



Published in final edited form as:

Child Dev. 2012 ; 83(1): 236–250. doi:10.1111/j.1467-8624.2011.01700.x.

Development in Children's Interpretation of Pitch Cues to Emotions

Carolyn Quam* and
University of California, San Diego

Daniel Swingley
University of Pennsylvania

Abstract

Young infants respond to positive and negative speech prosody (Fernald, 1993), yet 4-year-olds rely on lexical information when it conflicts with paralinguistic cues to approval or disapproval (Friend, 2003). This article explores this surprising phenomenon, testing 118 2- to 5-year-olds' use of isolated pitch cues to emotions in interactive tasks. Only 4- to 5-year-olds consistently interpreted exaggerated, stereotypically happy or sad pitch contours as evidence that a puppet had succeeded or failed to find his toy (Experiment 1) or was happy or sad (Experiments 2, 3). Two- and three-year-olds exploited facial and body-language cues in the same task. The authors discuss the implications of this late-developing use of pitch cues to emotions, relating them to other functions of pitch.

Intonation plays a special role in young infants' linguistic interactions with their parents, well before infants know any words. Parents speak to their infants in a distinctive way, using higher mean fundamental frequency (f_0) and a wider f_0 range than when they speak to adults. These pitch characteristics make infant-directed speech more interesting to infants, and easier for them to perceive, than adult-directed speech (Fernald, 1992). Infants prefer listening to infant-directed speech over adult-directed speech (Fernald, 1985), primarily because of its distinctive pitch features (Fernald & Kuhl, 1987; Katz, Cohn, & Moore, 1996). Adults express pragmatic functions (like comfort or prohibition) more clearly in the infant-directed register than in adult-adult speech (Fernald, 1989), and infants can group together infant-directed utterances by their pragmatic functions (Moore, Spence, & Katz, 1997). The intonation contours of infant-directed speech even help infants to segment words from the speech stream (Thiessen, Hill, & Saffran, 2005). Parents thus appear to shape the pitch patterns of their speech to attract the infant's attention and communicate emotional and pragmatic information.

In addition to attracting infants' attention to speech and conveying emotional and pragmatic information, pitch may drive infants' sensitivities to several linguistic features. Infants are sensitive to prosodic cues to phrase boundaries by 2 months (e.g., Mandel, Jusczyk, & Kemler Nelson, 1994). By 6 months, infants use pitch contours to parse utterances into clauses (e.g., Seidl, 2007), and by 10 months infants' recognition of a bisyllabic word is impaired when a prosodic break is inserted between its syllables (Gout, Christophe, & Morgan, 2004).

*Correspondence concerning this article should be addressed to Dr. Carolyn Quam, Center for Research in Language, UCSD, 9500 Gilman Drive, Dept. 0526, La Jolla, CA 92093-0526. cmquam@gmail.com.

Infants' early sensitivity to intonation in parents' speech does not automatically provide them with an adult-like *understanding* of the many functions of pitch contours, however. Pitch is exploited in language at several levels of structure, and different languages use pitch differently. For example, lexical-tone languages use pitch to contrast words, whereas English uses higher pitch as one of several correlated cues to word stress (which sometimes differentiates words, as in noun-verb pairs like CONduct–conDUCT). Other languages mark stress with pitch lowering rather than raising, or do not use pitch to mark stress at all (Pierrehumbert, 2003). A child learning English must identify the particular role of pitch in marking lexical stress, while a child learning Mandarin must identify the role of pitch in the tonal system; and both children must also attend to pitch for demarcation of phrase boundaries, marking of yes/no questions, and emotional and pragmatic information. Infants' early sensitivity to acoustic features of infant-directed speech—and even their apparent ability to respond appropriately to positive and negative prosody—do not necessarily entail the ability to *interpret* another person's vocal expressions of emotion.

Sensitivity to Vocal Cues to Emotions in Infancy

A distinction must be made between language-universal, direct effects of prosodic variations on infants' emotions, and a more reflective, interpretive capacity to integrate emotional prosody with the rest of the talker's linguistic message. It is in the former sense that young infants "understand" language before they know any words. Emotional prosody in parental speech induces infant emotion in appropriate and predictable ways, even if that speech was recorded in an unfamiliar language (Fernald, 1993). In one study, 12-month-olds showed reduced exploration of a toy and more negative affect upon hearing fearful-sounding speech, as opposed to more neutral speech, from a recorded actress (Mumme & Fernald, 2003; see also Friend, 2001). Though infants display some sensitivity to vocal expressions of emotions, they appear even more sensitive to facial expressions. D'Entremont and Muir (1999) found that even 5-month-olds smiled more in response to happy than to sad facial expressions, but the addition of vocal paralinguistic cues did not affect their responses, and they showed no differential responding to vocal paralinguistic cues alone.

Infants have some knowledge that certain facial expressions go with certain vocal expressions. In intermodal-preference tasks, in which infants see two faces conveying different emotions (such as *happy* and *sad*) and hear a voice that is more consistent with one face than the other, infants often gaze more at the matching face (Kahana-Kalman & Walker-Andrews, 2001; Walker, 1982; Soken & Pick, 1999).

Surprising Difficulty Interpreting Paralinguistic Cues to Emotions in Early Childhood

Early-developing reactions to emotional prosody, and the capacity to link facial and vocal affective signals appropriately, do not appear to provide young children with a ready appreciation of how emotional prosody affects talkers' linguistic messages. This has been shown in a number of studies, most of which have presented children with stimuli with discrepant linguistic versus paralinguistic content. When prosodic or facial cues conflict with lexical information, young children usually rely on the meaning of the words rather than on facial expressions or prosodic contours. Friend and Bryant (2000) asked children to place the emotion expressed by a disembodied voice on a 5-point scale ranging from "very happy" to "very mad." Four- and seven-year-olds relied more heavily on lexical information when it was in conflict with prosody, while ten-year-olds relied more on prosody (though in a similar experiment, even ten-year-olds relied more on lexical information than on paralinguistic cues; Friend, 2000). Friend (2003) examined a more naturalistic behavioral response—interaction with a novel toy—to consistent versus discrepant lexical and

paralinguistic (facial plus vocal) affective information from an adult face on a video-screen. Four-year-olds approached the toy faster and played with it longer when the adult's affect was consistently approving than when it was consistently disapproving. When the cues were discrepant, words trumped facial-and-vocal paralinguistic. Finally, Morton and Trehub (2001) examined 4- to 10-year-olds' and adults' ability to judge the speaker's happiness or sadness from vocal paralinguistic versus lexical cues. Four-year-olds relied on lexical information when the cues conflicted, while adults relied exclusively on paralinguistic. In between, there was a gradual increase in reliance on paralinguistic; only half of 10-year-olds relied primarily on paralinguistic. When 6-year-olds were primed to attend to paralinguistic, however, they successfully relied more on paralinguistic than on lexical information (Morton, Trehub, & Zelazo, 2003).

Like infants, preschoolers show some use of paralinguistic cues to emotion—when these cues are not pitted against lexical information. Friend (2000) found that 4-year-olds can identify the affect of happy versus angry reiterant speech, in which lexical content is replaced with repetitive syllables (e.g., “mama ma”; Friend and Bryant, 2000, found a similar result with 7- to 11-year-olds). They fail with low-pass-filtered speech, however, which preserves primarily fundamental frequency (f_0), suggesting that f_0 alone is not a sufficient cue. Still, this failure could be due to the unnaturalness of either low-pass-filtered speech or of the task, in which children listened to a sentence out of context. Morton and Trehub (2001) did find some success at age 4 with low-pass filtered speech, as well as with paralinguistic in Italian, though the Italian stimuli differed on many acoustic dimensions: happy sentences had higher pitch, a faster speaking rate, and greater pitch and loudness variability.

The present research reexamined children's sensitivity to vocal paralinguistic in the absence of conflicting lexical information, using interactive and age-appropriate tasks. In prior studies infants demonstrated the clearest sensitivity to vocal expressions of fear, which is likely evidence of a low-level, evolved behavioral response rather than interpretation of another person's emotions. Evidence from preschoolers is mixed, and we know very little about children's sensitivity to *pitch* cues in particular. Some previous studies have combined facial and vocal paralinguistic (e.g., Friend, 2003; Mumme, Fernald, & Herrera, 1996). Those studies that considered only vocal paralinguistic rarely disentangled pitch, speaking rate, loudness, and breathiness (e.g., Friend, 2000; Friend & Bryant, 2000; Morton & Trehub, 2001; Mumme & Fernald, 2003), except when using low-pass filtered speech, which introduces additional naturalness issues. Considering the arguments and evidence that pitch plays a crucial role in children's early language processing, it would be useful to isolate pitch cues to the speaker's emotions—by which we mean both pitch height and pitch contour—and identify when children can exploit them.

An important question concerns whether pitch cues to all levels of structure are equally accessible to the child, or whether cues to different levels of linguistic structure are acquired at different points in development, despite being carried by the same acoustic dimension. In line with Fernald's (1992) qualitative model of development, we argue that different levels of pitch structure are available to the child at different points, depending on the child's ability to access the cue in the signal, the reliability of the cue in the signal, and the developmental relevance of the cue (see also Werker & Curtin, 2005).

Access to a particular cue in the signal may require certain linguistic preconditions. For example, stressed syllables in English tend to be word-initial (Cutler & Norris, 1988), making stress a cue to word onsets. But learning this cue requires prior knowledge of at least some words (Swingley, 2005). The *reliability* of pitch cues may be reduced because pitch is used for multiple functions, so that realizations of categories like tones vs. intonational

meanings trade off (Papousek, Papousek, & Symmes, 1991). Finally, the *developmental relevance* of the cue, in addition to modulating children's attention to the cue, may impact the reliability of its realization in the signal. Mandarin-speaking mothers appear to reduce or neglect tone information in favor of producing simple intonation contours to 2-month-olds (Papousek & Hwang, 1991), but *exaggerate* tone categories in speech to 10- to 12-month-olds (Liu, Tsao, & Kuhl, 2007), much as parents exaggerate vowel categories in infant-directed speech (Burnham, Kitamura, & Vollmer-Conna, 2002). This difference could arise because intonational meaning is more relevant to younger infants, and tone and segmental information is more relevant to older infants (Kitamura & Burnham, 2003; Stern, Spieker, Barnett, & MacKain, 1983).

Infants acquiring tone languages appear to learn tones at about the same time that they learn consonant and vowel categories (e.g., Mattock & Burnham, 2006), though tones that are less consistently realized appear to be learned more slowly (Demuth, 1995; see also Ota, 2003). *Interpretation* of highly discriminable pitch variation may follow an even slower time-course; English-learning children learn to disregard potentially lexical pitch by 30 months (Quam & Swingley, 2010), possibly by detecting the variability of pitch contours of words across tokens (Quam, Yuan, & Swingley, 2008). If pitch cues to different levels of structure are indeed acquired at different time points, we might find relatively late development of successful interpretation of pitch cues to emotions, despite the early importance of intonation in infancy.

The present work addresses children's understanding of intonational cues to the emotions *happy* and *sad* compared with facial and body-language cues to the same emotions. We chose *happy* and *sad* to maximize the contrast between the emotions children had to distinguish, as they differ in both valence and activation/arousal (Russell, Bachorowski, & Fernández-Dolz, 2003).

Three experiments used interactive tasks to test preschoolers' interpretations of pitch cues to emotions in the absence of conflicting lexical information. Experiment 1 used a task inspired by Tomasello and Barton (1994, Experiment 4). Children had to interpret the emotions of a puppet, "Puppy," in order to infer which toy was the object of his search; Puppy was happy if he found his toy, and sad if he found a different toy. Children responded by giving the toy to Puppy if he was happy, and throwing it in a trashcan if he was sad. Experiments 2 and 3 used a simpler and more direct test of sensitivity to emotions. Puppy was again searching for toys, but this time children simply responded by pointing to a happy face (or saying "happy") if Puppy was happy, or pointing to a sad face (or saying "sad") if Puppy was sad.

Experiment 1

Method

Participants—Thirty-six children participated in Experiment 1 (twenty female, sixteen male): thirteen 3-year-olds, fifteen 4-year-olds, and eight 5-year-olds. Children were recruited by staff in preschools, via letters sent to parent addresses from a commercial database, and by word of mouth. Two children were Asian, two were African-American, five were of mixed race or reported to be "Other," and twenty-seven were Caucasian; three of the thirty-six children were Hispanic/Latino. These racial counts are estimates based on voluntary parental report for some children and observation for others. Parental SES was not evaluated. Two more 3-year-olds were excluded for failure to participate in enough trials, and one for experimenter error. Since many 3-year-olds needed some help with both pretrials (and many of these children still succeeded in the body-language trials), failure in the pretrials was not used as grounds for exclusion.

Apparatus and Procedure—Participants sat at a table across from the experimenter (the first author), either at the child’s preschool or in a university developmental-laboratory suite. A red cylindrical container (the “trashcan”) was placed to the child’s right, with a cardboard box behind it, closer to the experimenter. Cameras (with an external microphone) recorded the experimenter’s face, the child, and the table. See Figure 1 and Figure 2 for a photograph and diagram of the testing setup, respectively.

Children were told they would be playing a game in which they would see several toys and meet the experimenter’s friend Puppy, a puppet. In each trial, the experimenter put three toys in the box, using different toys for each trial. Children were first permitted to examine each of the three toys, and were then told that Puppy was looking for a particular toy (e.g., the *toma*). Puppy would be happy when he found the *toma*, and sad if he found a different toy. Children were instructed to give the *toma* to Puppy and throw the other toys in the trash. The experiment began with one or two *pretest trials*, intended to teach children the task and let them practice the giving/throwing-away response (if a child failed the first pretest trial, a second one was included). In pretest trials, the experimenter pulled each toy out of the box. Upon viewing the toy, Puppy was “feeling shy,” so he whispered in the experimenter’s ear whether each toy was the target, and the experimenter told the child explicitly. In experimental trials, after viewing each toy, Puppy turned to the child and responded with either body-language or pitch cues. The order of responses to the three toys was counterbalanced, so that the *happy* response occurred first, second, and third on different trials. After each response, the experimenter asked, “Is this the *toma*? Where should we put it?” If the child gave more than one toy to Puppy, the experimenter then said, “You gave two toys to Puppy, but only one toy is the *toma*. Can you tell me which one?” The child’s response was coded as either correct (the child either handed only the target toy to Puppy, or chose the target toy from among the toys she had handed to Puppy) or incorrect (the child chose a different toy).

In three *body-language trials*, Puppy expressed excitement by nodding and dancing side-to-side, and disappointment by shaking his head and slumping down. In the *pitch trials*, which followed the body-language trials, the experimenter, speaking for Puppy, produced excited pitch (high pitch with wide excursions) or disappointed pitch (low pitch with narrow excursions). The first seven children participated in three pitch trials, but for the remainder we added an extra pitch trial to permit a slightly finer-grained assessment of performance. Each child was considered to have succeeded in a given condition if he or she answered correctly in at least two of the first three trials. Otherwise, statistical tests used all completed trials.

Visual Stimuli—The toys used in the experiment were all intended to be novel. Appendix 1 displays four of the roughly two-dozen toys. Most were handmade from parts of kitchen appliances, dog toys, and electronics, though some toys were unmodified from their original form (e.g., an unusual-looking potato masher). Children occasionally recognized parts of toys, saying, e.g., “That’s a rolling pin!” The experimenter responded, “It looks kind of like a rolling pin, but it’s just a silly toy.” If the child asked, “What is that?” the experimenter responded, “I don’t know—it’s just a silly toy.”

The puppet was a plush, black-and-white spotted dog twelve inches high and six inches across (arm-span eleven inches). When the puppet was “talking,” the experimenter moved Puppy’s left hand once for each syllable so it was clear that Puppy was the one talking to the child. In all experimental trials, the puppet was placed between the experimenter’s face and the child’s face to prevent the experimenter from conveying any facial cues.

Auditory Stimuli—Auditory stimuli were produced live by the experimenter. The experimenter mostly talked directly to the child, but during the crucial part of the test trials, she said, “Look what I found! Puppy, is this the [toma]?” and then, keeping the puppet between her face and the child’s face, said “mmm, mm mm mmm” with stereotypical excited/happy pitch or disappointed/sad pitch. The happy pitch was high, with rise-fall contours and wide excursions, while the sad pitch was low, with falling contours and narrow excursions; see Figure 3 for waveforms and pitch tracks of two example contours. The experimenter reminded the child to listen before producing each pitch contour.

In addition to the differences in pitch height and contour that we intentionally produced, happy/excited speech is stereotypically higher in amplitude and faster than sad/disappointed speech; spectral (or timbre) differences can also result from these differences in amplitude and pitch as well as from differences in mouth shape (e.g., smiling versus frowning). The experimenter attempted to equate duration, amplitude, and mouth shape when producing the stimuli. To numerically compare the acoustics of the experimenter’s productions of happy versus sad pitches trial-by-trial, acoustic measurements from the two sad productions were averaged to produce a single value, which was then compared to the *happy* value from that trial. This analysis was conducted on the pitch trials from only those participants who had succeeded with the pitch cues. In this and all following acoustic analyses, duration and intensity values were log-normalized, and pitch measurements were converted from Hz to ERB ($ERB = 11.17 * \ln((Hz + 312)/(Hz + 14675)) + 43$; Moore & Glasberg, 1983). Results were comparable without these conversions, however, and means are given here, and in Appendix 2, in Hz, seconds, and dB for ease of interpretation.

Happy and sad productions differed significantly on all acoustic dimensions measured. Happy productions had higher pitch means (*happy*, 416.74 Hz; *sad*, 255.72 Hz; paired $t(23) = 57.74$), larger standard deviations of pitch samples (*happy*, 129.60 Hz; *sad*, 51.78 Hz; paired $t(23) = 80.59$), higher pitch maxima (*happy*, 745.61 Hz; *sad*, 386.86 Hz; paired $t(23) = 57.22$), higher pitch minima (*happy*, 210.80 Hz; *sad*, 163.33 Hz; paired $t(23) = 7.06$), greater intensities (*happy*, 72.72 dB; *sad*, 71.26 dB; paired $t(23) = 5.34$), and greater durations (*happy*, 3.30 seconds; *sad*, 3.24 seconds; paired $t(23) = 3.08$, all $p < .01$, all tests 2-tailed). Though all of these comparisons indicated statistically significant differences, it is unlikely that participants detected the between-condition differences in intensity and duration, because while these differences were consistent enough to be statistically significant, they were very small (e.g., in terms of JNDs, the mean pitch differences were probably *at least* 5 to 10 times greater than the amplitude or duration differences; Harris & Umeda, 1987; Miller, 1947; Abel, 1972). For example, the ratio of (log) durations for happy vs. sad productions was 1.00 (see Appendix 2).

Results and Discussion

Each participant gave responses in three body-language trials and three or four pitch trials. Table 1 reports the number of children from each age group who succeeded with each cue. A child was determined to have succeeded if he or she chose the correct toy (to which Puppy responded with happy body-language or pitch) in at least two of the first three trials. Across the full sample of children (and counting all available trials), performance was significantly better with the body-language cues (89.8%) than with the pitch cues (61.1%, paired $t(35) = 4.55$; $p < .001$, 2-tailed). On body-language trials, there was little developmental change in children’s rate of success (on the 2/3 criterion), though accuracy and age were correlated (using all available trials; $r = 0.36$, $p < .05$). At age 3, 11/13 children succeeded, as did all of the 4- and 5-year-olds (15/15 4-year-olds and 8/8 5-year-olds). By contrast, in the pitch condition, children showed marked improvement with age. At age 3, only 7/13 children succeeded with the pitch cue; at age 4, 10/15 succeeded; and by age 5, 7/8 children

succeeded. Age and accuracy were significantly correlated (using all available trials; $r = 0.46, p < .005$). Figure 4 plots accuracy against age. In sum, interpretation of pitch-contour cues to the expression of joy and sadness was poor at 3 years, middling at 4, and consistently successful only at 5, whereas interpretation of body-language cues was already successful at 3 years, the earliest age tested.

Experiment Two

Children's success with the body-language cue in Experiment 1 suggested that the task itself was not responsible for children's difficulty with the pitch cues. Still, several concerns motivated us to replicate the experiment. First, the body-language cues in Experiment 1 involved the puppet nodding his head and then dancing, to express excitement/happiness, or shaking his head and slumping down, to express disappointment/sadness. These physical cues map more closely onto the meanings *yes* and *no* than onto *excited* and *disappointed*. This alternative interpretation might lead children to succeed with the body-language cues without connecting the body-language cues to the emotions *happy* and *sad*; in addition, this interpretation might interfere with performance on pitch trials (which always followed body-language trials), where the pitch contours did not match *yes* and *no*.

The inherent difficulty of the task might also have blocked children's access to the pitch-contour cues. Though children did succeed with the body-language cues, they still might have performed better with the pitch cues had the task been simpler. The 3-year-olds sometimes struggled to remember their instructions. Finally, we were concerned that the task was not a direct test of interpretation of emotions. Children were not simply asked to tell the experimenter whether Puppy was happy or sad; they had to make the further inference that Puppy's emotional state indicated which toy he had lost. Simplifying the task might therefore reveal children's understanding of the emotions themselves (Baldwin & Moses, 1996).

Thus, Experiment 2 implemented a simpler and more direct test of interpretation of emotions. Again, Puppy was presented with toys—this time only two per trial—and responded to each toy with excitement or disappointment. Children were simply asked to tell the experimenter whether Puppy was happy or sad. The simplicity of this task relative to the task in Experiment 1 led us to include 2-year-olds. The body-language cue was also better matched to the pitch cue; both of them mapped onto the meanings *excited/happy* versus *disappointed/sad*. The pitch contrast was identical to that tested in Experiment 1, but was produced on the words “Oh, look at that,” which should be more familiar and natural than hummed speech. Here, the body-language and pitch cues were tested between subjects, to eliminate the possibility of transfer from one condition to the other.

Method

Participants—Sixty-two children participated (thirty-one girls): twelve 2-year-olds (six in each condition), twenty-six 3-year-olds (ten in the body-language condition, and sixteen in the pitch condition), twelve 4-year-olds (all in the pitch condition), and twelve 5-year-olds (all in the pitch condition). Participants were recruited as in Experiment 1. Of the children, one was Asian, ten were African-American, eleven were of mixed race or reported to be “Other,” and forty were Caucasian; five of the sixty-two children were Hispanic/Latino. Seven more children participated but were excluded from analysis: three 2-year-olds (two for failing both pretrials—i.e., not knowing the happy/sad faces—and one for having fewer than six usable trials), and four 3-year-olds (two for failing the pretrials, one for having fewer than six usable trials, and one because she was loudly singing along to the auditory stimuli).

Apparatus and Procedure—The experimental setup of Experiment 2 was similar to that of Experiment 1. A cardboard box was placed on the table to the child’s right, but there was no trashcan, and a laminated piece of paper depicting a smiley-face (on the left) and a frowny-face (on the right) was placed directly in front of the child on the table. See Figures 1 and 2 for the testing setup. At the beginning of the experiment, the child was told that Puppy was searching for his lost toys; that there were two toys in the box, one of which was Puppy’s lost toy; and that Puppy would be happy if he found his lost toy, and sad if he found the other toy. The child was taught to point to the happy face when Puppy was happy, and to the sad face when Puppy was sad.

Once the child was able to point correctly to each face, the experiment began. The experimenter pulled each toy out of the box one at a time and said, “Puppy, look what I found!” In the pretrials, Puppy was “feeling shy,” so he whispered in the experimenter’s ear, and the experimenter told the child directly how Puppy felt, and whether this was his lost toy. Then the experimenter asked the child, “Can you show me how Puppy feels?” If the child did not point immediately, the experimenter asked follow-up questions like “Can you point to the face?” or “Is Puppy happy or sad?” Verbal responses, e.g., “He’s happy/sad” were also accepted. In response to the second toy, Puppy expressed the opposite emotion. If the child was unable to point to the correct faces, the experimenter ran a second pretrial. After the first 14 participants were tested, the experimenter ran both pretrials regardless of children’s ability to point to the faces, in order to reduce the possibility of response bias by presenting examples of *both* the first and second toys being the target (throughout the experiment, target-first and target-second trials were intermixed). Children who were unable to point to the correct faces in the second pretrial were excluded from the analysis.

The 12 test trials had a similar structure to pretrials, except that each child was given either pitch-contour cues or facial and body-language cues to Puppy’s emotions. Before the first test trial, children were again asked to show the experimenter that they could point to the happy and sad faces. In the first test trial, children in the *pitch condition* were told that Puppy was “not feeling shy anymore, so he’s gonna talk this time!” Children were told they would have to listen carefully to tell if Puppy was happy or sad when he saw each toy. Then, the experimenter pulled each toy out of the box and again said, “Puppy, look what I found!” The experimenter reminded the child to listen, then kept the puppet between her face and the child’s face while saying “Oh, look at that.” The pitch contours were the same as those used in Experiment 1; see Figure 3.

In the *body-language condition*, children were told they would have to watch carefully to tell if Puppy was happy or sad when he saw each toy. After the experimenter pulled each toy out of the box, children were asked, “Are you ready to watch? Let’s see what Puppy does!” For a happy response, the experimenter smiled and raised her eyebrows, and she and Puppy danced side-to-side. For a sad response, the experimenter frowned and brought her eyebrows down, and she and Puppy slumped down (see Figure 5).

As children became more familiar with the task, they sometimes participated in the story, by filling in the words “happy” and “sad”:

Experimenter: “One of these toys is the one Puppy lost, so if he finds it he’ll be:”

Child: “Happy!”

Experimenter: “That’s right! But the other toy is *not* the toy Puppy lost, so if he finds it he’ll be:”

Child: “Sad!”

Children's productions of "happy" and "sad," either during this repetitive story or as verbal responses during test trials, were recorded and analyzed to see whether children produced a pitch contrast in their own productions of "happy" versus "sad."

In a few trials, children changed from the incorrect to the correct answer, apparently catching themselves making an error. In these cases, coders carefully analyzed the videos of both the child's face and the experimenter's face to determine whether the experimenter might have inadvertently shown surprise at the incorrect response, potentially biasing the child's response. Three trials were excluded from the analysis for this reason.

Auditory Stimuli—The experimenter attempted to equate duration and amplitude when producing the stimuli. Paired t-tests comparing acoustics of happy versus sad productions in each trial (again converting Hz to ERB, natural-log-normalizing duration and intensity, and including only those children who succeeded with the pitch cue) revealed significant differences between happy and sad productions on all acoustic dimensions measured except for duration. Happy productions had higher pitch means (*happy*, 379.80 Hz; *sad*, 230.01 Hz; paired $t(18) = 46.72$), larger standard deviations of pitch samples (*happy*, 143.55 Hz; *sad*, 49.96 Hz; paired $t(18) = 39.61$), higher pitch maxima (*happy*, 748.64 Hz; *sad*, 395.48 Hz; paired $t(18) = 41.95$), higher pitch minima (*happy*, 200.84 Hz; *sad*, 155.31 Hz; paired $t(18) = 11.23$), and greater intensities (*happy*, 71.89 dB; *sad*, 71.01 dB; paired $t(18) = 6.17$; all $p < .001$, all tests 2-tailed). As in Experiment 1, only the pitch measurements had ratios of *happy* to *sad* values that appeared to be meaningfully different from one (see Appendix 2), suggesting that the differences in intensity may not have been noticeable.

Results and Discussion

Each participant gave responses in either body-language trials or pitch trials. Participants were included if they completed at least six of the twelve trials. Table 2 reports the number of children at each age succeeding with each cue; success is defined as responding with the correct emotion ("happy" or "sad") for both toys in at least 2/3 of the trials the child completed. As in Experiment 1, there was no improvement in children's success rates with the body-language cues across development; 5/6 2-year-olds and 7/10 3-year-olds succeeded. For the pitch cues, none of the six 2-year-olds and only 5/16 3-year-olds succeeded; by age 4, just over half of children (7/12) succeeded (mean, 70.6%), and by age 5, 75% of children (9/12) succeeded (mean, 79.4%). Accuracy with the pitch cue and age were significantly correlated ($r = 0.59$, $p < .001$). Among 2-year-olds, children given the body-language cues performed significantly better (mean, 88.9%) than children given the pitch cues (mean, 9.7%; $t(7.56) = 7.35$; $p < .001$). This pattern also held for 3-year-olds (body-language, 77.8%; pitch, 52.6%, $t(20.28) = 2.31$; $p < .05$; all tests 2-tailed and assuming unequal variances). Figure 6 plots accuracy against age.

Children often produced their own pitch contours when saying the words "happy" and "sad" during the experiment. To determine whether children could produce the *happy* versus *sad* pitch contrast themselves, we analyzed children's productions. We included only children from the *pitch condition*, since we could relate their own pitch productions to their pitch interpretations. For both analyses, only cases in which the child produced both words were included, and t-tests were computed on data grouped by child. Children's productions of "happy" and "sad" during the trials differed on several acoustic dimensions. Productions of "happy" had higher pitch means (325.5 Hz) than productions of "sad" (280.8 Hz; $p < .001$), larger standard deviations of pitch samples (*happy*, 60.6 Hz; *sad*, 42.1 Hz; $p < .05$), higher pitch maxima (*happy*, 420.3 Hz; *sad*, 363.3 Hz; $p < .001$), and higher intensities (*happy*, 62.00 dB; *sad*, 60.47 dB; $p < .05$; all paired $t(35) > 2.0$, all tests 2-tailed). The two words did not differ significantly in pitch minima or durations. Children's ability to produce the *happy*

versus *sad* pitch contrast (operationalized as the happy – sad subtraction for each acoustic measurement in turn, averaged, within subject, across all productions made during the trials) was not predicted by age, success in interpreting pitch contours during the experiment, or their interaction in an analysis of covariance (ANCOVA). Since children who responded verbally in the task tended to be older children, who were also more likely to succeed in the task, there may not have been enough variance in either predictor to find an effect.

Children in both Experiments 1 and 2 showed improvement in use of pitch cues with age, whereas even the youngest children succeeded with the body-language cues. Children produced the happy/sad pitch contrast themselves, but they did not consistently interpret the emotional connotations of these sentences' pitch contours until at least 4 years of age.

Experiment Three

Experiment 2 was designed to be natural to perform, interactive, and easy for children to understand. However, the face-to-face design made the pitch and facial/body-language conditions different in some ways beyond the question of interest. In particular, in the pitch condition the experimenter's face was behind the puppet while the experimenter spoke on behalf of the puppet. By contrast, in the facial/body-language condition, the experimenter and the puppet produced facial and body-language cues side-by-side. Though two- and three-year-olds understand pretense in which an adult talks on behalf of a puppet (Friedman, Neary, Burnstein, & Leslie, 2010), it is possible that the presence of this form of pretense in the pitch condition—but not in the body-language condition—made the task in the pitch condition more taxing. With this in mind, Experiment 3 presented videotaped stimuli to children. Three-year-olds were tested, because children at this age showed a marked difference in performance between the two conditions in the prior experiments. If this difference was due to superficial features of the task, as opposed to better skill in interpreting body-language versus pitch cues to emotion, children in the two conditions of Experiment 3 should perform similarly.

Method

Participants—Twenty 3-year-olds participated: ten in the pitch condition (seven boys; mean age 3 years, 4 months, 14 days) and ten in the body-language condition (seven boys; mean age 3 years, 4 months, 13 days). Participants were recruited as in Experiment 1. Of the twenty children, one was Asian, seven were African-American, and twelve were Caucasian; one child was Hispanic/Latino. Parental educational attainment information was collected. Fifteen children had at least one parent with a college or advanced degree; four had at least one parent with some college experience; and one had two parents with high-school diplomas. Six more children participated but were excluded from the analysis: two for failing both pretrials—i.e., not knowing the happy/sad faces—three for having fewer than six usable trials, and one because the parent reported developmental and speech delays.

Apparatus and Procedure—The design of Experiment 3 was similar to that of Experiment 2, but stimuli were presented on a computer screen rather than by the live experimenter. The live experimenter sat next to the child. The child sat facing the computer screen with the smiley-faces placed on the table in front of the child. The introduction to the experiment was the same as in Experiment 2. All children participated in two pretrials and then twelve experimental trials (either body-language or pitch). On each trial, the computer screen displayed a film in which Puppy appeared on the left side of the screen. A gloved hand then raised a toy from the bottom of the screen up to Puppy's eye level.

In pretrials, the experimenter then told the child directly whether Puppy was happy or sad, and asked the child to point to the correct face. In experimental trials, the film continued:

Puppy turned toward the toy, faced the child again, and made his response (the same way he did in Experiments 1 and 2). In pitch trials, Puppy's mouth moved in synchrony with a recorded happy or sad voice, which was intended to be as similar as possible to the pitch patterns produced in Experiments 1 and 2. Two different recorded instances of the happy and sad pitches were used in different trials. Pitch tracks and waveforms for the happy and sad pitches are shown in Figure 3, and acoustic details of the auditory stimuli are presented in Appendix 2.

In body-language trials, unlike in Experiment 2, the experimenter did not contribute facial cues, and body-language cues were provided by the puppet alone. These cues were similar to Puppy's body-language cues from Experiment 2, but were embellished slightly to compensate for the lack of facial cues. For the sad response, Puppy put his hands up to his face and slouched down, facing away from the toy. For the happy response, Puppy danced side-to-side, and then clapped his hands (see Figure 7 for still images of Puppy producing these responses).

Results and Discussion

Each participant gave responses in either body-language trials or pitch trials. Table 3 reports the number of children succeeding with each cue; success is defined (as in Experiment 2) as responding with the correct emotion ("happy" or "sad") for both toys in at least 2/3 of the trials the child completed. Children given the body-language cues performed significantly better (9/10 children succeeded; mean, 86.7%) than children given the pitch cues (3/10 children succeeded; mean, 41.0%, $t(15.97) = 2.78$; p (two-tailed, unequal variances) = .013). Figure 8 plots accuracy in each condition against age. Performance was not significantly correlated with age in either condition (but recall that only 3-year-olds were tested).

We again found that children succeeded with the body-language cues but struggled with the pitch cues. This suggests that the increased difficulty children had with the pitch cues in Experiment 2 was not due to differences between the conditions introduced by our face-to-face design (specifically, the presence of pretense in the pitch condition).

General Discussion

Children did not consistently interpret happy- or sad-sounding pitch contours in accordance with the emotions they cue until about age 4. By 5 years, children's interpretations of our stereotyped pitch contours accorded with our own. This late development contrasts with young infants' sensitivity to prosodic cues to stress (Jusczyk & Houston, 1999) and phrase boundaries (Hirsh-Pasek et al., 1987), and with early acquisition of lexical-tone categories (e.g., Mattock & Burnham, 2006), which appear to be acquired synchronously with consonants and vowels (at least in some languages; Demuth, 1995). But these developments before the child's first birthday all concern perceptual categorization and generalization within the speech domain, before meaningful (semantic) interpretation of phrases or stress patterns is central—so they do not provide crucial connections between diverse types of phonetic variation and the *meanings* they convey. For example, infants and young children excel at distinguishing the consonants of their language, but do not reliably infer that a consonantal change in a familiar word yields another, different word, even well into the second year (e.g., Stager & Werker, 1997; Swingley & Aslin, 2007). The phonetic categories that compose the language's phonology do not come supplied with rules for their interpretation.

One might expect earlier sensitivity to pitch cues to emotional states, since emotional states in some cases appear to have universally recognized facial and vocal signatures (Bryant & Barrett, 2007; Sauter, Eisner, Ekman, & Scott, 2010; Pell, Monetta, Paulmann, & Kotz,

2009) and to evoke an innate response (e.g., Fernald, 1993; Mumme & Fernald, 2003). However, we find that, despite early sensitivity to pragmatic functions and emotions cued by prosody in maternal speech, preschoolers have trouble detecting the pitch contours that convey *happy* versus *sad*. One possible explanation is that the happy and sad contours we tested, though among the set of emotional expressions recognized by adults across cultures (e.g., Sauter et al., 2010), are not among the set of universal mood-inducing contours accessible to infants (Fernald, 1992). The contours tested by Fernald (1992) are also produced with the goal of shaping the infant's behavior, rather than expressing the mother's emotional state, so they may be qualitatively different from *happy* and *sad*. The connection between the stereotypical intonational patterns we used and the emotions *happy* and *sad* may therefore need to be learned from experience.

Another explanation is that happy and sad pitch contours may be accessible in infancy (perhaps by inducing these emotions in infants, rather than actually signaling the talker's internal state), but may lose their iconicity through reinterpretation during language acquisition (Snow and Balog, 2002). Loss of iconicity through reinterpretation has been documented in other cases. For example, deaf children initially use pointing gestures much as hearing infants do. As they acquire American Sign Language, however—in which pointing is used both pronominally and for other functions—some stop using pointing for first- or second-person reference altogether for several months, then for several weeks actually make reversal errors, such as pointing to their interlocutor to mean “me” (Petitto, 1987). In general, sign-language words are not markedly easier for children to learn when they are more iconic (Orlansky & Bonvillian, 1984; see also Namy, 2008). As children acquire language, they seem to accept the possibility that their earliest hypotheses may be wrong. These types of reinterpretations, which lead to a U-shaped developmental function in children's performance, do not imply regression or loss of ability, but instead reflect a fundamental change in children's interpretation.

Regardless of whether meaningful interpretation of happy and sad contours occurs in early infancy, we still need to account for the consistently late understanding of these contours in our task and in previous conflict tasks (Friend & Bryant, 2000; Morton & Trehub, 2001). Children in our task were not puzzled by the semantic categories of *happy* and *sad*, readily linking them to nonlinguistic behavioral manifestations like joyful dancing or distraught slumping. Furthermore, intuition suggests that children are not deprived of real-world experience with joyful and sad emotions and their vocal expressions, which are on abundant display in daycares and playgrounds. Most likely, children's late learning of connections between the modeled intonational types and their associated emotions is due to the complexity of pitch-contour patterning in the language as a whole.

Intonation functions at both the paralinguistic and phonological levels in English (Ladd, 2008; Scherer, Ladd, & Silverman, 1984; Ladd, Silverman, Tolkmitt, Bergmann, & Scherer, 1985; Snow & Balog, 2002), which may make discovering the connections between specific intonational patterns and conveyed emotions more difficult for children. In addition, the prototypical intonational patterns for *happy* and *sad* are not produced every time someone feels happiness or sadness. Elated joy and quiet happiness have very different vocal signatures, for example (Scherer, Johnstone, & Klasmeyer, 2003; Sauter & Scott, 2007), and semantically distinct emotions can have similar pitch characteristics. Happiness, anger, and fear, for example, are all often characterized by elevated pitch (though other pitch characteristics may help differentiate these emotional expressions). These factors likely reduce the cue validity of these pitch patterns in speech, making them harder to learn. Pitch also typically cues emotions in combination with facial information. The reliability of facial signals of emotion might lead to overshadowing of pitch contour cues.

We have found that children have surprising difficulty interpreting pitch cues to the speaker’s emotions, despite the well-attested early accessibility of pitch cues at other levels of structure. This is consistent with Fernald’s (1992) suggestion that different functions of pitch in language are accessed by the child at different points depending on their developmental relevance, and—we would add—the cue validity in the signal. The present findings emphasize the importance of considering not just the perceptual availability of a particular acoustic dimension (like pitch, or vowel duration; see Dietrich, Swingley, & Werker, 2007), but the reliability and developmental relevance of each particular cue being conveyed by that dimension.

Acknowledgments

Funding for this research was provided by NSF Graduate Research Fellowship and NSF IGERT Trainee Fellowship grants to C.Q., NSF grant HSD-0433567 to Delphine Dahan and D.S., and NIH grant R01-HD049681 to D.S. Many thanks to Swingley lab members Jane Park, Katie Motyka, Allison Britt, Alba Tuninetti, Gabriella Garcia, Rachel Weinblatt, and Rachel Romeo. This research would not have been possible without the support of John Trueswell and his lab members Katie McEldoon, Ann Bunger, and Alon Hafri, who helped us run our experiments at Philadelphia preschools. We also thank members of the Institute for Research in Cognitive Science at the University of Pennsylvania and three anonymous reviewers for their suggestions. Finally, sincere thanks go to the children, parents, and preschool administrators and teachers who supported this research.

Appendix 1



Example toys used in the experiments

Appendix 2

Acoustic measurements of the experimenter’s happy versus sad speech for each acoustic dimension in each experiment, and *happy/sad* ratios

A ratio of one suggests no meaningful between-conditions difference for that dimension; values above one indicate a greater mean value for the happy stimuli. Pitch values are in Hz, intensity in dB, duration in seconds. Pitch ratios were computed over ERB values (see text) and intensity and duration ratios were computed over log-transformed values.

Cue	Experiment 1			Experiment 2			Experiment 3		
	<i>Happy</i>	<i>Sad</i>	Ratio	<i>Happy</i>	<i>Sad</i>	Ratio	<i>Happy</i>	<i>Sad</i>	Ratio
Pitch mean	417	256	1.41	380	230	1.44	348	243	1.30
Pitch SD	130	52	2.29	144	50	2.42	140	41	2.65
Pitch max.	746	387	1.50	749	395	1.50	704	324	1.64
Pitch min.	211	163	1.22	201	155	1.23	199	181	1.08
Intensity	72.7	71.3	1.00	71.9	71.0	1.00	70.1	70.1	1.00
Duration	3.30	3.24	1.02	3.15	3.12	1.01	3.40	3.41	1.00

References

- Abel SM. Duration discrimination of noise and tone bursts. *Journal of the Acoustical Society of America*. 1972; 51:1219–1223. [PubMed: 5032936]
- Baldwin DA, Moses LJ. The ontogeny of social information gathering. *Child Development*. 1996; 67:1915–1939.
- Bryant GA, Barrett HC. Recognizing intentions in infant-directed speech: Evidence for universals. *Psychological Science*. 2007; 18:746–751. [PubMed: 17680948]
- Burnham D, Kitamura C, Vollmer-Conna U. What's new, Pussycat? On talking to babies and animals. *Science*. 2002; 24:1435. [PubMed: 12029126]
- Cutler A, Norris D. The role of strong syllables for lexical access. *Journal of Experimental Psychology: Human Perception and Performance*. 1988; 14:113–121.
- D'Entremont B, Muir D. Infant responses to adult happy and sad vocal and facial expressions during face-to-face interactions. *Infant Behav & Development*. 1999; 22:527–539.
- Demuth, K. The acquisition of tonal systems. In: Archibald, J., editor. *Phonological Acquisition and Phonological Theory*. Hillsdale, NJ: Erlbaum; 1995. p. 111-134.
- Dietrich C, Swingley D, Werker JF. Native language governs interpretation of salient speech sound differences at 18 months. *Proceedings of the National Academy of Sciences of the USA*. 2007; 104:16027–16031. [PubMed: 17911262]
- Fernald A. Four-month-old infants prefer to listen to motherese. *Infant Behavior and Development*. 1985; 8:181–195.
- Fernald A. Intonation and communicative intent in mothers' speech to infants: Is the melody the message? *Child Development*. 1989; 60:1497–510. [PubMed: 2612255]
- Fernald, A. Meaningful melodies in mothers' speech to infants. In: Papousek, H.; Jurgens, U.; Papousek, M., editors. *Nonverbal vocal communication: Comparative and developmental approaches*. Cambridge: Cambridge University Press; 1992. p. 262-282.
- Fernald A. Approval and disapproval: Infant responsiveness to vocal affect in familiar and unfamiliar languages. *Child Development*. 1993; 64:657–674. [PubMed: 8339687]
- Fernald A, Kuhl P. Acoustic determinants of infant preference for motherese speech. *Infant Behavior and Development*. 1987; 10:279–93.
- Friedman O, Neary KR, Burnstein CL, Leslie AM. Is young children's recognition of pretense metarepresentational or merely behavioral? Evidence from 2- and 3-year-olds' understanding of pretend sounds and speech. *Cognition*. 2010; 115:314–319. [PubMed: 20185121]
- Friend M. Developmental changes in sensitivity to vocal paralinguistic. *Developmental Science*. 2000; 3:148–162.
- Friend M. The transition from affective to linguistic meaning. *First Language*. 2001; 21:219–243.
- Friend M. What should I do? Behavior regulation by language and paralinguistic in early childhood. *Journal of Cognition and Development*. 2003; 4:161–183.
- Friend M, Bryant JB. A developmental lexical bias in the interpretation of discrepant messages. *Merrill-Palmer Quarterly*. 2000; 46:342–369.
- Gout A, Christophe A, Morgan JL. Phonological phrase boundaries constrain lexical access II. Infant data. *Journal of Memory and Language*. 2004; 51:548–567.
- Harris MS, Umeda N. Difference limens for fundamental frequency contours in sentences. *Journal of the Acoustical Society of America*. 1987; 81:1139–1145.
- Hirsh-Pasek K, Kemler Nelson DG, Jusczyk PW, Wright Cassidy K, Druss B, Kennedy L. Clauses are perceptual units for young infants. *Cognition*. 1987; 26:269–286. [PubMed: 3677573]
- Jusczyk PW, Houston DM. The beginnings of words segmentation in English-learning infants. *Cognitive Psychology*. 1999; 39:159–207. [PubMed: 10631011]
- Kahana-Kalman R, Walker-Andrews AS. The role of person familiarity in young infants' perception of emotional expressions. *Child Development*. 2001; 72:352–369. [PubMed: 11333071]
- Katz GS, Cohn JF, Moore CA. A combination of vocal f0 dynamic and summary features discriminates between three pragmatic categories of infant-directed speech. *Child Development*. 1996; 67:205–17. [PubMed: 8605829]

- Kitamura C, Burnham D. Pitch and communicative intent in mothers' speech: Adjustments for age and sex in the first year. *Infancy*. 2003; 4:85–110.
- Ladd, DR. *Intonational Phonology*. Second Edition. Cambridge: Cambridge University Press; 2008.
- Ladd DR, Silverman KEA, Tolkmitt F, Bergmann G, Scherer KR. Evidence for the independent function of intonation contour type, voice quality, and F0 range in signaling speaker affect. *Journal of the Acoustical Society of America*. 1985; 78:435–444.
- Liu H-M, Tsao F-M, Kuhl PK. Acoustic analysis of lexical tone in Mandarin infant-directed speech. *Developmental Psychology*. 2007; 43:912–917. [PubMed: 17605524]
- Mandel DR, Jusczyk PW, Kemler Nelson DG. Does sentential prosody help infants organize and remember speech information? *Cognition*. 1994; 53:155–180. [PubMed: 7805352]
- Mattock K, Burnham D. Chinese and English infants' tone perception: Evidence for perceptual reorganization. *Infancy*. 2006; 10:241–265.
- Miller GA. Sensitivity to changes in the intensity of white noise and its relation to masking and loudness. *Journal of the Acoustical Society of America*. 1947; 19:609–619.
- Moore BCJ, Glasberg BR. Suggested formulae for calculating auditory-filter bandwidths and excitation patterns. *Journal of the Acoustical Society of America*. 1983; 74:750–753. [PubMed: 6630731]
- Moore DS, Spence MJ, Katz GS. Six-month-olds' categorization of natural infant-directed utterances. *Developmental Psychology*. 1997; 33(6):980–89. [PubMed: 9383620]
- Morton JB, Trehub SE. Children's understanding of emotion in speech. *Child Development*. 2001; 72:834–843. [PubMed: 11405585]
- Morton JB, Trehub SE, Zelazo PD. Sources of inflexibility in 6-year-olds' understanding of emotion in speech. *Child Development*. 2003; 74:1857–1868. [PubMed: 14669900]
- Mumme DL, Fernald A. The infant as onlooker: Learning from emotional reactions observed in a television scenario. *Child Development*. 2003; 74:221–237. [PubMed: 12625447]
- Mumme DL, Fernald A, Herrera C. Infants' responses to facial and vocal emotional signals in a social referencing paradigm. *Child Development*. 1996; 67:3219–3237. [PubMed: 9071778]
- Namy LL. Recognition of iconicity doesn't come for free. *Developmental Science*. 2008; 11:841–846. [PubMed: 19046152]
- Orlansky MD, Bonvillian JD. The role of iconicity in early sign language acquisition. *Journal of Speech and Hearing Disorders*. 1984; 49:287–292. [PubMed: 6748624]
- Ota M. The development of lexical pitch accent systems: An autosegmental analysis. *Canadian Journal of Linguistics*. 2003; 48:357–383.
- Papousek M, Hwang SC. Tone and intonation in Mandarin babytalk to presyllabic infants: Comparison with registers of adult conversation and foreign language instruction. *Applied Psycholinguistics*. 1991; 12:481–504.
- Papousek M, Papousek H, Symmes D. The meanings of melodies in motherese in tone and stress languages. *Infant Behavior and Development*. 1991; 14:415–440.
- Pell MD, Monetta L, Paulmann S, Kotz SA. Recognizing emotions in a foreign language. *Journal of Nonverbal Behavior*. 2009; 33:107–120.
- Petitto LA. On the autonomy of language and gesture: Evidence from the acquisition of personal pronouns in American Sign Language. *Cognition*. 1987; 27:1–52. [PubMed: 3691016]
- Pierrehumbert JB. Phonetic diversity, statistical learning, and acquisition of phonology. *Language and Speech*. 2003; 46:115–154. [PubMed: 14748442]
- Quam C, Swingley D. Phonological knowledge guides two-year-olds' and adults' interpretation of salient pitch contours in word learning. *Journal of Memory and Language*. 2010; 62:135–150. [PubMed: 20161601]
- Quam, C.; Yuan, J.; Swingley, D. Relating intonational pragmatics to the pitch realizations of highly frequent words in English speech to infants. In: Love, BC.; McRae, K.; Sloutsky, VM., editors. *Proceedings of the 30th Annual Conference of the Cognitive Science Society*. Austin, TX: Cognitive Science Society; 2008. p. 217-222.
- Russell JA, Bachorowski J-A, Fernández-Dolz J-M. Facial and vocal expressions of emotions. *Annual Review of Psychology*. 2003; 54:329–349.

- Sauter DA, Scott SK. More than one kind of happiness: Can we recognize vocal expressions of different positive states? *Motivation and Emotion*. 2007; 31:192–199.
- Sauter DA, Eisner F, Ekman P, Scott SK. Cross-cultural recognition of basic emotions through nonverbal emotional vocalizations. *Proceedings of the National Academy of Sciences of the United States of America*. 2010; 107:2408–2412. [PubMed: 20133790]
- Scherer, KR.; Johnstone, T.; Klasmeyer, G. Vocal expression of emotion. In: Davidson, RJ.; Scherer, KR.; Goldsmith, HH., editors. *Handbook of Affective Sciences*. New York, New York: Oxford University Press; 2003. p. 433-456.
- Scherer KR, Ladd DR, Silverman KEA. Vocal cues to speaker affect: Testing two models. *Journal of the Acoustical Society of America*. 1984; 76:1346–1356.
- Seidl A. Infants' use and weighting of prosodic cues in clause segmentation. *Journal of Memory and Language*. 2007; 57:24–48.
- Snow D, Balog HL. Do children produce the melody before the words? A review of developmental intonation research. *Lingua*. 2002; 112:1025–1058.
- Soken NH, Pick AD. Infants' perception of dynamic affective expressions: Do infants distinguish specific expressions? *Child Development*. 1999; 70:1275–1282. [PubMed: 10621956]
- Stager CL, Werker JF. Infants listen for more phonetic detail in speech perception than in word-learning tasks. *Nature*. 1997; 388:381–382. [PubMed: 9237755]
- Stern DN, Spieker S, Barnett RK, MacKain K. The prosody of maternal speech: Infant age and context related changes. *Journal of Child Language*. 1983; 10:1–15. [PubMed: 6841483]
- Swingley D. Statistical clustering and the contents of the infant vocabulary. *Cognitive Psychology*. 2005; 50:86–132. [PubMed: 15556130]
- Swingley D, Aslin RN. Lexical competition in young children's word learning. *Cognitive Psychology*. 2007; 54:99–132. [PubMed: 17054932]
- Thiessen ED, Hill EA, Saffran JR. Infant-directed speech facilitates word segmentation. *Infancy*. 2005; 7:53–71.
- Tomasello M, Barton ME. Learning words in nonostensive contexts. *Developmental Psychology*. 1994; 30:639–650.
- Walker AS. Intermodal perception of expressive behaviors by human infants. *Journal of Experimental Child Psychology*. 1982; 33:514–535. [PubMed: 7097157]
- Werker JF, Curtin S. PRIMIR: A developmental framework of infant speech processing. *Language Learning and Development*. 2005; 1:197–234.



Figure 1. Photographs of the experimental setup for Experiments 1 (left) and 2 (right)
The Experiment 3 setup was similar to Experiment 2, but the live experimenter, puppet, and toy-box were replaced with a computer screen that displayed just the puppet and each toy.

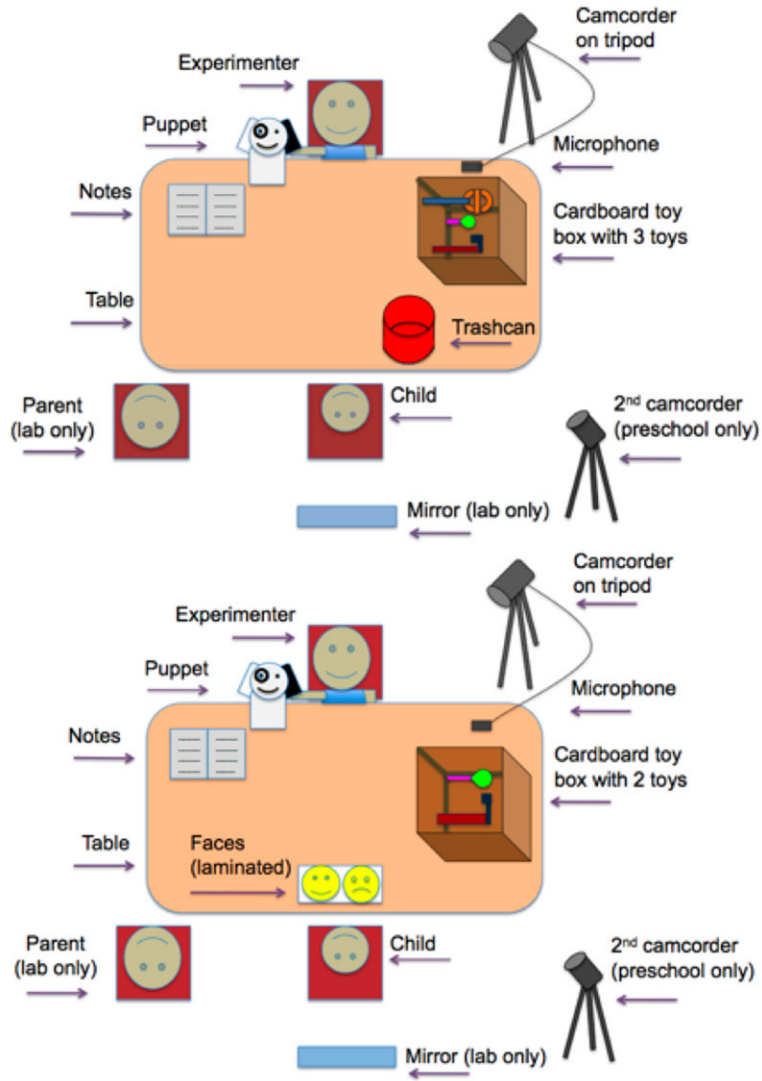


Figure 2. Diagrams of the experimental setup for Experiments 1 (top) and 2 (bottom)
 The Experiment 3 setup was similar to Experiment 2, but the live experimenter, puppet, and toy-box were replaced with a computer screen that displayed just the puppet and each toy.

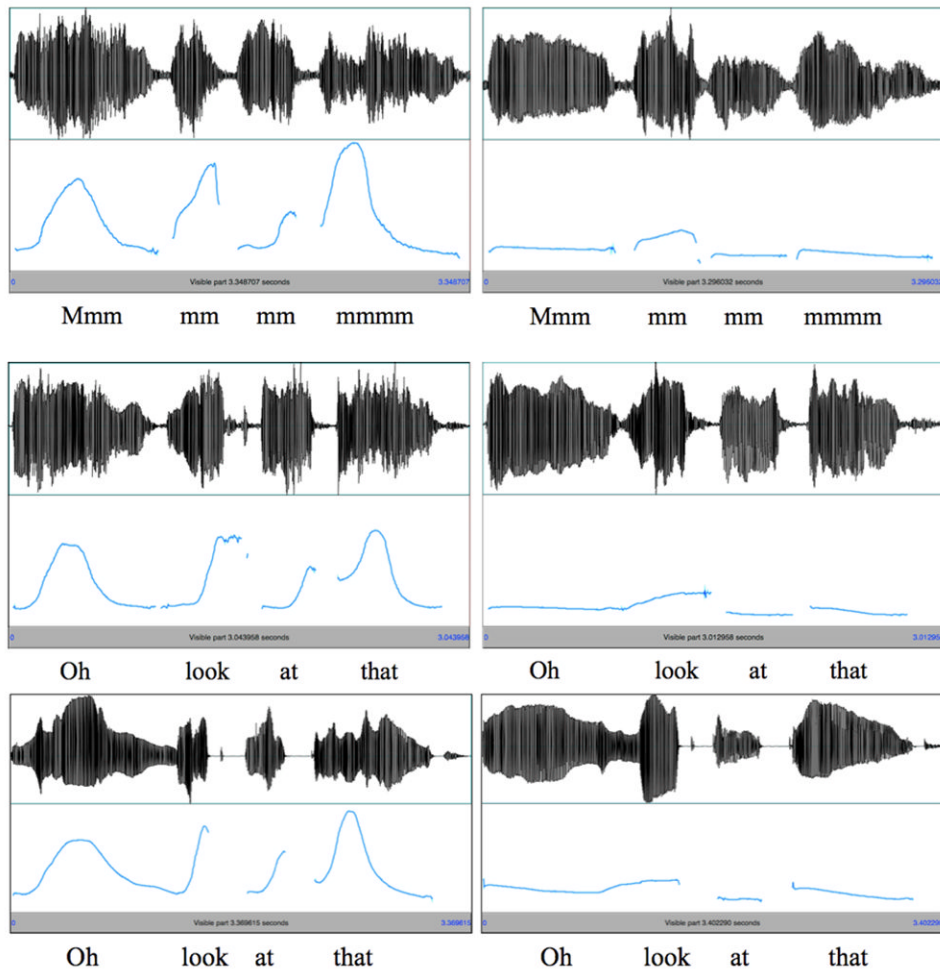


Figure 3. Waveforms and pitch contours for examples of the *happy* (left) and *sad* (right) pitch contours used in Experiments 1 (top), 2 (middle), and 3 (bottom)

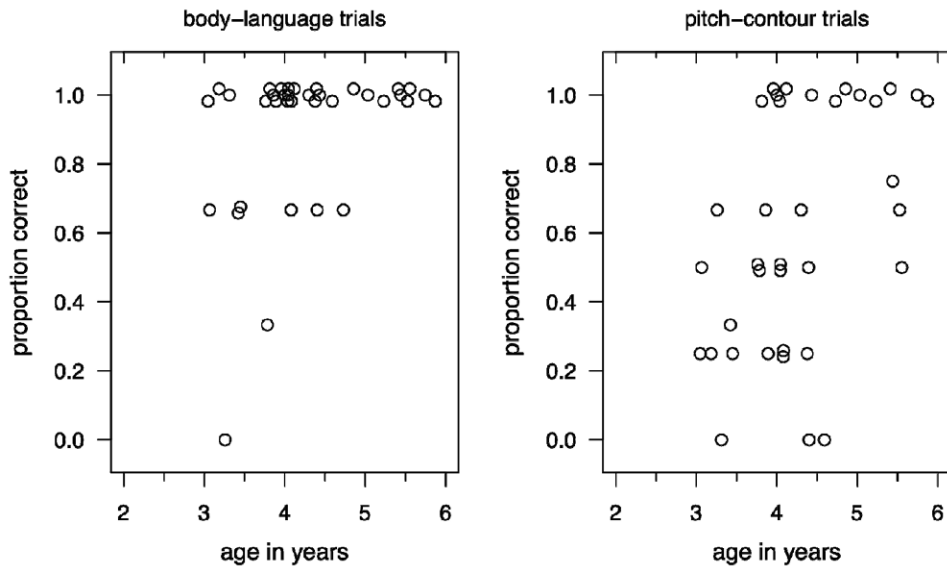


Figure 4. Scatterplots of accuracy with body-language (left) and pitch (right) cues across age in Experiment 1



Figure 5. Happy and sad facial expressions produced during Experiment 2 facial / body-language condition

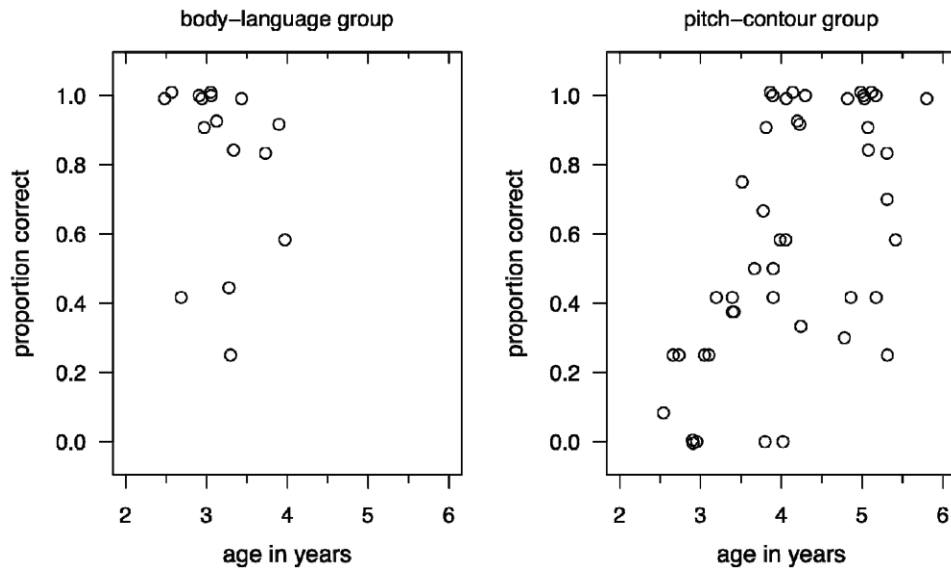


Figure 6. Scatterplots of accuracy with body-language / facial (left) and pitch (right) cues across age in Experiment 2



Figure 7. Happy and sad body language from Experiment 3

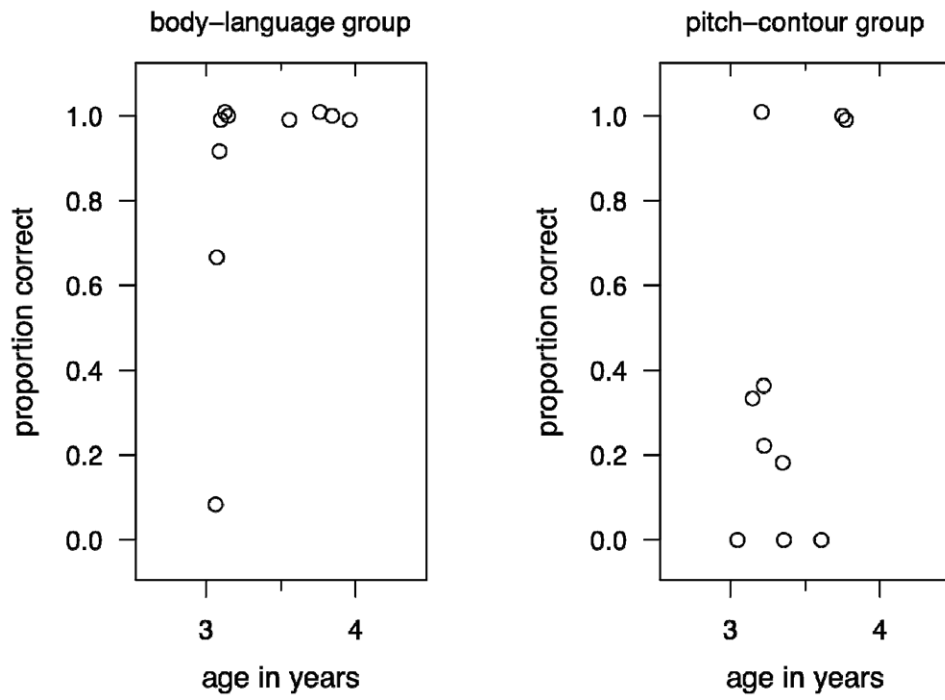


Figure 8. Scatterplots of accuracy with body-language (left) and pitch (right) cues across age in Experiment 3

Table 1

Children succeeding at each age with body-language versus pitch cues in Experiment 1 (at least two of first three trials correct)

Age	Body-language	Pitch
3	11 / 13 (85%)	7 / 13 (54%)
4	15 / 15 (100%)	10 / 15 (67%)
5	8 / 8 (100%)	7 / 8 (88%)

Table 2

Children succeeding at each age with facial/body-language versus pitch cues in Experiment 2 (at least two thirds of completed trials correct)

Age	Facial/Body-language	Pitch
2	5 / 6 (83%)	0 / 6 (0%)
3	7 / 10 (70%)	5 / 16 (31%)
4	NA	7 / 12 (58%)
5	NA	9 / 12 (75%)

Table 3

Children succeeding at each age with body-language versus pitch cues in Experiment 3 (at least two thirds of completed trials correct)

Age	Body-language	Pitch
3	9 / 10 (90%)	3 / 10 (30%)