Identifying Prosodic Contrasts in Utterances Produced by 4-, 7-, and 11-Year-Old Children

Rupal Patel
Julie T. Brayton
Northeastern University, Boston

Purpose: Acquisition of prosodic control appears to evolve across development with younger children relying on durational cues and older children utilizing a broader spectrum of cues including fundamental frequency, intensity, and duration. This study aimed to determine whether unfamiliar listeners could identify prosodic contrasts produced by 4-, 7-, and 11-year-olds despite differences in acoustic cues used by each age group.
Method: Thirty-six adult monolingual speakers of American English participated as listeners. A previous study yielded speech recordings from 12 children (2 male, 2 female from each age group) producing 2 linguistic contrasts, question–statement and contrastive stress, which served as listening stimuli.
Results: In both tasks, listener accuracy ranged from 39.7% to 100% with significant differences between 4-year-olds and both older age groups. Listeners had difficulty deciphering the 4-year-olds' questions compared with statements and were more accurate in identifying contrastive stress placed on sentence-initial words compared with sentence-final words across all age groups.
Conclusion: Although listeners identified prosodic contrasts produced by all 3 age groups, accuracy was significantly higher for 7- and 11-year-old productions. Findings are consistent with production studies that suggest relative stabilization of prosodic control between ages 4 and 7. Parallels between prosodic and segmental acquisition are discussed.

KEY WORDS: prosodic control, linguistic contrasts, children, listener perception

Speech is made up of numerous acoustic cues, including individual sound units and prosodic features. Individual sound units can be sequenced to create words. Meaning, however, is conveyed not only by combinations of sound units but also by *modulating prosody,* which refers to the manner in which words are said (Bolinger, 1958; Majewski & Blasdell, 1969; Uldall, 1961). Acoustic parameters associated with prosody include fundamental frequency (F0), intensity, and syllable duration, which are perceived as pitch, loudness, and length, respectively (cf. Bolinger, 1989; Lehiste, 1970, 1976; Netsell, 1973; Shattuck-Hufnagel & Turk, 1996). Perceptually, prosodic information serves many functions, including conveying attitudes and emotions, marking linguistic contrasts such as signaling questions versus statements, and placing contrastive stress within a phrase (Baltaxe, 1984; Eady & Cooper, 1986; Grant & Walden, 1996; Klatt, 1976; McClean & Tiffany, 1973).

Whereas prosodic patterns in adults tend to be relatively stable, the development of prosodic patterning begins early in childhood and matures over time (Allen & Hawkins, 1980; Patel & Grigos, 2006; Pierrehumbert, 1979; Snow, 1995, 1998). Even the earliest communicative gestures, such

1092-4388/09/5203-0790

as infant cries, modulate prosody (Gilbert & Robb, 1996; Lind & Wermke, 2002; Protopapas & Eimas, 1997; Wermke, Mende, Manfredi, & Bruscaglioni, 2002). It has been noted that some prosodic features begin to appear prior to word combinations, suggesting that typical development of prosodic control precedes and may facilitate segmental control (Bloom, 1973; Crystal, 1978; Katz, Beach, Jenouri, & Verma, 1996; MacNeilage & Davis, 1993; Menyuk & Bernholtz, 1969; Snow, 1994). Relative to the production of prosodic contrasts, less is known about the perception of contrasts produced by children across development. The present study addresses this gap by examining listener perception of two prosodic contrasts—question–statement and contrastive stress—as produced by 4-, 7-, and 11-year-old children. In English, these tasks provide a means for controlling the segmental units while assessing prosodic modulation. The goal is to map a developmental trajectory of prosodic development and establish a closer link between production and perception in language acquisition and use.

The course of acquiring prosodic control may vary across languages and may depend on a number of physiological and cognitive–linguistic factors. Physiologically, an infant has a shorter vocal tract, a shorter pharyngeal cavity, an anterior tongue mass, a gradual bend of oralpharyngeal channel, a higher larynx position, and a closer approximation of the velopharynx and epiglottis (Kent & Murray, 1982; Zemlin, 1998). Thus, maturation of these anatomical structures and the related motor control required to coordinate movements between these structures may impact the acquisition of control of prosodic features (Stathopoulos & Sapienza, 1997; Tingley & Allen, 1975). For example, Grigos and Patel (2007) found that lip and jaw movements associated with prosodic control continued to be refined throughout childhood. In terms of cognitive–linguistic factors, increasing facility with spoken and written language may influence the comprehension and thereby production of prosodic contrasts (Crary & Tallman, 1993; Cruttenden, 1985; Local, 1980; van der Meulen, Janssen, & den Os, 1997; Wells, Peppe, & Goulandris, 2004). These developmental factors that impact production may in turn influence listener perception of prosodic contrasts produced by children. In other words, if children use different cues or cue combinations to signal prosodic contrasts compared with adults, unfamiliar listeners may not be as adept in deciphering the linguistic intent in children's productions.

Crystal (1986) proposed five stages of prosodic acquisition. He noted that as early as 6 months of age, children's babbles resemble language-specific prosodic characteristics. In the second half of the first year, children imitate prosodic patterns produced by their parents. For example, if a parent says "all gone" after each meal, the child will accurately imitate the intonation but only approximate the phonological segments. Prosodic patterns that children hear more frequently are incorporated in reduplicative babbling strings (Whalen, Levitt, & Wang, 1991). As linguistic complexity increases to two-word productions and beyond, prosodic contrasts have been noted to become more inconsistent (Atkinson-King, 1973; Klein, 1984), which has obvious consequences on listener perception.

The present study examined two linguistic contrasts in which prosody is modulated to signal differences in meaning: question–statement and contrastive stress. In English, adults typically mark the yes/no question–statement contrast using a falling F0 contour to denote statements and a rising terminal F0 contour to indicate questions (Cruttenden, 1981; Eady & Cooper, 1986; Lieberman, 1960). Previous work suggests that falling contours impose fewer motor demands than rising contours and thus may be mastered earlier (Patel & Grigos, 2006; Snow, 1998; Xu & Sun, 2002). Falling F0 contours have been observed as early as 3 months of age in infant cries (Kent & Bauer, 1985; Kent & Murray, 1982). Although children 17 months of age have been observed to use rising versus falling intonation to signal statements versus requests (Galligan, 1987), rising contours do not appear to be mastered until later in development (Clumeck, 1980; Cruttenden, 1981; Crystal, 1986; Loeb & Allen, 1993; Patel & Grigos, 2006; Snow, 1995, 1998; Wells et al., 2004). Snow (1998) found that 4-year-olds could not imitate rising contours with the same precision as falling contours. He noted that children used a narrower F0 range than adults and tended to have longer word durations when imitating rising contours. Similarly, Wells et al. found that functional intonation was largely established by age 5 but not mastered until approximately 8 years of age. Patel and Grigos further examined the developmental course of question–statement acquisition across three different age groups. They noted that while 4-year-olds relied on durational cues to contrast questions from statements, 7-year-olds and 11-year-olds signaled the contrast in more adultlike ways using F0, intensity, and duration (similar to Allen and Arndorfer's [2000] findings on normal-hearing 8- to 12-year-old children). Given these age-related differences, listeners may need to attend to different patterns of acoustic consistencies in order to decipher questions from statements produced by children of different ages.

While the question–statement contrast requires global phrase-level control of prosody, marking contrastive stress within a phrase requires local word-level control. Thus, this task may pose even greater demands on the developing motor, cognitive, and linguistic systems of young children. In English, adults signal contrastive stress using F0, intensity, and durational cues (Cutler, 1984; O'Shaughnessy, 1979). Stressed syllables are typically higher in pitch as well as louder and longer in duration relative to unstressed syllables; however, individual speakers often vary in the cue combinations they use to signal stress (Eady & Cooper, 1986; Fry, 1955, 1958; Klatt, 1976; Kochanski, Grabe, Coleman, & Roser, 2005; Lieberman, 1960; McClean & Tiffany, 1973). Marking contrastive stress also appears to mature across development. Pollock, Brammer, and Hageman (1993) found that although 3- and 4-year-olds marked stressed syllables by increasing F0, intensity, and duration, 2-year-olds were only able to lengthen the duration of stressed syllables. As children mature, not only do they mark stressed syllables, but they also appear to have improved control over reducing the duration of unstressed syllables (Allen & Hawkins, 1980; Pollock et al., 1993), which may further enhance listener comprehension.

Grammatical role and sentential position of the stressed word also seems to impact the cue(s) used to signal contrastive stress (Cooper, Eady, & Mueller, 1985; Eady & Cooper, 1986). In adults, the extent of F0 rise and syllable lengthening is greater for stressed words in the sentence-initial position compared with the sentence-final position (Cooper et al., 1985; Eady & Cooper, 1986). In children, Snow (1998) noted greater difficulty in raising F0 in sentence-final positions than in nonfinal positions. These findings are further corroborated by Baltaxe (1984) and Hornby and Hass's (1970) findings that children make more errors when stressing objects compared with agents, given that objects tend to occur more often in sentence-final positions compared with agents.

The prosodic cues that listeners rely on to decipher linguistic contrasts appear to vary on an individual basis, just as speakers vary in how they use each cue (Howell, 1993; Peppe, Maxim, & Wells, 2000). In adult productions, the relative importance of each prosodic cue on listener perception has been heavily debated. Many argue that F0 is the prominent cue for signaling yes/no questions in English (Baum, 1998; Fry, 1955, 1958; Lieberman, 1960; Morton & Jassem, 1965), with temporal fluctuations serving as secondary cues (Baum, 1998; Klatt, 1976; Lyberg, 1979). Similarly, Beach (1991) and Beach, Katz, and Skowronski (1996) identified duration and pitch as the most important prosodic markers for identifying contrastive stress. In contrast, Kochanski et al. (2005) argued that changes in loudness may actually be the primary cue that listeners rely on to determine syllabic stress, despite Fry's (1955) findings that duration was more effective than loudness. Additionally, the absolute differences in prosodic cues needed to perceive a word as being stressed may differ based on its position within an utterance. Pierrehumbert (1979) found that listeners perceived F0 peaks later in the sentence as being the same pitch as those earlier in the sentence even if the absolute frequency value was lower at the end of the sentence. It appears that listener expectations about pitch declination within an utterance play a role in their ability to estimate F0 peaks in adult productions.

Given that children at different stages of development may signal prosodic contrasts using different prosodic cue combinations, it remains unclear whether unfamiliar listeners can attune to individual and age-related differences when making classification judgments. The present study aimed to determine whether 36 unfamiliar adult listeners could accurately identify question–statement and contrastive stress produced by 4-, 7-, and 11-year-old children despite variations in the acoustic cues used across development. Subjective impressions of the set of salient cues used to identify question–statement and contrastive stress within a phrase were also assessed. It was hypothesized that listeners would be more adept at accurately identifying the 7- and 11-year-old productions compared with those of the 4-year-old group, given that younger children's productions may be more variable and may rely on nonstandard acoustic cues, whereas the older children's productions would more closely approximate adult acoustic patterning.

Method

Speaker Database

A database collected in a previous study (Patel & Grigos, 2006) yielded question–statement recordings from 12 typically developing children ages 4, 7, and 11 years old. Contrastive stress recordings were also made at that time from the same 12 children. Each age group included 2 male and 2 female children (4-year-olds: M = 4 years, 4 months; SD = 3.5 months; 7-year-olds: M = 7 years, 4 months; SD = 5.1 months; 11-year-olds: M = 11 years, 3 months; SD = 3.3 months). All participants were monolingual speakers of American English with normal hearing based on observed hearing thresholds that fell at or below 25 dB for 500, 1000, 2000, and 4000 Hz. Prior to completing the task, parental interviews and informal play sessions were conducted to assess adequacy of speech and language abilities.

Speech recordings were collected in an audiometric booth using a MiniDisc recorder (HHB 500 PortaDisc[1]) and a unidirectional head-mounted cardiod dynamic microphone (Shure, SM10A[2]) placed 1 in. from the corner of the child's mouth. The researchers (Patel & Grigos, 2006) introduced each child to a set of four pictures mounted on popsicle sticks representing the target words in the elicited phrases. Pictures included the character SpongeBob SquarePants (*Bob*), cartoon images of a robot (*bot*), a grandfather (*Pop*), and a cooking pot (*pot*). Each child produced multiple repetitions of two phrases "Show Bob a bot" and "Show Pop a pot" in the form of either a

[1]Frequency response: 10 Hz–20 kHz; wow and flutter below measurable limits.

[2]Frequency response: 50 Hz–15000 Hz.

question or statement or with contrastive stress place on the nonfinal word (*Bob/Pop*) or on the final word (*bot/pot*).

The Appendix provides the contextual scenarios and prompts that were used to elicit the different forms of each phrase. Commands or requests issued by the child were enacted by the researcher using the popsicle stick characters as a means for creating a naturalistic, communicative interaction. For example, when eliciting contrastive stress on the agent (nonfinal word) in "Show BOB a bot," the researcher would place the "bot" character on the table in front of the child. The "pot" character would be placed out of view. She would then provide the contextual scenario while holding the BOB and POP characters in view of the child and ask, "Who should I show the bot to? Tell me what to do." If the child accurately produced the phrase "Show BOB a bot," the researcher would place the "Pop" character out of view and proceed to enact the appropriate command with the "Bob" character. If the child produced the wrong command, the researcher would enact the child's request and then follow up with probes to elicit the intended phrase.

In many instances, especially for children in the 4-year-old group, additional cues and sometimes models[3] were required to elicit the question form. In such instances, the response was modeled by one of the investigators, and the contextual scenario was repeated in order to reduce the effects of imitation. The 4-year-old group also required some modeling[4] of the contrastive stress conditions. Once again, one of the investigators modeled the accurate form. The contextual scenario and appropriate prompt were repeated prior to eliciting the child's response. Tokens produced immediately following a direct model were not analyzed.

Selection of Stimuli for the Listening Experiment

A subset of the original data collected by Patel and Grigos (2006) was used in the present study. The original datasets consisted of up to 15 tokens of each phrase and condition. Selection of the subset for the present study was based on the judgment of two independent research assistants who heard each production and selected tokens that were free of acoustic errors (e.g., F0 tracking errors, interjections or noise, multiple talkers) and produced without direct imitation. Productions of each individual speaker were further analyzed to select the three most representative tokens of each phrase and condition. This selection process entailed acoustic analysis of the eligible error-free productions.[5]

The Praat speech analysis software package (Boersma & Weenink, 2005) was used to derive estimates of F0, intensity, and duration for each word within utterances in a multistep process. First, the beginning and end of individual words within the eligible spoken utterances were manually labeled (r = .996 interlabeler reliability for 10% of the data) in order to generate a series of relative intensity values (in dB) and frequency values (in Hz) across the duration of each word. Manual correction of automatically generated F0 values was required on 37 of the 708 productions because of faulty pitch tracking that could not be verified auditorily. Adjusting the upper and lower F0 limits and frame duration parameters in Praat typically led to improved tracking. These new F0 values were verified through visual and auditory inspection and were confirmed using direct calculation of the pitch period from the waveform. When Praat-derived F0 values continued to be judged as errors (this occurred in 21 productions), these values were replaced by manually derived values obtained from the waveform. A customized program operated on the Praat generated values to calculate the average F0, average intensity, and duration of target words in each utterance.

The acoustic results were used to further prune the set of samples used in the perceptual experiment. Mean statistics of each prosodic feature were calculated for each phrase and condition per speaker. Tokens that had outlier values in one or more prosodic parameter (F0, intensity, or duration) in comparison to the mean values were eliminated. Of the remaining recordings, a random selection of three tokens was utilized. Interrater agreement for the selected tokens was 96.8%. The two research assistants did not agree on 9 out of 288 possible tokens. For these recordings, the first author selected one of the two samples chosen by the research assistants based on auditory and visual inspection of the acoustic data.

Tables 1 and 2 provide descriptive statistics on average F0 (in Hz), average intensity (in dB), and duration (in ms) of the final set of chosen trails per age group (2 phrases × 2 prosodic distinctions × 3 repetitions × 4 children) for the question–statement and contrastive stress listening tasks, respectively.

Listener Participants

Thirty-eight adult monolingual speakers of American English were recruited as listeners (19 men and 19 women), 21–50 years of age (M = 27.8 years). Listeners were

[3]Across all 4-year-olds, models were provided a total of 17 times, 15 of which were for question tokens. Given that tokens produced directly after models were not included in the analysis, the 4-year-olds produced more repetitions than the other age groups. In contrast, only three question models were required across all 7-year-olds, and the 11-year-old group did not require a model to elicit the question or statement forms.

[4]Across all 4-year-olds, a total of nine models were provided to elicit contrastive stress. The 7- and 11-year-olds only required two models each.

[5]There were a total of 708 eligible error-free productions across age groups and the question–statement and contrastive stress tasks.

Table 1. Means and standard deviations (in parentheses) of F0 (Hz), intensity (dB), and duration (ms) across 4-, 7-, and 11-year-olds when producing the question–statement contrast across both phrases.

Acoustic cue	Age	Contrast	Show	Bob/Pop	a	bot/pot
F0 (Hz)	4	Q		253.8 (29.6)		252.7 (34.8)
		S		263.3 (34.1)		224.1 (18.1)
	7	Q		256.3 (42.9)		281.7 (33.8)
		S		250.8 (26.2)		229.3 (20.1)
	11	Q		225.9 (21.0)		260.1 (24.4)
		S		212.1 (12.9)		191.1 (15.4)
Intensity	4	Q		59.9 (5.6)		57.9 (6.1)
		S		60.6 (4.5)		56.5 (5.1)
	7	Q		56.1 (7.1)		56.5 (8.2)
		S		54.9 (7.3)		50.8 (8.9)
	11	Q		62.9 (4.5)		64.4 (3.6)
		S		65.0 (3.8)		60.7 (3.1)
Duration	4	Q		241.9 (48.6)		336.4 (48.7)
		S		217.6 (23.4)		231.2 (36.2)
	7	Q		234.7 (40.7)		345.6 (23.1)
		S		257.3 (39.5)		237.7 (22.5)
	11	Q		188.1 (19.5)		280.5 (21.6)
		S		208.9 (22.6)		203.5 (21.5)

Note. Q = question; S = statement.

Table 2. Means and standard deviations (in parentheses) of F0 (Hz), intensity (dB), and duration (ms) across 4-, 7-, and 11-year-olds when producing contrastive stress on nonfinal versus final words across both phrases.

Acoustic cue	Age	Stress location	Show	Bob/Pop	a	bot/pot
F0	4	Nonfinal		266.1 (28.1)		240.7 (17.5)
		Final		255.3 (18.7)		232.8 (13.8)
	7	Nonfinal		307.3 (32.5)		224.6 (24.8)
		Final		235.5 (31.2)		293.53 (30.4)
	11	Nonfinal		234.4 (21.1)		183.7 (10.2)
		Final		207.2 (22.3)		236.9 (11.1)
Intensity	4	Nonfinal		68.7 (4.5)		64.6 (2.9)
		Final		67.4 (2.4)		64.5 (3.6)
	7	Nonfinal		69.7 (7.2)		60.7 (16.9)
		Final		65.5 (7.2)		67.4 (7.7)
	11	Nonfinal		77.5 (4.5)		69.4 (3.1)
		Final		74.0 (3.6)		75.5 (2.8)
Duration	4	Nonfinal		220.2 (12.4)		161.3 (16.1)
		Final		186.2 (8.6)		156.2 (10.6)
	7	Nonfinal		252.1 (16.4)		176.2 (10.3)
		Final		202.7 (12.4)		214.7 (14.2)
	11	Nonfinal		216.8 (8.6)		165.1 (8.9)
		Final		181.7 (8.2)		221.3 (10.1)

recruited from the greater Boston area using flyers, Internet postings, and word of mouth. All listeners were assessed to have hearing thresholds at or below 25 dB in at least one ear for 500, 1000, 2000, and 4000 Hz, and reported having vision within correctable limits.

Procedure

The study took place in a sound-treated booth with listeners seated in front of a computer displaying a custom-designed graphical interface that guided them through two perceptual tasks: (a) identifying questions versus statements (QS) and (b) identifying contrastive stress (CS) within an utterance. Prior to beginning the experiment, listeners were familiarized with the tasks using screen shots of each graphical interface. Listeners were also told that they would be probed at the end of each listening phase about how they made their classification judgments. The graphical interface was used to play each recording and to log listener responses as correct or incorrect based on the intended contrast. Speech recordings were presented through over-ear headphones (AKG K240 Studio[6]), and listeners responded using a standard computer mouse. In the QS task, listeners judged whether the phrase was produced as a question or a statement. In the CS task, listeners judged whether stress was placed on the nonfinal (*Bob* or *Pop*) or final (*bot* or *pot*) word in the sentence. In each task, listeners heard 144 recordings made up of 12 tokens (2 phrases × 2 prosodic distinctions × 3 repetitions) from each of the 12 children. An additional 36 tokens were repeated to assess listener reliability. The final analysis included those listeners with greater than 80% reliability on the repeated tokens across both tasks. Two listeners were excluded based on this criterion; thus, 36 listeners were included in the final analysis.

The experiment consisted of six distinct phases in which listeners heard recordings of the 4-, 7-, or 11-year-old groups producing the QS or CS tokens. In each phase, listeners were not told which age group they were listening to. Additionally, the order of tasks (QS and CS) and the order in which the age groups were presented were randomized across listeners.

At the completion of each phase, listeners were asked to verbally respond to a set of questions that probed for feedback about task ease and subjective impressions as to which acoustic cues they felt they used to make their classification decisions. Task ease was measured by providing listeners with the following nominal rating scale: "Was identifying question vs. statement 1 = *very easy,* 2 = *easy,* 3 = *hard,* or 4 = *very hard*" and "Was identifying stress placement 1 = *very easy,* 2 = *easy,* 3 = *hard,* or

[6]Frequency response: 15 Hz–25000 Hz.

4 = *very hard*." Subjective impressions regarding salient acoustic cues were elicited using open-ended questions such as "What did you listen to when making your decision for this age group?" At the completion of experiment, the experimenter defined the terms *pitch, loudness,* and *duration* and then asked listeners to rate which of the three cues they felt was most important in making their classification decisions.

Results

For each task, listener accuracy scores were analyzed using a repeated measures analysis of variance (ANOVA) with two within-subject variables. For the QS task, the within-subject variables were phrase type, which had two levels (question [Q] or statement [S]), and age, which had three levels (4-, 7-, or 11-year-old group). For the CS task, the within-subject variables were stress location, which had two levels (stress on the nonfinal word or on the final word), and age, which had three levels (4-, 7-, or 11-year-old group). Given unequal variance in listener performance across age groups, and possible ceiling effects when listening to the 7- and 11-year-olds' productions, the data were subjected to an arcsin square root transformation prior to conducting the statistical analyses. The Mauchly Test[7] indicated that the transformation satisfied the sphericity assumption for both the QS (p = .140 for age; p = .397 for Age × QS) and the CS (p = .261 for age; p = .781 for Age × CS) datasets. In the ANOVA analyses, the F statistic was used to test the null hypothesis at α = .05. Within the QS and CS tasks, nine pairwise t tests were conducted on the transformed data to examine differences in accuracy between question–statement or stress location both within and across age groups. To account for multiple comparisons, the Bonferroni correction factor was used to adjust the alpha level to .0056. For each task (QS and CS), additional t tests were used to compare average listener ratings of task ease across age groups using an adjusted alpha level of p < .016. All data analyses were conducted using SPSS (Version 15.0).

Question–Statement Task

Statistically significant main effects were found for phrase type, $F(1, 35) = 248.13$, $p < .0001$, and age, $F(2, 70) = 330.91$, $p < .0001$. The two-way interaction between phrase type and age was also statistically significant, $F(2, 70) = 204.02$, $p < .0001$; see Table 3). On average, listeners were 78.4% (*SD* = 2.0%) accurate when judging 4-year-olds' productions, compared with 98.7% (*SD* = 1.9%) accurate for 7-year-olds' productions and 98.5% (*SD* = 3.7%) accurate for 11-year-olds' productions (see Figure 1).

Table 3. Within-subject effects of the sphericity assumed repeated measures ANOVA for the question–statement task based on arcsin (square root transformation) transformed data.

Effect	*df*	Sums of squares	Mean square	*F*	*p*
Age group (4-, 7-, and 11-year-old)	2	5.20	2.60	330.91	< .0001
Error	70	0.55	0.01		
Phrase type (Q, S)	1	3.77	3.77	248.13	< .0001
Error	35	0.53	0.02		
Age Group × Phrase Type	2	3.91	1.95	204.02	< .0001
Error	70	0.67	0.01		

Note. ANOVA = analysis of variance.

Although listeners were more accurate in identifying questions produced by 7- and 11-year-olds compared with those produced by 4-year-olds (p < .0001 for 4- vs. 7-year-olds; p < .0001 for 4- vs. 11-year-olds), there were no significant differences between 7- and 11-year-olds (p = .98). The production data reflect these perceptual findings in that 7- and 11-year-olds tended to signal questions using increased F0 and duration of the final word, whereas the 4-year-olds were only able to increase duration (see Table 1). On the 4-year-olds' productions, listeners were more accurate at identifying statements than questions (p < .0001). The production data indicate that 4-year-olds decreased F0 for the final word of statements, which is consistent with how adults and older children signal statements. In contrast, the 4-year-olds marked questions with only increased duration of the final word, which may have led to listener errors. In contrast, listeners were equally successful at identifying both questions and statements produced by 7- and 11-year-olds.

Contrastive Stress Task

Statistically significant main effects were found for stress location, $F(1, 35) = 104.01$, $p < .0001$, age, $F(2, 70) = 300.29$, $p < .0001$, and the two-way interaction between stress location and age, $F(2, 70) = 16.90$, $p < .0001$; see Table 4). On average, listeners were 50.3% (*SD* = 4.8%) accurate when listening to 4-year-olds' tokens, compared with 84.4% (*SD* = 7.5%) accurate on 7-year-olds' productions and 84.3% (*SD* = 9.2%) accurate on 11-year-olds' productions (see Figure 2).

[7]This test assesses the equality of the variances of the differences between levels of the repeated measures factor. Sphericity is assumed for significance levels greater than .05.

Figure 1. Average listener accuracy by age group for the question–statement task. Error bars represent standard deviation.

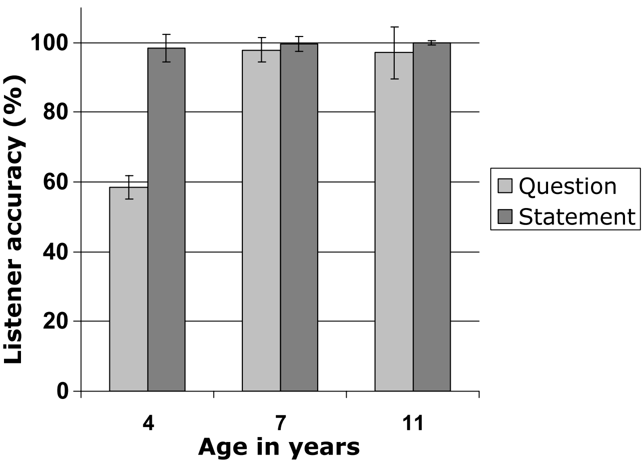
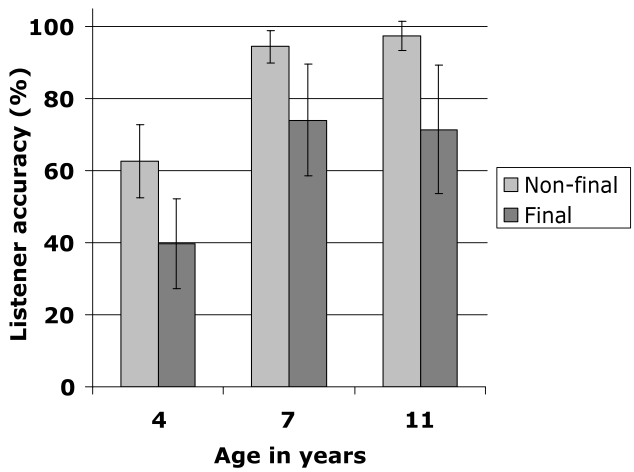


Figure 2. Average listener accuracy by age group for the contrastive stress task. Error bars represent standard deviation.



Across all age groups, listeners were more successful at identifying contrastive stress placed on the nonfinal position compared with the final position. There were also significant differences in listener perception of contrastive stress at each stress location between 4- versus 7-year-olds ($p < .0001$) and 4- versus 11-year-olds ($p < .0001$). Furthermore, while a statistically significant difference was noted for the nonfinal location between 7- and 11-year-old productions ($p < .0001$), the difference was not significant for the final position ($p = .38$).

The production data indicate that all three age groups were able to increase at least one prosodic cue for marking stress on the nonfinal word but were more variable on how they marked stress on the final word (see Table 2). Although the 4-year-olds primarily used duration to mark contrastive stress on the nonfinal word, the 7- and 11-year-olds used all three cues of F0, intensity, and duration. Strategies for indicating contrastive stress on the final word varied across age groups. Although the 4-year-olds were unable to use reliably any of the three prosodic cues, the 7-year-olds relied primarily on increased F0, and the 11-year-olds used both increased F0 and duration.

Table 4. Within-subject effects of the sphericity assumed repeated measures ANOVA for the contrastive stress task based on arcsin (square root transformation) transformed data.

Effect	*df*	Sums of squares	Mean square	*F*	*p*
Age group (4-, 7-, 11-year-old)	2	9.01	4.50	300.29	< .0001
Error	70	1.05	0.02		
Stress location (nonfinal, final)	1	6.06	6.06	104.01	< .0001
Error	35	2.04	0.06		
Age Group × Stress Location	2	0.39	0.20	16.90	< .0001
Error	70	0.81	0.01		

Subjective Impressions of Listener Perception

Listener ratings of classification ease varied by prosodic task and age group. On average, listeners found the QS task easier than the CS task. In both tasks, listeners rated 4-year-olds' productions as most difficult to discern and 11-year-olds' productions as easiest to identify. For the QS task, average listener ease ratings (on a scale of 1–5, with responses ranging from 1 [*extremely easy*] to 5 [*extremely hard*]) were as follows: 2.0 for 4-year-olds, 1.64 for 7-year olds, and 1.42 for 11-year-olds. Statistically significant differences were noted between ratings on the 4- versus 7-year-olds' tokens ($p = .002$) and the 4- versus 11-year-olds' tokens ($p < .0001$) but not for the 7- versus 11-year-olds' productions ($p = .044$). For the CS task, average listener ease ratings were as follows: 2.92 for 4-year-olds, 2.33 for 7-year-olds, and 2.25 for 11-year-olds. Statistically significant differences were noted between ratings on the 4- versus 7-year-olds' productions and the 4- versus 11-year-olds' productions ($p < .0001$) but not between the 7- versus 11-year-old groups ($p = .499$).

When probed about which acoustic cues they relied on for making their classification judgments, most listeners indicated that pitch was the main factor. For the

QS task, all 36 listeners reported relying foremost on pitch, especially toward the end of the phrase. They stated that "if (she or he) heard a rise at the end (she or he) would classify it as a question." Self-assessments of salient prosodic features in the CS task were more variable. Twenty-one of the 36 listeners reported attending to pitch, whereas 8 reported attending to loudness, and 7 stated that they relied on duration differences.

Discussion

The present study examined whether unfamiliar listeners could accurately identify two prosodic contrasts (viz., QS and CS) produced by 4-, 7-, and 11-year-old children despite possible variations in the acoustic cues used across development.

Listeners were more accurate on the QS task than on the CS task for all age groups. Although the QS contrast requires a global, sentence-level change in prosodic features, the CS task demands more precise word-level changes within an utterance. Thus, the latter may be more difficult to produce. These production difficulties may in turn impact listener accuracy.

On the QS task, listener errors were primarily on the question tokens produced by 4-year-olds. Furthermore, listener accuracies between question and statement tokens differed only for the 4-year-olds' productions. The production data provide some insights into these findings in that 4-year-olds used standard adultlike cues to signal statements (phrase-final F0 and intensity declination) but nonstandard cues (i.e., only using increased final-word duration in the absence of increased F0) to signal questions. Perhaps this resulted in a mismatch between the child's signaling strategy and listener expectations. This explanation is supported by listeners' subjective impressions that they were attending to the end of the phrase and classifying the phrase as a question if they heard a rise in pitch (similar to that documented by Cruttenden, 1981; Eady & Cooper 1986; Hadding-Koch & Studdert-Kennedy, 1964; Majewski & Blasdell, 1969). These perceptual findings reinforce production studies documenting that the rising intonation contour required to signal questions in English is difficult for young children and does not seem to be mastered until age 7 or 8 (Clumeck, 1980; Cruttenden, 1981; Crystal, 1986; Loeb & Allen, 1993; Patel & Grigos, 2006; Wells et al., 2004). Additionally, as Whalen et al. (1991) noted, given that American English speakers utilize falling contours more than rising contours, children learning to imitate prosodic contrasts may be more adept at producing what they hear most often (i.e., the falling contour of statements). Future cross-linguistic investigations may shed light on different patterns of prosodic acquisition across languages.

On the CS task, regardless of the child's age, listeners had more difficulty identifying stress placed on the final word (object name) than on the nonfinal word (agent name) in an utterance. Perhaps this relates to the findings of previous production studies (Baltaxe, 1984; Hornby & Hass, 1970) noting that children make more errors when stressing objects, which tend to be in sentence-final positions, than when stressing agents, which tend to occur more often in nonfinal positions. Similar to previous findings, the production data indicate that sentence position influences the extent of F0 rise and syllable lengthening on stressed words with increased contrasts for sentence-initial words compared with sentence-final words (Baltaxe, 1984; Cooper et al., 1985; Eady & Cooper, 1986; Hornby & Hass, 1970; Pierrehumbert, 1979; Snow, 1998). Additionally, we noted that each age group marked the two stress locations with different acoustic cues. Similar to Pollock et al. (1993), we found that younger children (4-year-olds) relied on duration to mark contrastive stress, whereas the older children used multiple cues. Interestingly, some authors have noted that in the course of English phonological development, children may use vowel length in a contrastive manner (Ingram, 1976; Renfrew, 1966; Smith, 1973). In response to open-ended questions about salient acoustic cues, 8 of the 36 listeners in the present study reported attending to pauses within the phrase and/or the clarity of the final consonant of the target word as cues for identifying stress location. Overall, listeners indicated that identifying stress location was more difficult than discerning between QS tokens, which could be attributed in part to variations in the cues used to mark contrastive prosody (Eady & Cooper, 1986; Fry, 1955, 1958; Klatt, 1976; Kochanski et al., 2005; Lieberman, 1960; McClean & Tiffany, 1973).

In general, listeners were less accurate and more variable at identifying 4-year-old productions in both tasks compared with identifying 7- and 11-year-old productions. The only significant difference between the 7- and 11-year-old productions was for identifying stress on nonfinal words. These findings may reflect differences in production abilities among the age groups studied (Allen & Hawkins, 1980; Patel & Grigos, 2006; Snow, 1995), suggesting that the development of prosodic patterning matures over time. There are numerous physiological changes that co-occur, which may facilitate more consistent and reliable control of prosody across development such as gradual lengthening of the vocal tract and pharyngeal cavity, lowering of the larynx, and widening of the gap between the velopharynx and epiglottis. In addition, advancing cognitive and linguistic abilities may play an integral role in achieving greater precision of prosodic control in older children, which in turn benefits listener comprehension.

Listener accuracy in identifying prosodic contrasts is dependent not only on acoustic consistencies in the speech signal but also on a priori biases about which cues are salient. Thus, if children in a particular age group use an acoustic cue(s) that differs from what is expected, listeners may not be attuned to it, even if the cue is used consistently and reliably. In the QS task, for example, Patel and Grigos (2006) noted that 4-year-olds marked questions with elongated final-syllable duration rather than the conventional rising intonation noted in the 7- and 11-year-olds' productions. In the present study, all 36 listeners indicated that they were attending to pitch differences at the end of the utterance when making their classification decisions; thus, differences in duration, although present in the acoustic signal, may not have been perceptually salient. Similarly in the CS task, children in the 7- and 11-year-old groups marked contrastive stress on nonfinal words with two or more cues, whereas the 4-year-olds relied primarily on increased duration. This mismatch between children's productive abilities and listener attunement to prosodic cues may impact communicative success. For example, a young child attempting to signal contrastive emphasis on a particular word within an utterance may be misunderstood despite consistent use of one or more prosodic cues. The literature on segmental development parallels our findings in that when children use nonstandard or unusual phonetic cues (i.e., non-adult-like) to signal phonological distinctions, listeners are unable to decipher the child's intention (cf. Gierut & Dinnsen, 1986; Macken & Barton, 1980; Weismer, Dinnsen, & Elbert, 1981). Thus, communication breakdowns, which result from mismatches in productive abilities and listener expectations, may impact a child's ability to grasp linguistic distinctions as well as a listener's perception of the child's communicative and cognitive competence. It is important to note that an assumption of an adult underlying form may lead us to erroneously underestimate a young child's prosodic phonology or to overlook ways in which they are in fact signaling differences within the constraints of their maturing speech production system. Elucidating the various ways in which prosodic contrasts can be signaled across the age span is essential for designing interventions that focus on improving speaker abilities as well as tuning listener expectations.

Limitations and Future Directions

While the present study suggests that listener accuracy of prosodic contrasts improves as production stabilizes across development, these initial findings are limited to English speakers in the three age groups studied. Further inquiry with a larger sample size of speakers in each age group and finer grained age levels is required to specify with greater precision the timeframe in which production and perception of prosodic contrasts are mastered. Examining the production and perception of additional prosodic contrasts such as verb–noun syllabic contrasts as well as task-related differences in read versus spontaneous speech in English and across languages is also warranted. Lastly, it may be fruitful to compare and contrast the perceptions of age-matched peers with those of adults to see if a priori biases also shift with age. Perhaps children may be more apt at identifying subtle prosodic changes produced by their peers that adults may not be attuned to.

Acknowledgments

This work was supported in part by National Institute on Deafness and Other Communication Disorders Grant DC-06118 and the Charles H. Hood Foundation. We would like to express our gratitude to Michael Everett for developing the graphical interface for the experiment; Howard Cabral, Michael Epstein, and Kevin Reilly for their assistance with the statistical analysis; and the participants for their time.

References

Allen, G. D., & Arndorfer, P. M. (2000). Production of sentence-final intonation contours by hearing-impaired children. *Journal of Speech, Language, and Hearing Research, 43,* 441–455.

Allen, G. D., & Hawkins, S. (1980). Phonological rhythm: Definition and development. In G. H. Yeno-Komshian, J. F. Kavanagh, & C. A. Ferguson (Eds.), *Child phonology: Vol. 1. Production* (pp. 227–256). New York: Academic Press.

Atkinson-King, K. (1973). Children's acquisition of phonological stress contrasts. *UCLA Working Papers in Phonetics, 25,* 1–162.

Baltaxe, C. (1984). Use of contrastive stress in normal, aphasic, and autistic children. *Journal of Speech and Hearing Research, 27,* 97–105.

Baum, S. (1998). The role of fundamental frequency and duration in the perception of linguistic stress by individuals with brain damage. *Journal of Speech, Language, and Hearing Research, 41,* 31–40.

Beach, C. M. (1991). The interpretation of prosodic patterns at points of syntactic structures ambiguity: Evidence of cue trading relations. *Journal of Memory and Language, 30,* 644–663.

Beach, C. M., Katz, W. F., & Skowronski, A. (1996). Children's processing of prosodic cues for phrasal interpretation. *The Journal of the Acoustical Society of America, 99,* 1148–1160.

Bloom, L. (1973). *One word at a time*. Mouton: The Hague, The Netherlands.

Boersma, P., & Weenink, D. (2005). *Praat: Doing phonetics by computer* (Version 4.3) [Computer program]. Retrieved July 30, 2005, from http://www.praat.org.

Bolinger, D. (1958). A theory of pitch accent in English. *Word, 14,* 109–149.

Bolinger, D. (1989). *Intonation and its uses: Melody in grammar and discourse*. Stanford, CA: Stanford University Press.

Clumeck, H. (1980). The acquisition of tone. In G. H. Yeno-Komshian, J. F. Kavanagh, & C. A. Ferguson (Eds.), *Child phonology: Vol. 1. Production* (pp. 257–275). New York: Academic Press.

Cooper, W. E., Eady, S. J., & Mueller, P. R. (1985). Acoustical aspects of contrastive stress in question-answer contexts. *The Journal of the Acoustical Society of America, 77,* 2142–2156.

Crary, M. A., & Tallman, V. L. (1993). Production of linguistic prosody by normal and speech-disordered children. *Journal of Communication Disorders, 26,* 245–262.

Cruttenden, A. (1981). Falls and rises: Meanings and universals. *Journal of Linguistics, 17,* 77–91.

Cruttenden, A. (1985). Intonation comprehension in ten-year-olds. *Journal of Child Language, 12*, 643–661.

Crystal, D. (1978). The analysis of intonation in young children. In F. D. Minifie & L. L. Lloyd (Eds.), *Communication and cognitive abilities: Early behavioral assessment* (pp. 257–271). Baltimore: University Park Press.

Crystal, D. (1986). Prosodic development. In P. Fletcher & M. Garman (Eds.), *Language acquisition* (pp. 33–48). Cambridge, England: Cambridge University Press.

Cutler, A. (1984). Stress accents in language production and understanding. In D. Gibbon & H. Richter (Eds.), *Intonation, accent, and rhythm: Studies in discourse phonology* (pp. 77–90). Berlin: deGruyter.

Eady, S. J., & Cooper, W. E. (1986). Speech intonation and focus location in matched statements and question. *The Journal of the Acoustical Society of America, 80,* 402–415.

Fry, D. (1955). Duration and intensity as physical correlates of linguistic stress. *The Journal of the Acoustical Society of America, 22,* 765–768.

Fry, D. (1958). Experiments in the perception of stress. *Language and Speech, 1,* 126–152.

Galligan, R. (1987). Intonation with single words: Purposive and grammatical use. *Journal of Child Language, 14,* 1–21.

Gierut, J. A., & Dinnsen, D. A. (1986). On word-initial voicing: Converging sources of evidence in phonologically disordered speech. *Language and Speech, 29,* 97–114.

Gilbert, H., & Robb, M. (1996). Vocal fundamental frequency characteristics of infant hunger cries: Birth to 12 months. *International Journal of Pediatric Otorhinolaryngology, 34,* 237–243.

Grant, K. W., & Walden, B. E. (1996). Spectral distribution of prosodic information. *Journal of Speech, Language, and Hearing Research, 39,* 228–238.

Grigos, M., & Patel, R. (2007). Articulator movement associated with the development of prosodic control in children. *Journal of Speech, Language, and Hearing Research, 50,* 119–130.

Hadding-Koch, K., & Studdert-Kennedy, M. (1964). An experimental study of some intonation contours. *Phonetica, 11,* 175–185.

Hornby, P. A., & Hass, W. A. (1970). Use of contrastive stress by preschool children. *Journal of Speech and Hearing Research, 13,* 395–399.

Howell, P. (1993). Cue trading in the production and perception of vowel stress. *The Journal of the Acoustical Society of America, 94,* 2063–2072.

Ingram, D. (1976). *Phonological disability in children*. London: Edward Arnold.

Katz, W. F., Beach, C. M., Jenouri, K., & Verma, S. (1996). Duration and fundamental frequency correlates of phrase boundaries in productions by children and adults. *The Journal of the Acoustical Society of America, 99,* 3179–3191.

Kent, R. D., & Bauer, H. R. (1985). Vocalizations of one-year-olds. *Journal of Child Language, 12,* 491–526.

Kent, R. D., & Murray, A. D. (1982). Acoustic features of infant vocalic utterances at 3, 6, and 9 months. *The Journal of the Acoustical Society of America, 72,* 353–365.

Klatt, D. (1976). Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *The Journal of the Acoustical Society of America, 59,* 1208–1221.

Klein, H. B. (1984). Learning to stress: A case study. *Journal of Child Language, 11,* 375–390.

Kochanski, G., Grabe, E., Coleman, J., & Roser, B. (2005). Loudness predicts prominence: Fundamental frequency lends little. *The Journal of the Acoustical Society of America, 118,* 1038–1051.

Lehiste, I. (1970). *Suprasegmentals*. Cambridge, MA: MIT Press.

Lehiste, I. (1976). Suprasegmental features of speech. In N. J. Lass (Ed.), *Contemporary issues in experimental phonetics* (pp. 225–239). New York: Academic Press.

Lieberman, P. (1960). Some acoustic correlates of word stress in American English. *The Journal of the Acoustical Society of America, 32,* 451–454.

Lind, K., & Wermke, K. (2002). Development of the vocal fundamental frequency of spontaneous cries during the first 3 months. *International Journal of Pediatric Otorhinolaryngology, 64*(2), 97–104.

Local, J. (1980). Modeling intonational variability in children's speech. In S. Romaine (Ed.), *Sociolinguistic variation in speech communities* (pp. 85–103). London: Edward Arnold.

Loeb, A. D., & Allen, G. D. (1993). Preschoolers' imitation of intonation contours. *Journal of Speech and Hearing Research, 36,* 4–13.

Lyberg, B. (1979). Final lengthening—Partly a consequence of restrictions on the speed of fundamental frequency change? *Journal of Phonetics, 7,* 187–196.

Macken, M. A., & Barton, D. (1980). The acquisition of the voicing contrast in Spanish: A phonetic and phonological study of word-initial stop consonants. *Journal of Child Language, 7,* 433–458.

MacNeilage, P. F., & Davis, B. L. (1993). Acquisition of speech production: Frames then content. In M. Jeannerod (Ed.), *Attention and performance XIII: Motor representation and control* (pp. 453–475). Hillsdale, NJ: Erlbaum.

Majewski, W., & Blasdell, R. (1969). Influence of fundamental frequency cues on the perception of some synthetic intonation contours. *The Journal of the Acoustical Society of America, 45,* 450–457.

McClean, M., & Tiffany, W. (1973). The acoustic parameters of stress in relation to syllable position, speech loudness, and rate. *Language and Speech, 16,* 283–290.

Menyuk, P., & Bernholtz, N. (1969). Prosodic features and children's language production. *MIT Quarterly Progress Report, 93,* 216–219.

Morton, J., & Jassem, W. (1965). Acoustic correlates of stress. *Language and Speech, 8,* 159–181.

Netsell, R. (1973). Speech physiology. In F. Minifie, T. J. Hixon, & F. Williams (Eds.), *Normal aspects of speech, hearing, and language* (pp. 211–234). Englewood Cliffs, NJ: Prentice Hall.

O'Shaughnessy, D. (1979). Linguistic features in fundamental frequency pattern. *Journal of Phonetics, 7,* 119–145.

Patel, R., & Grigos, M. (2006). Acoustics of question-statement contrast in 4, 7, and 11 year-old children. *Speech Communication, 48,* 1308–1318.

Peppe, S., Maxim, J., & Wells, B. (2000). Prosodic variations in Southern British English. *Language and Speech, 43,* 309–334.

Pierrehumbert, J. (1979). *The phonology and phonetics of English intonation*. Unpublished doctoral dissertation, Massachusetts Institute of Technology.

Pollock, K. E., Brammer, D. M., & Hageman, C. F. (1993). An acoustic analysis of young children's productions of word stress. *Journal of Phonetics, 21,* 183–203.

Protopapas, A., & Eimas, P. D. (1997). Perceptual differences in infant cries revealed by modifications of acoustic features. *The Journal of the Acoustical Society of America, 102,* 3723–3734.

Renfrew, C. E. (1966). Persistence of the open syllable in defective articulation. *Journal of Speech and Hearing Disorders, 31,* 370–373.

Shattuck-Hufnagel, S., & Turk, A. E. (1996). A prosody tutorial for investigators of auditory sentence processing. *Journal of Psycholinguistic Research, 25,* 193–247.

Smith, N. V. (1973). *The acquisition of phonology: A case study.* Cambridge, England: Cambridge University Press.

Snow, D. (1994). Phrase-final syllable lengthening and intonation in early child speech. *Journal of Speech and Hearing Research, 37,* 831–840.

Snow, D. (1995). Formal regularity of the falling tone in children's early meaningful speech. *Journal of Phonetics, 23,* 387–405.

Snow, D. (1998). Children's imitations of intonation contours: Are rising tones more difficult than falling tones? *Journal of Speech, Language, and Hearing Research, 41,* 576–587.

Stathopoulos, E. T., & Sapienza, C. M. (1997). Developmental changes in laryngeal and respiratory function with variations in sound pressure level. *Journal of Speech, Language, and Hearing Research, 40,* 595–614.

Tingley, B. M., & Allen, G. D. (1975). Development of speech timing control in children. *Child Development, 46,* 186–194.

Uldall, E. T. (1961, September). Ambiguity: Question or statement? Or "Are you asking me or telling me?" In *Proceedings of the Fourth International Congress of Phonetic Sciences* (pp. 779–783; Helsinki, Finland). The Hague, the Netherlands: Mouton.

van der Meulen, S., Janssen, P., & den Os, E. D. (1997). Prosodic abilities in children with specific language impairment. *Journal of Communication Disorders, 30,* 155–170.

Weismer, G., Dinnsen, D., & Elbert, M. (1981). A study of the voicing distinction associated with omitted, word-final stops. *Journal of Speech and Hearing Disorders, 46,* 320–328.

Wells, B., Peppe, S., & Goulandris, N. (2004). Intonation development from five to thirteen. *Journal of Child Language, 31,* 749–748.

Wermke, K., Mende, W., Manfredi, C., & Bruscaglioni, P. (2002). Developmental aspects of infant's cry melody and formants. *Medical Engineering & Physics, 24*(7–8), 501–514.

Whalen, D. H., Levitt, A. G., & Wang, Q. (1991). Intonational differences between the reduplicative babbling of French- and English-learning infants. *Journal of Child Language, 18,* 501–516.

Xu, Y., & Sun, X. (2002). Maximum speech of pitch change and how it may relate to speech. *The Journal of the Acoustical Society of America, 111,* 1399–1413.

Zemlin, W. R. (1998). *Speech and hearing science: Anatomy and physiology* (4th ed.). Boston: Allyn & Bacon.

Received June 21, 2007

Revision received March 10, 2008

Accepted September 23, 2008

DOI: 10.1044/1092-4388(2008/07-0137)

Contact author: Rupal Patel, Department of Speech Language Pathology and Audiology, Northeastern University, 106 Forsyth Building, Boston, MA 02115. E-mail: r.patel@neu.edu.

Appendix. Elicitation prompts.

Question–Statement Contrast

Target	Prompt
Show Bob a bot.	Bob is feeling lonely. He wants to play with someone. *What do you think he's like to play with? What should we show Bob?*
Show Bob a bot?	Bob is hungry. He wants to make soup. *What should we show Bob?* Experimenter 2 interjects: Show Bob a bot. Experimenter 1: Can Bob make soup in a bot? Ask Maria what she is going to do. Maybe we didn't hear her correctly.
Show Pop a pot.	Pop is hungry. He wants to make soup but he can't find anything in the kitchen. *What do you think he would need? What should we show Pop?*
Show Pop a pot?	Pop is tired. He needs help cleaning the house. *What should we show Pop?* Experimenter 2 interjects: Show Pop a pot. Experimenter 1: Can Pop clean the house with a pot? Ask Maria what she is going to do. Maybe we didn't hear her correctly.

Contrastive Stress
Scenario: Bob is home sick from school. His grandpa, Pop is taking care of him.

Target	Prompt
Show BOB a bot.	Bob wants to play but Pop is busy making lunch. *Who should I show the bot to? Tell me what to do.*
Show Bob a BOT.	Bob wants to play but Pop is busy making lunch. *What should I show Bob? Tell me what to do.*
Show POP a pot.	Pop wants to make soup for lunch but he can't find anything in the kitchen. *Who should I show the pot to? Tell me what to do.*
Show Pop a POT.	Pop wants to make soup for lunch but he can't find anything in the kitchen. *What should I show Pop? Tell me what to do.*

Note. All caps are used to denote contrastive stress.