from https://www.scopus.com/inward/record.uri?eid=2-s2.0-0034338064&partnerID=40&md5=afb1e6b4aebfc23d537559d380f546a6

Gagnon, Mathieu, Pierre Gosselin, Isabelle Hudon-Ven Der Buhs, Karine Larocque & Karine Milliard. 2010. Children's recognition and discrimination of fear and disgust facial expressions. *Journal of Nonverbal Behavior* 34(1). 27–42. doi:10.1007/s10919-009-0076-z

Gagnon, Mathieu, Pierre Gosselin & Reem Maassarani. 2014. Children's ability to recognize emotions from partial and complete facial expressions. *Journal of Genetic Psychology* 175(5). 416–430. doi:10.1080/00221325.2014.941322

Gao, Xiaoqing & Daphne Maurer. 2010. A happy story: Developmental changes in children's sensitivity to facial expressions of varying intensities. *Journal of Experimental Child Psychology* 107(2). 67–86. doi:10.1016/j.jecp.2010.05.003

Grossmann, Tobias, Regine Oberecker, Stefan Paul Koch & Angela D. Friederici. 2010. The Developmental Origins of Voice Processing in the Human Brain. *Neuron* 65(6). 852–858. doi:10.1016/j.neuron.2010.03.001

Harrigan, Jinni A. 1984. The effects of task order on children's identification of facial expressions. *Motivation and Emotion* 8(2). 157–169. doi:10.1007/BF00993071

Herba, Catherine. M., Sabine Landau, Tamara Russell, Christine Ecker & Mary L. Phillips. 2006. The development of emotion-processing in children: Effects of age, emotion, and intensity. *Journal of Child Psychology and Psychiatry and Allied Disciplines* 47(11). 1098–1106. doi:10.1111/j.1469-7610.2006.01652.x

Hoehle, Barabara. 2009. Bootstrapping mechanisms in first language acquisition. *Linguistics* 47(2), 359–382. doi:10.1515/LING.2009.013

Hoicka, Elena & S. H. Wang. 2011. Fifteen-month-old infants match vocal cues to intentional actions. *Journal of Cognition and Development* 12(3). 299–314. doi:10.1080/15248372.2010.542215

Hübscher, Iris, Núria Esteve-Gibert, Alfonso Igualada & Prieo Prieto. 2017. Intonation and gesture as bootstrapping devices in speaker uncertainty. *First Language* 37(1). 24–41. Retrieved from https://www.scopus.com/inward/record.uri?eid=2-s2.0-85011392086&doi=10.1177%2f0142723716673953&partnerID=40&md5=9c2d4b2ff417d2015caff0e78961e382

Izard, Carrol E. 1971. *The face of emotion*. New York: Appleton–Century–Crofts.

Jhang, Yuna & D. Kimbrough Oller 2017. Emergence of functional flexibility in infant vocalizations of the first 3 months. *Frontiers in Psychology* 8(MAR). doi:10.3389/fpsyg.2017.00300

Jun, Sun-Ah. 2010. *Prosodic typology: The phonology of intonation and phrasing*. Oxford: Oxford University Press.

Kahana-Kalman, Ronit & Arlene. S. Walker-Andrews. 2001. The role of person familiarity in young infants' perception of emotional expressions. *Child Development* 72(2). 352–369. doi:10.1111/1467-8624.00283

Kayyal, Mary. H. & Sherri. C. Widen. 2015. What made sahar scared?: Imaginary and realistic causes in Palestinian and American children's concept for fear. *Journal of Cognition and Culture* 15(1–2). 32–44. doi:10.1163/15685373-12342139

Kelly, Spencer. D. 2001. Broadening the units of analysis in communication: Speech and nonverbal behaviours in pragmatic comprehension. *Journal of Child Language* 28(2). 325–349. Retrieved from https://www.scopus.com/inward/record.uri?eid=2-s2.0-0035376342&doi=10.1017%2fS0305000901004664&partnerID=40&md5=f902c2e2adcc804f4acc7638af49ca7a

Khu, Melanie, Craig Chambers & Susan. A. Graham. 2017. When you're happy and I know it: Four year-olds' emotional perspective-taking during on-line language comprehension. *Child Devevelopment* 89. 2264–2281.

Langlotz, Andreas. & Miriam. A. Locher. 2013. The role of emotions in relational work. *Journal of Pragmatics* 58. 87–107. doi:10.1016/j.pragma.2013.05.014

Langlotz, Andreas. & Miriam. A. Locher. 2017. (Im)politeness and emotion. In *Palgrave Handbook of Linguistic (Im)politenes* 287–322. London.

Lawrence, L. L. & Anne Fernald. (1993). *When prosody and semantics conflict: infants' sensitivity to discrepancies between tone of voice and verbal content* Poster session presented at the biennial meeting of the Society for Research in Child Development New Orleans, LA.

Locher, Miriam. A. & Regula Koenig. 2014. 'All I could do was hand her another tissue' – Handling emotions as challenge in reflective writing texts by medical students. *SPELL: Swiss papers in English language and literature* 30. 215–236.

Locher, Miriam. A. & Andreas Langlotz. 2008. Relational work: at the intersection of cognition, interaction and emotion. *Swiss Association of Applied Linguistics* 88. 165–191.

Markham, Roslyn & Kym Adams. 1992. The effect of type of task on children's identification of facial expressions. *Journal of Nonverbal Behavior* 16(1). 21–39. doi:10.1007/BF00986877

Mastropieri, Diane. & Gerald Turkewitz. 1999. Prenatal experience and neonatal responsiveness to vocal expressions of emotion. *Developmental Psychobiology* 35(3). 204–214.

Mcclure, Erin B. 2000. A meta-analytic review of sex differences in facial expression processing and their development in infants, children, and adolescents. *Psychological*

Bulletin 126(3). 424–453. Retrieved from https://www.scopus.com/inward/record.uri?eid=2-s2.0-0034185772&partnerID=40&md5=3719b627961e46b85f3b54b345873073

Mccluskey, Ken. W. & Daniel. C. Albas. 1981. Perception of the emotional content of speech by Canadian and Mexican children, adolescents, and adults. *International Journal of Psychology* 16. 119–132.

Morton, J. Bruce & Sandra E. Trehub. 2001. Children's understanding of emotion in speech. *Child Development* 72(3). 834–843. Retrieved from https://www.scopus.com/inward/record.uri?eid=2-s2.0-0035348611&partnerID=40&md5=8a1c-c662ac0a6f90d4121e6d10341871

Morton, J. Bruce, Sandra E. Trehub & Philip David Zelazo. 2003. Sources of inflexibility in 6-year-olds' understanding of emotion in speech. *Child Development* 74(6). 1857–1868. doi:10.1046/j.1467–8624.2003.00642.x

Mumme, Donna L., Anne Fernald & C. Herrera. 1996. Infants' responses to facial and vocal emotional signals in a social referencing paradigm. *Child Development* 67(6). 3219–3237. Retrieved from https://www.scopus.com/inward/record.uri?eid=2-s2.0-0030348025&partnerID=40&md5=13bdeb6acfcba7d688e05a1e2ee0b0f0

Nelson, Nicole. L. & James A. Russell. 2011. Preschoolers' use of dynamic facial, bodily, and vocal cues to emotion. *Journal of Experimental Child Psychology* 110(1). 52–61. Retrieved from https://www.scopus.com/inward/record.uri?eid=2-s2.0-79957624786&doi=10.1016%2fj.jecp.2011.03.014&partnerID=40&md5=8a34211a0f607bdd0478c3291ac42bc2

Nelson, Nicole. L. & James. a. Russell. 2012. Children's understanding of nonverbal expressions of pride. *Journal of Experimental Child Psychology* 111(3). 379–385. doi:10.1016/j.jecp.2011.09.004

Nelson, Nicole. L. & James. a. Russell. 2015. Children distinguish between positive pride and hubris. *Developmental Psychology* 51(11). 1609–1614. doi:10.1037/dev0000048

Nowichi, Stephan. & Marshall P. Duke. 1994. Individual differences in the nonverbal communication of affect: The diagnostic analysis of nonverbal accuracy scale. *Journal of Nonverbal Behavior* 18. 9–35.

Oller, D. Kimbrough, Eugene. H. Buder, Heather. L. Ramsdell, Anne. S. Warlaumont, Lesya Chorna & Roger Bakeman. 2013. Functional flexibility of infant vocalization and the emergence of language. *Proceedings of the National Academy of Sciences of the United States of America* 110(16). 6318–6323. doi:10.1073/pnas.1300337110

Papaeliou, Christina, George Minadakis & D. Cavouras. 2002. Acoustic patterns of infant vocalizations expressing emotions and communicative functions. *Journal of Speech, Language, and Hearing Research* 45(2). 311–317.

Pierrehumbert, Janet. 1980. *The phonology and phonetics of English intonation*. PhD thesis, MIT: Indiana University Linguistics Club.

Quam, Carolyn & Daniel Swingley. 2012. Development in children's interpretation of pitch cues to emotions. *Child Development* 83(1). 236–250. Retrieved from https://www.scopus.com/inward/record.uri?eid=2-s2.0-84856262408&doi=10.1111%2fj.1467-8624.2011.01700.x&partnerID=40&md5=1b3445f440a261e6cf4a998c01b5f2f0

Roberson, Debi, Ljubica Damjanovic & Mariko Kikutani. 2010. Show and tell: The role of language in categorizing facial expression of emotion. *Emotion Review* 2(3). 255–260. doi:10.1177/1754073910361979

Scheiner, Elisabeth, Kurt Hammerschmidt, Uwe Jürgens & Petra Zwirner. 2003. Unterschiede im vokalen Ausdruck von Emotionen bei hörenden und hochgradig schwerhörigen

Säuglingen. In M. Kruse & E. Gross (Eds.) *Aktuelle phoniatrisch-pädaudiologische Aspekte 2003/2004*. Vol. 11, 282–286. Heidelberg: Median Verlag

Shochi, Takaaki, Donna Erickson, Kaoru Sekiyama, Albert Rilliard & Véronique Aubergé. 2009. *Japanese children's acquisition of prosodic politeness expressions*. Paper presented at the Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH.

Spencer-Oatey, Helen. 2011. Conceptualising 'the relational' in pragmatics: Insights from metapragmatic emotion and (im)politeness comments. *Journal of Pragmatics* 43(14). 3565–3578. doi:10.1016/j.pragma.2011.08.009

Székely, Eszter, Henning Tiemeier, Lidia. R. Arends, Vincent W. V. Jaddoe, Albert Hofman, Frank. C. Verhulst & Catherine. M. Herba. 2011. Recognition of facial expressions of emotions by 3-year-olds. *Emotion* 11(2). 425–435. doi:10.1037/a0022587

Szumilas, Magdalena. 2010. *Explaining Odds Ratios. Journal of the Canadian Academy of Child and Adolescent Psychiatry* 19. 227.

Tracy, Jessica L., Richard. W. Robins & Kristin. H. Lagattuta. 2005. Can children recognize pride? *Emotion* 5(3). 251–257. doi:10.1037/1528-3542.5.3.251

Vaillant-Molina, Mariana, Lorraine. E. Bahrick & Ross Flom. 2013. Young infants match facial and vocal emotional expressions of other infants. *Infancy* 18(SUPPL.1). doi:10.1111/infa.12017

Vaish, Amrisha & Tricia Striano. 2004. Is visual reference necessary? Contributions of facial versus vocal cues in 12-month-olds' social referencing behavior. *Developmental Science* 7(3). 261–269. doi:10.1111/j.1467-7687.2004.00344.x

Wagner, Laura, Shari. R. Speer, Leslie. C. Moore, Elizabeth. A. Mccullough, Kiwako Ito, Cynthia G. Clopper & Kathryn Campbell-Kibler. 2015. Linguistics in a science museum: integrating research, teaching, and outreach at the language sciences research lab. *Linguistics and Language Compass* 9(10). 420–431. doi:10.1111/lnc3.12164

Walker-Andrews, Arlene. S. & Wendy Grolnick. 1983. Discrimination of vocal expressions by young infants. *Infant Behavior and Development* 6(4). 491–498. doi:10.1016/S0163-6383(83)90331-4

Waxer, Matthew & J. Bruce Morton. 2011. Children's judgments of emotion from conflicting cues in speech: Why 6-year-olds are so inflexible. *Child Development* 82(5). 1648–1660. doi:10.1111/j.1467-8624.2011.01624.x

Widen, Sherri C. 2013. Children's interpretation of facial expressions: The long path from valence-based to specific discrete categories. *Emotion Review* 5(1). 72–77. doi:10.1177/1754073912451492

Widen, Sherri C. & James A. Russell. 2003. A closer look at preschoolers' freely produced labels for facial expressions. *Developmental Psychology* 39(1). 114–128. doi:10.1037//0012-1649.39.1.114

Widen, Sherri C. & James A. Russell. 2008. Children acquire emotion categories gradually. *Cognitive Development* 23(2). 291–312. doi:10.1016/j.cogdev.2008.01.002

Wilutzky, Wendy. 2015. Emotions as pragmatic and epistemic actions. *Frontiers in Psychology* 6(OCT). doi:10.3389/fpsyg.2015.01593

Zhang, Dandan, Yunzh Liu, Xinlin Hou, Guoyu Sun, Yawei Cheng & Yuejia Luo. 2014. Discrimination of fearful and angry emotional voices in sleeping human neonates: A study of the mismatch brain responses. *Frontiers in behavioral neuroscience* 8(DEC). 1–10. doi:10.3389/fnbeh.2014.00422

Appendix

DCT Instructions:

Imagine yourself being a pre-school teacher, in charge of a small group of three- to four-year-old children. The task consists of 10 situations. Imagine yourself being in each of these situations and then respond to them as spontaneously as possible.

Non-polite stance condition	Polite stance condition
Scene 1a. The child is playing with your smartphone and dropped it. You have already asked him/her to stop playing with your smartphone before. You're getting impatient. Tell the child to give you the smartphone.	**Scene 1b.** You have both your hands full of plates on your way to the kitchen and you have just dropped a fork. Ask the child nicely to give you the fork.
Scene 2a. The child is scribbling in a book from the class. You have told the child many times before to not scribble in the picture books. You're getting annoyed and tell the child to give you the book.	**Scene 2b.** The child is reading a picture book with a pen close by. You need to write something down. Ask the child nicely to give you the pen.
Scene 4a. The child is very excited and plays continuously with a noisy toy but you want him/her to be quiet. You're quite annoyed. Tell the child to give you the toy.	**Scene 4b.** You bring some toys to the table. A huge frog is still in front of the child. In order to have room for the new toys you ask the child nicely to give you the frog.
Scene 5a. The child is playing with a ball in the kitchen and you have told the child several times before not to play with the ball there. You get annoyed and tell the child to give you the ball.	**Scene 5b.** You're sitting at a table next to the children. You can't reach the bread and you ask the child who sits closest to the bread nicely to give you the bread.

Bionotes

Iris Hübscher

Iris Hübscher has recently received her PhD in Linguistics at the Universitat Pompeu Fabra in Barcelona, Catalunya and since then she has been working as a postdoctoral researcher at the URPP Language and Space at the University of Zurich in Switzerland. Her main research interest lies in multimodal communication, with a main focus on the interplay of prosody, gesture and lexicon in children's pragmatic language development.

Laura Wagner

Laura Wagner has an MA in Psychology and a PhD in Linguistics, both from the University of Pennsylvania. She is currently professor in the Ohio State University Psychology department. Her research specialty is children's language development. She is the director of the Language Sciences Research Lab, a working research lab embedded inside a science museum that is committed to engaging with the public about how language works.

Pilar Prieto

Pilar Prieto is an ICREA Research Professor at the Department of Translation and Language Sciences at UPF (Universitat Pompeu Fabra), Barcelona, Catalunya. Her main research interests focus on the communicative role of prosody and gesture in language, as well as their significance in language development and language training for both impaired and non-impaired populations. Currently she is an Associate Editor of Language and Speech and is coediting a book on Prosodic Development (2018, John Benjamins).