J. Child Lang. **43** (2016), 1174–1191.
doi:10.1017/S0305000915000458

BRIEF RESEARCH REPORT

Children's identification of questions from rising terminal pitch

MATHIEU R. SAINDON, SANDRA E. TREHUB*, E. GLENN SCHELLENBERG AND PASCAL VAN LIESHOUT

University of Toronto, Canada

(*Received* 22 *August* 2014 – *Revised* 31 *March* 2015 – *Accepted* 4 *August* 2015 – *First published online 16 September 2015*)

ABSTRACT

Young children are slow to master conventional intonation patterns in their *yes/no* questions, which may stem from imperfect understanding of the links between terminal pitch contours and pragmatic intentions. In Experiment 1, five- to ten-year-old children and adults were required to judge utterances as questions or statements on the basis of intonation alone. Children eight years of age or younger performed above chance levels but less accurately than adult listeners. To ascertain whether the verbal content of utterances interfered with young children's attention to the relevant acoustic cues, low-pass filtered versions of the same utterances were presented to children and adults in Experiment 2. Low-pass filtering reduced performance comparably for all age groups, perhaps because such filtering reduced the salience of critical pitch cues. Young children's difficulty in differentiating declarative questions from statements is not attributable to basic perceptual difficulties but rather to absent or unstable intonation categories.

INTRODUCTION

In contrast to *wh*-questions, which are marked by words such as *what, why, who*, and *how*, and typical *yes/no* questions, which are marked by subject/verb inversion, declarative or echoic questions (*It's snowing?*) are marked exclusively by prosodic cues. The principal cue to declarative questions is a pronounced rise in terminal fundamental frequency (F_o) in contrast to falling F_o for statements (Cruttenden, 1981; Eady & Cooper, 1986;

[*] Address for correspondence: University of Toronto – Psychology, 3359 Mississauga Rd, N Mississauga, Ontario L5L 1C6, Canada. e-mail: sandra.trehub@utoronto.ca

Gårding & Abramson, 1965; Studdert-Kennedy & Hadding, 1973). Secondary cues include increased intensity (Peng, Lu & Chatterjee, 2009) and final-syllable lengthening (Patel & Brayton, 2009; Patel & Grigos, 2006).

The ability to perceive the relevant acoustic distinctions is apparent in infancy. For example, 5-month-olds differentiate the intonation contours of European Portuguese statements from those of *yes/no* questions in the context of single (two-syllable) words (Froda, Butler & Vigario, 2014). English-learning infants 5–24 months of age exhibit greater attention to uninverted *yes/no* questions (i.e. declarative questions) than to statements (Soderstrom, Ko & Nevzorova, 2011), perhaps because of the attention-getting properties of rising terminal pitch (Papoušek, Bornstein, Nuzzo, Papoušek & Symmes, 1990) and the frequent use of prosodic contours in infant-directed speech (Snow, 1977). Evidence of discrimination and differential attention does not imply categorical representations of such acoustic forms. Children must go beyond detecting the differences between rising and falling pitch, reflecting the salience of terminal pitch contours, to categorizing these contours and associating them with questioning or declarative intentions. In fact, young children's productions suggest protracted acquisition of stable intonational categories that map onto specific meanings (Patel & Grigos, 2006; Snow, 1994).

Although preverbal infants produce vocalizations with rising as well as falling pitch contours (Whalen, Levitt & Wang, 1991), young language users are inconsistent in their use of a terminal F_0 rise for questions (Snow, 1994, 1998), and their imitations of declarative, monotone, and interrogative patterns are not clearly differentiated until five years of age (Loeb & Allen, 1993). Moreover, when declarative questions are elicited from four-year-olds, the utterances are often marked by final-syllable lengthening rather than F_0 changes (Patel & Grigos, 2006). Unfortunately, little is known about children's spontaneous use of declarative questions or their understanding of the contextual restrictions that guide their use (Gunlogson, 2003). Nevertheless, the available production data imply that five-year-old children have yet to acquire distinct intonational categories for terminal rise and fall that can be mapped reliably onto question versus statement functions, respectively.

The present study investigated five- to ten-year-old children's ability to interpret utterances as questions or statements on the basis of intonation alone, and compared children's performance with that of adults. The goal was to document developmental progression toward adult-like efficacy in mapping a terminal rise onto a question and a terminal fall onto a statement when all other variables are held constant. To this end, we used decontextualized declarative questions (e.g. *Bob is funny?*) rather than standard *yes/no* questions (e.g. *Is Bob funny?*).

Although five-year-old children differentiate declarative questions from statements in a same–different task (Doherty, Fitzsimons, Asenbauer & Staunton, 1999), their long-term representations of the contrasting pitch contours (i.e. the relevant intonational categories) may be insufficiently robust to support stable mapping onto pragmatic and non-linguistic functions. For example, preschoolers more readily remember a cartoon character's favorite melody from its timbre (i.e. instrument) than from its rising or falling pitch contour (Creel, 2014). Moreover, five- and six-year-olds readily discriminate pitch directional changes (e.g. rising vs. falling), but they do not typically apply labels such as 'higher', 'lower', 'up', and 'down' to pitch direction (Andrews & Madeira, 1977; Costa-Giomi & Descombes, 1996) unless they receive targeted training (Stalinski, Schellenberg & Trehub, 2008).

Weak categorical representations of pitch contours may lead children to accord less attention to prosody than adults do, especially in the context of conflicting cues. For example, when four- to nine-year-olds are asked to judge a speaker's feelings (happy or sad) from the sound of her voice, ignoring what she says, they focus on lexical or semantic cues rather than prosodic cues (Morton & Trehub, 2001). When situational cues are available, five- and seven-year-olds judge a speaker's feelings (good or bad) from situational rather than prosodic cues (Aguert, Laval, Bigot & Bernicot, 2010). In the absence of conflicting or distracting cues, young children succeed in distinguishing happy from sad expressiveness (Morton & Munakata, 2002; Morton & Trehub, 2001). Their success is facilitated by the availability of multiple acoustic cues to these emotion categories (e.g. pitch level, pitch contours, speaking rate, amplitude) as well as familiar, concrete response categories (happy, sad).

In the present study, the acoustic distinctions between statements and declarative questions were less pronounced than those of happy- and sad-sounding utterances, and the response categories, QUESTION and STATEMENT, were less familiar to young children, more abstract, and less readily amenable to visual depiction. In Experiment 1, adults and children five to ten years of age were required to identify each of several utterances as questions or statements. To counter younger children's potential unfamiliarity with terms such as 'statements' and 'questions', 'asking' and 'telling' were used as response labels along with supporting photographs. Children as young as five understand the meaning of *ask* and *tell*, although their responses are dominated, at times, by contextual and interpersonal factors (Warden, 1981). In principle, the verbal content of utterances, although irrelevant and non-conflicting, could prove distracting, as in previous research (Aguert *et al.*, 2010; Morton & Trehub, 2001) because of children's prepotent bias for message content (Waxer & Morton, 2011).

Accordingly, Experiment 2 featured the same task with the same utterances low-pass filtered to obscure the content.

EXPERIMENT 1

METHOD

Participants

The final sample consisted of 122 participants, including 30 five- and six-year-olds (14 girls, 16 boys; $M = 6;0$, range = 5;0–6;11), 31 seven- and eight-year-olds (10 girls, 21 boys; $M = 8;1$, range = 7;0–8;11), 32 nine- and ten-year-olds (17 girls, 15 boys; $M = 10;1$, range = 9;0–10;11), and 29 adults (21 women, 8 men; $M = 18{\cdot}52$ years, $SD = 1{\cdot}27$). Children were recruited from the community. Adults were college students who received partial course credit for their participation. Children had normal hearing and overall development, according to parental report. Inclusion criteria for adults were normal hearing and Canadian birth or arrival in Canada by eight years of age. An additional nine participants were tested but excluded because of technical errors (one five-year-old), parent-reported developmental delay (one five-year-old, one seven-year-old, one eight-year-old), failure to meet the criterion during the training phase (one six-year-old), and scores that were more than 2 *SD*s below the mean for their age group (a common predetermined criterion that affected two seven-year-olds and two ten-year-olds). The exclusion of children with atypically low scores had no effect on the findings.

Apparatus and stimuli

Stimulus recording and testing took place in a sound-attenuating booth (Industrial Acoustics Corporation, Bronx, NY) with loudspeakers (Electro-Medical Instrument Co., Mississauga, ON) mounted in two corners of the sound booth at 45° azimuth to the participant. Interactive software created with Affect4 (Spruyt, Clarysse, Vansteenwegen, Baeyens & Hermans, 2010) for a Windows 7 computer (outside the booth) presented instructions and stimuli and recorded participants' responses. Participants entered their responses on a 17-in touch-screen monitor (Elo LCD TouchSystems, Berwyn, PA) that faced them.

Two men and two women recorded declarative question and statement versions of each of ten sentences (see Table 1) using a microphone (Sony T) connected to the computer. They generated natural-sounding utterances while minimizing distinctive prosodic cues until the final syllable, which featured a rising F_0 glide for questions and a falling glide for statements. The stimulus set consisted of eighty utterances (10 sentences × 4 speakers × 2 versions). High-quality digital sound files (44·1 kHz, 16-bit, mono) created

TABLE 1. *Sentences used in Experiments 1 and 2*

The cat ran away
She lost her shoes
Mom made it
It's snowing
You found it
Mom went to the store
It's bedtime
He's watching TV
He's in the car
You're staying home

with a digital audio editor (Sound Forge Pro version 10·0; Sony, Tokyo, Japan) were amplitude normalized and cleaned for superfluous noise with the Sound Forge Noise Reduction plug-in. Stimuli were presented at approximately 65 dB SPL. The F_0 contours of a typical question and statement from the stimulus set are illustrated in Figure 1, which confirm the contour shape as relatively flat until the terminal pitch rise or fall. Audio samples are provided in supplementary materials (available at <http://www.journals.cambridge.org/JCL>). Digital photographs of a man and woman smiling and posing neutrally served as 'telling' pictures; photographs with a quizzical facial expression and pose served as 'asking' pictures (see Figure 2).

Procedure

Participants were tested individually. The experimenter remained in the booth only for children's testing. Children sat facing the touch-screen, and the experimenter was seated to one side, controlling trial presentations with a keypad. The task was described as a game in which children would hear a man or lady asking or telling them something. They were instructed to touch the 'asking' picture if the person was asking about something and the 'telling' picture if the person was telling them something. Adults controlled the presentation of trials and indicated whether each utterance was a question or statement by selecting the appropriate picture.

Pilot testing indicated that many of the five- and six-year-olds had difficulty with the task. Accordingly, all children were required to meet a training criterion – indicating their understanding of the task – before proceeding to the practice and test phases. First, children confirmed that they could correctly identify the 'asking' and 'telling' pictures. (e.g. "Which one is the *asking* picture?") The subsequent TRAINING phase consisted of a maximum of four blocks of four trials, with feedback on all trials. Each block consisted of declarative statements and standard *yes/no*

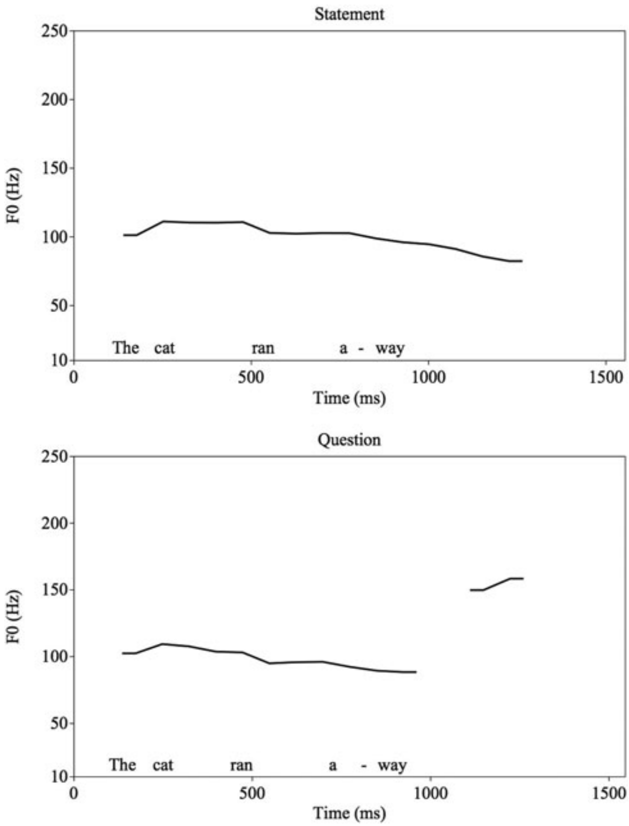

Fig. 1. F_o contour of a typical question and statement from Experiment 1.

questions (e.g. "Is the coat in the closet?"). Children who made errors on the last block of training trials were excluded from the final sample.

As soon as children achieved a perfect score on any of the training blocks, they proceeded directly to the PRACTICE phase, which featured statements and declarative questions that differed from those in the TEST phase. There were a maximum of four blocks of four practice trials, which were designed to clarify that utterances other than standard *yes/no* questions could still be asking something. If children obtained a perfect score in any practice block, they proceeded directly to the test phase. Thus, there were one to four blocks of training trials and one to four blocks of practice trials, depending on the individual child's understanding and performance. Although children received different numbers of practice trials depending on their understanding and performance, no children were excluded from the experiment on the basis of their performance on these trials.

Fig. 2. Black-and-white versions of 'asking' (left) and 'telling' photos (right) for the male talker.

Adults completed a four-trial familiarization phase with statements and declarative questions (not those used in testing) before the onset of the test phase. The actual 'test' phase comprised eighty trials, one for each stimulus utterance. Children were told at the start of the test phase that four pieces of a large smiley face could be exchanged for a prize. One piece of the smiley face appeared after each set of twenty trials regardless of children's performance. The order of trials in the test phase was randomized with the constraint that the same sentence content did not appear on successive trials. Adults and children received feedback on all trials (training, practice, and test), consisting of a 1 s presentation of one of ten cartoon characters for correct answers, and a 1 s blank screen for incorrect answers. The provision of continuous feedback ensured that age-related differences in performance were not attributable to differential memory of the task requirements.

RESULTS AND DISCUSSION

For each child, we summed the number of training and practice blocks as an index of initial difficulty with the task. As can be seen in Figure 3, younger children required more blocks than did older children, indicating greater initial difficulty in differentiating questions from statements. Data from the test phase were analyzed by converting responses to d' (d-prime) scores using proportions of hits and false alarms. Whereas d' scores provide an

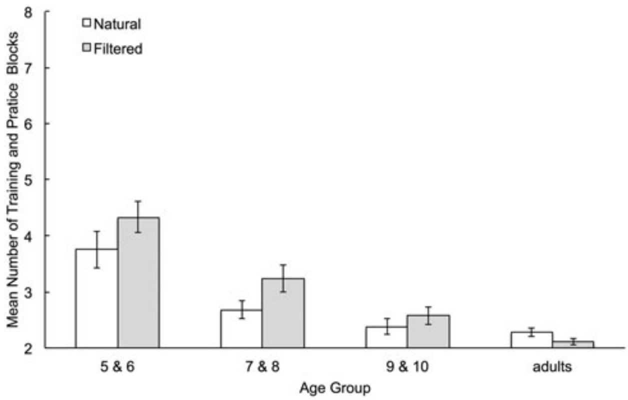

Fig. 3. Number of training and practice blocks in Experiment 1 (natural utterances) and Experiment 2 (low-pass filtered utterances) as a function of age group. Error bars are standard errors.

index of listeners' SENSITIVITY to question intonation (i.e. the issue of interest in the present investigation), percent correct scores reflect BIAS as well as sensitivity. For the present purposes, question responses to question stimuli constituted hits, and question responses to statement stimuli constituted FALSE ALARMS. A score of $d' = 1$ corresponds to 69% correct.

The stimulus set consisted of forty questions and forty statements, which meant that the maximum number of hits or false alarms was forty. Because d' scores cannot be computed when the proportion of hits is 1·0 or the proportion of false alarms is 0 (i.e. statistically infinite scores), proportions of hits and false alarms were calculated by adding 0·5 to the number of hits and also to the number of false alarms and dividing those numbers by 41 (total possible hits or false alarms + 1), as in previous developmental studies (e.g. Thorpe, Trehub, Morrongiello & Bull, 1988). These proportions were converted to z-scores and then to d' scores ($d' = z$[hit rate] − z[false-alarm rate]). The maximum d' score was 4·5. Raw data (percent correct) and standard errors are illustrated in Figure 4 separately for each age group.

One-sample t-tests conducted separately for each age group revealed that performance was significantly better than chance (i.e. chance or $d' = 0$ results from an equal number of hits and false alarms) for each age group (ps < .001). A one-way analysis of variance (ANOVA), with age (5–6, 7–8, 9–10, and adults) as the between-subjects variable and d' as the dependent variable, revealed a significant effect of age ($F(3,118) = 35{\cdot}78$, $p < .001$, $\eta^2 = {\cdot}48$). Follow-up pairwise comparisons (Tukey HSD) revealed that the performance of five- to six-year-olds was significantly poorer than all other

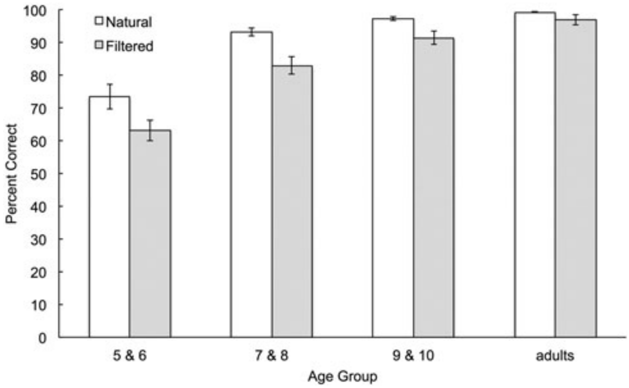

Fig. 4. Performance in the test sessions of Experiment 1 (natural utterances) and Experiment 2 (low-pass filtered utterances) as a function of age group. Error bars are standard errors.

age groups (ps < .001), and that seven- to eight-year-olds performed more poorly than adults (p = ·001). The performance of nine- to ten-year-olds did not differ significantly from adults or from seven- to eight-year-olds (ps > .1).

We also examined the percentage of individuals in each group whose correct responses (i.e. correct identification of questions and statements) exceeded chance levels. According to the normal approximation to the binomial test (one-tailed, correcting for continuity), 48 or more correct responses out of 80 is significantly better than chance. Only 63% (19 of 30) of five- and six-year-olds obtained a score of 48 or more, in contrast to 100% in each of the three older groups. Fisher's exact tests confirmed that the proportion of individuals performing at chance levels was significantly higher in the youngest group of children compared to any other group (ps < .001).

We then asked whether participants had a RESPONSE SET, specifically a bias to respond 'telling' (or 'statement') more or less often than 'asking' (or 'question'). For each participant, we calculated the total number of 'telling' responses. Comparisons with 50% (40 of 80) revealed that the youngest children had no response bias (p > .5), but each of the three older groups tended to respond 'telling' more than half of the time (ps < .03), presumably because of the syntax of the stimulus sentences. Nevertheless, the mean number of 'telling' responses was under 42 in each instance.

The variance in the number of 'telling' responses was considerably higher for the youngest children (SD = 15·79) compared to the other three groups (all SDs < 4·10), which motivated us to examine the number of participants who exhibited a bias to respond EITHER 'telling' OR 'asking.'

Using the criterion described above (i.e. 48 or more 'telling' responses, or 48 or more 'asking' responses), we identified 11 participants who exhibited such a bias: 10 five- to six-year-olds and 1 seven- to eight-year-old. Fisher's exact tests confirmed that this bias was greater for the youngest group of children than for any of the other three groups ($ps < .003$).

Finally, we examined whether children who required more training and practice blocks performed more poorly in the test phase than those who required fewer blocks. Positive skewness of the training/practice variable prompted the use of a Spearman's correlation. Test scores were negatively correlated with number of training and practice blocks ($r_s(n = 92) = -0{\cdot}48$, $p < .001$).

In sum, the results indicated that children and adults identified questions based on prosodic information alone. Nevertheless, five- to six-year-old children performed more poorly than older children and adults. In fact, 37% of children in the youngest age group performed at chance levels, but no participant in any other group did so, and one-third of the youngest children tended to respond consistently with 'telling' or 'asking' regardless of the stimuli. The performance of seven- to eight-year-old children was also less accurate than that of adults. Finally, children who required more training and practice trials had poorer outcomes on test trials.

In light of young children's propensity to focus on irrelevant verbal content or situational context when judging a speaker's feelings (Aguert *et al.*, 2010; Friend, 2000; Morton & Trehub, 2001), the irrelevant verbal content of declarative questions may have distracted them from the critical prosodic cues. Just as young children achieve greater success in identifying a speaker's feelings when the verbal content is obscured (Morton & Trehub, 2001), they may achieve greater success in identifying declarative questions when the verbal content is unintelligible. This possibility was examined in Experiment 2.

EXPERIMENT 2

The goal of this experiment was to examine whether young children's identification of declarative questions would approach the performance of older children when the verbal content of utterances was obscured by low-pass filtering.

METHOD

Participants

The final sample consisted of 126 participants: 33 five- and six-year-olds (18 girls, 15 boys; $M = 6;1$, range $= 5;0–6;9$), 34 seven- and eight-year-olds (15 girls, 19 boys; $M = 7;11$, range $= 7;0–8;10$), 31 nine- and ten-year-olds

(14 girls, 17 boys; $M = 10;1$, range = 9;2–10;9), and 28 adults (20 women, 8 men; $M = 18{\cdot}43$ years, $SD = 1{\cdot}26$). Recruiting and inclusion criteria were the same as in Experiment 1. An additional nineteen participants were tested but excluded because of parent-reported developmental delays (one ten-year-old), failure to meet the criterion during the training phase (7 five-year-olds and 1 six-year-old), failure to pay attention to the task (3 five-year-olds and 4 six-year-olds), and scores that were more than 2 *SD*s below the mean for their age group (1 eight-year-old, 1 nine-year-old, and 1 ten-year-old). Inclusion of these 'outliers' distorted the group means but did not alter the outcome of the analyses.

Apparatus and stimuli

The apparatus was the same as in Experiment 1. The stimuli were created by low-pass filtering the sentences from Experiment 1 at 400 Hz (following Friend, 2000; Knoll, Uther & Costall, 2009) using Praat version 5.3·68 (Boersma & van Heuven, 2001) and normalizing the amplitude of filtered utterances, which were presented at 65 dB SPL. Because low-pass filtered speech sounds unnatural, children were told that they would hear robots speaking and that they had to indicate whether the robots were asking or telling them something. The pictures of the man and woman were replaced with male and female robot cartoon characters, which were drawn as standing in either a neutral (arms down) pose or a questioning (arms raised, palms upward) pose. The same pictures were used with adults, who were told that the task was designed for children.

Procedure

The procedure was similar to that used in Experiment 1. The experimenter remained in the booth for the testing of children but not adults. Children were told they were going to play a game in which they would hear a girl or boy robot asking or telling them something. They were told that they would not understand what the robots were saying but they had to listen carefully to decide whether they were asking or telling them something. They were shown the picture of the girl and boy robot and told that they should touch the 'asking' picture if the robot was asking something and the 'telling' picture if the robot was telling them something. Adults also indicated whether each utterance was a question or statement by selecting the appropriate picture. They were also told that the utterances had been modified to make the words incomprehensible and that they needed to listen carefully to judge whether the utterances were questions or statements.

As in Experiment 1, children were required to complete training and practice phases prior to the test phase. The four blocks of the training phase were identical to Experiment 1. The four blocks of the practice phase were also identical to Experiment 1 except that the utterances were low-pass filtered. As soon as children obtained a perfect score in one of the four training blocks, they proceeded directly to the practice phase. Children who failed to achieve a perfect score on the fourth block of training trials with unfiltered speech and *yes/no* questions were excluded from the final sample. The practice phase comprised four blocks of four trials, but if children obtained a perfect score in a block, the practice phase was terminated and they began the test phase. Children and adults received feedback after each response in the training, practice, and test phases, as in Experiment 1. As before, children were told about gathering pieces of the smiley face for subsequent prizes. Children and adults had the option of hearing each utterance for a second time.

RESULTS AND DISCUSSION

As shown in Figure 3, children required more training and practice blocks in the present experiment ($M = 3{\cdot}40$, $SD = 1{\cdot}49$) than in Experiment 1 ($M = 2{\cdot}91$, $SD = 1{\cdot}33$) ($t(187{\cdot}55) = 2{\cdot}37$, $p = {\cdot}019$) (unequal variances test). The data from the test phase were converted to d' scores, as in Experiment 1. Raw data (percent correct) are illustrated in Figure 4. One-sample t-tests for each age group revealed that performance was significantly better than chance in all cases (ps < .001). An ANOVA with age (5–6, 7–8, 9–10, and adults) as the between-subjects variable and d' as the dependent variable revealed a significant effect of age ($F(3,122) = 33{\cdot}13$, $p < .001$, $\eta^2 = {\cdot}45$). The performance of five- to six-year-olds was significantly worse than that of all other age groups (ps < .001), and the performance of seven- to eight-year-olds was worse than that of nine- to ten-year-olds ($p = {\cdot}038$) and adults ($p < .001$). The performance of nine- to ten-year-olds and adults did not differ ($p > .2$).

We then examined the percentage of individuals in each group whose correct responses exceeded chance levels (48 or more correct responses). Because performance was not as consistently good as it was in Experiment 1, we used a 4 (age [5–6, 7–8, 9–10, adults]) by 2 (above chance, chance) chi-square test of independence to examine age-related differences in the percentage of individuals who successfully distinguished questions from statements (all cells had expected frequencies > 5). Successful performance varied across age groups ($\chi^2(3, N = 126) = 54{\cdot}70$, $p < .001$, $\Phi = {\cdot}66$). All adults performed above chance levels, as did all but one of the nine- to ten-year-olds, and all but three of the seven- to eight-year-olds. For the five- to six-year-olds, far fewer children – just slightly more than a third (36%, 12 of 33) – exceeded chance levels of performance.

Performance on the natural utterances from Experiment 1 and the low-pass filtered utterances from the present experiment was compared by means of a two-way ANOVA with stimulus type (natural, filtered) and age group (5–6, 7–8, 9–10, adults) as between-subjects variables and d' as the dependent variable. There were main effects of age ($F(3,240) = 67{\cdot}92$, $p < .001$, $\eta^2 = {\cdot}43$) and stimulus type ($F(1,240) = 26{\cdot}86$, $p < .001$, $\eta^2 = {\cdot}06$). Overall, performance improved with age, and participants could more easily identify natural rather than low-pass filtered questions and statements. The interaction between age and stimulus type was not significant ($F < 1$). In other words, the decrement in performance for filtered compared to natural stimuli was similar across age groups.

If the unusual acoustic quality resulting from low-pass filtering contributed to the unexpected reduction in performance, then performance may have improved from the first to the second half of the test session, after listeners adapted to the spectral degradation. Because individual participants had different numbers of questions and statements in the first and second half due to randomization of order, we analyzed the number of correct responses rather than d' scores. A mixed-design ANOVA, with age group (5–6, 7–8, 9–10, adults) as the between-subjects variable and test phase (first or second half) as the within-subjects variable revealed a main effect of age ($F(3,122) = 37{\cdot}14$, $p < .001$, $\eta^2_p = {\cdot}48$), reflecting age-related improvement, and a small but significant effect of test phase ($F(1,122) = 8{\cdot}40$, $p = {\cdot}004$, $\eta^2_p = {\cdot}06$). Performance was better during the second half of the test session ($M = 33{\cdot}63$, $SD = 7{\cdot}49$) than in the first half ($M = 32{\cdot}62$, $SD = 7{\cdot}73$), but the improvement represented only one additional correct answer on forty trials. There was no interaction between age group and test phase ($F < 1$), which indicated that all age groups had comparable adaptation to the low-pass filtered stimuli. In the second half of trials, overall performance with filtered stimuli (84% correct) remained substantially below overall performance with natural stimuli in Experiment 1 (91%) ($t(237{\cdot}04) = 3{\cdot}19$. $p = {\cdot}002$) (unequal variances test).

Examination of a possible response set (i.e. responding 'telling' more than 50% of the time) revealed no such bias among the three groups of children (ps > .1), but a small bias for adults ($p = {\cdot}011$) ($M = 40{\cdot}93$). Categorization of participants into those with or without a systematic bias to respond EITHER 'telling' OR 'asking' revealed 21 participants with such a bias: 14 five- to six-year-olds, 6 seven- to eight-year-olds, and 1 nine- to ten-year-old. The proportion of participants with this bias varied reliably across age groups ($\chi^2(3, N = 126) = 25{\cdot}42$, $p < .001$, $\Phi = {\cdot}45$).

As in Experiment 1, we examined the relation between number of training and practice blocks and subsequent test scores for the child participants. A Spearman's correlational analysis revealed a significant negative correlation between test scores and number of training and practice blocks ($r_s(n = 98) = -0{\cdot}50$ $p < .001$).

In sum, children differentiated questions from statements in low-pass filtered utterances at better than chance levels, with the two oldest age groups (9–10, adults) performing significantly better than the two youngest (5–6, 7–8). Contrary to expectations, all age groups performed worse on the filtered utterances than on the original versions, which indicates that young children's difficulty in differentiating question from statement intonation cannot be attributed to interference from the verbal content. As in Experiment 1, a substantial portion of children in the youngest group tended to respond 'telling' or 'asking' in general, and children who required more training and practice trials tended to perform poorly in the actual test session.

GENERAL DISCUSSION

The present study examined the identification of declarative questions and statements by English-speaking children (five to ten years of age) and adults. Age-related differences in performance were similar for natural utterances (Experiment 1) and low-pass filtered utterances (Experiment 2), but performance in general was poorer for the filtered utterances. Although five- and six-year-olds performed above chance levels, they performed more poorly than all other age groups despite having numerous training and practice trials and continuous feedback about response accuracy throughout the test session. The seven- and eight-year-olds performed no differently than nine- and ten-year-olds on natural utterances but they performed more poorly on filtered utterances. Finally, the nine- and ten-year-olds performed no differently than adults.

The findings from the five- and six-year-olds are consistent with limited knowledge of the relevant intonation categories and, consequently, poor mapping between intonational contours and communicative functions. These results are in line with preschool children's challenges in linking cartoon characters with rising or falling melodic sequences (Creel, 2014) and six-year-olds' difficulty with conventional verbal labels for rising and falling pitch (Costa-Giomi & Descombe, 1996).

It is likely that young children's difficulties were exacerbated by limitations in attention allocation and working memory (Cowan, Morey, AuBuchon, Zwilling & Gilchrist, 2010). To succeed on the present task, children had to focus on the terminal pitch contour, determine if it was rising, and designate it as a question if it was or as a statement otherwise. Young children's habitual focus on lexical cues at the expense of prosodic cues (Morton & Trehub, 2001; Snedeker & Trueswell, 2004) and limited cognitive flexibility (Munakata, Snyder & Chatham, 2012), even in the face of continuous feedback, may have interfered with consistent allocation of attention to the relevant cues in Experiment 1, which featured natural utterances.

Although all participants were required to meet the same training criterion before proceeding to the test phase, they did not achieve comparable understanding, as reflected in persistent performance differences in the test phase. In fact, the number of trials required to meet the training criterion – one index of initial comprehension – was predictive of subsequent comprehension, as indexed by performance in the test phase, which also featured feedback on every trial. Fully 37% of the five- to six-year-olds failed to identify utterance type, as reflected in their chance-level performance.

Low-pass filtering that made the verbal content unintelligible was expected to facilitate young children's attention to the relevant acoustic cues, in line with previous studies of emotional prosody (Morton & Trehub, 2001). Instead, such filtering had the opposite effect, reducing performance comparably for all age groups. Low-pass filtering decreases the salience of the component pitches that are relevant for differentiating statements from questions (Cruttenden, 1981; Eady & Cooper, 1986; Gårding & Abramson, 1965; Peng *et al*., 2009; Studdert-Kennedy & Hadding, 1973). Young children's difficulty with the filtered utterances confirmed that their problems with the unfiltered utterances in Experiment 1 did not stem from lexical biases. Although 63% of the five- and six-year-old children performed above chance levels on the natural utterances, only 36% performed above chance on the filtered utterances. Presumably, their problems with categorizing pitch contours were exacerbated by decreased salience of the relevant cues.

Low-pass filtering preserves the pitch directional differences of the original utterances, but the resulting speech is perceived as being reduced in pitch range (i.e. compressed pitch contours) and pitch variability relative to unfiltered speech (Scherer, Koivumaki & Rosenthal, 1972; van Bezooijen & Boves, 1986). It is likely, then, that reduced pitch salience in the filtered utterances contributed to the uniform reduction in performance across age.

The unusual sound quality of the filtered speech could have impaired performance even though listeners adapt to distorted speech after limited exposure (Hervais-Adelman, Davis, Johnsrude & Carlyon, 2008; Hervais-Adelman, Davis, Johnsrude, Taylor & Carlyon, 2011). Indeed, performance on the filtered utterances improved modestly (i.e. one additional item correct) during the second half of the test session although it remained below the levels achieved with unfiltered stimuli. Presumably, adaptation to the filtered speech began during the training phase and continued during the test phase. Further exposure, either during the training phase or in a subsequent session, could result in performance levels approaching those attained with unfiltered utterances.

If young children have difficulty discerning the communicative intent of declarative questions, one would expect caregivers to avoid such questions. There is evidence, however, that parents make regular use of declarative questions in their conversations with very young children (Estigarribia, 2010). One might further expect the acoustic cues to the question/ statement distinction to be exaggerated in child-directed speech, as are other cues to the speaker's intentions (Foulkes, Docherty & Watt, 2005; Jacobson, Boersma, Fields & Olson, 1983). Moreover, declarative questions in parent–child discourse are likely to be restricted to face-to-face contexts that provide supplementary visual cues such as raised eyebrows and head movements (Ekman, 1976, 1979; Srinivasan & Massaro, 2003). Most importantly, such questions would not occur in isolation, as in the present study. Instead, parents would adhere to the contextual constraints on declarative questions (Gunlogson, 2003), which would highlight their communicative intentions. For example, a child's request for a cookie just before dinner might receive an incredulous reply such as "You want a cookie?" Despite the limitations of five- and six-year-olds, as observed in the present study, it is likely that they would comprehend the gist of the message.

In conclusion, children between five and six years of age have considerably greater difficulty than older children and adults in differentiating isolated declarative questions from statements on the basis of intonation alone. We contend that this difficulty stems primarily from absent or unstable intonational categories and secondarily from immature working memory (Cowan *et al.*, 2010), which precluded successful mapping of terminal pitch contours onto communicative functions. An important challenge for future research is to ascertain the factors that support children's transition from continuous representations of pitch contours to categorical representations of rising and falling contour.

SUPPLEMENTARY MATERIALS

To view supplementary material for this article, please visit <www.journals.cambridge.org/JCL>.

REFERENCES

Aguert, M., Laval, V., Bigot, L. & Bernicot, J. (2010). The role of prosody and situational context in French-speaking 5- to 9-year-olds. *Journal of Speech, Language, and Hearing Research* **53**, 1629–41.

Andrews, M. L. & Madeira, S. S. (1977). The assessment of pitch discrimination ability in young children. *Journal of Speech and Hearing Disorders* **42**, 279–86.

Boersma, P. & van Heuven, V. (2001). Speak and unSpeak with PRAAT. *Glot International* **5**, 9–10.

Costa-Giomi, E. & Descombes, V. (1996). Pitch labels with single and multiple meanings: a study with French-speaking children. *Journal of Research in Music Education* **44**, 204–14.

Cowan, N., Morey, C. C., AuBuchon, A. M., Zwilling, C. E. & Gilchrist, A. L. (2010). Seven-year-olds allocate attention like adults unless working memory is overloaded. *Developmental Science* **13**, 120–33.

Creel, S. C. (2014). Tipping the scales: auditory cue weighing changes over development. *Journal of Experimental Psychology: Human Perception and Performance* **40**, 1146–60.
Cruttenden, A. (1981). Falls and rises: meanings and universals. *Journal of Linguistics* **17**, 77–91.
Doherty, C. P., Fitzsimons, M., Asenbauer, B. & Staunton, H. (1999). Discrimination of prosody and music by normal children. *European Journal of Neurology* **6**, 221–6.
Eady, S. J. & Cooper, W. E. (1986). Speech intonation and focus location in matched statements and questions. *Journal of the Acoustical Society of America* **80**, 402–15.
Ekman, P. (1976). Movements with precise meanings. *Journal of Communication* **26**, 14–26.
Ekman, P. (1979). About brows: emotional and conversational signals. In M. von Cranach, K. Foppa, W. Lepenies & D. Ploog (eds), *Human ethology: claims and limits of a new discipline*, 169–249. Cambridge: Cambridge University Press.
Estigarribia, B. (2010). Facilitation by variation: right-to-left learning of English yes/no questions. *Cognitive Science* **34**, 68–93.
Foulkes, P., Docherty, G. & Watt, D. (2005). Phonological variation in child-directed speech. *Language* **81**, 177–206.
Froda, S., Butler, J. & Vigario, M. (2014). Infants' perception of intonation: Is it a statement or a question? *Infancy* **19**, 194–213.
Friend, M. (2000). Developmental changes in sensitivity to vocal paralanguage. *Developmental Science* **3**, 148–62.
Gårding, E. & Abramson, A. S. (1965). A study of the perception of some American English intonation contours. *Studia Linguistica* **19**, 61–79.
Gunlogson, C. (2003). *True to form: rising and falling declaratives as questions in English*. New York: Routledge.
Hervais-Adelman, A., Davis, M. H., Johnsrude, I. S. & Carlyon, R. P. (2008). Perceptual learning of noise vocoded words: effects of feedback and lexicality. *Journal of Experimental Psychology: Human Perception and Performance* **34**, 460–74.
Hervais-Adelman, A. G., Davis, M. H., Johnsrude, I. S., Taylor, K. J. & Carlyon, R. P. (2011). Generalization of perceptual learning of vocoded speech. *Journal of Experimental Psychology: Human Perception and Performance* **37**, 283–95.
Jacobson, J. L., Boersma, D. C., Fields, R. B. & Olson, K. L. (1983). Paralinguistic features of adult speech to infants and small children. *Child Development* **54**, 436–42.
Knoll, M. A., Uther, M. & Costall, A. (2009). Effects of low-pass filtering on the judgment of vocal affect in speech directed to infants, adults and foreigners. *Speech Communication* **51**, 210–16.
Loeb, D. F. & Allen, G. D. (1993). Children's imitation of intonation contours. *Journal of Speech and Hearing Research* **36**, 4–13.
Morton, J. B. & Munakata, Y. (2002). Are you listening? Exploring a developmental knowledge–action dissociation in a speech interpretation task. *Developmental Science* **5**, 435–40.
Morton, J. B. & Trehub, S. E. (2001). Children's understanding of emotion in speech. *Child Development* **72**, 834–43.
Munakata, Y., Snyder, H. R. & Chatham, C. H. (2012). Developing cognitive control: three key transitions. *Current Directions in Psychological Science* **21**, 71–7.
Papoušek, M., Bornstein, M. H., Nuzzo, C., Papoušek, H. & Symmes, D. (1990). Infant responses to prototypical melodic contours in parental speech. *Infant Behavior and Development* **13**, 539–45.
Patel, R. & Brayton, J. T. (2009). Identifying prosodic contrasts in utterances produced by 4-, 7-, and 11-year-old children. *Journal of Speech, Language, and Hearing Research* **52**, 790–801.
Patel, R. & Grigos, M. I. (2006). Acoustic characterization of the question–statement contrast in 4-, 7- and 11-year-old children. *Speech Communication* **48**, 1308–18.
Peng, S. C., Lu, N. & Chatterjee, M. (2009). Effects of cooperating and conflicting cues on speech intonation recognition by cochlear implant users and normal hearing listeners. *Audiology and Neurotology* **14**, 327–37.

Scherer, K. R., Koivumaki, J. & Rosenthal, R. (1972). Minimal cues in the vocal communication of affect: judging emotions from content-masked speech. *Journal of Psycholinguistic Research* **1**, 269–85.

Snedeker, J. & Trueswell, J. C. (2004). The developing constraints on parsing decisions: the role of lexical biases and referential scenes in child and adult sentence processing. *Cognitive Psychology* **49**, 238–99.

Snow, C. E. (1977). Mothers' speech research: from input to interaction. In C. E. Snow & C. A. Ferguson (eds), *Talking to children* (pp. 31–49). Cambridge: Cambridge University Press.

Snow, D. (1994). Phrase-final syllable lengthening and intonation in early child speech. *Journal of Speech, Language, and Hearing Research* **37**, 831–40.

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**, 576–87.

Soderstrom, M., Ko, E.-S. & Nevzorova, U. (2011). It's a question? Infants attend differently to yes/no questions and declaratives. *Infant Behavior and Development* **34**, 107–10.

Spruyt, A., Clarysse, J., Vansteenwegen, D., Baeyens, F. & Hermans, D. (2010). Affect 4·0: a free software package for implementing psychological and psychophysiological experiments. *Experimental Psychology* **57**, 36–45.

Srinivasan, R. J. & Massaro, D. W. (2003). Perceiving prosody from the face and voice: distinguishing statements from echoic questions in English. *Language and Speech* **46**, 1–22.

Stalinski, S. M., Schellenberg, E. G. & Trehub, S. E. (2008). Developmental changes in the perception of pitch contour: distinguishing up from down. *Journal of the Acoustical Society of America* **124**, 1759–63.

Studdert-Kennedy, M. & Hadding, K. (1973). Auditory and linguistic processes in the perception of intonation contours. *Language and Speech* **16**, 293–313.

Thorpe, L. A., Trehub, S. E., Morrongiello, B. A. & Bull, D. (1988). Perceptual grouping by infants and preschool children. *Developmental Psychology* **24**, 484–91.

van Bezooijen, R. & Boves, L. (1986). The effects of low-pass filtering and random splicing on the perception of speech. *Journal of Psycholinguistic Research* **15**, 403–17.

Warden, D. (1981). Children's understanding of *ask* and *tell*. *Journal of Child Language* **8**, 139–49.

Waxer, M. & Morton, J. B. (2011). Children's judgments of emotion from conflicting cues in speech: why 6-year-olds are so inflexible. *Child Development* **82**, 1648–60.

Whalen, D., Levitt, A. G. & Wang, Q. (1991). Intonational differences between the reduplicative babbling of French- and English-learning infants. *Journal of Child Language* **18**, 501–16.