



Young children's sensitivity to polite stance expressed through audiovisual prosody in requests

Iris Hübscher¹, Laura Wagner², Pilar Prieto^{1,3}

¹ Departament de Traducció i Ciències del Llenguatge, Universitat Pompeu Fabra, Barcelona, Spain

² Department of Psychology, Ohio State University, Columbus, Ohio

³ Institució Catalana de Recerca i Estudis Avançats, ICREA, Barcelona, Spain

iris.huebscher@upf.edu, laura.602@osu.edu, pilar.prieto@upf.edu

Abstract

While children's acquisition of lexically encoded politeness formulas has been investigated to a certain extent, little is known about their sensitivity to prosody and facial expressions as cues to politeness. The goal of this paper is to test the ability of 3-year-old American English-speaking children to recognize a speaker's polite stance in their native language based on prosodic and facial gestural politeness cues. A total of 36 3-year-olds were presented with recordings of an adult performing child-directed polite and non-polite requests, where the videos consisted of either audio-only (AO), visual-only (VO) or audio-visual (AV) modalities. Polite and non-polite requests differed only in terms of prosody and facial expressions, though all spoken requests followed the same structure and included the word *please*. Each request asked the child to place a specified object in the bucket in front of the twin that had asked more nicely. Analysis of the results suggests that at age three children are clearly able to access a speaker's polite stance through both prosody and facial gesture cues.

Index Terms: audiovisual prosody development, multimodal comprehension, acquisition of politeness

1. Introduction

In US American parent-child interactions, parents often use explicit linguistic socialisation formulas in the form of routinized prompts such as *What's the magic word?* to elicit politeness from their children. Research in child language has thus far tended to focus, firstly, on how children perceive and then produce such lexically (and syntactically) encoded politeness formulas and the role parents play in this process [1,2], and, secondly, how children develop sensitivity to distance and power by hearing and using polite speech acts [1, 3, 4, 5].

Yet politeness is not only encoded by lexical cues; facial gestures and prosody also play crucial roles in the perception of an utterance as polite or non-polite (see, e.g., [6, 7]). One of the questions is therefore what role audiovisual prosody plays in guiding children to meaning. To our knowledge, only [8] has experimentally tested children's perception of politeness as encoded through prosody. [8] investigated the production and perception of politeness in spontaneous requests in 60 Italian 3- to 6-year-old children. The children were asked to judge which frog puppet made a nicer request while the

experimenter varied the puppet's use of different lexical cues and tone of voice. Tone of voice varied from harsh to gentle and the verbal message *dammi un dolce* 'give me a candy' was accompanied by *grazie* 'please' or not. The results showed that the children had acquired *grazie* '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, [8] argued that the younger children may have judged the harsh tone as 'nicer' because they found it amusing, indicating an earlier sensitivity to intonation after all. [9] and [10] replicated Bates' [8] study with 3-, 5- and 7-year-old US English-speaking children, although neither of them tested for responses to prosodic or gestural cues. Overall, [10] agreed with [8] and [9] that while children seemed to improve their perception of politeness as expressed in lexical cues over the years, certain contrasts proved particularly difficult for 3-year-olds, such as '*Give me the candy*' (imperative) vs. '*Can you give me the candy?*' (interrogative).

More recently, a growing body of research has begun to examine children's ability to use co-speech gestures and prosodic cues in interpreting pragmatic meanings other than politeness. On the one hand, the role of hand gestures in helping children to map form to meaning has been established (see, e.g., [11, 12]) as well as beneficial presence of facial gestures [13, 14]. Focusing on the marking of speakers' disbelief and uncertainty, [13, 14] have shown that facial gestures seem to also scaffold children's acquisition of epistemic belief states. On the other hand, explorations of children's ability to access pragmatic meanings through prosody have yielded inconsistent results. While pre-school children (≤ 4 years) only seem to be able to accurately judge a speaker's emotional state based on prosody when the lexical content is either neutral or filtered out [15], this ability seems to be compromised when there are competing cues regarding the relevant emotion [16]. Contrary to this, [17] showed that when using a more implicit methodology, such as eye-tracking, young children already at age 3 were able to link speech bearing different acoustic cues to emotion. Furthermore, [14] showed that children had a particular sensitivity to intonational cues such as a rising intonation (L* H%) encoding a speaker's uncertainty (a belief-based emotion) at an age when they have not yet acquired the meaning of lexical cues to uncertainty like *perhaps*. These results led the authors to formulate the prosodic bootstrapping hypothesis whereby intonation scaffolds the acquisition of pragmatic meaning.

Following up on the results showing children being sensitive to audio-visual cues to a speakers' epistemic stance [13, 14], it would seem to be of interest to further test children's acquisition of pragmatic meaning encoded by

prosody and gestures in the presence of a lexical cue. Thus, the present study tested 3-year-old children's sensitivity to a speaker's polite stance as it is encoded through audiovisual prosody in requests such as '*Can you give me the ball, please*'. While the lexicon was kept constant in both polite and non-polite conditions, pitch contours and facial expressions were varied, ranging from falling L* L% (non-polite) to rising L+H* H% (polite) and frown (non-polite) to smile (polite) respectively. In order to investigate the children's sensitivity to pitch/prosodic patterns in comparison with gestures in the perception of politeness, the materials were presented in three different modalities: (a) audio-visual, with both prosody and non-verbal cues available; (b) visual-only, just non-verbal cues available; and (c) audio-only, with just prosody available. We hypothesized that if facial gestures provide clear scaffolding to the pragmatic meaning of politeness, then 3-year-old children should be able to distinguish a polite from a non-polite stance when they have both modalities at hand in the audio-visual condition. We furthermore hypothesized that the lack of one modality would weaken children's performance, but their performance would be better in the visual-only condition as compared to the audio-only condition. Regarding the audio-only condition, if children solely rely on the lexical information rather than intonational cues, the presence of *please* in the request should lead to a 'polite' interpretation despite the presence of gestural cues to non-politeness. However, if the prosodic bootstrapping hypothesis also holds for a social pragmatic meaning such as politeness, then children should be able to perform above the mean also in this condition, that is, purely on the basis of prosody.

2. Experimental setup

2.1. Participants

Thirty-six American English speaking children (18 female and 18 male) participated in a politeness comprehension task. Their ages ranged from 2;10,22 to 4;0,16 (mean age: 3.51 years). The number of participants was evenly distributed within each of the three conditions (see below). The participants were all children visiting the Center of Science and Industry, a science education center in Columbus, Ohio. Parental consent was obtained before the experiment. The children were given a hand stamp as a reward for participation.

2.2. Materials

For all three conditions in each experiment, audio-only (AO), video-only (VO) and audio-visual (AV), four different PowerPoint presentations were prepared in order to be able to counterbalance the presentation order of the stimuli within each trial (left/right) and the order of 6 toy items. Each PowerPoint presentation contained eight slides (see Figure 1 for an example) one for familiarisation, six for the test trials and one for a re-familiarisation trial.



Figure 1. Example of a slide used for a test trial in still position.

2.2.1. Stimuli

2.2.1.1. Elicitation study with adults

In order to investigate the prosodic and facial gestural characteristics of politeness in American English, a free discourse elicitation task was carried out with 10 US English native speakers. The participants were asked to imagine themselves in a situation like those described in (1) and (2), where the context prompts the speaker to make a request to a child. In (1) the context prompts the speaker to ask for help in a friendly way (i.e., a polite request) while context (2) prompts the speaker to produce a command rather than a question (i.e., a non-polite request). Next to the description of the context, the participants were provided with a picture of a 3-year-old child.

(1) POLITE 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 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.

A total of 100 requests (10 requests x 10 speakers) were analysed prosodically and gesturally. The intonation contours displayed in the requests used by adults can be seen in Table 1.¹

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

Table 1. Prosodic cues displayed in the requests

The main results showed that there was a clear preference for falling (L* L%) nuclear configuration with a lower pitch range in the non-polite condition, and a rising (L+H* H%) pitch accent and a higher pitch range in the polite condition. Furthermore, while in the non-polite condition participants displayed a stern facial expression, in the polite condition they displayed a friendly smile.

¹ The most commonly used lexical structures were the conventional indirect requests *Can you give me xy?* (with and without *please*) or *Could you give me xy?* (with and without *please*) in the polite condition, and the imperative *Give me xy* (with and without *please*) in the non-polite condition.

2.2.1.2. Stimuli creation

Six female psychology or linguistics students and researchers from Ohio State University (not those who recorded the discourse elicitation task) volunteered to take part. They were individually videotaped while producing 12 tokens (two for each toy) of the target structure ‘*Can you give me the + [toy] + please*’ in which they replicated the non-polite and polite conditions described above. Two stimuli (one for each condition and experiment) for each participant were selected and acoustically analysed in PRAAT in order to check that they matched the prosodic characteristics corresponding to the two conditions. The typical intonation contours used for the command vs. question request intonation can be seen in Table 2.

Condition	Pitch register	Nuclear configuration	Facial gestures
Polite	High	L+H* H%	Smile
Non-polite	Low	L* L%	Stern expression with furrowed eye-brows

Table 2. Facial gestural and prosodic cues displayed in the polite vs. non-polite conditions.

The two panels in Figure 2 illustrate the typical pitch nuclear configurations found in the two conditions.

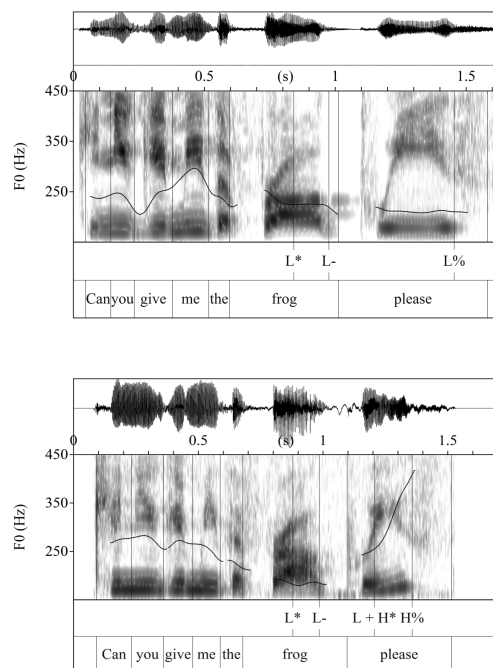


Figure 2. Pitch track, spectrogram and waveform for the polite/interrogative request (top panel) and non-polite/imperative request (bottom panel) stimuli used in Experiment 1.

In order to create the VO condition, the audio track was eliminated from the video. Thus, the children would be

exposed only to facial gestural cues without any audio cues. The facial gestures that the children typically saw were a smile in the left-hand panel and a stern expression in the right-hand panel (see Figure 3).



Figure 3. Facial expressions displayed in the polite (left panel) and non-polite (right panel) conditions.

In order to make sure that the stimuli would be interpreted as polite vs. non-polite by adult American English speakers, we performed a preliminary control survey with the online survey platform SurveyGismo. The survey results confirmed these adult interpretations.

2.3. Procedure

Each child was seated next to the experimenter, a native American English speaker, on the floor in a quiet room. As a warm up, the child was asked to name/identify each of the six toy items that would occur in the experimental trials: a *ball*, *shark*, *frog*, *bear*, *horse* and *duck*. For the experiment proper, the child faced a laptop on which the politeness comprehension task was presented with two empty buckets placed between the child and the computer. To familiarise 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 child listened to the left-hand star, which 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 inset video segments of a speaker and one of the objects named above (e.g., the ball in Figure 1). Because it was the same speaker in both videos, the experimenter referred to them as ‘sisters that look very much alike’. The experimenter then directed the child’s attention to the object at the top right corner of the screen (e.g., the ball) and asked ‘*What do you think they want? Both of them are going to ask you to give you the ball. You have to listen/watch very carefully. (after playing the trials) Can you give the ball to the person who asks more nicely?*’ The two inset videos were then played consecutively, one in which the speaker made a request replicating the polite cues and the other in which the speaker produced an imperative while displaying the non-polite cues. The child then showed which of the two twins had asked more nicely by placing the named object in the buckets that was in front of that person. In each trial a new set of ‘twins’ and object was displayed.

After the first three test trials, the children were again shown the familiarization activity with the stars. The child then performed the remaining three test trials. In half of the

trials, the polite stimulus recording was presented first and was located on the left side of the screen; in the other half the polite choice was presented second and located on the right side of the screen. In total each child saw 6 test trials and 2 familiarization trials. The whole experiment was video-taped.

3. Results

First, the results of the test trials were coded as ‘correct’ or ‘incorrect’. The child’s response was ‘correct’ if the child placed the object in the bucket of the ‘twin’ that had replicated the ‘polite’ cues appropriate for each modality, that is, in the AV condition a verbal request (e.g., *Can you give me the ball please?*) with a L+H* H% intonation contour and a polite smile; in the AO condition a L+H* H% intonation contour; and in the VO condition a polite smile. Failure to do this was counted as an ‘incorrect’ response. A total of 216 trials were analysed (6 test trials x 36 participants). A bar graph illustrating the mean ratios of correct to incorrect responses for each modality is shown in Figure 5.

First, in order to test the actual performance in each condition, a binominal test was applied to the data. The results indicated that the ratio of correct interpretations of politeness was .78 in conditions AV and VO, greater than the expected chance ratio of .50 ($p < .001$, 1-sided). Also the ratio of correct interpretations of politeness in the AO condition was .71, also greater than chance ($p = .001$, 1-sided). Thus, the performed very significantly better than at chance in all three conditions, indicating that children are sensitive to audiovisual prosody in classifying politeness.

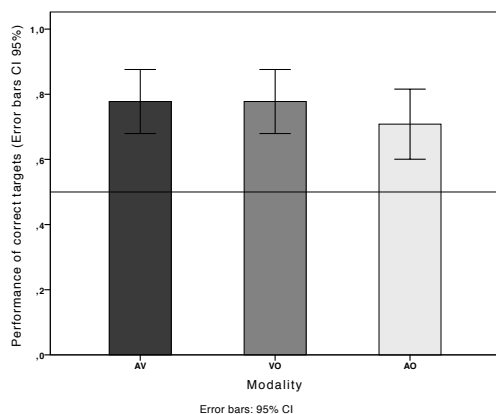


Figure 5. Mean ratio of correct responses by modality (audio-visual, visual-only and audio-only)

Second, a Generalized Linear Mixed Model 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 was not found to be significant, with $F(2, 213) = 0.315$, $p = .730$. Independent of the condition, the children performed about equally well.

4. Discussion and Conclusions

The results of the current study showed that 3-year-old children overall performed significantly above chance and thus were clearly able to access a polite stance encoded through audiovisual prosody in sentences which had a lexical marker of politeness (in this case *please*).

Interestingly, not only did the children perform significantly above chance in the audio-visual and visual-only condition, but they also showed themselves able to access prosodic cues marking a speaker’s polite stance very well in the audio-only condition. Therefore, it seems that neither the combination of cues nor the lack of any one cue hindered the children from performing well. While these results are consistent with other studies suggesting that (facial) gestures play a scaffolding role when children are accessing more complex pragmatic meanings [11, 12, 13, 14], they stand out compared to [15, 16], since at 3 years of age the children in the present experiment were already perfectly capable of accessing a speaker’s polite stance through prosody. Furthermore, compared to [14], the current results suggest that children may be able to access a polite stance through prosody-only cues better and probably earlier than they can an epistemic stance.

Comparing these results with previous reports on politeness acquisition such as [8, 10], it can be seen that children seem to be able to distinguish a polite from a non-polite request based on prosody around the same age as they are able to make the same distinction on the basis of different lexical cues. Yet, while [10] showed that American English-speaking children at 3 years of age could not yet successfully distinguish an interrogative from an imperative as being more polite, this experiment shows that they are able to make such a distinction based on audiovisual prosody. Furthermore, very importantly in contrast to studies reporting prosody to be a weak cue for reading emotional state [16], children actively use prosodic as well as facial gestural features as powerful indicators of politeness. Specifically, children use prosodic and gestural cues to override the lexical meaning of the word *please*, which was present and thus controlled for in both polite and non-polite conditions, and ultimately gets its final interpretation depending on the way it is uttered.

What is still unanswered and should be addressed in future research is whether also 2-year-old children or even younger are successfully able to process audio-visual prosodic cues to children’s acquisition of politeness. In addition, it would be of interest to further investigate the relative weight of each individual cue, be it lexicon, prosody or facial gesture, in order to test the possible bootstrapping role of prosody and facial gesture in children’s early acquisition of politeness.

5. Acknowledgements

We would like to express our gratitude to Faith Stagge, who helped us collect the data at the Center of Science and Industry in Columbus. Many thanks also go to students and researchers at Ohio State University who performed in our stimulus recordings. We are grateful to the linguistics team at Ohio State University who helped along the way with some native speaker inputs and to Dr. Shari Speer and her team at the Psycholinguistics Lab (Kiwako Ito and Kathryn Kimberley in

particular), where some preliminary results were presented. We are also grateful to the Prosodic Studies Group in Barcelona, where we presented the study, and to Joan Borràs for helping with the statistics. This research has been funded by a research grant awarded by the Spanish Ministry of Science and Innovation (FFI2012-31995 “Gestures, prosody and linguistic structure”), and by a grant awarded by the Generalitat de Catalunya (2014 SGR-925) to the Prosodic Studies Group.

6. References

- [1] J. Becker, “Processes in the acquisition of pragmatic competence”, in G. Conti-Ramsden and C. Snow (eds.), *Children’s Language*, vol. 7. Hillsdale, NJ. Lawrence Erlbaum Associates, 1990, pp. 7–24.
- [2] C. E. Snow, R. Perlmann, J. B. Gleason and N. Hooshya, “Developmental perspectives on politeness: Sources of children’s knowledge”, *Journal of Pragmatics*, vol. 14, pp. 289–305, 1990.
- [3] W. Corsaro, “‘We’re friends, right?’: Children’s use of access rituals in a nursery school”, *Language in Society*, vol. 8, pp. 315–336, 1979.
- [4] S. Ervin-Tripp and D. Gordon, “The development of requests”, in R. L. Schiefelbusch (ed.), *Language competence: Assessment and intervention*. San Diego: College Hill, 1986, pp. 61–95.
- [5] T. Schöll, *The Acquisition of Politeness in the Language of Children*, Seminar paper. University of Hannover, 1997.
- [6] J. Culpeper, “It’s not what you said, it’s how you said it!’: Prosody and impoliteness”, *Discursive Approaches to Politeness*, pp. 1–18, 2011.
- [7] S. Mckinnon and P. Prieto, “The role of prosody and gesture in the perception of mock impoliteness”, *Journal of Politeness Research*, vol. 10, no. 2, pp. 185–219, 2014.
- [8] E. Bates, “Acquisition of Polite Forms: Experimental Evidence”. In Bates, E (ed.). *Language and Context: The Acquisition of Pragmatics* (pp. 255–326). New York, NY: Academic Press, 1976.
- [9] E. Bates and L. Silvern, “Social adjustment and politeness in pre-schoolers”, *Journal of Communication*, vol. 27, pp. 104–111, 1977.
- [10] B. M. A. Nippold, L. B. Leonard and A. Anastopoulos, “Development in the use and understanding of polite forms”, *Journal of Speech, Language and Hearing Research*, vol. 25, pp. 193–202, 1982.
- [11] S. D. Kelly, “Broadening the units of analysis in communication: speech and nonverbal behaviours in pragmatic comprehension”, *Journal of Child Language*, 28(2), pp. 325–349, 2001.
- [12] C. Butcher and S. Goldin-Meadow, “Gesture and the transition from one- to two-word speech: When hand and mouth come together”, in D. McNeill (ed.), *Language and gesture*. Cambridge: Cambridge University Press, pp. 235–258, 2000.
- [13] M. Armstrong, N. Esteve-Gibert and P. Prieto, “The acquisition of multimodal cues to disbelief”, *Proceedings of the Speech Prosody* 2014. ISSN: 2333-2042. Dublin, Ireland, May 20-23, 2014.
- [14] I. Hübscher, N. Esteve-Gibert, A. Igualada and P. Prieto, “Intonation and gesture as bootstrapping devices to speaker uncertainty”, *First Language*, (accepted).
- [15] C. Quam and D. Swingley. “Development in children’s interpretation of pitch cues to emotions.” *Child Development*, 83(1), 236–250, 2012.
- [16] M. Aguert, L. Le Bigot, A. Lacroix and V. Laval, “Understanding expressive speech acts: The role of prosody and situational context in French-speaking 5- to 9-year-olds.” *Journal of Speech, Language, and Hearing Research*, 53, 1629–1641, 2014.
- [17] J. M. J. Berman, C. G. Chambers and S. A. Graham, “Preschoolers’ real-time coordination of vocal and facial emotional information”, *Journal of Experimental Child Psychology*, 142, 391–399, 2016.