Available online at www.sciencedirect.com
ScienceDirect
Speech Communication 48 (2006) 1308–1318
SPEECH COMMUNICATION
www.elsevier.com/locate/specom

Acoustic characterization of the question–statement contrast in 4, 7 and 11 year-old children

Rupal Patel [a,*], Maria I. Grigos [b]

[a] Northeastern University, Department of Speech Language Pathology and Audiology, 360 Huntington Avenue, Room 102 FR, Boston, MA 02115, USA
[b] New York University, Department of Speech Language Pathology and Audiology, Steinhardt School of Education, 719 Broadway, Suite 200, New York, NY 10003, USA

Received 12 December 2005; received in revised form 16 June 2006; accepted 22 June 2006

Abstract

Prosodic features of the speech signal include fundamental frequency (F0), intensity and duration. In order to study the development of prosody independent from segmental aspects of speech, we considered the question–statement contrast. In English, adults mark the contrast using changes in fundamental frequency, duration and intensity, with F0 being the most prominent cue. Declarative questions are marked by rising intonation whereas statements are marked by falling intonation. While previous studies have noted that young children can signal this contrast in imitative paradigms, little is known about the acoustic cues children use at different stages in development. The present study sought to provide an acoustic characterization of prosodic cues used by 12 children from three age groups, 4-year-olds, 7-year-olds and 11-year-olds, for elicited productions of declarative statements and questions. Results indicated that 4-year-olds were unable to reliably signal questions using rising fundamental frequency contour. Instead, they used increased final syllable duration to mark questions. Children in the 7-year-old group used all three cues, fundamental frequency, intensity and syllable duration, to contrast questions from statements. The oldest group relied primarily on changes in fundamental frequency and less so on intensity and duration cues. An age-related pattern is evident in that children employ different combinations of acoustic cues to mark the question–statement contrast across development. The impact of motor and cognitive-linguistic complexity on the development of prosodic control is discussed.

Keywords: Prosody; Children; Acoustics; Speech; Development; Acquisition; Questions; Statements; Intonation

1. Introduction

Prosody, the variations in pitch, stress, and timing of speech, can be quantified in terms of fundamental frequency (F0), intensity and duration (Bolinger, 1989; Lehiste, 1976; Netsell, 1973). These acoustic cues are employed individually or in concert to play various linguistic and communicative roles such as to attract attention, to signal given versus new information, to distinguish kinds of speech acts (e.g., questions versus statements), to contrast the meaning of an utterance, to convey

* Corresponding author. Tel.: +1 617 373 5842; fax: +1 617 373 2239.
E-mail address: r.patel@neu.edu (R. Patel).

0167-6393/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.specom.2006.06.007

affective state, and to achieve many other functions. Prosodic manipulations have been noted in the earliest communicative gestures, such as infant cries (Gilbert and Robb, 1996; Lind and Wermke, 2002; Protopapas and Eimas, 1997; Wermke et al., 2002). Recent findings indicate that people with severely impaired speech can also control prosody despite little or no segmental (speech sound) clarity (Patel, 2002, 2003, 2004; Vance, 1994; Whitehill et al., 2001). While there was a time when prosody was thought of as merely an overlaid signal on top of the "meaningful" segmental units, the interconnections between prosodic cues and speech segments are now widely acknowledged. In fact, over the past few decades, some researchers have suggested that typical development of prosodic control precedes and may facilitate segmental control (Bloom, 1973; Crystal, 1978; Katz et al., 1996; MacNeilage and Davis, 1993; Menyuk and Bernholtz, 1969; Snow, 1994).

The prosody of questions varies across languages and question types (Bolinger, 1989; Geluykens, 1988; Hirst and Di Cristo, 1998; House, 2002; Ladd, 1996). In English, the meaning of declarative statements (e.g., "He plays piano.") can be contrasted from declarative questions (e.g., "He plays piano?") using prosodic cues alone; declarative statements are marked by falling intonation whereas declarative questions are generally[1] marked by rising phrase-final intonation (Cruttenden, 1981; Lieberman, 1967; Hirst and Di Cristo, 1998). In this paper, we focus on children's production of this grammatical (or attitudinal[2]) distinction because it provides a lens into the development of prosody independent from segmental aspects of speech. Previous studies on acquisition of intonation have focused on young children, noting that by age five children can signal question phrase-final intonation in imitative paradigms (Allen and Hawkins, 1980; Loeb and Allen, 1993; Snow, 1994, 1998). It is acknowledged, however, that prosodic control for a variety of linguistic and affective tasks continues to develop beyond age five and into the early teens (Cruttenden, 1985; Crystal, 1986; Local, 1980; Wells et al., 2004) suggesting that the acoustic profile of the question–statement contrast may undergo change across childhood.

Acoustic correlates of phrase-final intonation for declarative questions versus statements in English speaking adults include changes in F0, duration, and intensity, with F0 being the most prominent cue (Cruttenden, 1986; Lieberman, 1967). Lieberman (1967) found that read statements such as "Joe ate the soup" were characterized by a falling terminal F0 whereas the question form "Joe ate the soup?" was marked by rising terminal F0. These changes in F0, however, may not be restricted to the phrase-final syllable. O'Shaughnessy (1979) found that questions were marked by a rising F0 throughout the entire sentence. Recently, Srinivasan and Massaro (2003) noted that statements and questions can be discriminated on the basis of F0 contour, amplitude, and duration cues. Thus, even within the adult literature there appear to be individual differences in the acoustic cues that speakers use in order to mark the question–statement contrast. This transfer of informational cues between prosodic features has been referred to as *cue trading* (Howell, 1993; Lieberman, 1960). Given that the acquisition of motor control of speech happens in parallel with general motor development, cognitive development and language learning, we hypothesized that children at different stages in development may also employ cue trading strategies that rely on different cue combinations to signal prosodic contrasts.

While the acquisition of phrase-final intonation has been studied to some extent, the predominant methodology has been comparative studies across development with imitation protocols being used to elicit contrasts from young children. In terms of falling intonation, Snow (1994) conducted a longitudinal analysis of nine girls between 16 and 25 months of age to examine phrase-final falling of F0 for declarative statements and phrase-final lengthening. Snow's study sought to determine whether phrase-final falling and lengthening resulted from one another or if they were independent. Results indicated that younger children acquired control of intonation earlier than duration. He found that phrase-final lengthening did not appear until the two-word phase (early syntax) and was judged to be a learned behavior.[3] In fact,

[1] While it is commonly assumed that declarative questions have a rising pitch contour, Geluykens (1988) did not find empirical evidence of this pattern in spontaneous speech.

[2] Some argue that this distinction between declarative statements and questions is attitudinal in that the statement conveys certainty while the question form conveys surprised doubt (Cruttenden, 1981).

[3] It should be noted that there are cross linguistic differences in phrase-final lengthening, with English having more extreme lengthening compared to other languages (Oller and Smith, 1977).

Allen and Hawkins (1980) found that final lengthening may be exaggerated by older toddlers. If tone and timing develop asynchronously, might there be differences in which acoustic cues children use to signal declarative statements and questions across different stages in development?

Different motor demands have been associated with the production of rising versus falling F0 contours. In adults, Xu and Sun (2002) found that rising F0 contours impose greater motoric demands in that the speed of change in F0 is slower for rising F0 contours versus falling contours. Loeb and Allen (1993) studied the ability of 3 and 5-year-old children to imitate adult models of declarative (statement), interrogative (question) and monotone intonation contours. They found that 5-year-old children were able to imitate the contours but that 3-year-old children could only partially imitate the contours. The 3-year-old group had difficulty with the rising intonation contour for interrogatives. The authors attributed these age related differences to motoric complexity associated with the production of the rising contour.

Snow (1998) further examined the issue of motor complexity by studying the ability of ten 4-year-olds to imitate rising and falling intonation contours modeled by an adult. Perceptual evidence demonstrated that young children did not imitate rising contours as well as falling contours. He compared the duration and F0 range of imitated utterances, as well as a small sample of spontaneous productions and found that children had more difficulty with rising contours. Rising contours produced by children had a narrower F0 range than adults and tended to have longer word durations. Further, the children had more difficulty raising F0 in the word-final position than the non word-final position. Snow attributed the latter finding to markedness given that the utterance final position was argued to be a marked position along the lines of Lieberman's breath group theory (Lieberman, 1967). Lieberman's theory posits that phrase-final falling is a consequence of the rapid decline of subglottal pressure that coincides with the end of breath groups and thus is unmarked. In contrast, rising contours require muscular (laryngeal and respiratory) effort and are thus are considered to be "marked".

The issue of motoric complexity for rising F0 contours may be related to physiological constraints. Most studies on question–statement prosody to date report on the abilities of children below the age of 7. The requisite changes in respiratory and laryngeal physiology for achieving this acoustic difference, however, are complex and may continue to evolve within the context of a developing speech motor control system. For example, speech timing which is a necessary component of prosody, improves with age from 5 years to 11 years (Tingley and Allen, 1975).

Similarly, respiratory function for varying sound pressure level continues to develop throughout childhood and does not become adult-like until 12–14 years of age (Stathopoulos and Sapienza, 1997). Thus, even 11-year-olds who may be able to raise F0 to mark questions, may have more variability in control of intensity. While previous studies have focused on F0 and duration, we sought to determine whether intensity cues were salient for signaling questions versus statements and if so, which age group made use of these cues.

To summarize, further attention is required to better understand the acoustic features used to signal the question–statement contrast across different stages in development through tasks that go beyond imitation. While many researchers believe that a rising F0 contour is the primary acoustic cue for marking declarative questions (cf. Cruttenden, 1986; Eady and Cooper, 1986; Hadding-Koch and Studdert-Kennedy, 1964; Majewski and Blasdell, 1969; Morton and Jassem, 1965; O'Shaughnessy, 1979; Shattuck-Hufnagel and Turk, 1996), young children may use different acoustic cues and/or cue combinations than older children to signal the contrast in light of their developing motor system. The present study sought to provide an acoustic characterization of the question–statement contrast across different stages of development. It is unclear whether control of F0, intensity and duration develop asynchronously and if so, which cue combinations are used to convey the question–statement contrast across different stages in development. It is hypothesized that while 11-year-olds may have mastered the contrast using predominantly rising F0 to mark questions, 4-year-olds and 7-year-olds may rely on duration cues given that rising F0 may impose greater motoric demands. Furthermore, while 7-year-olds may be able to raise F0, the extent and speed of change in terminal F0 may be reduced in comparison to 11-year-olds.

2. Method

2.1. Participants

Twelve monolingual English-speaking children, ages 4, 7 and 11 years, participated in this study.

Four subjects were included in each age group (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). Within each age group, there were two male and two female children. These three age groups were selected to provide an initial account of the acoustic differences in question–statement prosody at different stages in development. Children younger than 4 years of age were not included given the cognitive-linguistic demands of the speaking task. Snow's (1998) previous work demonstrates that by 4 years of age children can imitate question–statement pairs to some extent. In order to minimize confounding variables, the oldest group was comprised of 11-year-old boys and girls to limit this group to prepubescent children. All children passed a pure-tone hearing screening presented bilaterally at 25 dB at 500, 1000, 2000 and 4000 Hz. Prior to data collection, parental interviews and informal play sessions were conducted to assess whether the children had adequate speech and language abilities necessary for completing the task. All participants had no reported histories of speech, language, or hearing problems and/or developmental or neurological impairments.

2.2. Materials and apparatus

Speech recordings were collected in an audiometric booth using a MiniDisc recorder (HHB 500 PortaDisc) and a unidirectional head-mounted cardiod dynamic microphone (Shure, SM10A) placed 1 inch from the corner of the child's mouth.

2.3. Procedure

Each child produced multiple repetitions of two phrases spoken as statements and the same two phrases spoken as questions. Bonvillian et al. (1979) have shown that 4-year-old children are better at imitating sentences with six or fewer words compared to longer sentences. Thus, stimuli in this experiment were limited in length to four monosyllabic words each. Target phrases were designed to accommodate kinematic and acoustic analyses. Target words in the first phrase began with /b/ ("Show Bob a bot") and in the second phrase began with /p/ ("Show Pop a pot"). Each phrase began with the word "show" for consistency. Syllable structure and vowel height were controlled for across the target phrases. Voiced and unvoiced cognates of the bilabial plosive were selected to explore interactions between laryngeal and respiratory subsystems in prosodic control.

A naturalistic discourse elicitation technique was used to collect spoken utterances (see Allen and Arndorfer, 2000 for a similar technique used to elicit question–statement tokens in children with hearing impairment). Specifically, each child was introduced to popsicle stick puppets consisting of colored drawings[4] of the four target words. "Pop" was depicted as a grandfather, "pot" was a cartoon drawing of a black pot, "Bob" referred to the character SpongeBob SquarePants, and "bot" was the name of a robot. The child's task was to instruct the experimenter to perform an action using one character (e.g., "Bob" or "Pop") and one object (e.g., "bot" or "pot"). A short contextual scenario was provided to elicit the appropriate phrase and tone. For example, to elicit "Show Bob a bot", the child was told that Bob was lonely and he wanted something to play with. The experimenter would ask "What should I show Bob?" The child would respond "Show Bob a bot". If the child responded with just the object name (e.g., "bot"), the experimenter would encourage the production of a complete sentence. The scenario used to elicit the question form "Show Bob a bot?" was that Bob was hungry and he needed something for making soup. One of the investigators would act as though they were helping the child by volunteering the response "Show Bob a bot". The other investigator would then look puzzled and ask the child "Does that make sense? Ask her what she wants to do". The child would then ask "Show Bob a bot?" In many instances, especially for children in the 4-year-old group, additional cues and sometimes models[5] were required to elicit the question form. In instances where the question form was modeled by one of the investigators, the contextual scenario was repeated in order to reduce the effects of imitation. Only those tokens produced without direct models were analyzed. While this protocol does not yield truly volitional productions,

[4] Pictures were used to account for differences in reading skills among age groups, to make the task more engaging, and to better approximate discourse level interactions.

[5] 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 a model were not included in the analysis, the 4-year-olds produced more repetitions than the other age groups. In contrast, only 3 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.

we felt that elicited utterances may better approximate the child's prosodic control abilities compared to directly imitated productions used in most previous studies of child prosody.

To control for order effects, phrase type (the "Bob" vs. "Pop" phrase) and sentence type (question vs. statement) orders were randomized across speakers. At least 15 repetitions of each phrase produced as a statement and 15 repetitions of each phrase produced as a question were requested. Of these tokens, 10 statement productions and 10 question productions of each phrase were acoustically analyzed. Selection of these tokens was based on the judgments of two independent research assistants who heard each production and selected tokens that were free of acoustic errors (e.g., F0 tracking errors, interjections of noise, multiple talkers) and produced without imitation.

2.4. Acoustic analyses

In total, 480 utterances (40 productions * 12 speakers) were acoustically analyzed. All utterances were sampled at 22,050 Hz. The Praat speech analysis software package (Boersma and Weenink, 2004) was used to mark the beginning and end of each syllable within each phrase. Cursor placements were marked by listening to the waveform and using the intensity envelope as a guide. Praat was also used to calculate four acoustic features: syllable duration, average F0 (F0ave), change in F0 within a syllable (F0slope), and average intensity (INTave). Multiple features were collected in order to best account for the ways in which children may be achieving the prosodic contrasts. While previous work has focused on average F0 and its range, examining the change in F0 within a syllable provides insights into the shape of the underlying F0 contour and allows for comparisons of the speed of change in terminal F0 between questions and statements. Along the lines of Lieberman (1960) and Howell (1993), we sought to understand whether trading relationships existed in children's productions. In other words, were children in one age group relying on one of the three prosodic cues more than another and if so, how did the cues and cue combinations used to mark questions, change over the course of development?

2.4.1. Syllable duration

The duration of each syllable within each utterance was calculated using a customized program that used the Praat generated labels for demarking the beginning and end of each syllable. Four duration measures were calculated for each phrase: S1_dur, S2_dur, S3_dur, and S4_dur, the duration of syllable 1, 2, 3, and 4, respectively.

2.4.2. Fundamental frequency

For each syllable within each phrase, the Praat system was used to estimate F0 values (Hz) at 10 ms intervals within the utterance. Manual correction of the automatically generated F0 values was required on 16 of the 480 utterances. A majority of these F0 tracking errors occurred for samples produced by children in the 4-year-old group and one of the 11-year-old female speakers. Pitch-tracking errors consisted of octave jumps and periods of devoicing that could not be verified auditorily. Manually adjusting the upper and lower F0 limits, and frame duration parameters in Praat typically led to improvements in F0 tracking. These new F0 values were verified through visual and auditory inspection and confirmation using direct calculation of the pitch period from the waveform. In total, eight F0 measures were calculated for each phrase; (a) S1_F0ave, S2_F0ave, S3_F0ave, and S4_F0ave; the average of F0 values within each syllable, and (b) S1_F0slope, S2_F0slope, S3_F0slope and S4_F0slope; the rate of change in F0 within each syllable.

2.4.3. Intensity

The Praat system generated relative intensity values (dB) across the duration of each syllable within each phrase. Four measures were calculated for each phrase (S1_INTave, S2_INTave, S3_INTave, and S4_INTave), corresponding to the average of intensity values within each syllable.

2.4.3.1. Reliability of acoustic measures. To evaluate intra-labeler reliability, 10% of each child's utterances were randomly selected and relabeled. Intra-labeler reliability of syllable duration measures across these two points in time was $r = 0.986$ (mean difference across the two measurement points = 0.004 s, SD across the measurement points = 0.012 s). All intensity and F0 values were recomputed based on these new duration labels. The mean difference between the first and second measurements in intensity and F0 were minimal (3.8 Hz (SD = 2.6 Hz) for F0ave, −42.8 Hz/s (SD = 5.6 Hz) for F0slope; and 1.3 dB (SD = 0.7 dB) for INTave).

3. Results

Acoustic analyses were conducted on all 240 statement productions and 240 question productions. Separate repeated measures analyses of variance were conducted for each of the dependent measures (F0ave, F0slope, INTave and duration). In each analysis, the effect of one between-subject factor (group; 4 year, 7 year, vs. 11 year old) and three within-subject factors were examined. Factor 1 represented the sentence type and had two levels (question; statement). Factor 2 represented phrase type and had two levels (/b/; /p/ phrase). Factor 3 represented syllables and had four levels (syllable 1 (S1), syllable 2 (S2), syllable 3 (S3), and syllable 4 (S4)). The response variables were all continuous; fundamental frequency in Hz, intensity in dB, and duration in seconds. The *F* statistic was used to test the null hypothesis with $\alpha = 0.05$. Considering each acoustic parameter separately, up to 20 pair-wise contrasts were conducted to examine meaningful differences across syllables, sentence type, phrase type and age group. To account for multiple comparisons in these post-hoc tests, a Bonferroni correction factor was applied and only values of $p < 0.0025$ were considered to be statistically significant.

3.1. Syllable duration

A statistically significant main effect in duration was found for syllables ($F = 76.66$; df = 3, 21; $p < 0.0001$) (Fig. 1). Two-way interactions between syllable by phrase type ($F = 5.86$; df = 3, 21; $p = 0.0045$), syllable by sentence type ($F = 19.42$; df = 3, 21; $p < 0.0001$), and phrase type by sentence type ($F = 5.62$; df = 1, 7; $p = 0.049$) were also significant. All three age groups elongated the final syllable for questions compared to statements. These differences in final syllable duration were more pronounced for the /p/ phrase than the /b/ phrase.

3.2. Fundamental frequency

Two F0 measures were examined for each syllable within each phrase: F0ave and F0slope. In each analysis, gender was initially included as a second between-subjects factor to account for inherent differences in F0 among female and male children. While these differences in F0 may not be present in young children, they may have been evident in the 11-year old group. There were, in fact, no statistically significant main effects or interaction effects of gender for any age group in F0ave or F0slope.

Statistically significant main effects in F0ave were found for syllable ($F = 26.36$; df = 3, 21; $p < 0.0001$), phrase type ($F = 5.99$; df = 1, 7; $p = 0.0442$), and sentence type ($F = 7.34$; df = 1, 7; $p = 0.0302$) (Fig. 2). Interactions between syllable by age group ($F = 6.41$; df = 6, 21; $p = 0.0006$), syllable by sentence type ($F = 69.46$; df = 3, 21; $p < 0.0001$) and syllable by sentence type by age group ($F = 2.74$; df = 6, 21; $p = 0.0401$) were all statistically significant. While the relative patterning of F0ave used by 4-year-olds was similar to that used by 7 and 11-year-old children for making statements, the change in F0ave from S3 to S4 for questions was considerably lower than that employed by the older children. In fact, for the /b/ phrase, 4-year-olds were unable to raise F0ave for the question form. These results suggest that 4-year-olds were having more difficulty with rising intonation than

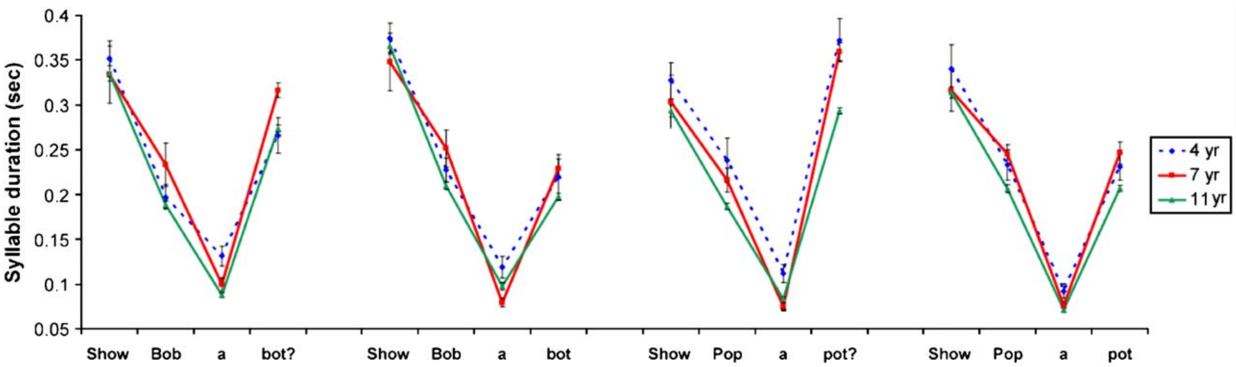

Fig. 1. Mean syllable duration (s) for each syllable of statement and question tokens produced by 4, 7 and 11-year-olds for "Show Bob a bot" and "Show Pop a pop".

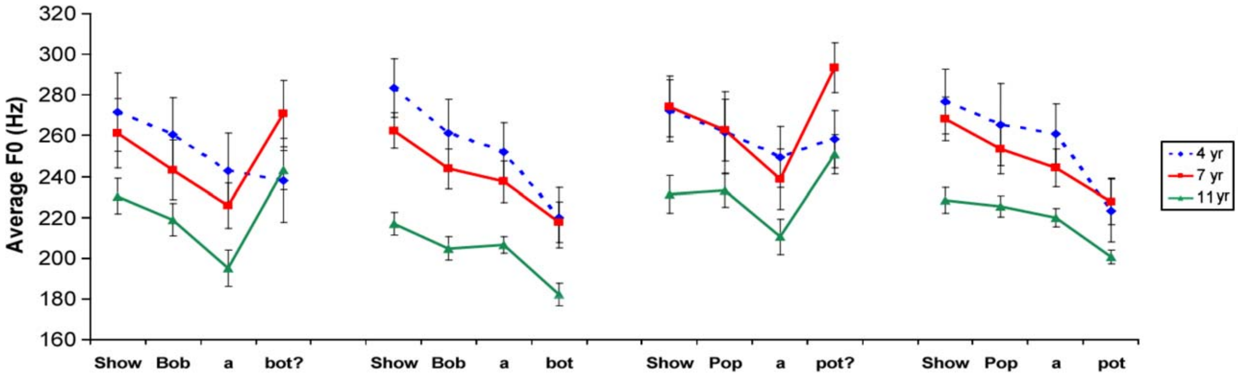

Fig. 2. Mean of the average fundamental frequency (Hz) for each syllable of statement and question tokens produced by 4, 7 and 11-year-olds for "Show Bob a bot" and "Show Pop a pop".

falling intonation. It is interesting to note that 4-year-olds dropped F0ave for the final syllable of statement to a greater extent than 7-year-olds and 11-year-olds. Both the 7-year-old and 11-year-old groups signaled questions using rising F0ave on the final syllable.

Statistically significant main effects for F0slope were found for sentence type ($F = 11.37$; df = 1, 7; $p = 0.0150$) (Fig. 3). Interactions between syllable by age group ($F = 4.67$; df = 6, 18; $p = 0.005$), and syllable by sentence type ($F = 21.59$; df = 3, 18; $p < 0.0001$) were also significant. For all three age groups, within syllable changes in F0 were most salient for the final syllable. While 7 and 11-year-old children marked questions with a sharply rising F0 slope on the final syllable, 4-year-old children were unable to achieve this rapid change in F0 for the phrase-final position. It should be noted, however, that the 4-year-olds were slightly better able to raise F0 for the final syllable of the /p/ phrase compared to the /b/ phrase. Also noteworthy is the fact that statements were marked by a more steeply falling F0slope on the final syllable for 4-year-olds compared to 7 and 11-year-olds.

3.3. Intensity

A statistically significant main effect in INTave was also only found for syllable ($F = 30.87$; df = 3, 21; $p < 0.005$). Only the two-way interactions between syllable and age group ($F = 4.96$; df = 6, 21; $p = 0.0026$), syllable and phrase type ($F = 5.09$; df = 3, 21; $p = 0.0084$), and syllable and sentence type ($F = 23.36$; df = 3, 21; $p < 0.0001$) were significant. Table 1 shows mean INTave values by syllable for each phrase by age group. While questions were marked by a slightly higher INTave than statements, the difference was only statistically significant for the 7-year-old group. The difference in mean INTave for questions versus statements

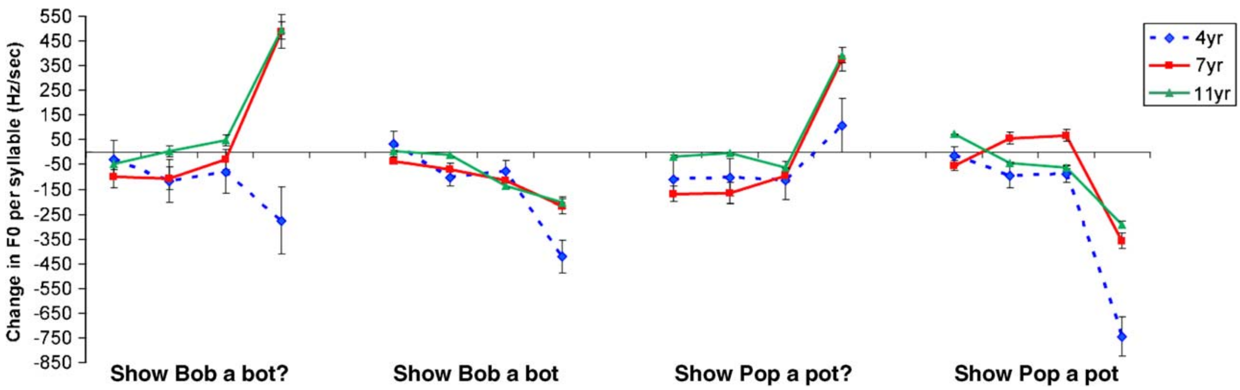

Fig. 3. Mean change in fundamental frequency within each syllable (Hz/s) of statement and question tokens produced by 4, 7 and 11-year-olds for "Show Bob a bot" and "Show Pop a pop".

Table 1
Mean and standard deviation of INTave in statement and question forms of "Show Bob a bot" and "Show Pop a pot" by age group.

		INTave (dB)			INTave (dB)	
		Statement	Question		Statement	Question
4 year olds	Show	56.7 (3.8)	56.1 (4.4)	Show	55.6 (4.1)	54.7 (4.3)
	Bob	63.1 (4.0)	63.5 (6.1)	Pop	58.1 (4.0)	57.7 (4.5)
	a			a		
	bot	57.7 (3.5)	59.2 (4.1)	pot	53.5 (4.1)	56.8 (4.6)
7 year olds	Show	48.7 (5.1)	48.7 (5.1)	Show	47.6 (5.1)	47.2 (6.2)
	Bob	55.7 (5.4)	56.2 (5.9)	Pop	52.1 (5.3)	53.7 (5.2)
	a			a		
	bot	50.7 (5.2)*	63.4 (6.6)*	pot	49.3 (4.9)*	62.0 (6.7)*
11 year olds	Show	62.4 (2.7)	62.9 (3.3)	Show	61.9 (3.0)	60.7 (3.5)
	Bob	66.9 (3.7)	67.2 (2.8)	Pop	63.4 (3.6)	64.9 (3.9)
	a			a		
	bot	62.1 (3.0)	64.2 (2.9)	pot	60.8 (2.7)	63.9 (3.7)

was highest in the 7-year-old group; however, this group also had the highest standard deviation in INTave.

4. Discussion

Acoustic analyses of question and statement productions of children aged 4, 7 and 11 years highlight the motoric and cognitive-linguistic complexity of prosodic control. While adults contrast questions from statements using F0, intensity, and duration, rising F0 is thought to be the most prominent acoustic cue for signaling declarative questions (Cruttenden, 1986; Lieberman, 1967). Results of the present study suggest that between the ages of 7 and 11, children begin to use adult-like patterns of prosodic cues to signal the question–statement contrast. In fact, 7 years olds in this study used more exaggerated differences in F0, intensity and duration between question and statement tokens than the 11-year-olds.

With regard to the issue of which acoustic cues children use over the developmental course to mark questions, the present data suggest that young children rely on syllable duration to contrast questions from statements, while older children additionally rely on F0 to mark the contrast. Similar to previous studies (Loeb and Allen, 1993; Snow, 1994, 1998), the 4-year-old group had difficulty with phrase-final rising of F0. Instead, children in the 4-year-old group employed final syllable lengthening to mark questions (similar to that noted by Allen and Hawkins, 1980). Although one may argue that the 4-year-olds' performance may have been affected by their understanding of the task, these children consistently modified duration to signal the contrast. Thus, the 4-year-olds appear to rely on temporal features to mark the prosodic contrast. It should also be noted that despite being unable to raise F0 for questions, 4-year-olds were able to lower F0 for statements. The 4-year-olds may have been attempting to optimize the contrast by exploiting their limited F0 range. They also did so with a much steeper slope than the older children. This finding is consistent with Xu and Sun's (2002) conclusions that rising F0 is more motorically complex and associated with a slower change in F0 than falling F0. Collectively, these findings suggest that the 4-year-olds were attempting to mark the question–statement contrast even though their developing speech motor control and/or their maturing cognitive-linguistic systems had not yet mastered the use of F0 and intensity cues for this purpose. Future extensions of this work may explore whether listeners can make use of the acoustic information in the 4-year-olds productions to accurately distinguish among questions and statements.

The 7-year-old group appeared to be using all three prosodic cues, namely, F0, intensity and duration to convey the difference between questions and statements. The extent and speed of change in terminal F0 was comparable between the 7-year-olds and 11-year-olds. Although 7-year-olds marked the contrast using greater differences in average intensity compared to 4-year-olds and 11-year-olds, they also exhibited more variability suggesting that physiological control of intensity continues to be refined beyond age 7 (Stathopoulos and Sapienza, 1997). In contrast, the 11-year-old group relied primarily on manipulations of F0 and less so on intensity and duration changes. This pattern of relying on F0 cues to mark the contrast is similar to that noted in adults (Crutenden, 1986; Lieberman, 1967; O'Shaughnessy, 1979). Perhaps the laryngeal and respiratory control and coordination available to 11-year-olds enabled them to rely on F0 manipulations to signal questions. On the other hand, the 7-year-olds altered F0, intensity, and duration, to ensure that a question was signaled. It is unclear whether the F0 cues within the 7-year-old productions may have been sufficient to convey the contrast to listeners.

We found that rising intonation was difficult for 4-year-olds but mastered by the 7-year-old and 11-year-old groups. As expected, the variability in F0 control decreased considerably with age. These patterns are consistent with previous studies that have shown that children continue to refine acoustic parameters until adolescence (Cruttenden, 1985; Crystal, 1986; Eguchi and Hirsch, 1969; Kent, 1976; Kent and Forner, 1980; Koenig, 2000; Smith and Kenney, 1998; Tingley and Allen, 1975; Wells et al., 2004). While an adult control group was not used in the present study, it may be fruitful in future work to compare and contrast the productions of older children and adults.

An unanticipated finding relates to differences in prosodic control among the 4-year-old group for the voiceless plosive phrase "Show Pop a pot" versus the phrase with voiced plosive targets "Show Bob a bot". The former yielded more acoustically contrastive productions with respect to F0 control compared to the latter. One possible explanation may be related to the breath control and laryngeal constriction necessary for raising F0 and intensity during voiced plosive production. This phrase requires vocal fold adduction throughout most of utterance while maintaining sufficient subglottal pressure and simultaneously increasing tension on the folds at the end of the utterance to raise F0 for questions. Alternatively, a cognitive-linguistic explanation may be related to word versus non-word differences. Perhaps "Show Pop a pot?" was linguistically simpler or more familiar for the 4-year-old group than "Show Bob a bot?" given that "pot" is a word and "bot" is a non-word. It should be noted that although "bot" is a non-word, children were introduced to a robot character called "bot" and thus this explanation does not seem sufficiently satisfying. These phoneme specific differences were found only for the 4-year-old group. While further investigation with a larger sample of children and target phrases is required to shed light on this issue, our data highlight the complex interconnections between prosodic and segmental units which cannot easily be teased apart.

With regard to insights into the motor control of prosody, an interesting parallel can be drawn between the productions of the 4-year-olds in this study and previous findings on adults with severe speech impairment (Patel, 2003; Vance, 1994). Despite difficulties in raising F0 for questions, both groups signaled the contrast with other acoustic cues that were easier for them to control. Questions were marked by increased final syllable duration and statements with sharply falling F0 contour. These compensations are predicted by previous work which suggests that falling contours impose fewer motor demands than rising contours and thus may be mastered earlier (Snow, 1998; Xu and Sun, 2002). It is interesting how both of these motor systems, the maturing and the disordered, arrive at similar cue combination strategies. Further research is required to inform our understanding of the interactions between physiologic constraints and the acoustic manifestation of prosodic cues.

While the findings reported herein are limited in that they provide an acoustic characterization of only 12 children, these initial insights may spark further inquiry into the acoustic realization of prosodic contrasts throughout development. The present findings suggest that control of prosodic cues, primarily F0, may not be mastered until at least 7 years of age (see Wells et al., 2004 for similar findings). Moreover, the patterns used by 7-year-olds continue to be refined with increasing age so as to produce contrastive yet natural prosodic contrasts. Larger scale investigations with a greater variety of prosodic tasks and a larger sample size are necessary to generalize these initial findings. While imitation of prosodic contrasts may be mastered earlier, volitional control may require more motor practice and further maturation of the child's linguistic and cognitive abilities. Furthermore, the profile of acoustic cues used to mark prosodic contrasts such as question versus statement appears to evolve with increasing motor skill and maturation.

5. Conclusions

This study sought to provide initial insights into the acoustic characterization of the question–statement contrast in 12 children aged 4, 7, and 11 years. The results indicate that the 4-year-olds were unable to reliably signal questions using rising fundamental frequency contour. Instead, they used increased final syllable duration to mark questions. In contrast, the 7-year-olds used a combination of fundamental frequency, intensity and syllable duration to mark the contrast. Productions of 7-year-old children were exaggerated compared to the 11-year-old group. The oldest group relied primarily on changes in fundamental frequency and less so on intensity and duration cues. These findings suggest that even a simple linguistic contrast such as question versus statement requires considerable motor coordination of prosodic features and access to linguistic and cognitive resources that may not be mastered and adult-like until at least 7 years of age. These findings provide developmental data that address prosodic development alongside acquisition of segmental control.

Acknowledgements

This research was conducted in the Department of Biobehavioral Sciences at Teachers College Columbia University and the Department of Speech Language Pathology and Audiology at Northeastern University. The authors are grateful to the children and their families for their time in participating. The authors would also like to thank Mariam Syeda and Nicole Seaman for labeling the acoustic data, Deepak Chakravadhanula and Elyes Yaich for developing various programs and interfaces that facilitated data labeling and analysis, Howard Cabral for assistance with the statistical analysis, Kenneth Stevens for input on interpretation of the results, as well as Kris Tjaden and Christopher Dromey for comments on the manuscript. This work is supported in part by NIH Grant No. DC-06118 from the National Institute of Deafness and other Communication Disorders.

References

Allen, G.D., Arndorfer, P.M., 2000. Production of sentence-final intonation contours by hearing-impaired children. J. Speech, Language, Hearing Res. 43, 441–455.

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

Bloom, L., 1973. One Word at a Time. Mouton, The Hague.

Boersma, P., Weenink, D., 2004. Praat, a system for doing phonetics by computer, version 4.231. Technical Report 132, Institute of Phonetic Sciences of the University of Amsterdam. Available from: <www.praat.org>.

Bolinger, D., 1989. Intonation and its Uses: Melody in Grammar and Discourse. Stanford University Press, Stanford.

Bonvillian, J.D., Raeburn, V.P., Horan, E.A., 1979. Talking to children: the effects of rate, intonation, and length on children's sentence imitation. J. Child Language 6 (3), 459–467.

Cruttenden, A., 1981. Falls and rises: meanings and universals. J. Linguist. 17, 77–91.

Cruttenden, A., 1985. Intonation comprehension in ten-year-olds. J. Child Language 12, 643–661.

Cruttenden, A., 1986. Intonation. Cambridge University Press, Cambridge.

Crystal, D., 1978. The analysis of intonation in young children. In: Minifie, F.D., Lloyd, L.L. (Eds.), Communication and Cognitive Abilities – Early Behavioral Assessment. University Park Press, Baltimore, pp. 257–271.

Crystal, D., 1986. Prosodic development. In: Fletcher, P., Garman, M. (Eds.), Language Acquisition: Studies in First Language Development, second ed. Cambridge University Press, Cambridge.

Eady, S.J., Cooper, W.E., 1986. Speech intonation and focus location in matched statements and questions. J. Acoust. Soc. Amer. 80, 402–415.

Eguchi, S., Hirsch, I.J., 1969. Development of speech sounds in children. Acto Oto-Laryngologica (Suppl. 257), 1–51.

Geluykens, R., 1988. On the Myth of Rising Intonation in Polar Questions. J. Pragmat. 12, 467–485.

Gilbert, H., Robb, M., 1996. Vocal Fundamental Frequency Characteristics of Infant Hunger Cries: Birth to 12 Months. Internat. J. Pediatric Otorhinolaryngol. 34, 237–243.

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

Hirst, D., Di Cristo, A., 1998. A survey of intonation systems. In: Hirst, D., Di Cristo, A. (Eds.), Intonation Systems. Cambridge University Press, Cambridge, pp. 1–45.

House, D., 2002. Intonational and visual cues in the perception of interrogative mode in Swedish. In: Proc. Internat. Conf. on Spoken Language Processing, 2002, Denver, Colorado, 1957–1960.

Howell, P., 1993. Cue trading in the production and perception of vowel stress. J. Acoust. Soc. Amer. 94 (4), 2063–2073.

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. J. Acoust. Soc. Amer. 99, 3179–3191.

Kent, R.D., 1976. Anatomical and neuromuscular maturation of the speech mechanism: evidence from acoustic studies. J. Speech Hearing Res. 19, 421–447.

Kent, R.D., Forner, L.L., 1980. Speech segment durations in sentence recitations between children and adults. J. Phonetics 8, 157–168.

Koenig, L.L., 2000. Laryngeal factors in voiceless consonant production in men, women and 5-year-olds. J. Speech, Language Hearing Res. 43 (5), 1211–1228.

Ladd, D.R., 1996. Intonation Phonology. Cambridge University Press, Cambridge.

Lehiste, I., 1976. Suprasegmental features of speech. In: Lass, N.J. (Ed.), Contemporary Issues in Experimental Phonetics. Academic Press, New York, pp. 225–239.

Lieberman, P., 1960. Some acoustic correlates of word stress in American English. J. Acoust. Soc. Amer. 32, 451–454.

Lieberman, P., 1967. Intonation, Perception, and Language. MIT Press, Cambridge.

Lind, K., Wermke, K., 2002. Development of the vocal fundamental frequency of spontaneous cries during the first 3 months. Internat. J. Pediatric Otorhinolaryngol. 64 (2), 97–104.
Local, J., 1980. Modeling intonational variability in children's speech. In: Romaine, S. (Ed.), Sociolinguistic Variation in Speech Communities. Edward Arnold, London.
Loeb, D.F., Allen, G.D., 1993. Preschoolers' imitation of intonation contours. J. Speech Hearing Res. 36, 4–13.
MacNeilage, P.F., Davis, B.L., 1993. Acquisition of speech production: frames then content. In: Jeannerod, M. (Ed.), Attention and Performance XIII: Motor Representation and Control. Lawrence Erlbaum, Hillsdale, New Jersey, pp. 453–475.
Majewski, W., Blasdell, R., 1969. Influence of fundamental frequency cues on the perception of some synthetic intonation contours. J. Acoust. Soc. Amer. 45 (2), 450–457.
Menyuk, P., Bernholtz, N., 1969. Prosodic features and children's language production. MIT Quart. Progress Rep. 93, 216–219.
Morton, J., Jassem, W., 1965. Acoustic correlates of stress. Language Speech 8, 159–181.
Netsell, R., 1973. Speech Physiology. In: Minifie, F., Hixon, T.J., Williams, F. (Eds.), Normal Aspects of Speech, Hearing, and Language. Prentice-Hall, Englewood Cliffs, New Jersey, pp. 211–234.
Oller, D.K., Smith, B.L., 1977. Effect of syllable-final position on vowel duration in infant babbling. J. Acoust. Soc. Amer. 62 (4), 994–997.
O'Shaughnessy, D., 1979. Linguistic features in fundamental frequency pattern. J. Phonetics 7, 119–145.
Patel, R., 2002. Prosodic control in severe dysarthria: preserved ability to mark the question–statement contrast. J. Speech, Language Hearing Res. 45, 858–870.
Patel, R., 2003. Acoustic differences in the yes–no question–statement contrast between speakers with and without dysarthria. J. Speech, Language Hearing Res. 46, 1401–1415.
Patel, R., 2004. Contrastive Prosody in adults with cerebral palsy. J. Med. Speech Pathol. 12 (4), 189–193.
Protopapas, A., Eimas, P.D., 1997. Perceptual differences in infant cries revealed by modifications of acoustic features. J. Acoust. Soc. Amer. 102 (6), 3723–3734.
Shattuck-Hufnagel, S., Turk, A.E., 1996. A prosody tutorial for investigators of auditory sentence processing. J. Psycholinguist. Res. 25 (2), 193–247.
Smith, B.L., Kenney, M.K., 1998. An assessment of several acoustic parameters in children's speech production development: longitudinal data. J. Phonetics 26, 95–108.
Snow, D., 1994. Phrase-final syllable lengthening and intonation in early child speech. J. Speech Hearing Res. 37, 831–840.
Snow, D., 1998. Children's imitations of intonation contours: are rising tones more difficult than falling tones? J. Speech Hearing Res. 41, 576–587.
Srinivasan, R.J., Massaro, D.W., 2003. Perceiving prosody from the face and voice: distinguishing statements from echoic questions in English. Language Speech 46 (1), 1–22.
Stathopoulos, E.T., Sapienza, C.M., 1997. Developmental changes in laryngeal and respiratory function with variations in sound pressure level. J. Speech, Language Hearing Res. 40, 595–614.
Tingley, B.M., Allen, G.D., 1975. Development of speech timing control in children. Child Develop. 46, 186–194.
Vance, J.E., 1994. Prosodic deviation in dysarthria: a case study. Eur. J. Disorders Comm. 29 (1), 61–76.
Wells, B., Peppe, S., Goulandris, N., 2004. Intonation development from five to thirteen. J. Child Language 31, 749–778.
Wermke, K., Mende, W., Manfredi, C., Bruscaglioni, P., 2002. Developmental aspects of infant's cry melody and formants. Med. Eng. Phys. 24 (7–8), 501–514.
Whitehill, T., Ciocca, V., Lam, S., 2001. Fundamental frequency control in connected speech in Cantonese speakers with dysarthria. In: Maassen, B., Hulstijn, W., Kent, R., Peters, H., van Lieshout, P. (Eds.), Speech Motor Control in Normal and Disordered Speech. University of Nijmegen Press, Nijmegen, pp. 228–231.
Xu, Y., Sun, X., 2002. Maximum speech of pitch change and how it may relate to speech. J. Acoust. Soc. Amer. 111, 1399–1413.