Coherence in children's speech perception

Susan Nittrouer and Court S. Crowther
Boys Town National Research Hospital, 555 North 30th Street, Omaha, Nebraska 68131

(Received 4 October 2000; accepted for publication 26 July 2001)

Studies with adults have demonstrated that acoustic cues cohere in speech perception such that two stimuli cannot be discriminated if separate cues bias responses equally, but oppositely, in each. This study examined whether this kind of coherence exists for children's perception of speech signals, a test that first required that a contrast be found for which adults and children show similar cue weightings. Accordingly, experiment 1 demonstrated that adults, 7-, and 5-year-olds weight *F*2-onset frequency and gap duration similarly in "spa" versus "sa" decisions. In experiment 2, listeners of these same ages made "same" or "not-the-same" judgments for pairs of stimuli in an AX paradigm when only one cue differed, when the two cues were set within a stimulus to bias the phonetic percept towards the same category (relative to the other stimulus in the pair), and when the two cues were set within a stimulus to bias the phonetic percept towards different categories. Unexpectedly, adults' results contradicted earlier studies: They were able to discriminate stimuli when the two cues conflicted in how they biased phonetic percepts. Results for 7-year-olds replicated those of adults, but were not as strong. Only the results of 5-year-olds revealed the kind of perceptual coherence reported by earlier studies for adults. Thus, it is concluded that perceptual coherence for speech signals is present from an early age, and in fact listeners learn to overcome it under certain conditions. [DOI: 10.1121/1.1404974]

PACS numbers: 43.71.Es, 43.71.Ft [KRK]

I. INTRODUCTION

The acoustic signal of speech consists of isolable properties, or cues as they are commonly called. Until recently, prevailing views of speech perception have all proposed that each linguistic element (defined either as a phonemic segment or as a feature) is signaled by specific settings of one or several critical cues. Accordingly, much research in the area of speech perception focused on trying to identify what settings of which cues signal each linguistic element. To be sure, differences of opinion persisted concerning whether the cue settings that define linguistic units would prove to be invariant or not. While some investigators held hope that specific and stable settings of certain properties would be found that signal each linguistic element (e.g., Blumstein and Stevens, 1980; Kewley-Port, 1983; Stevens and Blumstein, 1978), others postulated that there are no invariants, but instead the settings of any one property that signal specific linguistic elements vary depending on the settings of other acoustic properties (e.g., Liberman, 1957; Liberman *et al.*, 1967; Mann and Repp, 1980; Studdert-Kennedy, 1983). Nonetheless, both approaches focused on understanding the relation between isolated cues and linguistic segments.

The question of *how* cues combine was never explicitly addressed in that early work. With hindsight, however, we can suggest that the assumptions underlying all that work clearly fit one of two models, what Best *et al.* (1989) called "cue extraction" and "cue integration." The line of work searching for invariant cues to linguistic identity fits under the rubric of cue extraction: listeners extract a specific property, and make a linguistic judgment based on that property. The work attempting to describe the sets of cues that define linguistic elements adheres to notions of cue integration: listeners extract a set of cues, and then sum them to derive the most probable estimate of the linguistic element represented.

Among the predominance of studies seeking to make the connection between isolated acoustic cues and phonetic categories can be found the rare, early study interested in perceptual organization of speech signals (e.g., Broadbent and Ladefoged, 1957). More recently, however, more investigators have focused their work specifically on the question of how acoustic properties are organized in speech perception. One point on which these investigators seem to agree is that acoustic properties cohere in speech perception such that it is difficult, if not impossible, to distinguish stimuli based on the auditory qualities of separate signal components (e.g., Best *et al.*, 1989; Bregman, 1990; Remez *et al.*, 1994). A variety of results support that position. For example, duplex perception experiments demonstrate that when an isolated acoustic property (the cue) is presented to one ear and the rest of the signal (the context) is presented to the other ear, two percepts are reported: a nonspeech percept (consisting of the cue only) and a speech percept (consisting of the context plus the cue). The critical finding of these experiments is that only for the nonspeech percept can listeners describe the auditory quality of the cue (e.g., whether a transition is rising or falling). In other words, even though the cue is needed for labeling the speech percept, it remains inaccessible as a separate perceptual entity (Liberman, Isenberg, and Rackard, 1981; Mann and Liberman, 1983; Whalen and Liberman, 1987).

The question of whether acoustic properties cohere in children's speech perception, according to any principles, has been largely unaddressed, but is important for several reasons. For one, answers to this question can extend our understanding of the general nature of speech perception by informing us as to whether the perceptual organization revealed in studies with adults is present from a young age (possibly birth), or is something that emerges through extensive experience with language. Also, knowing how acoustic properties cohere in children's speech perception will provide some fundamental information for studying other developmental phenomena, such as how perceptual attention changes with age and how access to phonological structure emerges. Thus, the focus of the current investigation was on whether or not acoustic properties cohere in children's speech perception such that stimuli heard as speech cannot be discriminated based on auditory qualities of the signal.

Originally we asked whether children can make discrimination decisions based on the auditory qualities of the signal as well as or better than adults when listening to speech. If children were found to discriminate stimuli better than adults based strictly on auditory qualities, our thinking went, support would be garnered for the position that the perceptual coherence of speech signals demonstrated by adults is a learned phenomenon. This position is represented by the statement of Best *et al.* (1989), ''The infant's task is to discover the phonetic coherence of phonological categories in the surrounding language by focusing attention on recurrent auditory contrasts that signal changes of meaning in that language.'' (p. 249) On the other hand, if children were found to demonstrate the same inability to discriminate stimuli based on auditory qualities that adults demonstrate, then we would suggest that such coherence is an integral aspect of speech perception, present from birth, or at least from a very young age. At the outset of this work, however, we never anticipated the possibility that the strict coherence of signal components reported for adults' speech perception might be even stronger for children. Of course, we never anticipated the possibility that adults would show anything other than strict coherence for speech signals.

From the start we did know that children's perceptual strategies for speech differ in at least one way from those of adults. Experiments have reliably shown that the weights assigned to the several cues involved in a phonetic decision can differ for children and adults. In one of the earliest demonstrations of this difference, Parnell and Amerman (1978) presented various combinations of stop-vowel syllable portions (the release burst, the burst plus aspiration noise, the burst plus aspiration and transition, the transition, the transition plus vowel) for labeling to 4- and 11-year-old children and adults. Four-year-olds labeled stops as accurately as older listeners only when syllable portions included the transitions. Thus, it seemed that children weight formant transitions more, or all other relevant cues less, than adults. Then, three studies investigating differences between young children (3 and 6 years) and adults in the use of vowel duration and syllable-final transitions in decisions of voicing for final stops all concluded that young children's decisions were based more than those of adults on the transitions, and less on vowel duration (Greenlee, 1980; Krause, 1982; Wardrip-Fruin and Peach, 1984). Most recently, experiments on labeling of syllable-initial fricatives have shown both that children weight formant transitions more and fricative-noise spectra less than adults (Nittrouer, 1992, 1996; Nittrouer and Miller, 1997a, 1997b; Nittrouer *et al.*, 2000; Nittrouer and Studdert-Kennedy, 1987). Thus, there is evidence that the relative contributions made to linguistic decisions by the various properties of the acoustic speech signal change as children get older, and gain experience with a native language. However, no study has suggested either that children attend perceptually to properties that are ignored by adults or that children ignore properties to which adults attend. Regardless of the age of the listener, linguistic decisions are based on the same sets of properties for each linguistic decision. Consequently, it was reasonable for us to ask how strongly these signal properties cohere in children's speech perception.

The approach used in this study was pioneered by Fitch *et al.* (1980). They constructed synthetic stimuli that approximated the words ''slit'' and ''split'' by varying each of two cues: the silent interval between the [s] noise and the vocalic portion (which is longer for natural split than for natural slit) and the onset frequencies of the first three formants (which are lower for split than for slit). Specifically, the silent interval varied from 8 to 144 ms in steps of 8 ms, and the transitions at the onset of the vocalic portions were appropriate for either [lɪt] or [plɪt]. Using an oddity paradigm, listeners were asked to choose the stimulus, out of three, that was different. Three conditions were included. In the ''one-cue'' condition, both stimuli in the pair had the same setting for one cue, but had different settings for the other cue. Fitch *et al.* chose to keep the silent interval constant across stimuli within a pair, and varied formant onset frequencies (rather than holding formant onset frequencies constant, and varying the silent interval). In both ''two-cue'' conditions, stimuli within a pair differed on both cues. In the ''two-cooperating-cues'' condition, the settings of both cues within a stimulus biased perception towards either slit or split, relative to settings in the other stimulus: the stimulus with the [lɪt] formant onset frequencies always had the shorter silent interval; the stimulus with the [plɪt] formant onset frequencies always had the longer silent interval. In the ''two-conflicting-cues'' condition, cues were set to oppose each other, such that they biased perception towards different categories: the stimulus with the [plɪt] formant onset frequencies always had the shorter silent interval; the stimulus with the [lɪt] formant onset frequencies always had the longer silent interval. The difference in silent interval between the two members of each pair (in the two-cue conditions) was always 24 ms, a value derived from a labeling experiment showing that 24 ms was the separation in functions for stimuli with the [lɪt] and [plɪt] formant onset frequencies (i.e., the difference in formant onset frequencies was ''worth'' 24 ms of silent interval). Results showed that all listeners were most accurate at discriminating stimuli in the two-cooperating-cues condition: these stimuli showed the familiar peak in discrimination accuracy at the category boundary. Following the two-cooperating-cues condition in discrimination accuracy was the one-cue condition, with discrimination in the two-conflicting-cues condition the poorest. Thus, even though acoustic differences between the two stimuli were identical for the cooperating- and conflicting-cues conditions, whether or not these differences facilitated or inhibited discrimination depended on whether or not cues biased responses towards the same or different categories.[1]

Best, Morrongiello, and Robson (1981) replicated the findings reported by Fitch *et al.* (1980) using synthetic versions of "say" and "stay," where the cues were the silent interval between the [s] noise and the vocalic portion, and the onset frequency of the first formant ($F1$): $F1$ is lower at voicing onset for burstless stay than for say. Presumably by coincidence, labeling data showed that the difference in formant frequencies was worth 24 ms of silent interval for these stimuli, just as it had been in Fitch *et al.* When stimuli were arranged as they had been by Fitch *et al.*, the same ordering of discrimination sensitivity was revealed: two cooperating cues > one cue > two conflicting cues. Noteworthy is the finding of both studies that discrimination for the two-conflicting-cues conditions was barely above chance performance.

Fitch *et al.* (1980) and Best *et al.* (1981) both concluded that the two acoustic properties manipulated in their experiments must be "perceptually equivalent," meaning that both biased responses equally towards specific phonetic labels. Furthermore, it seemed that the auditory qualities of these properties were unavailable to the listeners; instead, listeners were obliged to hear the composite signals according to the phonetic labels they were assigned. In 1989, Best *et al.* explicitly tested four models of how acoustic properties might combine in perception: the "cue extraction" and "cue integration" models that have already been discussed, and two models of coherence termed "auditory" and "phonetic" coherence. Both of these latter models suggest that linguistic structure emerges from the speech signal due to principles that mandate signal coherence, but in the case of auditory coherence, these principles guide all of auditory perception. The notion of phonetic coherence, on the other hand, suggests that the components of the speech signal cohere specifically because they arise from the coordinated pattern of articulatory movement that produced them. According to this account, we would not expect the same kind of coherence for a signal with all the properties of speech, if it were not heard as speech. That is, if principles specifically of phonetic coherence explain why speech signals are heard as unitary percepts, then this coherence would be absent for the same signals when they are not heard as speech. According to auditory explanations, these signals should be heard in the same manner regardless of whether they are perceived as speech or nonspeech.

Best *et al.* (1989) tested this hypothesis by comparing the labeling and discrimination of speech-like stimuli when listeners heard these stimuli as music, and when they heard them as speech. Findings across the music and speech conditions led these authors to conclude that "The results of these three experiments are inconsistent with the claim that speech perception entails the simple extraction, or the extraction and integration, of discrete information-bearing elements or cues." (p. 248) Instead, differences in response patterns across the two groups led the authors to conclude that only the model of phonetic coherence could explain how listeners combine separate acoustic properties to derive unitary and distinctive percepts representing linguistic categories. In particular, the listeners in the music group were unable to recover the isolated property that defined the categories, nor could they integrate this property with the rest of the spectral array to derive distinctive percepts. Thus, the general conclusion of this study was that the separate properties of the speech signal cohere according to principles that are unique to speech (i.e., phonetic coherence), and that once a signal is heard in this way, it is difficult, if not impossible, either to recover the individual properties or to have them cohere in a manner that would not be unique to speech (i.e., auditory coherence).

Only two studies have even tried to address the question of signal coherence in children's speech perception. Morrongiello *et al.* (1984) measured discrimination abilities of 5-year-olds for stimuli differing by one cue, two cooperating cues, and two conflicting cues using the same stimuli as those used by Best *et al.* (1981). Unlike the adults of Best *et al.*, children discriminated stimuli in the two-conflicting-cues condition as well as stimuli in the one-cue condition: discrimination abilities were clearly above chance in the two-conflicting-cues condition. Although those authors did not offer possible explanations for this age-related difference in results, we may speculate that it showed that children can hear the auditory qualities of a speech stimulus, independent of its phonetic label, better than adults can. Thus, we may further suggest that the phonetic coherence of speech signals demonstrated by adults may be a phenomenon that emerges as one learns a language. However, there is one caveat to this suggestion. For labeling results, children's data in the Morrongiello *et al.* study showed less separation between functions for the two $F1$-onset conditions than adults' data did in the Best *et al.* study. In other words, the difference in formant onset frequencies was worth less than 24 ms in children's perception. Nonetheless, the difference in silent interval between members of each pair was set according to adults' labeling results, at 24 ms. Therefore, the effective perceptual difference between stimuli in a pair may have been greater for children than for adults. As a result, the Morrongiello *et al.* experiment leaves unanswered the question of whether or not cues cohere in children's speech perception as they do in adults' speech perception.

Eilers *et al.* (1989) asked if infants can discriminate speech stimuli based on their auditory qualities. These investigators manipulated vowel duration and consonant periodicity as cues to voicing for utterance-final alveolar stops, and trained 9-month-old infants to perform a discrimination task. As Fitch *et al.* (1980) and Best *et al.* (1981) reported for adults, Eilers *et al.* found that infants were able to discriminate stimuli better in the two-cooperating-cues condition than in the one-cue condition.[2] As Morrongiello *et al.* (1984) reported for 5-year-olds, Eilers *et al.* found that infants were able to discriminate stimuli as well in the two-conflicting-cues condition as in the one-cue condition. Thus, it might be concluded that infants were making these discriminations based on the auditory qualities of the signal, either because they could recover individual properties of the acoustic signal or because these properties cohered according to principles of general audition. However, a couple findings of the Eilers *et al.* study encourage caution in interpretation. First, different groups of infants participated in the two-cooperating-cues and the two-conflicting-cues conditions, even though both groups heard the same one-cue stimuli. Overall, infants in the two-cooperating-cues condition performed more accurately than infants in the two-conflicting-cues condition, even in the one-cue condition. Furthermore, adults performed the same discrimination task with the same stimuli as the infants, and they discriminated stimuli as well in the two-conflicting-cues condition as in the two-cooperating-cues condition, a result that conflicts with those of Fitch *et al.* (1980) and Best *et al.* (1981). It might be that results across studies cannot validly be compared because Eilers *et al.* used vowel–stop-consonant syllables as stimuli, whereas Fitch *et al.* and Best *et al.* used stimuli consisting of an [s] noise, followed by a silent interval and then a vocalic portion. However, it is puzzling to have different results based only on the content of the stimuli. Most likely, acoustic properties either cohere according to the principles of phonetic coherence, or they don't. There seems to be no satisfactory explanation for why results might fit this model for some stimuli, but not for others.

For the current study the decision was made to use stimuli that consisted of an [s] noise followed by a silent interval (as appropriate for a stop closure) and a vocalic portion, as Fitch *et al.* (1980) and Best *et al.* (1981) had done. Those are the studies that reported that adults cannot discriminate stimuli based on auditory qualities of the speech signals, and that was the phenomenon under investigation for children in this study. However, the exact choice of stimuli was tricky because, as described earlier, labeling experiments have shown that children and adults weight differently the acoustic properties upon which at least some linguistic decisions rest. For such contrasts, stimuli within a pair would differ by a different perceptual amount across listener age, even though they differed by the same acoustic amount. Thus, the first step of the current study had to be identifying a set of stimuli of the form ([s] noise)–(stop closure)–(vocalic portion) that adults and children label similarly. Only then could we appropriately compare discrimination by children and adults of stimuli differing by one cue, two cooperating cues, and two conflicting cues.

II. EXPERIMENT 1: LABELING

The goal of this experiment was to find a set of stimuli consisting of an [s] noise, followed by a silent interval and then a vocalic portion that children and adults label similarly. In particular, the duration of the silent interval (i.e., the gap duration) and the onset frequency of one or more formant transitions needed to serve as the cues to be manipulated if we were to retain consistency in procedures with those of Fitch *et al.* (1980) and Best *et al.* (1981). While numerous studies (reviewed in the Introduction) have demonstrated that the relative weights assigned to the various signal components for speech can change as children get older (and so gain experience with language), one study has shown that this developmental shift in weighting strategies does not occur for all phonetic contrasts. Nittrouer and Miller (1999) showed that the relative weights assigned to the noise and to formant transitions in decisions of fricative identity are similar for adults and children for /f/-vowel and /θ/-vowel syllables. Nittrouer and Miller argued that this lack of developmental change can be found for any contrasts for which the "optimal" strategy (i.e., the one that most effectively facilitates recovery of phonetic structure) is the one used by young children. Consequently, we were hopeful that we would find a contrast for which adults and children would show similar labeling results.

Three studies (Morrongiello *et al.*, 1984; Nittrouer, 1992; Nittrouer, Crowther, and Miller, 1998) have shown that adults and children weight differently the acoustic properties upon which "say" versus "stay" decisions rest. Therefore, those stimuli could not be used. At the same time, we wanted to avoid using "slit" and "split," as Fitch *et al.* had used. Best *et al.* appropriately point out that the production of these syllables is complex articulatorily, and so the resulting signal is complex acoustically. The syllables "sa" and "spa" were found to meet our requirements.

A. Method

1. Participants

Three groups of listeners participated: 11 adults between the ages of 20 and 40 years; 11 children between the ages of 6 years, 11 months and 7 years, 5 months (the 7-year-olds); and eleven children between the ages of 4 years, 11 months and 5 years, 5 months (the 5-year-olds). All participants were required to pass hearing screenings of the frequencies 0.5, 1.0, 2.0, 4.0, and 6.0 kHz presented at 25 dB HL (ANSI, 1989). All participants were monolingual speakers of American English, had no histories of speech or language problems, and were free from significant early histories of otitis media, defined as six or more episodes before the age of 2 years. In addition, all children participating scored better than the 40th percentile for their age groups on the Goldman-Fristoe Test of Articulation (Goldman and Fristoe, 1986). Participating adults demonstrated a reading level of at least the 11th grade on the Wide Range Achievement Test-Revised [WRAT-R (Jastak and Wilkinson, 1984)]. In addition to these participants, two 5-year-olds and one adult participated, but their data were excluded from the final analysis because they failed to label reliably best exemplars of each category during testing (see Procedures).

2. Equipment

Testing took place in a soundproof booth. Hearing was screened with a Welch Allyn TM262 Auto Tymp tympanometer/audiometer using TDH-39 headphones. Stimuli were stored on a computer, and presented with a Data Translation 2801A digital-to-analog converter, a Frequency Devices 901-F filter, a Crown D-75 amplifier, and AKG-K141 headphones. Cartoon drawings were shown on a color-graphics monitor.

3. Stimuli

Stimuli were created with a SenSyn Laboratory Synthesizer, Version 1.1. They were synthesized at a 10-kHz sampling rate, and presented with low-pass filtering below 4.9 kHz. The [s] noise was a single-pole noise, centered at 3.8 kHz. This noise has been used extensively in labeling experiments in the past, and is known to be a good exemplar of [s]. Two vocalic portions were created. Both were 270 ms long: 40-ms transitions followed by 230-ms steady-state regions. The fundamental frequency of both began at 120 Hz, and fell throughout the vocalic portion to an ending frequency of 100 Hz. For both, $F1$ started at 450 Hz and rose to a steady-state frequency of 650 Hz. $F3$ started at 2100 Hz, and rose to a steady-state frequency of 2500 Hz. $F2$ onset was either 1175 Hz (the ''high $F2$ onset,'' most appropriate for ''sa'') or 950 Hz (the ''low $F2$ onset,'' most appropriate for ''spa''). Steady-state $F2$ was 1130 Hz in both portions. Each vocalic portion was combined with the [s] noise at each of ten gap durations: 0 to 36 ms, varying in 4-ms steps. Thus, there were 20 stimuli: two $F2$ onsets $\times$ ten gap durations.

4. Procedures

The hearing screening was done first, followed by either the articulation screening (children) or the reading screening (adults). For the 5-year-olds, recorded stories about each response label (i.e., the animals they named) were presented next. ''Sa'' was a sea creature, and ''spa'' was a horse. These stories were presented both by recorded, natural speech and by synthetic speech. Thus, they served both to teach children the labels and to provide experience listening to synthetic speech. Past experience in this laboratory has shown that children older than 5 years of age do not benefit from these stories, and so 7-year-olds did not hear them. Next, the pictures to be used were introduced. These pictures were 8×10 in. Participants were instructed to point to the picture representing the syllable heard, and say the syllable after hearing a stimulus. The experimenter then entered the response into the computer. Two kinds of practice were provided before testing: natural tokens of sa and spa, and the best exemplars of the synthetic stimuli. The natural tokens were five samples each of sa and spa spoken by the second author. The synthetic tokens were the low $F2$ onset with the 36-ms gap (best exemplar of spa) and the high $F2$ onset with the 0-ms gap (best exemplar of sa), presented five times each. Listeners had to respond correctly to nine out of the ten practice items (for both kinds of practice) to move onto either the next practice set or the test set. Testing consisted of ten blocks of the 20 stimuli presented in random order, and children were shown cartoon drawings on the graphics monitor at the end of each block. Participants had to respond correctly to 80% of the best exemplars (i.e., those presented during practice with the synthetic stimuli) during testing for their data to be included in the final analysis. Because all participants demonstrated the ability to label these best exemplars correctly during practice, failure to do so during testing was taken as evidence of a general decrease in attention. A lack of general attention of this sort would diminish the reliability of all responses.

Labeling functions were derived for each $F2$ onset, and were the proportion of sa responses given at each level of gap duration. These proportions were transformed to probit functions (i.e., cumulative normal distributions, represented as probit scores). From these probit functions, distribution means and slopes were computed. Distribution means were the points (given in milliseconds of gap duration) at which the probabilities of sa and spa responses were equal. These values are traditionally termed the ''phoneme boundaries,'' and will be here too. Slope is given as the change in probit units per millisecond of change in gap duration. The difference in location of the two functions at the phoneme boundaries is generally taken as an index of the weight assigned to the dichotomous property (in this case, $F2$ onset), and slope is an index of the weight assigned to the continuous property (in this case, gap duration).

B. Results

Figure 1 shows the labeling functions for all three listener groups, and Tables I and II display the phoneme boundaries and slopes, respectively. As can be seen, all groups performed similarly. Two-way analyses of variance (ANOVAs), with age as the between-subjects factor and transition as the within-subjects factor, done on these data support that conclusion. The main effect of age was not significant, either for phoneme boundaries, $F(2,30)=1.14$, $p=0.333$, or for slopes, $F(2,30)=2.78$, $p=0.078$. Only the main effect of transition was significant, both for phoneme boundaries, $F(1,30)=228.27$, $p<0.001$, and for slopes, $F(1,30)=7.20$, $p=0.012$. The interaction of age $\times$ transition was not significant, either for phoneme boundaries, $F(2,30)=0.26$, $p=0.774$, or for slopes, $F(2,30)=0.24$, $p=0.789$. Therefore, adults, 7-year-olds, and 5-year-olds may be said to have weighted the transition and the gap to the same extent in making this phonetic decision, and so these stimuli met our criterion for the discrimination task to be completed in experiment 2.

III. EXPERIMENT 2: DISCRIMINATION

Once stimuli were identified that children and adults label with similar weights assigned to each acoustic property, the next step was to examine discrimination of these stimuli when they differed by only one cue, by two cooperating cues, and by two conflicting cues. Based on the work of Fitch *et al.* (1980) and Best *et al.* (1981), we anticipated that adults would show enhanced sensitivity for stimuli in the two-cooperating-cues condition and diminished sensitivity in the two-conflicting-cues condition, compared to the one-cue condition. However, the most critical condition in this experiment was the two-conflicting-cues condition: this is the one condition in which stimuli cannot be perceived according to the principles of phonetic coherence, if they are to be discriminated. Listeners must either recover the separate properties of the signal or form unique and unitary percepts with these separate properties, percepts that do not neatly correspond to phonetic categories. Because the results of earlier studies (e.g., Best *et al.*, 1981, 1989; Fitch *et al.*, 1980; Remez *et al.*, 1994) suggest that adults do not accomplish either of these perceptual maneuvers (instead forming unitary percepts according to phonetic principles), discrimination in the two-conflicting-cues condition should be poor. We should find a similar response pattern for children, if phonetic coherence is intrinsic to their speech perception.[3] Alternatively, if the phonetic coherence of speech signals found

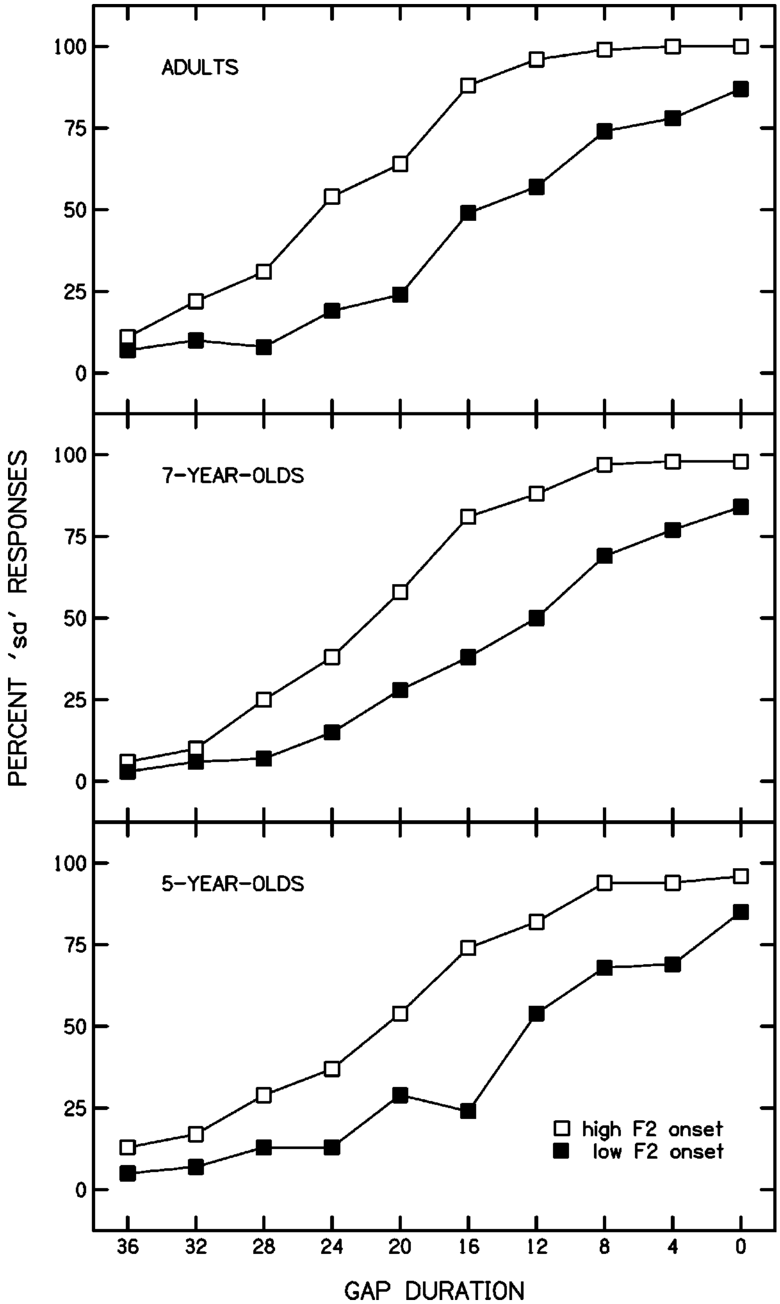

FIG. 1. Mean labeling functions for each age group from experiment 1. Open squares represent results for the high-*F*2 onset condition, and filled squares represent results for the low-*F*2 onset condition.

for adults is a learned phenomenon, then we could find a different pattern of response across the three conditions for children. In particular, we might expect children to demonstrate enhanced sensitivity to signal differences in both two-cue conditions, regardless of whether the cues cooperate or conflict in terms of the category they signal. That is, these children may not have had sufficient experience with speech signals to have discovered phonetic coherence.

TABLE I. Mean phoneme boundaries in ms of gap duration for each age group (with standard deviations in parentheses), for the high *F*2 onset and the low *F*2 onset. Number of participants in each group is given in italics.

	Adults *11*	7-year-olds *11*	5-year-olds *11*
High *F*2 onset	24.8 (3.4)	22.1 (5.1)	22.0 (4.3)
Low *F*2 onset	14.4 (5.3)	12.7 (6.8)	11.6 (4.0)

TABLE II. Mean slopes (in probit units per ms of gap duration) for each age group (with standard deviations in parentheses), for the high *F*2 onset and the low *F*2 onset. Number of participants in each group is given in italics.

	Adults *11*	7-year-olds *11*	5-year-olds *11*
High *F*2 onset	0.14 (0.04)	0.16 (0.05)	0.11 (0.05)
Low *F*2 onset	0.11 (0.06)	0.12 (0.05)	0.09 (0.04)

A. Method

1. Participants

Thirteen adults, and 13 children of each of the ages of 7- and 5 years participated in this experiment. All participants met the same criteria as those described for participants in the first experiment. Four additional 5-year-olds attempted the task, but were unable to perform this discrimination task with natural tokens of spa and sa, and eight additional 5-year-olds and one 7-year-old were unable to discriminate the most different stimuli to be used in any of the three test conditions (see Procedures).

2. Equipment and stimuli

The same equipment was used as in experiment 1. The stimuli used in the first experiment were used in this experiment, but they were presented in pairs in an AX format where A was a constant standard. Three sorts of stimulus pairings were used: those in which stimuli differed by only one cue (the one-cue condition), those in which stimuli differed by two cues such that the settings of both cues biased responses towards the same category label (the two-cooperating-cues condition), and those in which stimuli differed by two cues such that the settings of both cues biased responses towards different category labels (the two-conflicting-cues condition). The ‘‘standard’’ stimulus (i.e., the one that remained constant across pairs within any one condition) always had a 36-ms gap. The ‘‘comparison’’ stimuli (i.e., those that varied across pairs within any one condition) had gaps varying between 0 and 36 ms. Of course, we could have arranged stimuli so that the standard had a 0-ms gap and the comparisons had longer gaps. However, listeners in experiment 1 were slightly more consistent labeling stimuli with the 36-ms gap and the high $F2$ onset as spa than they were labeling stimuli with the 0-ms gap and the low $F2$ onset as sa. (See Fig. 1: the open squares at the 36-ms gap are closer to 0% than the filled squares at the 0-ms gap are to 100%.) Thus, using stimuli with 36-ms gaps as the standards ensured that the standard was a good exemplar of one of the categories (in this case spa) in every condition, even when cues conflicted. Once the decision was made to use this end of the continuum for standard stimuli, the selection of stimuli (standard and comparisons) for each condition was mandated by the desired arrangement of cues across stimuli within the condition.

In the one-cue condition, all stimuli (standard and comparisons) had the low $F2$ onset, as appropriate for spa; only the duration of the gap varied. In this way, the standard was the clearest exemplar of spa possible. The first stimulus in the pair (the standard) always had the 36-ms gap, and it was followed by a stimulus having another gap duration. Having formant onset frequency remain constant for stimuli within a pair, and the gap duration vary, was one change from the procedures of Fitch *et al.* (1980) and of Best *et al.* (1981). In their one-cue conditions, gap duration remained the same for stimuli within a pair, and formant onset frequencies varied. However, those investigators provided no reason, either procedural or theoretical, for why they did it this way (or why it might matter). Because we wanted our standard to remain constant across pairs, gap duration had to vary. We had no reason, and still do not, to suspect that the arrangement of cues (i.e., which one remains constant and which one varies) would affect listeners' abilities to discriminate speech stimuli based on the auditory qualities of the signals. In any event, if the phenomenon previously reported (that adults have extremely limited access to acoustic properties during speech perception) can only be observed for specific arrangements of cues, then the conclusions reached by those studies need to be reconsidered.

In the two-cooperating-cues condition, the standard had the low $F2$ onset, so both gap-(36 ms) and $F2$ onset biased responses towards spa. All comparison stimuli had the high $F2$ onset, so both gap and $F2$ onset biased responses towards sa (compared to the standard). In the two-conflicting-cues condition, the standard needed to have the high $F2$ onset, to bias responses towards sa. The comparison stimuli in this condition had low $F2$ onsets. Thus, gap and $F2$ onset conflicted in terms of which category they biased responses toward.

3. Procedures

Five-year-olds were provided with a preliminary task that 7-year-olds and adults did not have. Because of this extra task, and their generally shorter attention spans, 5-year-olds were tested over 2 days. Seven-year-olds and adults participated in just one session.

The screening measures were administered first. Next, 5-year-olds were provided with practice labeling pictures as ‘‘same’’ or ‘‘not-the-same.’’[4] These were hand-drawn pictures of pairs of simple objects, such as flowers and cars. Five of the pictures showed the same object twice, and five showed two different objects. Normally developing children understand the concept of same by age 5 years; this extra practice simply helped familiarize children with the task.

The procedure used here differed from that of Fitch *et al.* (1980) and Best *et al.* (1981), where the stimuli within a pair differed acoustically by the same amount across pairs. We made the decision to use a fixed standard with varied step sizes between stimuli because the method of constant difference (used by Fitch *et al.* and Best *et al.*) leads to the situation where none of the pairs of stimuli has very large acoustic differences between members. As a result, the discriminations to be made are all fairly difficult perceptually, and young children do not tolerate long strings of difficult discriminations. The procedure of using a fixed standard should minimize memory load and decrease stimulus uncertainty, both goals generally viewed as desirable in work with children.

For all discrimination tasks, the response was to point to a picture of two red squares and say ‘‘same’’ if the stimuli within a pair were judged to be the same, and to point to a picture of a red square and a black triangle and say ‘‘not-the-same’’ if the stimuli within a pair were judged to be different. All participants received practice with this procedure using recorded, natural tokens of sa and spa (five same trials and five different trials) before any testing started at all. If a participant was unable to recognize four of the five ‘‘different’’ trials as different (while recognizing that all same trials

TABLE III. Numbers of participants of each age who were unable to hear the most different comparison stimulus as different from the standard in each condition. Total numbers of participants are given in italics.

	Adults *13*	7-year-olds *13*	5-year-olds *13*
Two cooperating cues	0	0	2
One cue	4	2	3
Two conflicting cues	3	3	5

were the same), this natural practice set was presented a second time. If the participant was still unable to meet the criterion, the participant was dismissed. Then, as each stimulus condition was introduced, practice was provided with the most acoustically different stimuli in that condition. Again, five same trials (with both members of the pair being the standard stimulus) and five different trials (with the standard and the most acoustically different stimulus) were provided. If a participant was unable to recognize four of the five different trials as different (while recognizing that all same practice trials were the same), the practice set was presented a second time. If the participant was still unable to meet the criterion, the participant was not tested in that condition. For testing, the standard was paired with all comparison stimuli in that condition (including itself) ten times each (i.e., ten blocks of ten pairs). The order of presentation of conditions was randomized across participants. During testing, participants needed to perform at 80% accuracy for the most different stimuli. Data were discarded for any participant who did not attain this level of performance. Discrimination functions were derived for each condition, and are the percentage of not-the-same responses at each level of gap duration.

B. Results

Table III shows the numbers of participants of each age who were unable to meet the practice or test criteria in each condition, out of the 13 participants of each age who could do the task in at least one condition. Two findings are noteworthy. First, out of all participants, only two 5-year-olds were unable to meet the criteria for the two-cooperating-cues condition. In other words, if a listener was able to discriminate stimuli in only one condition, it was likely to be the two-cooperating-cues condition. Second, there was no participant who was able to do only the two-conflicting-cues condition.

Figure 2 displays the percentage of not-the-same responses at each level of gap duration, for each condition. Table IV lists mean percentages of not-the-same responses for each condition. For each age group, matched *t*-tests were done comparing the percentages of not-the-same responses for the one-cue versus two-cooperating-cues conditions, the two-conflicting-cues versus two-cooperating-cues conditions, and the one-cue versus two-conflicting-cues conditions. For these statistical analyses, missing data were replaced using generally accepted procedures. If a participant was unable to meet the practice or the test criterion in just one condition, that missing value was estimated by regressing that variable on the listener's other two measures (using regression equations derived from the group as a whole). This method of replacing missing data is fairly standard, and did not affect the overall outcome of the study because the estimated values did not change relative percentages across conditions within groups, the result of most interest. If a participant was unable to meet the practice or test criterion in two conditions, those missing values were replaced with the group means for each of those conditions. Again, this method is generally accepted and should not affect overall outcome. The computed *t*-ratios and associated *p*-values are given in Table V.

Looking at adults' responses, the first trend we notice in Fig. 2 and from the percentages provided in Table IV is that adults were much better at discriminating pairs of stimuli in the two-conflicting-cues condition than would have been predicted from Fitch *et al.* (1980) and Best *et al.* (1981): Adults' discrimination was better in the two-conflicting-cues condition than in the one-cue condition. In fact, adults in this experiment discriminated stimulus pairs more readily whenever two cues distinguished the members of the pairs than when only one cue did, regardless of whether the two cues cooperated or conflicted in terms of the category they signaled. This result is commensurate with that of Eilers *et al.* (1989). Looking at the statistical results in Table V we see that the percentage of stimulus pairs discriminated by adults differed across all three conditions, as indicated by the finding that all three *t*-tests were significant. Thus, for adults, the pattern of response was clearly two cooperating cues > two conflicting cues > one cue.

Results for 7-year-olds mirror results of adults in that the order of discrimination functions is similar (two cooperating cues > two conflicting cues > one cue), although the trend is not as strong. In particular, discrimination in the one-cue condition was not as poor as that of adults: the function for the one-cue condition is closer to those of the other two functions than is the case for adults. For 7-year-olds, the only *t*-test to reach statistical significance was the one-cue versus two-cooperating-cues conditions, the best- and the poorest discriminated stimuli.

For 5-year-olds, the order of discrimination functions is different from that of the two older groups: Performance was similar for the two-cooperating-cues and one-cue conditions, but poorer for the two-conflicting-cues condition. Looking at the statistical results in Table V, both *t*-tests involving the two-conflicting-cues condition were statistically significant, indicating that 5-year-olds really were worse at discriminating these stimuli than the stimuli in the other two conditions. The one-cue versus two-cooperating-cues *t*-test was not statistically significant.[5]

Although differences among age groups in discrimination for individual conditions was not the main focus of this study, we did perform ANOVAs on percentages of not-the-same responses for each condition, with age as the factor. Pairwise *t*-tests were also computed. Results of these analyses are shown in Table VI, and indicate that a significant effect of age was found for all conditions. The pairwise *t*-tests confirm impressions from Fig. 2 and Table IV: For the one-cue condition, 7-year-olds showed greater sensitivity to

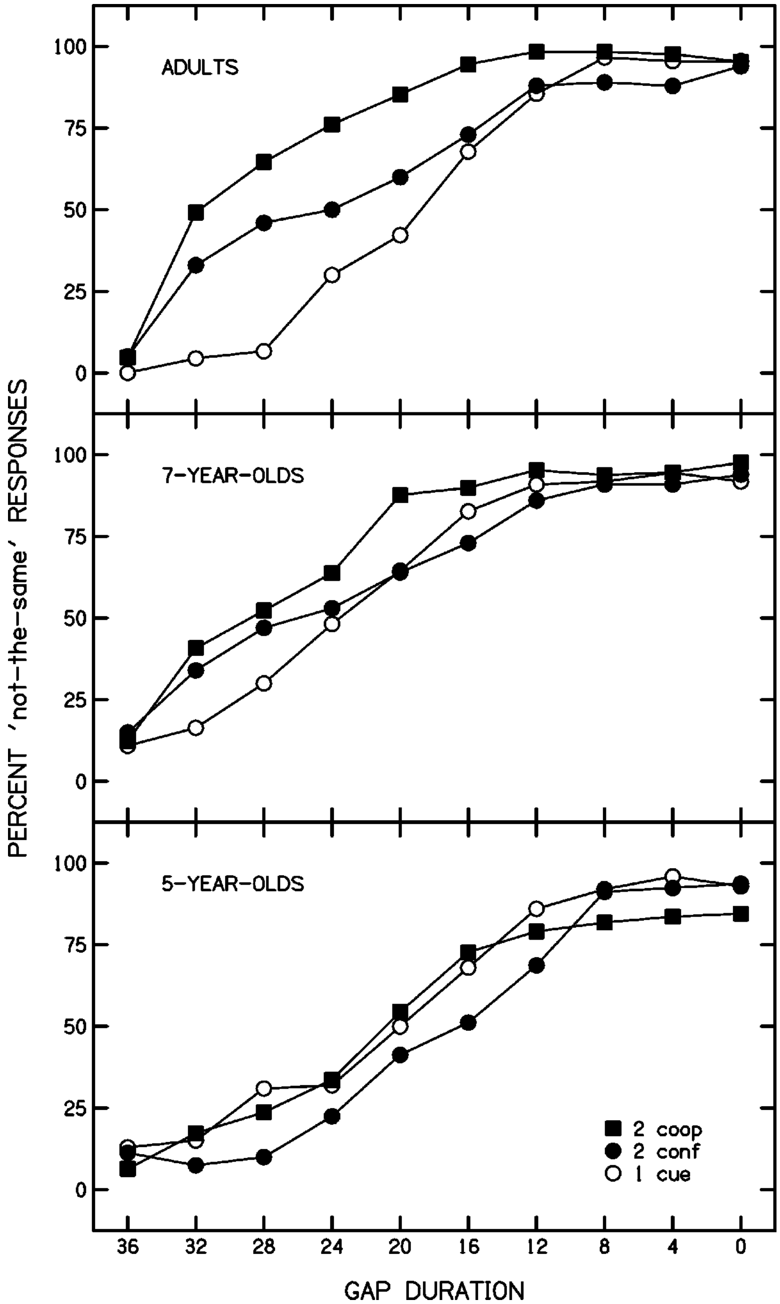

FIG. 2. Mean discrimination functions for each age group from experiment 2. Filled squares represent results for the two-cooperating-cues condition; filled circles represent results for the two-conflicting-cues condition; open circles represent results for the one-cue condition.

TABLE IV. Mean percentages of "not-the-same" responses for each age group (with standard deviations in parentheses), for all three conditions. Number of participants in each group included in computations is given in italics.

	Adults	7-year-olds	5-year-olds
	13	*13*	*11*
Two cooperating cues	76.5	72.8	60.9
	(13.4)	(10.4)	(15.9)
	12	*12*	*10*
One cue	52.4	61.9	57.6
	(7.8)	(9.1)	(10.8)
	12	*12*	*8*
Two conflicting cues	62.6	64.3	49.0
	(12.6)	(18.1)	(9.5)

differences between stimuli than adults, but adults and 5-year-olds showed similar sensitivities. For both conditions involving two cues, 5-year-olds showed poorer sensitivities than listeners in the other two groups.

IV. DISCUSSION

This study was originally undertaken to examine whether children would show the same pattern of perceptual coherence for speech that adults showed in studies by Fitch *et al.* (1980) and Best *et al.* (1981). In those studies, adults demonstrated enhanced discrimination for pairs of stimuli in which two cues cooperated in terms of which category they signaled, but a reduction in discriminability when those same cues conflicted in terms of which category they signaled. To use the notation of Best *et al.*, the pattern of results found in

TABLE V. t-ratios and p-values (given in parentheses) for each matched t-test, for each age group. Degrees of freedom are 12 for each group.

	Adults	7-year-olds	5-year-olds
One/ two cooperating	−5.70 (<0.001)	−3.75 (0.003)	−0.72 (0.484)
Two conflicting/ two cooperating	−3.29 (0.007)	−1.77 (0.102)	−3.06 (0.010)
One/ two conflicting	−3.27 (0.007)	−0.57 (0.576)	3.58 (0.004)

TABLE VI. F-ratios and p-values (given in parentheses) for overall tests of age effect for each condition given in the first row, with Bonferroni significance level of each t-test in subsequent rows. Degrees of freedom are 2,36 for the F-ratios, and 36 for each t-test.

	One cue	Two cooperating	Two conflicting
F-ratio	4.07 (0.026)	5.16 (0.011)	5.53 (0.008)
Adults vs 7 years	0.05	⋯	⋯
Adults vs 5 years	⋯	0.05	0.05
7 vs 5 years	⋯	0.10	0.05

those studies was two cooperating cues > one cue > two conflicting cues. Unlike those reports, however, the adults in the present study showed enhanced discrimination abilities any time two cues differed between members of the pair, regardless of whether cues were set to cooperate or conflict. Using Best *et al.*'s notation again, the pattern of results was two cooperating cues > two conflicting cues > one cue. This finding indicates that adults could either recover the separate acoustic properties and do a simple summation to obtain a measure of auditory difference, or could derive unitary, distinctive percepts of the stimuli that were not tied to linguistic labels. Deciding between these two possibilities is beyond the scope of this study, but in either case, adults were clearly not obliged to hear these signals strictly according to the principles of phonetic coherence.

It is, of course, tempting to suggest that our conclusions differ from those of Fitch *et al.* (1980) and Best *et al.* (1981) because our procedures differed such that standard stimulus and most comparison stimuli varied more in how acoustically different they were from each other than the two stimuli in any one triad of those earlier experiments. However, the major difference between results of this study and that of both Fitch *et al.* and Best *et al.* (that adults discriminated between stimuli in the two-conflicting-cues condition) was found when the standard and comparison did not differ much in gap duration. In other words, this result was not obtained only for large interstimulus differences. Besides, our results largely replicate those of Eilers *et al.* (1989). Of course, some aspects of the procedures used by Eilers *et al.* differed from those of Fitch *et al.*, of Best *et al.*, and of this study. For example, Eilers *et al.* used a repeating background stimulus that was interrupted for brief periods by comparison stimuli. Nonetheless, both that study and this study found that adults' discrimination was better when two cues differed across stimuli than when only one cue did, regardless of whether the two cues cooperated or conflicted regarding which linguistic category they signaled. Apparently, the procedures of this study and of Eilers *et al.* were simply more sensitive than those of Fitch *et al.* and Best *et al.*

At the same time, adults' discrimination of stimuli in the two-conflicting-cues condition was not as good as in the two-cooperating-cues condition in this study. The proportions of not-the-same responses given to stimuli in the two-conflicting-cues condition were not as high as those in the two-cooperating-cues condition. Also, the shapes of the discrimination functions were different for the two-cooperating-cues and the two-conflicting-cues conditions. Adults were better than expected at judging that two stimuli in the two-conflicting-cues condition were probably not the same when standard and comparison differed in gap durations by only a small amount: The function is never close to "0% not-the-same responses" (except of course when the stimuli are physically the same), as it was in the one-cue condition. However, adults never attained the same level of accuracy in judging that stimuli were different from each other in the two-conflicting-cues condition that they attained in the two-cooperating-cues condition: Even at the shortest gap durations (0 to 16 ms), where the largest differences between standard and comparisons existed, adults did not discriminate comparison stimuli from the standard 100% of the time in the two-conflicting-cues condition.

Results for 7-year-olds mirror those of adults, but differences across conditions were not as well-defined. The pattern of results was two cooperating cues > two conflicting cues > one cue, as it had been for adults. However, only the t-test for the one-cue versus two-cooperating-cues conditions reached statistical significance. Failure to find a significant difference between the one-cue and the two-conflicting-cues conditions (as was found for adults' data) probably reflects two facts: variability in discrimination performance was high for 7-year-olds in the two-conflicting-cues condition, and the mean function for 7-year-olds in the one-cue condition was at a longer gap duration than that of adults. In other words, a large part of the reason for the difference in statistical findings for 7-year-olds and adults is that 7-year-olds showed *better* discrimination in the one-cue condition, not that they showed poorer discrimination in the two-conflicting-cues condition. The variability in discrimination performance for the two-conflicting-cues condition can probably be offered as the major reason that a significant difference was not found between the two-cooperating-cues and the two-conflicting-cues conditions.

Five-year-olds were the one group that performed as predicted based on Fitch *et al.* (1980) and Best *et al.* (1981), at least with regard to the finding that discrimination was hindered when the two acoustic properties differed across stimuli in terms of which linguistic category they signaled. Unlike the adults in those studies, however, 5-year-olds did not discriminate stimuli any better when the two acoustic properties covaried appropriately in terms of which linguistic category they signaled than when only one property varied across stimuli. As a result, it might be suggested that there was no coherence of signal components for 5-year-olds, that instead decisions were based solely on gap duration. However, results from the labeling experiment contradict that suggestion: the separation between labeling functions depending on $F2$ onset was exactly the same for 5-year-olds as it was for adults and 7-year-olds, indicating that these children were sensitive to and used both the $F2$-onset cue with the gap-duration cue. Thus we suggest that, in fact, 5-year-olds categorized stimuli, and made their discrimination judgments strictly based on those categories.

It might also be suggested that 5-year-olds were simply poor at discriminating stimuli, perhaps due to general difficulty with the task. However, there is no evidence of that for the 5-year-olds included in the analyses for each condition: Their discrimination thresholds were not at particularly brief gap durations (i.e., far from the gap duration of the standard), and variability was not much greater for 5-year-olds in any condition than it was for the other groups. In sum, we have every reason to believe that these 5-year-olds were discriminating these stimuli with no particular difficulty.

What then are we to conclude about how acoustic properties are integrated in children's speech perception? We suggest that speech perception from a very young age promotes coherence of signal properties. Apparently, it is only with experience that listeners are able to discriminate stimuli in which acoustic properties do not covary together to specify linguistic categories, and so come to discriminate pairs of stimuli that receive the same category labels. In other words, human listeners learn to overcome the coherence of individual cues that normally characterizes speech perception to make the discrimination judgments asked of them in laboratory experiments. Our suggestion that none of the general (i.e., non-speech-specific) principles of auditory organization investigated largely by Bregman and colleagues (Bregman, 1990), and summarized by Remez *et al.* (1994), would explain the signal coherence observed for 5-year-olds' responses is based on the finding that the adults in this study performed better than expected for the two-conflicting-cues condition. From that we conclude that adults were able either to recover the separate acoustic properties or to derive unitary percepts that did not depend on phonetic coherence. In other words, it is possible that the adults were using one of these general-purpose processing strategies. The response patterns of 5-year-olds across conditions did not resemble those of adults, and so we conclude that the perceptual coherence observed in their data was based on different principles, and the principles of phonetic coherence seem the best candidates. Again, principles of phonetic coherence suggest that signal properties cohere when they arise from the same articulatory event. Children must be attentive to these events because they need to recover information that allows them to learn how to move their own vocal tracts in order to produce the sounds of their native language.

The finding that 5-year-olds actually demonstrated evidence of stronger signal coherence than older listeners has implications not only for developmental theories of speech perception, but for general theories of perceptual organization, as well. One view of perceptual coherence holds that multiple attributes come to be perceived as a group following experience with systematic covariation, and this account has been used to explain the role of multiple cues in speech perception (e.g., Holt, Lotto, and Kluender, 2001; Kluender *et al.*, 1998). Earlier descriptions of phonetic coherence made a distinction between coherence in the perception of speech signals and coherence in the perception of other signals largely by suggesting that covariation of signal attributes for speech is specifically a consequence of articulatory movement. Nonetheless, the suggestion has commonly been that phonetic coherence results from extended experience with covariation among acoustic properties (again, see Best *et al.*, 1989). In this experiment, however, the greatest evidence of coherence for these speech signals was demonstrated by the least-experienced listeners. The more experienced listeners in fact showed evidence of having learned how to separate components of the signal.

In summary, we found evidence of strong coherence of separate acoustic cues in the speech perception of young children. With hindsight, perhaps this finding should not have come as a surprise. An important developmental task facing young children is learning how to produce the articulatory gestures required of their native languages, and so those gestures are extremely relevant ecologically. Learning to strip off the individual acoustic properties (which on their own are ecologically irrelevant) is a perceptual skill that may only be acquired later. Thus, children do not discover phonetic coherence; instead, they learn to overcome it when necessary.

ACKNOWLEDGMENTS

This work was supported by research Grant No. R01 DC 00633 from the National Institute on Deafness and Other Communication Disorders to the first author. We thank Donna L. Neff for help in the design of the task in experiment 2, and we are grateful to Donal G. Sinex, Michael Studdert-Kennedy, and Keith R. Kluender for comments on earlier drafts.

[1]It should be borne in mind that whether cues within a stimulus are described as ''cooperating'' or ''conflicting'' is only meaningful in relation to how cues were set for the other stimulus in the triad. For example, in the two-cooperating-cues condition, stimuli with [lit] formant onset frequencies sometimes had silent intervals much longer than would be found for [slit] in natural speech. However, in this condition, these stimuli always had the shorter interval of the two, and so cues are described as cooperating. As another example, in the two-conflicting-cues condition, stimuli with [plit] formant onset frequencies could have silent intervals that would very likely be found for [split] in natural speech. Nonetheless, as long as these stimuli had the shorter interval of the two in the pair, the cues are described as conflicting. Thus, the terms ''cooperating cues'' and ''conflicting cues'' have precise definitions in this context.

[2]Eilers *et al.* (1989) actually included two one-cue conditions by having stimuli in a pair differ only on consonant periodicity or only on vowel duration. However, some infants performed near chance in the one-cue condition where stimuli differed on consonant periodicity, and so we report here only on the one-cue condition where stimuli differed on vowel duration.

[3]Of course, finding similar response patterns for adults and children in this study might indicate something other than that phonetic coherence for speech signals is innate. It could indicate that children had acquired the perceptual strategies of adults by the age of 5 years. However, concern about this alternative explanation would only arise if adults and children were found to respond similarly.

[4]The verbal label not-the-same was used instead of different because the

notion of two items being different is more complex than simply recognizing that two items are not the same. Consequently the youngest children might have difficulty with it. For consistency, listeners in all age groups used the labels same and not-the-same. Morrongiello *et al.* (1984) used the same procedure.

[5]Because only eight 5-year-olds could do this discrimination task in all three conditions, these *t*-tests were also conducted with just those eight children. In that case, no *t*-test reached statistical significance, although the percentage of not-the-same judgments showed the same trend as for the full group of 13: for these eight 5-year-olds, the percentages of not-the-same responses were 57.8 for the two-cooperating-cues condition, 55.4 for the one-cue condition, and 49.0 for the two-conflicting-cues condition.

ANSI (**1989**). ''Specifications for Audiometers'' (American National Standards Institute, New York).

Best, C. T., Morrongiello, B. A., and Robson, R. C. (**1981**). ''Perceptual equivalence of acoustic cues in speech and nonspeech perception,'' Percept. Psychophys. **29**, 191–211.

Best, C. T., Studdert-Kennedy, M., Manuel, S., and Rubin-Spitz, J. (**1989**). ''Discovering phonetic coherence in acoustic patterns,'' Percept. Psychophys. **45**, 237–250.

Blumstein, S. E., and Stevens, K. N. (**1980**). ''Perceptual invariance and onset spectra for stop consonants in different vowel environments,'' J. Acoust. Soc. Am. **67**, 648–662.

Bregman, A. S. (**1990**). *Auditory Scene Analysis: The Perceptual Organization of Sound* (MIT Press, Cambridge, MA).

Broadbent, D. E., and Ladefoged, P. (**1957**). ''On the fusion of sounds reaching different sense organs,'' J. Acoust. Soc. Am. **29**, 708–710.

Eilers, R. E., Oller, D. K., Urbano, R., and Moroff, D. (**1989**). ''Conflicting and cooperating cues: Perception of cues to final consonant voicing by infants and adults,'' J. Speech Hear. Res. **32**, 307–316.

Fitch, H., Halwes, T., Erickson, D., and Liberman, A. (**1980**). ''Perceptual equivalence of two acoustic cues for stop-consonant manner,'' Percept. Psychophys. **27**, 343–350.

Goldman, R., and Fristoe, M. (**1986**). *Goldman Fristoe Test of Articulation* (American Guidance Service, Circle Pines, MN).

Greenlee, M. (**1980**). ''Learning the phonetic cues to the voiced-voiceless distinction: A comparison of child and adult speech perception,'' J. Child Lang. **7**, 459–468.

Holt, L. L., Lotto, A. J., and Kluender, K. R. (**2001**). ''Influence of fundamental frequency on stop-consonant voicing perception: A case of learned covariation or auditory enhancement?'' J. Acoust. Soc. Am. **109**, 764–774.

Jastak, S., and Wilkinson, G. S. (**1984**). *The Wide Range Achievement Test—Revised* (Jastak Associates, Wilmington, DE).

Kewley-Port, D. (**1983**). ''Time-varying features as correlates of place of articulation in stop consonants,'' J. Acoust. Soc. Am. **73**, 322–335.

Kluender, K. R., Lotto, A. J., Holt, L. L., and Bloedel, S. L. (**1998**). ''Role of experience in language-specific functional mappings for vowel sounds as inferred from human, nonhuman and computational models,'' J. Acoust. Soc. Am. **104**, 3568–3582.

Krause, S. E. (**1982**). ''Vowel duration as a perceptual cue to postvocalic consonant voicing in young children and adults,'' J. Acoust. Soc. Am. **71**, 990–995.

Liberman, A. M. (**1957**). ''Some results of research on speech perception,'' J. Acoust. Soc. Am. **29**, 117–123.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., and Studdert-Kennedy, M. (**1967**). ''Perception of the speech code,'' Psychol. Rev. **74**, 431–461.

Liberman, A. M., Isenberg, D., and Rakerd, B. (**1981**). ''Duplex perception of cues for stop consonants: Evidence for a phonetic mode,'' Percept. Psychophys. **30**, 133–143.

Mann, V. A., and Liberman, A. M. (**1983**). ''Some differences between phonetic and auditory modes of perception,'' Cognition **14**, 211–235.

Mann, V. A., and Repp, B. H. (**1980**). ''Influence of vocalic context on perception of the [ʃ]–[s] distinction,'' Percept. Psychophys. **28**, 213–228.

Morrongiello, B. A., Robson, R. C., Best, C. T., and Clifton, R. K. (**1984**). ''Trading relations in the perception of speech by five-year-old children,'' J. Exp. Child Psychol. **37**, 231–250.

Nittrouer, S. (**1992**). ''Age-related differences in perceptual effects of formant transitions within syllables and across syllable boundaries,'' J. Phonetics **20**, 1–32.

Nittrouer, S. (**1996**). ''Discriminability and perceptual weighting of some acoustic cues to speech perception by 3-year-olds,'' J. Speech Hear. Res. **39**, 278–297.

Nittrouer, S., and Miller, M. E. (**1997a**). ''Developmental weighting shifts for noise components of fricative-vowel syllables,'' J. Acoust. Soc. Am. **102**, 572–580.

Nittrouer, S., and Miller, M. E. (**1997b**). ''Predicting developmental shifts in perceptual weighting schemes,'' J. Acoust. Soc. Am. **101**, 2253–2266.

Nittrouer, S., and Miller, M. E. (**1999**). ''Developmental changes in perceptual weighting strategies for speech are contrast specific,'' J. Acoust. Soc. Am. **106**, S2246.

Nittrouer, S., and Studdert-Kennedy, M. (**1987**). ''The role of coarticulatory effects in the perception of fricatives by children and adults,'' J. Speech Hear. Res. **30**, 319–329.

Nittrouer, S., Crowther, C. S., and Miller, M. E. (**1998**). ''The relative weighting of acoustic properties in the perception of [s]+stop clusters by children and adults,'' Percept. Psychophys. **60**, 51–64.

Nittrouer, S., Miller, M. E., Crowther, C. S., and Manhart, M. J. (**2000**). ''The effect of segmental order on fricative labeling by children and adults,'' Percept. Psychophys. **62**, 266–284.

Parnell, M. M., and Amerman, J. D. (**1978**). ''Maturational influences on perception of coarticulatory effects,'' J. Speech Hear Res. **21**, 682–701.

Remez, R. E., Rubin, P. E., Berns, S. M., Pardo, J. S., and Lang, J. M. (**1994**). ''On the perceptual organization of speech,'' Psychol. Rev. **101**, 129–156.

Stevens, K. N., and Blumstein, S. E. (**1978**). ''Invariant cues for place of articulation in stop consonants,'' J. Acoust. Soc. Am. **64**, 1358–1368.

Studdert-Kennedy, M. (**1983**). ''Limits on alternative auditory representations of speech,'' in *Cochlear Prostheses: An International Symposium*, edited by C. W. Parkins and S. W. Anderson (The New York Academy of Sciences, New York), Vol. 405.

Wardrip-Fruin, C., and Peach, S. (**1984**). ''Developmental aspects of the perception of acoustic cues in determining the voicing feature of final stop consonants,'' Lang. Speech **27**, 367–379.

Whalen, D. H., and Liberman, A. M. (**1987**). ''Speech perception takes precedence over nonspeech perception,'' Science **237**, 169–171.