Differentiating Questions from Statements: The Role of Intonation

by

Mathieu R. Saindon

A thesis submitted in conformity with the requirements for the degree of Doctor of Philosophy
Department of Psychology
University of Toronto

© Copyright by Mathieu R. Saindon 2016
Differentiating Questions from Statements: The Role of Intonation

Mathieu R. Saindon

Doctor of Philosophy

Department of Psychology
University of Toronto

2016

Abstract

The most salient cues to yes/no questions and statements are rising and falling terminal pitch contours. Because young children readily differentiate rising from falling pitch contours, the presumption is that they can discern speakers’ questioning or declarative intentions from intonation alone. This thesis examined developmental changes in the use of intonation to identify questions and statements. In the first of three studies, participants from 5 years of age judged naturally produced utterances and low-pass filtered versions as questions or statements. Children correctly identified utterance type above chance levels and achieved adult accuracy levels at 9 years of age, highlighting younger children’s confusion about the links between intonation contours and pragmatic intentions. In the second study, participants from 8 years of age judged utterances with graded manipulations of terminal contours as questions or statements. Children’s question/statement judgments shifted more gradually than those of adults, reflecting greater uncertainty, but the location of the category shift was comparable across age. Adults’ discrimination of utterance pairs was best for the pair that crossed the category shift, implying categorical perception of the intonation contours. In the final study, participants from 7 years of age judged the statement or question status of child- and adult-directed utterances in a gating task with words added incrementally. After limited exposure to the speaker’s voice, adults and children from 9 years of age correctly identified questions and statements from the initial word, demonstrating for the first time that English-speaking adults and children perceive pre-terminal
cues to these utterance types. Identification was more accurate for child-directed than for adult-directed utterances, indicating that the former speech register exaggerates the distinctions between question and statements. Collectively, the findings reveal a protracted course of development for the questioning and declarative intentions signalled by intonation patterns and, more generally, for the pragmatic functions of intonation.
Dedication

This thesis is dedicated to my late mother, Suzanne Hébert, who did not live to see me complete my graduate studies. She always made it clear that she was proud of what I had accomplished and the person I had become. Merci maman. Je t'aime beaucoup.
Acknowledgments

I would like to thank my supervisors, Drs. Sandra E. Trehub and E. Glenn Schellenberg, for their guidance and support over the past seven years. I am especially grateful for their patience—I had much to learn during the course of this degree, and their demanding workloads and busy lives never prevented them from providing help whenever I needed it. I would also like to thank Dr. Pascal van Lieshout for his encouragement and constructive feedback during this process.

Many thanks also to the research and administrative assistants that have helped me over the years, including Deanna Feltracco, Lisa Hotson, Brad Marks, Ansha Thirumeny, Lily Zhou, Terilyn Harber, Sarah Lindsay, Sari Park, Snover Dhillon, and Sofia Rokerya. Thank you also to my fellow researchers and graduate students—Judy Plantinga, Michael Weiss, Stephanie Stalinski, and Patrick Hunter—for their help and guidance.

Lastly, I would like to thank my parents, Paul and Suzanne, and my wife, Lauren. My parents' selfless devotion enabled me to pursue doctoral studies, and my wife's love, support, and unending patience enabled me to survive the process.
# Table of Contents

General Abstract ........................................................................................................... ii

Dedication ................................................................................................................... iv

Acknowledgments ....................................................................................................... v

Table of Contents ....................................................................................................... vi

List of Tables ................................................................................................................ viii

List of Figures ............................................................................................................ ix

Introductory Comments .......................................................................................... 1

Study 1: Children's Identification of Questions from Rising Terminal Pitch ............. 6
  Introduction ............................................................................................................... 7
  Experiment 1 .......................................................................................................... 10
    Method ............................................................................................................... 10
    Results and Discussion ..................................................................................... 15
  Experiment 2 .......................................................................................................... 19
    Method ............................................................................................................... 19
    Results and Discussion ..................................................................................... 21
  General Discussion ............................................................................................... 24

Study 2: Children’s and Adults’ Categorization of Questions from Terminal Pitch Contours .... 28
  Introduction .......................................................................................................... 29
  Method ............................................................................................................... 31
  Results ............................................................................................................... 36
  Discussion ........................................................................................................... 41


List of Tables

Table 1. Sentences Used in Experiments 1 and 2 (Study 1) ............................................................. 12
Table 2. Sentences Used in the Identification and Discrimination Tasks (Study 2) ......................... 33
Table 3. Stimulus Sentences (Study 3) .................................................................................................. 51
Table 4. Mean Duration and Fundamental Frequency of Adult- and Child-Directed Utterances (Study 3) ........................................................................................................................................ 51
Table 5. Proportions of Participants Identifying Questions Above Chance After 1-5 Words (Study 3) ........................................................................................................................................ 59
Table 6. Mean Statement Responses and Age (Study 3) ........................................................................ 60
List of Figures

Figure 1. $F_0$ contour of a typical question and statement from Experiment 1 (Study 1) ............ 13

Figure 2. Black-and-white versions of asking (left) and telling photos (right) for the male talker (Study 1) ........................................................................................................................................ 14

Figure 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 (Study 1). Error bars are standard errors ........................................................................................................................................ 16

Figure 4. Performance in the test sessions of Experiment 1 (natural utterances) and Experiment 2 (low-pass filtered utterances) as a function of age group (Study 1). Error bars are standard errors ........................................................................................................................................ 17

Figure 5. Sample $F_0$ contours for the eight steps of one utterance with question stem (upper panel) and statement stem (lower panel) (Study 2) .................................................................................................................. 35

Figure 6. Probability of question responses as a function of step and age group (Study 2) ....... 37

Figure 7. Range of indecision for children and adults (Study 2) ................................................. 40

Figure 8. Performance in Experiment 1 (adults) and Experiment 2 (children) as a function of speech register and utterance increments (Study 3) ........................................................................................................... 54

Figure 9. Mean number of words required to identify questions and statements as a function of age and utterance number (Study 3) .............................................................................................................. 61

Figure 10. $F_0$ contours and ToBI annotations from sample adult- and child-directed utterances (Study 3) .............................................................................................................................................. 62
Introductory Comments

Melody characterizes somewhat different aspects of speech and music. Musical melodies are defined by relationships between consecutive tones in terms of their pitches and durations (Dowling & Fujitani, 1971; Gfeller et al., 2002; Halpern, 1984; Kidd, Boltz, & Jones, 1984; Kong, Cruz, Jones, & Zeng, 2004; Monahan & Carterette, 1985; Palmer & Krumhansl, 1987; Sturges & Martin, 1974). The available evidence indicates that pitch patterns (i.e., changes in fundamental frequency or F0) make greater contributions to the distinctiveness of melodies than do rhythmic patterns (Hébert & Peretz, 1997; White, 1960). For example, adults readily recognize melodies in the absence of rhythm or relative timing cues (Galvin, Fu, & Nogaki, 2007; Kang et al., 2009; Kong et al., 2004), but not in the absence of pitch cues (Hébert & Peretz, 1997).

Speech melodies commonly refer to intonation patterns, or the rising and falling pitch levels of phrases or utterances (e.g., Bolinger, 1986, 1989). Such melodies are defined primarily by their contours, or directional changes in pitch (upward, downward, or the same). For musical melodies, however, pitch relations are defined more precisely by intervals, or the exact pitch distance between consecutive tones. In Western music, intervals are quantified in logarithmic scales using semitones, each semitone corresponding to one twelfth of an octave.

Pitch is also used very differently in speech and music. In contrast to the discrete, sustained pitches of musical melodies, pitch varies almost continuously in speech (Bolinger, 1986, 1989; Cruttenden, 1997), with voices gliding upward and downward to a greater or lesser degree. Furthermore, musical melodies use specific pitch values from a prescribed pitch set (i.e., scales), as determined by cultural conventions, so that one pitch functions as the central or reference tone for all other pitches (Krumhansl, 1990). There are no comparable constraints on
the component pitches of speech, although directional changes in pitch are associated with the speaker’s intentions (Bolinger, 1986, 1989).

Changes from one sound to another also occur much more rapidly in speech than in music. For example, English has roughly 12.5 perceptible sounds per second (Miller, 1982), in contrast to 2–5 in music (Drake & Bertrand, 2001). The sustained pitches of music may account for the small range of acceptable variation relative to speech. For example, semitone deviations are perceived as errors in music (Drayna, Manichaikul, de Lange, Snieder, & Spector, 2001; Warrier & Zatorre, 2002) but are rarely meaningful in speech, even in tonal languages (Francis, Ciocca, Ma, & Fenn, 2008).

One example of the use of speech melody is to signal a speaker’s intention—to request rather than transmit information by manipulating the terminal pitch contours that distinguish statements from yes/no questions. Yes/no questions differ from wh-questions (i.e., those beginning with who, what, where, when, and why) in that they seek confirmation (i.e., yes or no) rather than information from the addressee. Yes/no questions usually invert the subject and verb (e.g., Is John going?), but the change in structure is not obligatory (e.g., John is going?; Gunlongson, 2002). Yes/no questions without such inversion are designated echoic or declarative. In yes/no questions, including declarative versions, the pitch contour rises at the end of the utterance, which contrasts with the falling pitch contour at the end of statements (Bolinger, 1986; Gårding & Abramson, 1965; Hadding-Koch & Studdert-Kennedy, 1964; Ladd, 1996; Studdert-Kennedy & Hadding, 1973).

Although adults typically interpret utterances with rising and falling terminal pitch contours as questions and statements, respectively (Cruttenden, 1981; Eady & Cooper, 1986; Hadding-Koch & Studdert-Kennedy, 1964; Gårding & Abramson, 1965; Studdert-Kennedy & Hadding, 1973), the situation is less clear for children. For example, 4-year-old children often
omit rising pitch contours when producing yes/no questions (Patel & Brayton, 2009; Patel & Grigos, 2006; Snow, 1998, 2001). Moreover, there is limited research on children’s use of intonation as a means of identifying the speaker’s questioning intentions. In the two studies that examined children’s perception of questions and statements, they were considerably less proficient than adults at differentiating questions and statements based on intonation alone (Doherty, Fitzsimons, Asenbauer, & Staunton, 1999; Gerard & Clément, 1998). Their difficulty interpreting utterances with rising pitch contours as yes/no questions is unlikely to stem from limitations in the discrimination of pitch patterns, because the differentiation of rising from falling pitch contours is evident in infancy (Frota, Butler, & Vigário, 2014; Karzon & Nicholas, 1989; Nazzi, Floccia, & Bertoncini, 1998). As children master the complexities of language use, they eventually learn to categorize those contours and link the categories to different intentions.

Although young children readily acquire conventional labels for objects and attributes (e.g., big, small), they often have difficulty with conventional spatial labels for pitch (e.g., high, low) and pitch direction (rising/falling, up/down; Fancourt, Dick, & Stewart, 2013; Hair, 1977). With relatively little training and feedback, however, even 5-year-olds succeed in labeling small directional changes in pitch (Stalinski, Schellenberg, & Trehub, 2008). Little is known, however, about whether they can use information about pitch contours in utterances to discern the speaker’s requesting or informing intentions.

The overall goal of the present investigation was to shed light on developmental changes in the identification of declarative questions on the basis of intonation or speech melody. The first of three studies aimed to ascertain the earliest age at which children could reliably identify naturally produced declarative questions and statements and the age at which they achieved adult-like performance. The second study examined the magnitude of the terminal pitch contour that was necessary for consistent categorization of yes/no questions and statements. Although
terminal contours provide the most salient acoustic cues to questions and statements (Bolinger, 1986; Ladd, 1996), there has been little descriptive detail about such contours other than their pitch directional differences (upward for questions; downward for statements). Accordingly, this study focused on the magnitude of the pitch excursion that is necessary to sustain reliable categorization in children and adults.

Once it is established that children are able to identify naturally produced questions and statements, it becomes possible to inquire, as in the third study, about developmental changes in the ability to use pre-terminal pitch cues to identify declarative questions and statements. The ability to use such cues has been documented for Dutch (van Heuven & Haan, 2000), Spanish (Face, 2005), and French (Vion & Collas, 2006) adults, but not for English adults. One would expect children to have more difficulty than adults with pre-terminal cues to questions, which are considerably less salient than terminal pitch cues.

There are enormous cognitive changes in early and middle childhood, which are likely to influence children’s ability to discern a speaker’s questioning intentions on the basis of intonation alone. For example, age-related improvements in cognitive flexibility (Davidson, Amso, Anderson, & Diamond, 2006; Deak, 2003; Ceci & Howe, 1978; Hermer-Vazquez, Moffet, & Munkholm, 2001) involve suitable adjustments in behaviour to various situations, and the ability to focus on relevant stimuli or specific aspects of stimuli (i.e., selective attention; Doyle, 1973; Maccoby & Konrad, 1966; Pearson & Lane, 1991; Plude, Enns, & Brodeur, 1994). Such improvements are bound to facilitate children’s use of intonation to interpret the intentions of speakers. Although one might reasonably expect younger children to perform more poorly on question-identification tasks because of their lesser cognitive capabilities and experience with language, it is nevertheless important to specify the time course of developmental improvements, ultimately relating them to concurrent changes in other skills, including other aspects of prosodic
perception such as the perception of emotion in speech. The goal here was to document age-related differences in (1) the ability to distinguish statements from questions on the basis of naturally occurring changes in pitch (Study 1), (2) the size of terminal prosodic cues that allow for such discrimination (Study 2), and (3) the ability to distinguish statements from questions on the basis of pre-terminal prosodic cues (Study 3).
Study 1: Children's Identification of Questions from Rising Terminal Pitch

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, 5- to 10-year-old children and adults were required to judge utterances as questions or statements on the basis of intonation alone. Children 8 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.
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 (e.g., It’s snowing?) are marked exclusively by prosodic cues. The principal cue to declarative questions is a pronounced rise in terminal fundamental frequency ($F_0$) in contrast to falling $F_0$ for statements (Cruttenden, 1981; Eady & Cooper, 1986; 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 (Fróta et al., 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 4-year-olds, the utterances are often marked by final-syllable lengthening rather than F₀ 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, 2002). Nevertheless, the available production data imply that 5-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 5- to 10-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 5-year-old children differentiate declarative questions from statements in a same–different task (Doherty et al., 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 favourite melody from its timbre (i.e., instrument) than from its rising or falling pitch contour (Creel, 2014). Moreover, 5- and 6-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 et al., 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 4- to 9-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, 5- and 7-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 5 to 10 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 utterance low-pass filtered utterances to obscure the content.
Experiment 1

Method

Participants. The final sample consisted of 122 participants, including 30 5- and 6-year-olds (14 girls, 16 boys; $M = 6;0$, range = 5;0–6;11), 31 7- and 8-year-olds (10 girls, 21 boys; $M = 8;1$, range = 7;0–8;11), 32 9- and 10-year-olds (17 girls, 15 boys; $M = 10;1$, range = 9;0–10;11), and 29 adults (21 women, 8 men; $M = 18.52$ years, $SD = 1.27$). 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 5-year-old), parent-reported developmental delay (one 5-year-old, one 7-year-old, one 8-year-old), failure to meet the criterion during the training phase (one 6-year-old), and scores that were more than two $SD$s below the mean for their age group (a common predetermined criterion that affected two 7-year-olds and two 10-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 x 4 speakers x 2 versions). High-quality digital sound files (44.1 kHz, 16-bit, mono) created 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. 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 5- and 6-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 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.

Table 1

*Sentences Used in Experiments 1 and 2 (Study 1)*

<table>
<thead>
<tr>
<th>Sentence</th>
</tr>
</thead>
<tbody>
<tr>
<td>The cat ran away</td>
</tr>
<tr>
<td>She lost her shoes</td>
</tr>
<tr>
<td>Mom made it</td>
</tr>
<tr>
<td>It’s snowing</td>
</tr>
<tr>
<td>You found it</td>
</tr>
<tr>
<td>Mom went to the store</td>
</tr>
<tr>
<td>It’s bedtime</td>
</tr>
<tr>
<td>He’s watching TV</td>
</tr>
<tr>
<td>He’s in the car</td>
</tr>
<tr>
<td>You’re staying home</td>
</tr>
</tbody>
</table>
Figure 1. $F_0$ contour of a typical question and statement from Experiment 1 (Study 1).
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.

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 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 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 ($d' = \text{z[hit rate]} - \text{z[false-alarm rate]}$). The maximum $d'$ score was 4.5. Raw data (percent correct) and standard errors are illustrated in Figure 4 separately for each age group.

![Figure 3](image)

**Figure 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 (Study 1). Error bars are standard errors.

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 ($F(3,118) = 35.78, p < .001, \eta^2 = .48$). Follow-up
pairwise comparisons (Tukey HSD) revealed that the performance of 5- to 6-year-olds was significantly poorer than all other age groups ($p < .001$), and that 7- to 8-year-olds performed more poorly than adults ($p = .001$). The performance of 9- to 10-year-olds did not differ significantly from adults or from 7- to 8-year-olds ($p > .1$).

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

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 5- and 6-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 choose the telling response 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 5- to 6-year-olds and 1 7- to 8-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 ($ps < .001$).

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 ($rs(n = 92) = -0.48, p < .001$).

In sum, the results indicated that children and adults identified questions based on prosodic information alone. Nevertheless, 5- to 6-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 7- to 8-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 5- and 6-year-olds (18 girls, 15 boys; $M = 6;1$, range = 5;0–6;9), 34 7- and 8-year-olds (15 girls, 19 boys; $M = 7;11$, range = 7;0–8;10), 31 9- and 10-year-olds (14 girls, 17 boys; $M = 10;1$, range = 9;2–10;9), and 28 adults (20 women, 8 men; $M = 18.43$ years, $SD = 1.26$). Recruiting and inclusion criteria were the same as in Experiment 1. An additional 19 participants were tested but excluded because of parent-reported developmental delays (one 10-year-old), failure to meet the criterion during the training phase (seven 5-year-olds and one 6-year-old), failure to pay attention to the task (three 5-year-olds and four 6-year-olds), and scores that were more than two $SD$s below the mean for
their age group (one 8-year-old, one 9-year-old, and one 10-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, 2002) 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 ($M = 3.40$, $SD = 1.49$) than in Experiment 1 ($M = 2.91$, $SD = 1.33$), $t(187.55) = 2.37$, $p = .019$ (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 ($F(3,122) = 33.13$, $p < .001$, $\eta^2 = .45$). The performance of 5- to 6-year-olds was significantly worse than that of all other age groups ($ps < .001$), and the performance of 7- to 8-year-olds was worse than that of 9- to 10-year-olds ($p = .038$) and adults ($p < .001$). The performance of 9- to 10-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 ($\chi^2(3, N = 126) = 54.70, p < .001, \Phi = .66$). All adults performed above chance levels, as did all but one of the 9- to 10-year-olds, and all but three of the 7- to 8-year-olds. For the 5- to 6-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 ($F(3,240) = 67.92, p < .001, \eta^2 = .43$) and stimulus type ($F(1, 240) = 28.86, p < .001, \eta^2 = .06$). 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 ($F(3,122) = 37.14, p < .001, \eta^2 = .48$), reflecting age-related improvement, and a small but significant effect of test phase ($F(1,122) = 8.40, p = .004, \eta^2 = .06$). Performance was better during the second half of the test session ($M = 33.63, SD = 7.49$) than in the first half ($M = 32.62, SD = 7.73$), 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%), $t(237.04) = 3.19, p = .002$ (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 ($M = 40.93; p = .011$). 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 5-to 6-year-olds, 6 7- to 8-year-olds, and 1 9- to 10-year-old. The proportion of participants with this bias varied reliably across age groups ($\chi^2(3, N = 126) = 25.42, p < .001, \Phi = .45$).

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 ($rs(n = 98) = -0.50, p < .001$).

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 (5 to 10 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 5- and 6-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 7- and 8-year-olds performed no differently than 9- and 10-year-olds on natural utterances but they performed more poorly on filtered utterances. Finally, the 9- and 10-year-olds performed no differently than adults.

The findings from the 5- and 6-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 6-year-olds’ difficulty with conventional verbal labels for rising and falling pitch (Costa-Giomi & Descombe, 1996).

It is likely 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 5- to 6-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 5- and 6-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, 2002), 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 5- and 6-year-olds, as observed in the present study, it is likely that they would comprehend the gist of the message.

In conclusion, children between 5 and 6 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.
Study 2: Children’s and Adults’ Categorization of Questions from Terminal Pitch

Abstract

The present study had two goals: (1) to compare children’s and adults’ identification of declarative questions and statements on the basis of terminal cues alone, and (2) to ascertain whether adults perceived questions and statements categorically. Children (8–11 years, n = 41) and adults (n = 21) judged utterances as statements or questions from sentences with natural statement and question endings and with manipulated endings that featured intermediate pitch values. Adults were also tested on their discrimination of the utterances. Children’s judgments shifted more gradually across categories than those of adults, but their category boundaries were comparable. Adults’ discrimination and identification performance was consistent with categorical perception of statements and questions. Although questions and statements are typically characterized by rising and falling terminal pitch contours, listeners in the present study identified utterances as statements even when the terminal pitch rose by as much as four semitones. The terminal contours of all utterances included falling and rising portions, but the duration of the rising portion was larger for questions and the duration of the falling portion was longer for statements. Parametric manipulations of pitch and duration are necessary to specify their joint contribution to the identification of statements and questions.
Prosody encompasses the rhythm, stress, and intonation (pitch variation) in utterances. It conveys information about a speaker’s emotional state (Banse & Scherer, 1996; Scherer, 1986), sincere or ironic intentions (Kreutz & Glucksberg, 1989; Pexman & Glenwright, 2007), focus of interest (Tom bought a car vs. Tom bought a car), and utterance form (e.g., statement vs. question). The present study was concerned with utterance form, specifically with children and adult listeners’ ability to differentiate questions from statements solely on the basis of the terminal pitch contour.

In English, the principal cue to yes/no questions, as demonstrated by perception and production studies, is considered to be rising terminal intonation, in contrast to falling terminal intonation for statements (Cruttenden, 1981; Gårding & Abramson, 1965; Eady & Cooper, 1986; Studdert-Kennedy & Hadding, 1973). Young children’s questions often lack the characteristic terminal pitch rise (Patel & Grigos, 2006; Snow 1994, 1998, 2001), which makes it difficult to identify their questions on the basis of prosodic cues alone (Patel & Brayton, 2009). In addition to the primary cues that distinguish declarative questions from statements, there are secondary cues that vary across languages (Face, 2005; Falé & Faria, 2006; Heeren, Bibyk, Gunlogson, & Tanenhaus, 2015; van Heuven & Haan, 2000; Vion & Colas, 2006), which demonstrates multiple cues can be used to identify questions in naturally produced speech.

Nevertheless, forcing listeners to rely on terminal cues alone can provide information about the specific terminal cues that differentiate questions from statements. It can also reveal whether listeners’ judgments shift gradually with increasing difference or abruptly after a certain criterion or threshold value. In other words, are the intonation patterns underlying questions and statements perceived categorically or continuously?

In general, categorical perception is established by demonstrating that the discrimination of acoustic stimuli along various acoustic continua (e.g., /ba/ to /pa/) is considerably more
difficult when the comparison stimuli are from the same category (e.g., different examples of /ba/) than from different categories (e.g., /ba/ to /pa/) even though the magnitude of acoustic differences (e.g., voice-onset time) may be identical for within- and between-category comparisons (e.g., Liberman, Harris, Hoffman, & Griffith, 1957). When the discrimination peak and labeling threshold converge, perception is considered to be categorical.

Speakers of tonal languages perceive level and rising tones categorically (Francis, Ciocca, & Ng, 2003; Hallé, Chang, & Best, 2004; Wang, 1976; Xu, Gandour, & Francis, 2006). In these languages, F0 height and contour signal lexical distinctions. In Mandarin, for example, the syllable /ma/ means hemp when produced with a rising tone but mother when produced with a level tone. To date, however, evidence about categorical perception of intonation in non-tonal languages is mixed. Stress or emphasis in English is labeled categorically but not perceived categorically (Ladd & Morton, 1997). In Swedish, emotional prosody is thought to be perceived categorically (Laukka, 2005). For English (Hutchins, Gosselin, & Peretz, 2010) and German (Meister, Landwehr, Pyschny, Walger, & Wedel, 2009), there are suggestions of a sharp statement–question boundary, which is also consistent with categorical perception.

The principal goal of the present study was to compare children’s and adults’ identification of declarative questions and statements on the basis of terminal pitch cues alone. Natural-sounding utterances were achieved by using naturally produced questions and statements and electronically altering the terminal contours. As noted, listeners can identify statements and questions from pre-terminal cues, but it is unclear whether such cues are used when more salient terminal pitch cues are available. The latter question was addressed by using pre-terminal cues from naturally produced questions for half of the stimuli and from naturally produced statements for the other half. In essence, the manipulated terminal pitch contour was appended to question or statement stems. If listeners profit from pre-terminal cues when terminal cues are present, they
should identify utterances more accurately when sentence stems and endings are congruent (e.g., question stems with question endings) rather than incongruent (e.g., question stems with statement endings).

Because young school-age children have lesser sensitivity than adults to the cues that distinguish declarative questions from statements (Study 1), one might expect them to require larger terminal pitch contrasts to distinguish questions from statements. In addition, children might exhibit a more gradual shift from statement to question labeling and a different location of the boundary between statement and question categories.

A secondary goal was to ascertain whether adults’ perception of statements and questions is categorical. To this end, adults participated in a discrimination task as well as an identification task. Children completed the identification task only because it was not feasible for them to complete both tasks in a single laboratory visit.

**Method**

**Participants**

Participants were 21 college students (14 women, 7 men; \( M = 21.29 \) years, \( SD = 4.05 \)) and 41 children between 8 and 11 years of age. A median split was used to divide the children into younger and older groups, resulting in younger children from 8;0 to 9;5 years (\( n = 20, 11 \) boys, 9 girls, \( M = 8;9 \) years, \( SD = 6 \) months). The older children ranged from 9;9 to 10;11 years (\( n = 21, 9 \) boys, 12 girls, \( M = 10;5 \) years, \( SD = 4 \) months). These age groupings were motivated by developmental changes in children’s understanding of intonational cues to statements and questions (Study 1).

Adults were born in Canada or living in Canada before 8 years of age and had normal hearing, according to self-report. Children were born in Canada and had normal hearing and development, according to parental report. An additional eight children were tested but excluded
because of failure to identify natural (i.e., unaltered) questions above chance levels (16 or fewer correct out of 24, \( n = 7 \)) or parent-reported developmental delay (\( n = 1 \)). Adults received partial course credit for participation; children received a small gift.

**Apparatus and Stimuli**

The stimuli consisted of recorded question and statement versions of each of four utterances (see Table 2) produced by a 27-year-old woman who was a native speaker of Canadian English. For all utterances, she attempted to maintain relatively constant loudness and speaking rate and relatively level pitch until the terminal fundamental frequency (\( F_0 \)) contour. A digital audio editor was used to divide each utterance into a stem (entire utterance except for the final two syllables) and ending (final two syllables). The amplitude of each sound clip was normalized to 75 dB SPL, and syllable durations were equalized across statement and question versions of each utterance. Sample stimuli are available in supplementary materials.

An examination of natural utterance endings revealed that the terminal pitch contours of questions and statements featured a final syllable with initially falling pitch followed by rising pitch. The pitch range of the entire set of question utterances was 168.6–397.5 Hz (\( M = 225.4 \) Hz) and the corresponding range of the entire set of statements was 162.5–316.7 Hz (\( M = 215.9 \) Hz). For natural statement endings, pitch fell initially by a substantial amount (\( M = 6.5 \) semitones), then rose much less (\( M = 1.5 \) semitones), with the falling portion being of longer duration (\( M = 305 \) ms) than the rising portion (\( M = 165 \) ms). For natural question endings, pitch fell very modestly (\( M = 0.5 \) semitones), then rose substantially (\( M = 12.5 \) semitones), with the falling portion being considerably shorter in duration (\( M = 105 \) ms) than the rising portion (\( M = 352 \) ms). In other words, statements and questions contrasted in the location, extent, and duration of the rising and falling portions, even though statement endings were primarily falling and question endings were primarily rising.
Manipulations of the final syllable alone resulted in unnatural-sounding utterances. Examination of the penultimate syllable of the natural utterances revealed falling pitch levels that were more pronounced for questions ($M = 4.1$ semitones) than for statements ($M = 2.3$ semitones). Thus, we manipulated the contour of the two final syllables to ensure that the utterances sounded natural (as in Hutchins et al., 2010).

Sentence endings were analyzed by means of Praat version 5.3.68 (Boersma, 2002) to reveal the $F_0$ contour with a series of time points separated by intervals of 0.01 s. Points not deemed crucial to the contour were deleted. New points were then created so that the statement and question versions had $F_0$ points at the same temporal locations. For example, after the deletion of superfluous pitch points, if the statement ending had an $F_0$ point at 0.84 s of the sample but the question ending for the same sentence did not, a new $F_0$ point was created at 0.84 s in the question ending. Subsequently, we calculated the difference in cents (1 semitone = 100 cents) between the matching points of the question and statement versions of each sentence. This difference was divided into eight steps, and the resulting values were used to create modified intonation contours. The manipulations based on these values created intonation contours that shifted gradually from statement to question contours.
Stimuli with the modified intonation contours were generated using the F0 Synchronous Overlap Add (PSOLA) method in Praat. Because large F0 manipulations can make utterances sound unnatural, the natural statement ending was used to create half of the stimuli (steps 1 through 4) and the natural question ending was used to create the other half (steps 5 through 8). Because amplitude and pacing were identical across the two versions, judgments were based solely on the intonation contours. Finally, to control for differences in prosodic cues that precede the penultimate syllable, two versions of each step were created, one using the question stem (the portion of the sentence prior to the penultimate syllable) and one using the statement stem. The resulting stimulus set consisted of 64 utterances (4 sentences x 8 steps x 2 stems). An example of the manipulations is shown in Figure 5.

The sentences were recorded in a quiet room with a tube microphone (Apex 460) connected directly to a MacBook Pro (OS 10.7). Testing was conducted in a double-walled, sound-attenuating chamber (Industrial Acoustics Corporation Co., Bronx, NY). A computer workstation and amplifier (Harmon-Kardon 3380, Stamford, CT) outside of the booth interfaced with a 17-in. (43.2 cm) touch-screen monitor (Elo LCD TouchSystems, Berwyn, PA) and two wall-mounted loudspeakers (Electro-Medical Instrument Co., Mississauga, ON) inside the booth. The touch-screen monitor was used to present instructions and record participants’ responses. The loudspeakers were mounted at the corners of the sound booth, each located at a distance of .76 m and 45° azimuth from the participant, with the touch-screen monitor placed at the midpoint. An interactive computer program created with Affect4 software (Spruyt et al., 2010) presented the sentences and recorded responses via the touch-screen. All stimuli were played at a comfortable listening level of approximately 65 dB SPL.
Figure 5. Sample F₀ contours for the eight steps of one utterance with question stem (upper panel) and statement stem (lower panel) (Study 2).

Procedure

Adults completed the identification and discrimination tasks in a single session, and children completed the identification task only. The identification task began with a short
familiarization phase during which listeners heard a sample question and statement (not from the stimulus set used in testing). In the subsequent test phase, each of the 64 stimulus sentences was presented three times in random order with the constraint that the same sentence content did not occur on successive trials. Adults and children were required to indicate whether each sentence was a question or statement, such that they made 192 (3 x 64) responses in total.

In the discrimination task, adults heard two sentences that were identical or that differed by a single step and judged whether they were the same or different. The sentences were arranged in three types of pairs: same (1–1, 2–2, and so on), lower-higher step (1–2, 2–3, 3–4, 4–5, 5–6, 6–7, and 7–8), and higher-lower step (2–1 to 8–7), resulting in a total of 22 pairs per sentence (8 same, 7 lower-higher, 7 higher-lower) and a grand total of 88 pairs. Each pair was presented twice, for a total of 176 trials. The discrimination task also began with a short familiarization phase in which listeners heard an example of a same and different pair (not from the test set).

The identification tasks for adults and children were nearly identical, but the children’s task was modified to make it more entertaining. Specifically, the children’s task was presented as a game in which the goal was to collect pieces of a puzzle (i.e., corresponding to different parts of an image that appeared on screen) that mirrored their progress toward task completion. A piece of the puzzle appeared after each set of 12 trials regardless of the children’s responses, and 16 pieces completed the puzzle. In addition, a short bell-like sound was presented after each response.

Results

Identification Task

For each participant, we calculated the proportion of times stimuli were labeled as questions separately for each step and each stem, such that each participant had 16 scores.
Standard repeated-measures analysis of variance (ANOVA) has the assumption of sphericity, which was violated for the main effect of step, \( p < .001 \), and the interaction between stem and step, \( p < .001 \). Thus, the first analysis was a mixed-design multivariate analysis of variance (MANOVA), which has no such assumption. The MANOVA had two repeated measures (two stems, eight steps) and one between-subjects variable (age: younger children, older children, adults). This analysis revealed no main effect of stem, \( p = .704 \), no two-way interaction between stem and age group, \( p = .178 \), or between stem and step, \( p = .082 \), and no three-way interaction, \( p = .730 \) (\( p \)-value for Wilk’s lambda), which meant that sentence stem had no discernible effect on response patterns. Consequently, stem was not considered further, and responses for stimuli with question and statement stems were combined in subsequent analyses. Descriptive statistics are illustrated in Figure 6.

*Figure 6.* Probability of question responses as a function of step and age group (Study 2).
The MANOVA revealed a robust main effect of step, $F(7, 53) = 1398.97, p < .001, \eta^2 = .995$, and an interaction between step and age group, $F(14, 106) = 2.48, p = .004, \eta^2 = .247$. As shown in Figure 6, compared to both groups of children, adults were more consistent at labeling stimuli at steps 6 to 8 as questions.

For each participant, we derived a new value that quantified how well defined their boundary was between statements and questions. Specifically, we used logistic regression to fit a logit curve to the participant’s responses (0: statement or 1: question) as a function of step number. The resulting slopes varied such that a steep slope indicated a clear or sharp boundary, whereas a shallower slope indicated that listeners were less consistent in the step at which they switched from statement to question responses across trials. A one-way ANOVA revealed significant differences in slope across age groups, $F(2, 59) = 7.92, p = .001, \eta^2 = .212$. Pairwise comparisons (Tukey HSD) indicated that adults’ average slope ($M = 2.17, SD = 1.01$) was significantly steeper than the slopes of older children ($M = 1.41, SD = 0.64$), $p = .006$, and younger children ($M = 1.29, SD = 0.59$), $p = .002$, but the two groups of children did not differ, $p = .875$. Thus, no developmental differences were evident in children between 8 and 11 years of age.

We then compared the location of the question/statement boundary in adults and children by calculating the point at which participants responded question for 50% or more of the trials (i.e., when the logit function of question responses crossed 0.5 on the $y$-axis). An ANOVA revealed no age-related differences in the location of the midpoint, $F < 1$. For adults, older children, and younger children, the average midpoints were 4.93 ($SD = 0.54$), 5.10 ($SD = 0.69$), and 5.19 ($SD = 0.68$), respectively. In sum, listeners in all age groups changed from statement to question responses at approximately step 5, but adults did so more consistently.
Acoustic analyses were conducted to examine the terminal intonation contour of steps 4 and 5 (i.e., the steps at which judgments shifted from statement to question). In step 4, the average pitch fall—from the pitch level at onset to the pitch level at the lowest portion of the fall—was 2.9 semitones, occurring over the course of 193 ms, and the average terminal pitch rise—from the lowest portion of the pitch declination to the highest pitch value at the final rising portion—was 4.9 semitones, occurring over 285 ms. In step 5, the average pitch fall and rise were 1.8 semitones and 6.2 semitones, respectively, with corresponding durations of 176 and 289 ms.

We also calculated a range of indecision for each participant, which represented the number of adjacent steps that were inconsistently labeled as statement or question. Because our sample included children, consistency was defined as a proportion of question responses .2 or less or .8 or more and inconsistency as a proportion of question responses greater than .2 but less than .8. The data are illustrated in Figure 7. Because the older and younger children did not differ in all previously reported analyses, the two groups of children were combined for further analyses. As shown in Figure 7, the range of indecision varied as a function of age (children vs adults), which was documented with a 2 x 4 chi-square test of independence with group (adults, children) and range of steps (0, 1, 2, 3 or more) treated as categorical variables, \( \chi^2(3, N = 62) = 8.07, p = .045, \phi = .361 \). The mean range of indecision was 1.47 steps (SD = .81) for adults and 2.00 (SD = .84) for children.

**Discrimination Task**

For the discrimination task conducted with adults, we calculated performance for each comparison (i.e., steps 1 and 2, steps 2 and 3, and so on) by subtracting the number of false alarms, or identifying a same pair as different, from the number of hits, or correctly identifying a different pair as different. The probability of identifying each pair was calculated by dividing the
resulting score (hits minus false alarms) by 16, which was the maximum number of hits for each pair.

![Figure 7](image-url)

**Figure 7.** Range of indecision for children and adults (Study 2).

The results revealed a peak at steps 4 and 5, indicating highest accuracy for that comparison. Moreover, adults’ identification data revealed that the probability of labeling a sentence as a question crossed the 0.5 threshold between steps 4 and 5. Taken together, these findings are consistent with categorical perception of statements and questions. To confirm the possibility of categorical perception, two paired-sample t-tests were conducted. The first test compared the mean probability of indicating that pairs 4 and 5 were different to the mean probability of indicating that the adjacent pairs were different (i.e., steps 3–4 and 5–6), and the second test compared the mean probability of discriminating steps 4–5 to the mean probability of discriminating all other pairs. The results indicated that the probability of discriminating steps 4 and 5 ($M = .52, SD = .20$) was higher than the probability of discriminating steps 3–4 and 5–6 ($M$
= .36, SD = 22), \( t(20) = 6.02, p < .001 \), as well as the probability of discriminating the other pairs \( (M = .41, SD = 22), \( t(20) = 3.58, p = .002 \). Thus, adults seemed to exhibit categorical perception of terminal pitch, even though they found it difficult to discriminate between the terminal contours. The average discrimination score (hits minus false alarms) across all sentence pairs was 42.2%, and the average difference between the final pitch values of the intonation contours was 1.7 semitones.

**Discussion**

The principal goal of the present study was to compare children’s and adults’ identification of declarative questions and statements on the basis of terminal pitch contours. A secondary question was to ascertain whether adults perceive terminal contours categorically. Because children are less sensitive than adults to the acoustic cues that distinguish questions from statements (Study 1), we expected children’s and adults’ perceptual boundary between statements and questions to differ in sharpness and location. Specifically, we expected children to transition more gradually than adults from one response category to another, regardless of the boundary location.

In fact, children shifted more gradually than adults from one category to another, reflecting their greater uncertainty about the distinctions between statements and questions, but the location of the category boundary was stable across age, presumably reflecting broad similarity in adults’ and children’s criteria for statements and questions. Children’s uncertainty about questions and statements in the region of the category boundary is consistent with previous evidence of children’s lesser accuracy than adults in identifying questions from natural utterances (Study 1).

On the basis of listeners’ ability to identify statements and questions from cues appearing early in an utterance (Face, 2005; van Heuven & Haan, 2000; Vion & Colas, 2006), the absence
of performance advantages for congruent sentence stems and endings (e.g., question stem, question ending) over incongruent stems and endings (e.g., question stem, statement ending) is surprising. It is likely, however, that pre-terminal cues, which are much less salient than terminal pitch cues, are useful primarily in the absence of terminal pitch cues (Majewski & Blasdell, 1968), as in gating procedures (Face, 2005; van Heuven & Haan, 2000; Vion & Colas, 2006).

Adults’ performance on the identification and discrimination tasks is consistent with categorical perception of statements and questions in that the boundary for statement/question identification and the discrimination peak occurred between the fourth and fifth steps. Although the discrimination peak was relatively small, it differed significantly from the discrimination of adjacent pairs (i.e., steps 3 vs. 4 and 5 vs. 6) as well as other pairs. These findings mirror the categorical perception of level and rising tones by tone-language speakers, for whom those tones have linguistic significance, but not the continuous perception by those who do not speak or understand a tone language (Francis et al., 2003; Hallé et al., 2004; Xu et al., 2006; Wang, 1976).

Acoustic analyses revealed that children as well as adults were more likely to label utterances as statements when the terminal pitch fall ranged from 6.3 to 2.9 semitones, and they were more likely to label them as questions when the terminal pitch rise ranged from 6.2 to 12.4 semitones. It was not surprising that listeners identified utterances as questions when the terminal pitch rise was six semitones. In two studies examining the perception of terminal pitch, adult listeners labeled a two-syllable word (popcorn) as a question approximately 80% of the time when the F0 rose by six or more semitones over the course of the word (Chatterjee & Peng, 2008; Peng et al., 2009).

An unexpected finding in the present study was that children and adults identified utterances as statements when the terminal pitch rose by as much as four semitones. In previous perceptual studies, terminal pitch in natural utterances was reported as falling 4 to 7 semitones
for statements and rising 10 to 12 semitones for questions (Patel, Wong, Foxton, Lochy, & Peretz, 2008; Patel, Foxton, & Griffiths, 2005). In other studies that manipulated terminal contour, listeners labeled utterances with level contours as statements and shifted their judgment from statement to question when the contour rose by as little as two semitones (Chatterjee & Peng, 2008; Peng et al., 2009). Hutchins et al. (2010) found, however, that listeners label some utterances with rising contours as statements. Specifically, four utterances labeled as statements exhibited a rise on the final portion of the last syllable, and the pitch offset of the last syllable was higher than its onset for two of these utterances.

Rising intonation contours are generally associated with yes/no questions, but they are also found in other types of utterances, such as statements with patronizing or self-justifying intentions (Cruttenden, 1985). It is possible, then, that our utterances with a small rise were not perceived as traditional statements or questions. However, the two-option, forced-choice task did not allow listeners to indicate degree of certainty about their judgment, and it did not allow them to indicate that the utterance was neither a question nor a statement. When listeners are required to label utterances as questions or statements, their default option is the statement response (e.g., Vion & Colas, 2006; see also Study 1). Thus, it is possible that listeners labeled some utterances with rising terminal pitch as statements because of uncertainty.

Acoustic analyses revealed systematic differences in the duration of the falling and rising portions of the terminal contour. Specifically, the most salient portion of the terminal contour—the fall in statements and the rise in questions—was extended. In utterances labeled as statements, the duration of the terminal fall ranged from 193 to 305 ms, and the terminal rise ranged from 165 to 285 ms. In utterances labeled as questions, the terminal fall ranged from 105 to 176 ms, and the terminal rise ranged from 289 to 352 ms. The relative duration of these portions of the terminal contour is likely to influence the perception of statements and questions,
although we cannot indicate the importance of duration and its impact on listeners’ judgments because we did not manipulate duration separately from intonation contour.

Another unexpected finding was adults’ modest performance on the sentence discrimination task. Although the terminal pitch of comparison stimuli differed by 1.7 semitones, performance (hits minus false alarms) averaged 42.2% across the seven pairs of stimuli. In the context of isolated tones, adults discriminate pitch differences as small as 0.1 semitones (Micheyl, Delhommeau, Perrot, & Oxenham, 2006; Olsho, Sakai, Turpin, & Sperduto, 1982; Spiegel & Watson, 1984). In the context of melodies, differences of a semitone are readily apparent in familiar melodies (Drayna et al., 2001; Warrier & Zatorre, 2002) and often in unfamiliar melodies as well.

Although some scholars claim that adults’ ability to discriminate pitch in vocal and non-vocal stimuli is comparable (Moore, Estis, Gordon-Hickey, & Watts, 2008), there is also evidence to the contrary. For example, Chinese and English speakers discriminate pitch differences of 1.4 semitones more readily in the context of pure tones than in speech stimuli, with discrimination scores improving by 12 to 18% for non-speech stimuli (Xu et al., 2006). Similarly, listeners require a lesser pitch difference to identify a rising contour in a tone sequence than in a speech sequence (Hutchins et al., 2010). There is additional evidence that irrelevant auditory information can impair selective attention, with adverse effects on cognitive performance, including memory (Banbury, Macken, Tremblay, & Jones, 2001; Neath, 2000). Moreover, children have difficulty identifying emotions conveyed in speech on the basis of prosody alone, especially when utterances have conflicting semantic content (Morton & Trehub, 2001). Thus, it is possible that acoustic information associated with message content interfered with listeners’ processing of the pitch information.
In short, when the terminal contours of utterances were altered electronically in successive steps with natural statements and questions as endpoints, the boundary between statement and question judgments was similar for children and adults, but the shift between response categories was more gradual for children. Although declarative questions and statements are typically described as ending with rising and falling pitch contours, respectively, all natural and manipulated endings in the present study began with a falling portion and ended with a rising portion. The relative duration of the falling and rising portions contributed to their relative salience and to their categorization as questions or statements. The present study systematically manipulated the final pitch excursions of utterances, but systematic manipulations of the duration of falling and rising portions are necessary to specify the joint contribution of terminal pitch changes and their durations to the identity of statements and questions. Greater understanding of these cues to the identification of sentence types could have implications for interventions with special populations such as cochlear implant users (Marx et al., 2015), who experience difficulty with various aspects of prosody (Meister et al., 2009; Nakata, Trehub, & Kanda, 2012), and people diagnosed with autism spectrum disorder (Filipe, Frota, Castro, & Vicente, 2014) and Parkinson's disease (Ma, Whitehill, & So, 2010).
Study 3: When is a Question a Question for Children and Adults?

Abstract

Terminal changes in fundamental frequency provide the most salient acoustic cues to declarative questions, but adults sometimes identify such questions from pre-terminal cues. In the present study, adults and 7- to 10-year-old children judged a single speaker’s adult- and child-directed utterances as questions or statements in a gating task with word-length increments. Listeners of all ages successfully used pre-terminal cues to identify utterance type, often only the initial word, and they were more accurate for child-directed than adult-directed utterances. There were age-related differences in identification accuracy and number of words required for correct identification. Age differences were already apparent on the initial (first five) utterances, confirming adults’ superior explicit knowledge of intonation patterns that signify questions and statements. Adults’ performance improved over the course of the test session, reflecting taker-specific learning, but children exhibited no such learning.
The most salient acoustic differences between declarative questions and statements are the terminal fundamental frequency ($F_0$) contours, which rise in yes/no questions and fall in statements (Bolinger, 1986; Ladd, 1996). Artificially increasing the pitch excursion of utterance-final syllables increases adults’ likelihood of identifying those utterances as questions (Gårding & Abramson, 1965). Even sine-wave patterns with rising pitch contours are identified as questions and those with falling pitch contours as statements (Studdert-Kennedy & Harding, 1973).

Production studies confirm the distinctiveness of terminal pitch contours, but they also reveal differences in pre-terminal positions. In English, for example, successive content words often exhibit rising pitch in questions but not statements, and stressed syllables are typically followed by a pitch decrease in statements but not questions (O’Shaughnessy, 1979). In the present study, we sought to document age-related changes in the ability to differentiate declarative questions from statements on the basis of pre-terminal cues, and to ascertain whether speech register—child- or adult-directed—affects such differentiation.

Gating tasks, which were developed to assess the speed of word identification (Grosjean, 1980), provide a means of determining the minimum information necessary for identifying statements and declarative questions. Listeners judge whether utterances differing only in prosodic cues are questions or statements from incremental increases in syllables or words. Findings from these tasks reveal a different time course of identification across languages. For example, Dutch listeners identify declarative questions after hearing the second accented syllable in an utterance (van Heuven & Haan, 2000). Portuguese listeners identify statements after the first stressed vowel, but they need the final stressed vowel to identify questions (Falé & Faria, 2006). Spanish listeners identify questions after hearing the first content word (Face, 2005), but French listeners do so after hearing the falling $F_0$ contour that precedes the final rise (Vion &
Colas, 2006). Surprisingly, there are no comparable studies with English-speaking children or adults.

Only one study used a gating task to examine children’s identification of statements and declarative questions (Gérard & Clément, 1998). French 5-, 7-, and 9-year-olds judged utterance type from increasing numbers of words (i.e., one word, two words, three, etc.). Even after hearing complete utterances, 5- and 7-year-olds failed to differentiate questions from statements, but 9-year-olds were able to do so. Although children’s performance was well below adult levels, adult gating studies suggest that there are fewer pre-terminal cues to question identity in French (Vion & Colas, 2006) than in Dutch (van Heuven & Haan, 2000) or Spanish (Face, 2005). Moreover, English-speaking children as young as 5 (Study 1) and Spanish- and Mandarin-speaking 4-year-olds (Armstrong, 2012; Zhou, Crain, & Zan, 2012) can differentiate questions from statements on the basis of prosody alone, although they are much less accurate than adults. Note, however, that Mandarin questions are cued by duration and intensity contrasts rather than pitch contrasts (Zhou et al., 2012).

There are also notable cross-language differences in young children’s perception and production of the relevant acoustic distinctions. For example, 21-month-old Spanish-speaking toddlers mark their declarative questions with rising intonation contours (Armstrong, 2012), but English-speaking preschoolers are unlikely to do so (Loeb & Allen, 1993; Patel & Grigos, 2006; Snow, 1994, 1998). Children’s difficulties, when evident, do not stem from perceptual limitations. For example, infants differentiate rising from falling pitch contours (Frota et al., 2014; Karzon & Nicholas, 1989; Nazzi et al., 1998; Soderstrom et al., 2011), and young children can learn to identify small directional differences (up, down) in pitch (Stalinski et al., 2008). Instead, their difficulties seem to arise from interpretive issues with sentence-level prosody, which are resolved progressively with increasing age and cognitive maturity (Cutler & Swinney,
It is also notable that young children perform much better in conversational contexts that provide contextual support (e.g., Doherty-Sneddon & Kent, 1996) than in laboratory contexts in which isolated utterances are presented or elicited (e.g., Patel & Grigos, 2006).

In line with our goal of ascertaining the minimum information required to differentiate question from statement intentions, we asked English-speaking adults (Experiment 1) and 7- to 10-year-old children (Experiment 2) to identify stimuli as questions or statements from the initial word of five-word utterances and from incremental additions of words. In contrast to previous gating studies, which used conventional or adult-directed speech, we used child- and adult-directed speech. Child-directed speech (or ‘clear’ adult-directed speech) features wider $F_0$ range, slower speaking rate, increased pitch and intensity accents, and longer pauses than conventional adult-directed speech (Foulkes et al., 2005; Jacobson et al., 1983; Smiljanić & Bradlow, 2009). These features enhance the intelligibility of messages for children (Fernald & Mazzie, 1991; Liu, Tsao, & Kuhl, 2009; Syrett & Kawahara, 2014), older adults (Caporael, 1981; Masataka, 2002), and non-native speakers (Uther, Knoll, & Burnham, 2007). Because suprasegmental aspects of child-directed speech typically exaggerate cues to the speaker’s intentions, this speaking style should highlight the distinctions between statements and questions. Accordingly, we expected listeners of all ages to differentiate questions from statements earlier (i.e., from fewer words) with child-directed than with adult-directed utterances. We also predicted age-related changes, in line with French-speaking children’s performance on a gating task (Gérard & Clément, 1998) and English-speaking children’s performance on full utterances presented in isolation (Study 1). The performance of adults was examined in Experiment 1 and that of children in Experiment 2.

Our use of a single talker provided an opportunity to examine attunement to her intonation patterns over the course of the test session. Adults often exhibit adaptation to
individual phonetic signatures (Theodore, Myers, & Lomibao, 2015) and conversational style (Pogue, Kurumada, & Tanenhaus, 2016). Children’s ability to perceive accented or atypical speech implies similar attunement, but younger children experience greater difficulty in this regard (Creel, Rojo, & Paullada, 2016). In the present study, adaptation to the speaker’s questioning and stating style would be reflected in progressive improvement in the use of pre-terminal cues.

Experiment 1

Method

Participants. Participants were 45 college students (36 women, 9 men; mean age = 18.82 years, SD = 2.85), who received partial course credit for their participation. Inclusion criteria were Canadian birth or arrival in Canada before 8 years of age and absence of hearing difficulty, according to self-report. Three additional adults were tested but excluded from the final data set because of inattentiveness, as reflected in their failure to identify questions and statements above chance levels (14 or more incorrect out of 40) after hearing the complete sentence.

Apparatus and stimuli. A woman with experience in vocal performance and early-childhood education produced four versions of 10 5-word utterances (see Table 3). The sentence-length utterances were selected on the basis of their prosodic form rather than their syntax or content. Specifically, each sentence began with a stressed syllable, and words in the same position across sentences had the same number of syllables (first word: two syllables; second word: one syllable; third word: two syllables; fourth word: one syllable; fifth word: two syllables). Sentences were produced as declarative questions or statements in an adult-directed or child-directed style. On average, child-directed versions had higher F0, greater F0 range and variability, and longer duration than adult-directed versions (see Table 4). Each sentence was amplitude-normalized at 75 dB SPL using Sound Forge Pro (Version 10.0; Sony, Tokyo, Japan).
Table 3

*Stimulus Sentences (Study 3)*

Judy knows Michael is going.

Lisa seems happy for Alex.

Maybe one apple is enough.

Someone went shopping for balloons.

People played music from Japan.

Puddles make rainy days better.

Donuts with sprinkles are tasty.

Brian is going home today.

Brendan is leaving town again.

Roses are growing on bushes.

---

Table 4

*Mean Duration and Fundamental Frequency of Adult- and Child-Directed Utterances (Study 3)*

<table>
<thead>
<tr>
<th>Speech register</th>
<th>Duration (ms)</th>
<th>$F_0$ mean (Hz)</th>
<th>$F_0$ min (Hz)</th>
<th>$F_0$ max (Hz)</th>
<th>$F_0$ SD (Hz)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adult-directed</td>
<td>2798.00</td>
<td>258.01</td>
<td>138.74</td>
<td>522.60</td>
<td>82.90</td>
</tr>
<tr>
<td>Child-directed</td>
<td>3074.48</td>
<td>276.61</td>
<td>164.08</td>
<td>588.00</td>
<td>103.83</td>
</tr>
</tbody>
</table>

The stimuli were recorded in a double-walled, sound-attenuating chamber (Industrial Acoustics Corporation Co., Bronx, NY) with a microphone (Sony T) connected directly to a Windows 7 workstation. Testing was conducted in the same sound-attenuating chamber. A computer workstation and amplifier (Harmon-Kardon 3380, Stamford, CT) outside
the chamber interfaced with a 17-in touch-screen monitor (Elo LCD TouchSystems, Berwyn, PA) and two wall-mounted loudspeakers (Electro-Medical Instrument Co., Mississauga, ON) inside the chamber. The touchscreen monitor facing the participant was used to present instructions and record responses. The loudspeakers were mounted at the corners of the sound chamber, each located at a distance of .76 m and 45° azimuth from the participant. A customized program created with Affect4 software (Spruyt et al., 2010) presented the auditory stimuli and recorded responses. All stimuli were played at a comfortable listening level of approximately 65 dB SPL.

**Procedure.** Listeners heard all 40 stimulus sentences presented in random order, constrained so that different versions of the same utterance (i.e., question or statement, adult- or child-directed) could not appear consecutively. (Due to computer malfunction, one listener heard 39 sentences.) Each sentence was presented in a block of five trials, beginning with the one-word stimulus, with successive trials adding a word, ending with the complete five-word sentence on the fifth trial. After each trial, listeners were required to select one of four options: *question*, *maybe a question*, *maybe a statement*, or *statement*, which combined choice of utterance type with confidence level. The task began with a short familiarization phase featuring one trial block with an adult-directed declarative question and another with an adult-directed statement, neither of which appeared in the test phase. The experimenter provided feedback at the end of each trial block (i.e., complete sentence) only during the familiarization phase.

**Results and Discussion**

Preliminary analyses revealed no meaningful differences as a function of whether responses indicated more confidence (i.e., *question*) or less confidence (i.e., *maybe a question*). Accordingly, each response was considered dichotomously as *statement* or *question*, which allowed us to transform these responses into $d'$ scores. For each participant, $d'$ scores were
computed separately for each number of words and for adult- and child-directed sentences. Thus, each participant had 10 $d'$ scores. A hit was a response in the question category (*question* or *maybe a question*) when the stimulus was a question. A false-alarm was a response in the question category when the stimulus was a statement. In terms of signal-detection theory, the “signal” consisted of the acoustic cues that distinguish questions from statements, and $d'$ scores reflected sensitivity to these cues.

The stimulus set consisted of 20 adult-directed and 20 child-directed utterances, half questions and half statements. Thus, the maximum number of successes for adult- and child-directed questions was 10 for each number of words. Because $d'$ scores cannot be computed when the hit rate equals 1.0 or the false-alarm rate equals 0 (i.e., infinite scores), proportions of correct responses were adjusted by using the following formula: \((\text{hit OR false-alarm rate} + .5) / (\text{maximum number of successes} + 1)\), as in previous developmental studies (e.g., Thorpe et al., 1988). These proportions were then converted to $z$-scores and subsequently to $d'$ scores ($d' = z[\text{hit rate}] - z[\text{false-alarm rate}]$). The maximum $d'$ score, indicating perfect performance (i.e., 10 hits, 0 false alarms), was 3.38, and chance responding (i.e., equal number of hits and false alarms) was 0. Descriptive statistics are illustrated in Figure 8 as a function of age group, speaking style, and number of words.

Initial analyses compared performance with chance levels using one-sample $t$-tests. Because there were 10 tests (two speaking styles x five different numbers of words), the alpha-level was lowered to .005. Nevertheless, performance exceeded chance levels in all cases, $ps < .001$, indicating that cues to statements or questions were discernible even from the first word, regardless of speaking style.
Figure 8. Performance in Experiment 1 (adults) and Experiment 2 (children) as a function of speech register and utterance increments (Study 3).

When the entire five-word utterance was presented, performance was at ceiling, as expected, with mean $d'$ scores of 3.20 ($SD = 0.36$) and 3.12 ($SD = 0.38$) for adult- and child-directed speech, respectively, including perfect performance by 34 of 45 listeners for adult-directed speech and 29 of 45 for child-directed speech. Moreover, performance on five-word utterances vastly exceeded performance on one-, two-, three-, and four-word stimuli for both
speaking styles, $ps < .001$, yet there was no difference between speaking styles for five-word stimuli, $p < .2$. Thus, further analyses were limited to stimuli with one to four words.

Response patterns were analyzed with an analysis of variance (ANOVA), with speaking style (adult- or child-directed speech) and number of words (1 through 4) as repeated measures and $d'$ scores as the dependent variable. A main effect of speaking style revealed that participants more readily differentiated questions from statements in child-directed than in adult-directed speech, $F(1, 44) = 24.43, p < .001, \eta^2 = .357$. A main effect of number of words stemmed from the fact that performance improved monotonically as the number of words increased, $F(3, 132) = 58.41, p < .001, \eta^2 = .570$. Both main effects were qualified, however, by a significant two-way interaction, $F(3, 132) = 9.29, p < .001, \eta^2 = .174$. Separate analyses of the two speaking styles revealed a linear improvement in performance for adult-directed speech, $F(1, 44) = 97.90, p < .001, \eta^2 = .690$, and for child-directed speech, $F(1, 44) = 27.59, p < .001, \eta^2 = .385$, with the interaction highlighting more dramatic improvement for adult-directed speech. As shown in Figure 8, performance on child-directed speech was relatively good even with one-word stimuli, which left less room for improvement as the number of words increased.

The present results confirmed the perceptibility of acoustic differences between statements and declarative questions that featured identical words. When the pitch contour of the penultimate and final words, whether rising or falling, was present, the identification of declarative questions and statements was at ceiling. Nevertheless, cues earlier in the utterances permitted successful identification, especially when the speaker used a child-directed style. Experiment 2 asked whether children are comparably successful at discriminating statements from questions when terminal pitch cues are present, whether they capitalize on subtle acoustic cues in pre-terminal position, and whether they benefit comparably from the child-directed speaking style.
Experiment 2

Although English-speaking children as young as 5 can differentiate questions from statements on the basis of intonation alone, 5- and 6-year-olds have much greater difficulty compared to older children, even after considerable training (Study 1). Accordingly, the present study was limited to children between 7 and 10 years of age.

Method

Participants. The participants, 32 children from the local community, included 15 7- and 8-year-olds (designated “younger children”: 7 girls, 8 boys; mean age = 7 years, 9 months, range = 7;2–8;10), and 17 9- and 10-year-olds (designated “older children”: 9 girls, 8 boys; mean age = 10 years, 1 month, range = 9;1–10;11). Inclusion criteria were Canadian birth and no personal or family history of hearing loss, according to parental report. An additional two 7-year-olds were tested but excluded from the final sample because they failed to identify questions and statements above chance levels (14 or more incorrect out of 40) after hearing the complete sentence, indicating poor understanding or attention. Children received a small gift for their participation.

Apparatus and stimuli. The apparatus and stimuli were the same as in Experiment 1.

Procedure. The adults’ task from Experiment 1 was modified to make it more engaging for children. On the basis of previous research indicating young children’s confusion with the terms statement and question (Study 1), the response options from Experiment 1 were simplified by changing them to asking, maybe asking, maybe telling, and telling. In addition, non-contingent feedback was provided during the test phase by presenting the image of one of several cartoon characters after completion of each trial block.

Results and Discussion
Responses were converted to $d'$ scores, as in Experiment 1, with *telling* and *maybe telling* considered as *statement* responses, and *asking* and *maybe asking* considered as *question* responses. Figure 8 illustrates descriptive statistics as a function of age group, speaking style, and number of words.

Initial analyses examined whether performance exceeded chance levels separately for both age groups, both speaking styles, and different numbers of words, with correction for multiple tests, as in Experiment 1. For the older children, performance exceeded chance levels for three- and five-word stimuli in adult-directed style, $ps < .005$, and was at the cusp of significance for four-word stimuli, $p = .005$. For all stimuli in child-directed style, performance exceeded chance levels, $ps < .005$. For the younger children and adult-directed utterances, performance exceeded chance only for the five-word stimuli, $p < .001$. For child-directed utterances, performance exceeded chance for three- and five-word stimuli, $ps < .005$.

For both speaking styles and age groups, performance for five-word stimuli was much better than performance for utterances with fewer words, $ps < .001$, as it was for adults in Experiment 1 (see Figure 8). Moreover, for both groups of children, performance with five-word stimuli did not vary as a function of speaking style, $ps > .6$. As in Experiment 1, the main analysis focused on stimuli with one to four words.

A mixed-design ANOVA was used to analyze $d'$ scores as a function of two repeated measures (speaking style, number of words) and one between-subjects factor (age group). A main effect of speaking style confirmed that children performed better with child-directed than with adult-directed speech, $F(1, 30) = 7.10, p = .012, \eta^2 = .191$. There was also a main effect of number of words, $F(3, 90) = 10.65, p < .001, \eta^2 = .262$, which was qualified by a two-way interaction between age group and number of words, $F(3, 90) = 5.19, p = .002, \eta^2 = .147$. There were no other main effects or interactions, $ps > .2$. Separate analyses of the older and younger
children revealed that although both age groups exhibited a monotonic increase in mean $d'$ scores as the stimuli increased from one to four words, the linear trend was significant for the older children, $F(1, 16) = 28.65, p < .001, \eta^2 = .642$, but not for the younger children, $p > .3$.

Subsequent analyses considered the present samples of children jointly with the adults tested in Experiment 1. We first examined how many individual participants could identify questions at above-chance levels. We conducted a series of 3 (age group) by 2 (above or at chance) chi-square tests of independence (one for each number of words) separately for adult- and child-directed utterances. According to the normal approximation to the binomial test (one-tailed, correcting for continuity), participants were significantly above chance if they had at least 15 out of 20 correct responses. The results are summarized in Table 5. For adult-directed utterances, age-related differences in the proportion of participants who successfully differentiated questions from statements were evident at three and four words. For child-directed utterances, age-related differences were evident at one, two, three, and four words. In each instance, adults had the greatest proportion of participants performing at above chance levels, followed by older children, and then younger children.

We also found a bias for responding statement rather than question for adult-directed utterances by way of one-sample $t$-tests, which examined whether the mean number of statement responses differed from the number expected in the absence of bias (50 out of 100), separately for each speaking style and age group (see Table 6). The results revealed a significant bias for statement responses for adult-directed ($ps < .005$) but not child-directed utterances ($ps > .06$) across age groups. Perhaps listeners expect more pitch variability in questions than in statements, choosing the latter category in the absence of obvious cues, as suggested by Vion and Colas (2006). This would also explain the absence of a comparable bias for child-directed utterances, which had more pitch variability and were therefore judged as questions more often than adult-
directed utterances. Moreover, because statements occur more commonly than questions in everyday discourse, a statement response is a reasonable default option in the context of ambiguous cues.

Table 5

<table>
<thead>
<tr>
<th></th>
<th>Adult-directed</th>
<th></th>
<th>Child-directed</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>7–8 9–10 Adults</td>
<td>(\chi^2)</td>
<td>7–8 9–10 Adults</td>
<td>(\chi^2)</td>
</tr>
<tr>
<td>1</td>
<td>.000 .059 .178</td>
<td>4.157 .125</td>
<td>.133 .353 .511</td>
<td>6.899 .032*</td>
</tr>
<tr>
<td>2</td>
<td>.067 .059 .244</td>
<td>4.415 .110</td>
<td>.200 .235 .667</td>
<td>15.071 .001*</td>
</tr>
<tr>
<td>3</td>
<td>.067 .235 .489</td>
<td>10.083 .006*</td>
<td>.200 .412 .778</td>
<td>18.141 .001*</td>
</tr>
<tr>
<td>4</td>
<td>.200 .471 .600</td>
<td>7.247 .027*</td>
<td>.267 .529 .867</td>
<td>20.590 .001*</td>
</tr>
<tr>
<td>5</td>
<td>1.00 1.00 1.00</td>
<td>n/a</td>
<td>.933 1.00 1.00</td>
<td>4.188 .123</td>
</tr>
</tbody>
</table>

Note: Each chi-square test of independence had \(df = 2\) and \(n = 77\).}

Because the stimuli were limited to a single speaker, listeners had an opportunity to profit from speaker-specific cues over the course of the test session. We tested this hypothesis with a multilevel model that had position of correct judgment (1–5 words) as the dependent variable, group (younger children, older children, adults) and speaking style (adult- or child-directed) as fixed factors, and utterance number (1–40, centered) as a fixed, continuous variable. A random intercept was included for each participant. Descriptive statistics are illustrated in Figure 9.

There was a main effect of speaking style, \(F(1, 2993.045) = 16.01, p < .001\), because performance on child-directed utterances exceeded that on adult-directed utterances, as noted above. Main effects of age group, \(F(2, 74.006) = 22.92, p < .001\), and utterance number, \(F(1,
\[ F(2, 2993.255) = 10.69, p = .001, \] were qualified by a two-way interaction between age group and utterance number, \[ F(2, 2993.237) = 4.61, p = .010 \] (see Figure 2). No other interactions were evident, \[ F_s < 1. \] Separate analyses of the three age groups revealed that performance improved with increasing exposure to the speaker (i.e., fewer words required for correct identification) for adults, \[ p < .001, \] but not for older, \[ p > .2, \] or younger, \[ F < 1, \] children (see Figure 9). Adults also outperformed children on the initial utterances (first 5 of 40), \[ t(75) = 2.07, p = .042. \]

Table 6

**Mean Statement Responses and Age (Study 3)**

<table>
<thead>
<tr>
<th>Age group</th>
<th>Adult-directed</th>
<th>Child-directed</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean ( t )</td>
<td>( p )</td>
</tr>
<tr>
<td>7–8</td>
<td>58.6 ( 3.503 ) ( .004^* )</td>
<td>54.7 ( 1.889 ) ( .080 )</td>
</tr>
<tr>
<td>9–10</td>
<td>58.6 ( 3.476 ) ( .003^* )</td>
<td>49.2 ( -.294 ) ( .773 )</td>
</tr>
<tr>
<td>Adults</td>
<td>62.2 ( 8.022 ) ( .001^* )</td>
<td>48.1 ( -1.890 ) ( .065 )</td>
</tr>
</tbody>
</table>

Note: * indicates values that significantly exceeded chance levels (50).

Descriptive analyses provided insight into the pre-terminal cues that enabled listeners to distinguish questions from statements. Qualitative annotation conventions from Tonal Breaks and Indices (ToBI) transcriptions (Pierrehumbert, 1980) depict pitch events associated with intonational boundaries and accented syllables. The transcriptions include two tone levels—high (H) and low (L). The * symbol denotes the location of an accented syllable, and the % symbol denotes the end of an intonational phrase. The + symbol is used when a nuclear accent is marked by a bitonal pitch accent, or an accent comprising a fall-rise (L+H) or rise-fall (H+L) contour.
Figure 9. Mean number of words required to identify questions and statements as a function of age and utterance number.

Sample question and statement contours are depicted in Figure 10. All questions in the set began with a low pitch accent followed by a rise (L*+H). By contrast, most statements began with a pitch rise followed by high pitch accent (L+H*), with the remaining statements having an initial high pitch accent without the preceding rise (H*). Statements tended to end with a low phrase accent and low boundary tone (L-L%), and questions tended to end with a high phrase accent and high boundary tone (H-H%). Some child-directed statements, however, ended with a low phrase accent and high boundary tone (L-H%). The boundary tone of questions was often preceded by a low pitch accent (L*), whereas the boundary tone of statements was often preceded by a high pitch accent (H*). Moreover, downstepped tones—a sequence of high pitch accents that are not as high as the preceding high (‘H’) accent—were more common in statements than in questions, reflecting the tendency for pitch to fall gradually over the course of
Figure 10. $F_0$ contours and ToBI annotations from sample adult- and child-directed utterances (Study 3).
the utterance. Finally, differences between adult- and child-directed utterances were inconsistent, with the patterns being sentence-dependent, presumably because of the location of nuclear tones.

**General Discussion**

The present study is the first to use a gating task to examine the identification of adult- and child-directed questions and statements by English-speaking children and adults. We established that children as young as 7 could use pre-terminal cues to identify questions and statements. Adults differentiated questions from statements after hearing the first word regardless of speech register. Older children (9- to 10-year-olds) performed better than younger children (7- to 8-year-olds), with both groups identifying child-directed utterances more readily than adult-directed utterances. Register also had more dramatic consequences for children than for adults. For example, older children identified child-directed utterances after hearing the first word, but they identified adult-directed utterances only after the first three words. Younger children correctly identified utterance type after hearing the first three words of child-directed utterances, but they required the complete adult-directed utterances to differentiate questions from statements. Finally, repeated exposure to the speaker’s voice enhanced performance for adults but not for children.

Performance differences between children and adults were evident early in the test session (the initial 5 of 40 utterances), which indicates that adults’ explicit knowledge of the intonation patterns that signify questions and statements exceeds that of children. In previous research that required listeners to judge full sentences, 7- and 8-year-olds identified questions and statements more poorly than 9- and 10-year-olds, who performed at adult levels (Study 1). The findings of both studies are consistent with the notion that pitch contour differences, although discriminable, are less memorable for children than for adults (Creel, 2014).
It is likely that adaptation to the demands of the gating task occurred relatively quickly. By contrast, adaptation to speaker-specific cues would be expected to unfold more slowly after exposure to a range of utterances from the single speaker. In fact, improvement in adults’ use of pre-terminal cues became evident at utterances 26–30, with no further improvement thereafter (see Figure 9). Children, by contrast, derived no benefit from such exposure within the time frame of the test session. Talker familiarity advantages are likely to be achieved more readily for word recognition than for intonation recognition. At times, however, familiarity effects in school-age children who receive protracted exposure to talkers are restricted to highly familiar words (Levi, 2015).

For questions in the present study, the pre-nuclear pitch accent was low and followed rapidly by a rising contour (L*+H). For statements, the contour tended to rise during the pre-nuclear pitch accent (L+H*). Although little is known about pre-terminal cues to questions and statements, there is evidence that intonation contours tend to fall during stressed syllables in statements but not in questions (O’Shaughnessy, 1979). This was not the case for the questions and statements in the present study, both of which featured a rising pitch contour during the first word. The main difference between our questions and statements involved the timing of the rise, which occurred during the primary stress for statements but after the primary stress for questions.

Our findings are consistent with adults’ ability to differentiate Spanish questions from statements after hearing the first content word, which features higher F0 peaks in questions than in statements (Face, 2005). There are notable differences, however, in cues to utterance type across languages and speakers. For example, adults correctly identify Dutch questions and statements after hearing the second F0 accent, which is larger in questions than in statements for female speakers but smaller for male speakers (van Heuven & Haan, 2000). Question
identification requires information in the penultimate syllable for French questions (Vion & Colas, 2006) and in the final syllable for Portuguese questions (Falé & Faria, 2006).

In principle, intensity and duration could influence the perception of utterance type. Although duration is not a reliable cue to question identity (Peng, Chatterjee, & Nelson, 2012), longer utterances and those with higher terminal intensity are more likely to be identified as questions than as statements (Ma et al., 2010; Patel, 2003; Peng et al., 2012). Intensity is highly correlated with F0 (Bolinger, 1986), but listeners are thought to use intensity and duration cues for question/statement identification only when F0 cues are unavailable (Ma et al., 2010; Patel, 2003; Peng & Chatterjee, 2009).

Our study is also the first to examine the impact of child-directed speech on listeners’ differentiation of questions and statements. Child-directed speech, which is characterized by slower speaking rate, more careful pronunciation, and wider pitch range relative to adult-directed speech (Foulkes et al., 2005; Jacobson et al., 1983), emphasizes the speaker’s affective intentions and focus of interest. The present findings indicate that it also emphasizes the speaker’s questioning or declarative intent, as one might expect. Acoustic analyses confirmed that the differences between questions and statements in adult- and child-directed speech were similar except for the greater magnitude of differences in child-directed speech. As can be seen in Figure 10, pitch accents were more prominent in child-directed than in adult-directed utterances.

The finding of age-related differences in overall question and statement identification is consistent with age-related differences in French-speaking children’s identification of utterance type (Gérard & Clément, 1998). In contrast to French 7-year-olds’ inability to distinguish declarative questions from statements and 9-year-olds’ ability to do so only after hearing the entire utterance, English-speaking children in the present study correctly identified statements and questions early in the utterance (except for 7- and 8-year-olds’ judgments of adult-directed
utterances). As noted, French includes fewer pre-terminal cues to questions and statements (Vion & Colas, 2006), whereas questions and statements in our English stimuli included distinct pre-terminal differences ($L^*$+$H$ tone in questions and $L$+$H^*$ tone in statements). It is unlikely, however, that the cross-language differences are attributable solely to differential cue distinctiveness or issues of incidence. For example, although the French study also featured a single talker, it was considerably more challenging because of five response alternatives—asking, telling, happy, sad, and ironic—rather the two general response classes—asking and telling (each qualified by maybe)—in the present study.

It is important to note the limitations of the gating paradigm for assessing listeners’ identification of the speaker’s pragmatic intentions based on intonation alone. In principle, the units of interest are intonation phrases rather than words, and studies with adults have used intonation rather than word gates (e.g., Face, 2005; Vion & Colas, 2006; van Heuven & Haan, 2000). Young children’s difficulty with the identification of complete declarative questions and statements in the absence of conversational context (Study 1) makes it unlikely that that they would be able to interpret the speaker’s intentions from partial words (i.e., gates based on intonation contour rather than words). Eye-tracking measures (e.g., Zhou et al., 2012) could provide a viable means of assessing the impact of intonation contour on question and statement identification, but the use of such measures necessitates utterances with clear visual referents, unlike those in the present study.

Although 9- and 10-year-olds exhibited considerable success in question identification, especially for child-directed utterances, their failure to capitalize on speaker-specific intonation patterns prevented them from achieving adult performance levels. Similar challenges may underlie the failure of 13-year-olds to match adults’ perception of emotional prosody (Aguert, Laval, Lacroix, Gil, & Le Bigot, 2013). Future research could examine parallels and differences
in children’s understanding of prosodic cues that signal questions and emotional tone by means of gating tasks involving familiar talkers.
Concluding Statements

The goal of this thesis was to examine developmental changes in the ability to use intonation cues to identify declarative questions and statements. In Study 1, 5- to 10-year-old children and adults were required to identify natural and low-pass filtered questions and statements. Children and adults were able to identify the utterance type based on intonation cues alone, but 5- and 6-year-olds performed more poorly than older listeners, and 7- and 8-year-olds performed more poorly than adults. Performance of the 9- and 10-year-olds was comparable to that of adults. Performance of all age groups was affected adversely by low-pass filtering of the stimuli, which removed lexical information and decreased the salience of pitch cues. In Study 2, 8- to 11-year-old children and adults judged utterances with electronically manipulated endings as questions and statements. Adults were also required to discriminate between pairs of these utterances. Children shifted their judgments more gradually between statement and question categories than adults did, but their category boundaries were similar. Moreover, adults’ performance on the discrimination and identification tasks was consistent with categorical perception of statements and questions. In Study 3, 7- to 10-year-old children and adults were required to identify adult- and child-directed questions and statements in a gating task that provided the first word of the utterance, the first two words, the first three words, and so on. Adults succeeded in identifying the utterance type after hearing the first word of adult- and child-directed utterances. For children, identification accuracy was greater for child-directed than for adult-directed utterances. Moreover, adults performed better than 9- to 10-year-olds, who performed better than 7- to 8-year-olds.

In contrast to the prevailing view that the rules of language are largely mastered by 4 or 5 years of age (e.g., Werker & Hensch, 2015), the present findings reveal that children’s association of prosodic cues with a speaker’s intentions has a protracted developmental
trajectory. Why do children, as late as 10 years of age, find it more difficult than adults to use prosodic information to infer a speaker’s intentions? Immature cognitive control is likely to be implicated, especially in younger children. For example, cognitive flexibility—the ability to adapt behaviour to different situations—improves between 4 and 13 years of age (Ceci & Howe, 1978; Davidson et al., 2006), with important gains occurring between 5 and 7 years of age (Hermer-Vazquez et al., 2001). Limited cognitive flexibility may affect children's ability to focus on prosodic information to interpret a speaker's intentions when their usual focus is elsewhere. For example, when children 4–10 years of age and adults are asked to judge a speaker's emotions from the way she speaks, ignoring what she says, the verbal content dominates the responses of 4- to 9-year-olds and continues to influence the responses of 10-year-olds (Morton & Trehub, 2001). Not surprisingly, adults’ responses are based exclusively on prosodic cues.

Similar biases may influence children’s perception of prosodic cues related to questions and statements. For example, 5- to 9-year-old children judge pairs of questions and statements as similar if they feature identical lexical information but different prosodic cues (e.g., rising and falling terminal pitch contours, respectively), and as different if they feature different lexical information but similar prosodic cues (Ploog, Bannerjee, & Brooks, 2009). In other words, their judgments of utterance pairs as similar or different is based on lexical rather than prosodic cues. It seems reasonable to conclude from the findings of other studies in conjunction with the present findings that prosodic cues are less salient in childhood than they are in adulthood.

The active-latent account (Morton & Munakata, 2002) posits that performance on cognitive tasks is influenced by the interaction between attentional biases and representations of task goals or instructions in working memory. Children between 5 and 7 years of age seem to be unaware that prosody can provide semantic information, which results in weak representations of a speaker’s true intentions and, at times, to prosody production preceding perception (Cutler &
Limited explicit knowledge of prosody coupled with a lexical bias could explain the youngest children’s difficulty with identification of questions and statements on the basis of prosodic information alone.

Lexical cues interfere with children’s use of prosodic cues to emotion whether or not the lexical cues have emotional implications (Morton & Trehub, 2001; Morton et al., 2003). For example, children often infer a speaker's feelings from emotionally neutral lexical information or contexts rather than prosodic information, inventing reasons to explain why the speaker is happy or sad (Aguert et al., 2013; Morton et al., 2003). In short, children may favour uninformative or even misleading cues (lexical) over informative cues (prosodic) because of immature cognitive flexibility or difficulty shifting from their usual strategies to ones that are more appropriate for the task at hand. Selective attention to auditory cues or filtering also improves between 5 and 14 years of age (Doyle, 1973; Maccoby & Konrad, 1966; Pearson & Lane, 1991; Plude et al., 1994), which may contribute further to children’s difficulty in focusing selectively on prosodic cues and ignoring irrelevant lexical cues.

With respect to question and statement identification, children’s difficulty is mirrored in production. For example, their production of questions with rising terminal pitch, which is the most salient prosodic cue to questions, increases between 4 and 11 years of age (Patel & Grigos, 2006; Snow, 1998; Wells et al., 2004). Children also have difficulty in interpreting other aspects of linguistic prosody such as the distinction between topic and comment and compound versus phrasal stress (hot dog vs. hot dog) (Cutler & Swinney, 1987; Hornby, 1971; Moore, Harris, & Patriquin, 1993; Quam & Swingley, 2014; Vogel & Raimy, 2002; Wells et al., 2004).

Prosodic comprehension is related to general language and cognitive abilities. For example, the receptive prosodic skills of 6- to 14-year-old children, such as the ability to identify sentence focus or compound nouns, are related to grammatical comprehension (Stonajovik,
Setter, & van Ewijk, 2007) and to verbal intelligence (Peppé, McCann, Gibbon, O'Hare, & Rutherford, 2007). Moreover, third graders who read quickly and accurately also produce shorter and more adult-like pauses, larger pitch falls at the end of statements, and larger pitch rises at the end of yes/no questions (Miller & Schwanenflugel, 2006). Those who use larger pitch falls and rises while reading statements and questions, respectively, also demonstrate greater reading comprehension. Further evidence of a link between prosodic abilities and general cognitive and linguistic abilities comes from adolescents with Williams Syndrome, who are markedly poorer at interpreting linguistic prosody than age-matched controls, with the performance gap narrowing for controls matched on language level (Catterall, Howard, Stojanovik, Szcerbinski, & Wells, 2006).

Children’s difficulty with identifying questions and statements solely on the basis of prosodic cues may be ameliorated in more natural situations. For example, face-to-face conversations provide contextual language cues (i.e., previous utterances) and visual cues in addition to prosodic cues. Even in the absence of some contextual cues, as in telephone conversations, prior utterances would be available to facilitate children’s comprehension of declarative questions.

Although there is some evidence that children use supplementary cues to interpret emotional prosody, there are important age-related differences. For example, 5- and 7-year-old children use contextual (i.e., situational) cues rather than prosody to interpret a speaker's emotions, 9-year-olds use both contextual and prosodic cues, and adults largely rely on prosody (Aguert et al., 2010). As noted, children between 5 and 13 years of age are more likely than adults to use contextual rather than prosodic cues to interpret a speaker's emotion even when the contextual cues are neutral, which provides additional evidence that prosodic cues pose challenges for children (Aguert et al., 2013). For example, when using contextual cues to explain
a speaker’s emotions, children tend to invent reasons why the context was happy or sad. Even when visual (e.g., facial expressions) and prosodic cues are available, children between 5 and 7 years of age use contextual (i.e., situational) cues to interpret a speaker’s emotions (Gil, Aguert, Le Bigot, Lacroix, & Laval, 2014).

Although adults prioritize prosodic cues over visual cues when interpreting a speaker’s intentions, they also use the visual information. For example, Dutch listeners give more weight to prosodic cues when interpreting the location of sentence focus, but they perceive the sentence focus as more prominent when visual cues (e.g., eyebrow movements, head nods, and articulatory movements) and prosodic cues are congruent (Krahmer & Swerts, 2008; Swerts & Krahmer, 2008). Moreover, speakers of American English can identify phrasal and lexical stress based on facial movements alone (Scarborough, Keating, Mattys, Cho, & Alwan, 2009).

Visual cues also help listeners differentiate questions from statements. For example, facial cues reinforce auditory cues to Dutch questions when the auditory cues are weak or ambiguous (House, 2002). Moreover, facial expressions have more influence than auditory cues when differentiating contrastive focus from echo questions in Catalan, but the combination of facial and acoustic cues produces the fastest and most accurate judgments (Borràs-Comes & Prieto, 2011). Visual cues also affect the identification of questions and statements in American English. For example, listeners rely primarily on auditory cues to differentiate declarative questions from statements, but they can also distinguish these utterance types based on visual cues (head tilt and eyebrow movements), and performance is best when auditory and visual cues are available (Srinivasan & Massaro, 2003). Future developmental research that combined auditory with visual cues would provide much needed information about age-related differences in the understanding of pragmatic intentions.
To date, however, there is little information about children’s use of visual cues to interpret linguistic prosody. A single study by Doherty-Sneddon and Kent (1996) indicated that 6-year-olds performed more poorly than 11-year-olds when the speaker was behind a screen but not when the speaker was visible. This finding implies that visual cues influence young children’s interpretation of linguistic prosody. If children have weak representations of prosodic cues, they may require support from supplementary cues. Future research could examine the role of visual cues in children’s interpretation of linguistic prosody, including the prosodic cues that signal questions and statements.

There is increasing interest in the interpretation of linguistic prosody, including question and statement identification, by special populations. For example, children with Autism Spectrum Disorder (Peppé et al., 2007), language impairment (Wells & Peppé, 2003), and Williams Syndrome (Catterall et al., 2006) can discriminate declarative questions and statements as well as typically developing children, but they have greater difficulty in identifying the speaker’s intentions. In other words, atypically developing children perceive the relevant prosodic cues but their challenge, like that of young, typically developing children, is mapping these cues onto meaning or pragmatic intentions.

Deaf individuals who use cochlear implants tend to have difficulty perceiving prosodic cues because their prostheses transmit degraded spectral information (Galvin, Fu, & Shannon, 2009), which impairs the perception of pitch contours (Gfeller et al., 2005; Gfeller et al., 2012; McDermott, 2004). As a result, their perception (Chatterjee & Peng, 2008; Meister et al., 2009; Most & Peled, 2007; Peng et al., 2009) and production (Peng, Tomblin, & Turner, 2008) of questions and statements is impaired relative to that of their normally hearing peers.

In sum, by 5 years of age, children can identify questions and statements on the basis of prosody alone, but they are much less proficient in this regard than older children and adults. By
9 years of age, children are as proficient as adults at identifying statements and questions from complete utterances, but they are less proficient than adults on more challenging tasks involving electronically altered terminal cues or pre-terminal cues. Children’s difficulties are attributable, in part, to limitations in their language and cognitive abilities. They can discriminate the requisite prosodic distinctions, but there are important age-related changes in attending to those cues and understanding the links between prosodic cues and meaning. Because the interpretation of linguistic prosody is related to general language abilities, it may be possible to improve children’s comprehension of prosodic information by focusing on language abilities generally rather than prosody specifically. Links between the perception and production of prosody (Peppé et al., 2007; Wells et al., 2004) raise the possibility that interventions designed to improve the perception of prosodic cues could have consequences for production as well as perception. For example, an intervention designed to improve the perception of prosodic cues by children with cochlear implants had favourable consequences on production (Klieve & Jeanes, 2001).

It is important to emphasize, however, that young children are likely to be more disadvantaged than older children and adults by decontextualized stimuli such as those used in the present studies. Children’s skills in ecologically valid contexts, or real environments, often differ substantially from those observed in laboratory environments (Neisworth & Bagnato, 2004). Future research could compare age-related changes in the identification of questions and statements in relatively natural contexts that include visual and contextual cues with contexts in which listeners must rely exclusively on prosody, as in the present dissertation.
References


Falé, I., & Faria, I. H. (2006). Time course of intonation processing in European Portuguese: A


doi:10.1016/j.rasd.2009.02.004


Stalinski, S. M., Schellenberg, E. G., & Trehub, S. E. (2008). Developmental changes in the


Trehub, S. E., Schellenberg, E. G., & Kamenetsky, S. B. (1999). Infants' and adults' perception


Copyright Acknowledgments

Study 1: Children's Identification of Questions from Rising Terminal Pitch is the manuscript of an article published by Cambridge University Press in the Journal of Child Language. The manuscript was accepted on 5 August 2015 and is available online at http://dx.doi.org/10.1017/S0305000915000458. Citations should refer to the published text.