Explaining coherence in coherence masking protection for adults and children

Eric Tarr[a] and Susan Nittrouer

Department of Otolaryngology, The Ohio State University, 915 Olentangy River Road, Suite 4000, Columbus, Ohio 43212

(Received 30 June 2012; revised 21 March 2013; accepted 22 March 2013)

Coherence masking protection (CMP) is the phenomenon in which a low-frequency target (typically a first formant) is labeled accurately in poorer signal-to-noise levels when combined with a high-frequency cosignal, rather than presented alone. An earlier study by the authors revealed greater CMP for children than adults, with more resistance to disruptions in harmonicity across spectral components [Nittrouer and Tarr (**2011**). Atten. Percept. Psychophys. **73**, 2606–2623]. That finding was interpreted as demonstrating that children are obliged to process speech signals as broad spectral patterns, regardless of the harmonic structure of the spectral components. The current study tested three alternative, auditory explanations for the observed coherence of target + cosignal: (1) unique spectral shapes of target + cosignal support labeling, (2) periodicity of target + cosignal promotes coherence, and (3) temporal synchrony across target + cosignal reinforces temporal expectancies. Adults, eight-year-olds, and five-year-olds labeled stimuli in five conditions: F1 only and F1 + a constant cosignal (both used previously) were benchmarks for comparing thresholds for F1 + 3 new cosignals. Children again showed greater CMP than adults, but none of the three hypotheses could explain their CMP. It was again concluded that children are obliged to recognize speech signals as broad spectral patterns.

© *2013 Acoustical Society of America*. [http://dx.doi.org/10.1121/1.4802638]

PACS number(s): 43.71.Ft, 43.71.An [EB] Pages: 4218–4231

I. INTRODUCTION

The term *coherence* as applied in speech perception describes the phenomenon in which separate spectral components of the signal coalesce perceptually such that the auditory qualities of those separate components cannot be recovered. Thus, when a formant transition that typically composes part of a syllable is isolated from that syllable, listeners can evaluate its frequency or direction of glide. However, when that same brief piece of the speech spectrum is presented in combination with the broader spectral array that arises in the course of speech production, those auditory qualities are largely lost to inspection (e.g., Best *et al*., 1989; Mann and Liberman, 1983; Remez *et al*., 2001). Under special circumstances, listeners can learn to tune their attention to those auditory qualities (e.g., Carney *et al*., 1977; McMurray *et al*., 2002), but it is not how listeners normally process signals heard as speech. Contrived means of signal presentation, a great deal of intensive training, or both are required before listeners can recover and inspect the separate elements of the speech signal. The more common finding in speech perception experiments is that listeners are unable to perform that sort of analysis of acoustic details, and the interpretation for that finding is that human listeners organize speech signals perceptually according to principles that promote the recovery of phonologically relevant form (Remez *et al*., 1994). Other terms that have been used for this phenomenon of signal coherence in speech perception include *phonetic coherence* (Best *et al*., 1989; Nygaard, 1993), *perceptual grouping* (Darwin, 1981), and *spectral integration* (Hall *et al*., 2008). However, investigations employing the last of these terms typically do not involve evaluating perception of signal components defined according to phonetically relevant attributes, such as individual formants; rather, that work typically involves filtering the signal into spectrally discrete bands, as is done in the first stage of processing for a cochlear implant. Furthermore, the notion of spectral coherence for speech signals might be cast in the broader light of auditory scene analysis (Bregman, 1990), in which case terms such as *perceptual fusion* apply, and the suggestion that humans organize speech signals in this manner fits the definition of schemas. In general, principles of auditory scene analysis get divided into those that are described as primitive—meaning that the perceptual processes are simple and driven by properties of the acoustic signal—and those described as schema based—meaning that the principles require some application of organizational strategies present in the perceiver. For the most part, primitive principles of perceptual organization are innate to listeners, while schema-based principles are learned, but not always.

A. Phonetic coherence in children's speech perception

When it comes to children, several studies have reported that they are more obliged than adults to integrate signal components across the spectrum so strongly that it is impossible to recover the auditory qualities of those individual components. For example, Nittrouer and Crowther (2001)

[a] Author to whom correspondence should be addressed. Electronic mail: tarr.18@osu.edu

used a paradigm in which two acoustic cues to a phonemic contrast either cooperated or conflicted in how they signaled that contrast. Results showed that 5-yr-olds were less able than older children and adults to recognize the qualities of those separate components. That outcome motivated another investigation of phonetic coherence by children and adults, using an experimental paradigm known as coherence masking protection (CMP). This paradigm was first applied to the study of speech perception by Gordon (1997b), and involves presenting stationary first formants (F1s) from two synthetic vowels (/ɪ/ and /ɛ/) in noise to listeners for labeling. The signal level required for those F1 targets to be labeled correctly is measured when each is presented alone, and when each is presented with a cosignal that is higher in frequency, lower in amplitude, outside the critical band of either F1 target, but constant across those targets. Underlying this procedure is the premise that the protection from masking observed for the F1 + cosignal stimuli compared to the F1-only stimuli is an indication that the low-frequency target (F1) has been integrated with the higher frequency cosignal. According to Gordon, the phenomenon demonstrates that this single spectral component primarily exerts its influence on vowel labeling as part of a broader spectral entity, rather than as an isolated element.

In applying CMP techniques, Gordon (1997b) found that thresholds for adults were 3.2 dB lower in the F1 + cosignal condition than in the F1-only condition; that is, adults showed 3.2 dB of masking protection. Nittrouer and Tarr (2011) replicated Gordon's procedures using adults and children as listeners in order to test the prediction arising from the results of Nittrouer and Crowther (2001), that children would show stronger coherence of spectrally diverse components in speech signals. Gordon's finding was replicated in that study for adults, with 3.3 dB of masking protection observed. For children, the effect was found to be even larger in magnitude: 6.2 and 9.2 dB for 8- and 5-yr-olds, respectively. The magnitude of the CMP effect was interpreted as indicating the strength of coherence across signal components: incorporating that low-frequency target as part of a broader spectral pattern associated with vowel quality evoked much greater protection from masking for children than for adults.

The finding that children demonstrate stronger coherence of spectrally disparate signal components than adults at first might seem surprising, because it suggests that children are better than adults at something that could be considered a perceptual skill: integrating separate spectral components that comprise a signal. That idea contradicts most accounts of development, which typically suggest that children are less skilled than adults at virtually all perceptual, cognitive, and motor activities. From a different perspective, however, that earlier outcome might be described as revealing how strongly children are obliged to process speech signals as integrated spectral patterns: It is not so much that children benefited more than adults from the high-frequency cosignal as it is that children were disadvantaged when it was not present. Cohen's *d*'s computed on CMP scores for adults and 5-yr-olds in Nittrouer and Tarr (2011) were 2.49 for the F1-only condition and 1.40 for the F1 + cosignal condition, meaning that age-related differences were greater for the F1-only condition than for the F1 + cosignal condition. Thus, it seems fair to suggest that children required the broad spectral pattern for judging vowel quality, whereas adults were able to make do with F1 alone, to some extent.

From still another perspective, the demonstration that children so strongly require a broad spectral pattern for making a phonemic judgment might be viewed as mirroring their poor abilities at making auditory judgments about narrow-frequency signals: School-age children have poorer discrimination capacities for tones than adults (Jensen and Neff, 1993), and infants show similar detection thresholds regardless of whether they are listening for a selected tone or for a variety of tones (Bargones and Werner, 1994). That latter result differs from what is found for adults, where selective listening to a narrow region of the spectrum leads to lowered thresholds compared to broad-spectral listening. Consequently, what might be viewed as an immature perceptual strategy where nonspeech signals are concerned (i.e., lack of ability to attend selectively to specific frequency regions) might be viewed as an adaptive strategy when it comes to speech (i.e., attending to broad spectral swatches of phonemically relevant signals).

The reason for this last suggestion is that the perceptual strategy of listening across the spectrum seems to serve an important role in acquisition. In addition to learning how to understand the speech of others, children need to learn how to produce speech themselves. The time-varying, broad spectral shape of the speech signal arises from the articulatory maneuvers of the speaker, so provides information about how to produce speech. Attending to that broad shape across the entire spectrum allows children to recover information about what to do with their own articulators (Best *et al.*, 1989; Boysson-Bardies *et al.*, 1986; Nittrouer and Crowther, 2001; Studdert-Kennedy, 2000). Thus, this spectrally broad perceptual strategy may represent one kind of schema that is innate, or at least learned early in life, and that suggestion provides a reason for why it is beneficial for children to integrate spectral components as strongly as they do. Another question concerns what underlies that strong signal coherence, and that question was the focus of the study reported here. Exploring answers to this question serves theoretical as well as clinical purposes. On the theoretical side, the current experiment explored whether the coherence displayed in children's (as well as adults') responses might be explained by either of two principles that fit the definition of primitive principles of auditory scene analysis (Bregman, 1990), and examined one other possible source of the effect. Examining the relationship between these principles and the organization of acoustic signals for children sheds light on ideas about the innateness of these principles. On the clinical side, understanding the basis of that broad-spectral listening in children might provide hints as to how signals could be processed in auditory prostheses for children with hearing loss. For example, combined electric-acoustic stimulation is an idea gaining credence. Because the acoustic signal is generally low frequency and the electric signal is higher frequency, examining basic mechanisms underlying integration of these spectral components should be worthwhile.

B. Hypotheses to be tested

The work on CMP of Gordon (1997a,b) was based on the premise that two auditory principles likely explain the strong spectral coherence observed across vowel formants: (1) all formants comprising a single vowel share a common fundamental frequency, so have the same harmonic structure and (2) all formants start and stop at the same time. In addition to replicating the basic CMP effect for adults and children, Nittrouer and Tarr (2011) tested whether the first of these principles accounted for that effect for each listener group. That was done primarily by creating stimuli in which the F1 target and the F2/F3 cosignal had different fundamental frequencies, or pitch, which meant they had different harmonic structures. Using those stimuli, it was found that children's CMP was unaffected by inconsistent harmonic structure across formants. For adults, on the other hand, CMP disappeared when the target and cosignal had different harmonic structures. That outcome would be predicted by auditory scene analysis (Bregman, 1990), which holds that one principle accounting for integration of separate spectral components is common harmonic structure across those components. Accordingly, it would seem this primitive principle of auditory grouping was eliminated as a potential account for the strong coherence observed in children's speech perception. The current study continued the search for explanations of that strong coherence, by proposing and testing three new hypotheses.

Before moving to a description of those new hypotheses, one other outcome of Nittrouer and Tarr (2011) is worth mentioning. In addition to measuring CMP using stimuli with different harmonic structures across the target and cosignal, a sine-wave target was used. That single sine wave was not recognizable as speech on its own, but was incorporated into a speech-like percept when combined with a harmonically structured cosignal. CMP was obtained in that condition. In fact, the lowest thresholds of any condition, in earlier experiments or in the one reported here, were observed in that sine-wave target + speech cosignal condition. The mean threshold (and SD) for adults in that condition was 55.6 dB (1.4 dB) with background noise of 62 dB. That finding is relevant because, as will be seen, it means that the amount of CMP exhibited by adults has not been constrained by sensory limits in previous or current work.

In the current study, five stimulus conditions were used to test the three hypotheses proposed. The two main stimulus conditions used in the earlier experiment were replicated in this experiment, and served as benchmarks against which to compare performance in the other three conditions. In the first of the two baseline conditions, the same two F1 targets (one for /ɪ/ and one for /ɛ/) as used by Nittrouer and Tarr (2011) were presented in noise filtered below 1 kHz. That condition is described as the *F1-only* condition. In the other condition being replicated, each of those F1 targets was presented synchronously with the same F2/F3 cosignal above the 1-kHz noise cutoff. In the current study, that condition is termed the *constant-formants* condition. It was the original condition incorporating a cosignal used by Gordon (1997b), and serves as a benchmark because it demonstrates the basic phenomenon. Listeners are more sensitive to energy in a target when it is recognized as part of a broader spectral pattern than as an independent percept. The paradigm used for evaluating this effect employs a labeling task, rather than a detection task, so it is imperative that the cosignal contribute no information that could help in identifying the target. If it did, interpretation of outcomes would be difficult because the effect would not be attributable solely to the idea that the target was being perceived as part of that broader pattern. Furthermore, the CMP paradigm derives from comodulation masking release paradigms in which manipulation of off-target maskers can affect thresholds. In both paradigms, the off-target component is traditionally uninformative. A first goal of the current experiment involved reliability, asking if labeling thresholds measured for each condition in the earlier experiment could be replicated for each age group with different listeners. Three new stimulus conditions provided an opportunity to test the three new hypotheses offered as possible explanations for the CMP found for adults and children, and those new conditions are described below.

1. Hypothesis 1: Unique spectral shape

Developmental psycholinguists have proposed that children attend as strongly as they do to shape across the entire spectrum because that structure provides information needed to learn how to produce speech (e.g., Studdert-Kennedy, 2000). Gross spectral structure arises from the actions of the vocal tract above the larynx during speech production. Strong evidence that children attend to this level of structure, and do so from a young age, is provided by Boysson-Bardies *et al.* (1986), who showed that the long-term, average spectra of babble produced by 10-months-old infants matched the long-term spectra of speech from adults in the language community of which the infants were a part. Thus, these infants already were sensitive to the postures and movements of the vocal tract that shape the gross spectral envelope of speech for their native language. The current study tested the possibility that children's enhanced CMP and its resilience in the face of disruptions in harmonicity are related to that strong attention to gross spectral shape.

In both Gordon (1997b) and Nittrouer and Tarr (2011), the cosignal remained constant in shape and spectral location across F1 targets in the constant-formants condition. Therefore, the cosignal was described as uninformative because it could provide no information in addition to F1 frequency to help distinguish the two stimuli. However, that description may not be quite accurate. In fact, when that constant cosignal is integrated with the F1 target, the shape of the entire spectrum varies depending on F1. It may be that having had two spectra that differed in overall shape meant that the F1 + cosignal stimuli were more distinctive than the F1-only stimuli, an effect that could have facilitated children's abilities to label stimuli in noise. That hypothesis was tested in the current experiment by shifting the location of the cosignal so that the spectral shape of the combined stimulus (F1 + cosignal) was constant across the two stimuli with different F1 frequencies. If CMP is facilitated by having two spectra that differ in overall shape, the effect should be diminished or eliminated in this condition.

2. Hypothesis 2: Periodicity

Nittrouer and Tarr (2011) tested the hypothesis that harmonicity explained the spectral coherence exhibited in the CMP of adults and children by constructing stimuli in which the F1 target and the F2/F3 cosignal had different fundamental frequencies. Using those stimuli, evidence was found that formants need to share a common harmonic structure in order for adults to demonstrate CMP; for children, no evidence was found to support the hypothesis that this principle explains their CMP. That finding is consistent with results of others demonstrating that children require greater frequency differences than adults to discriminate pitch (e.g., Jensen and Neff, 1993), or at least are less attentive to pitch changes (Moore *et al.*, 2008). In the study of Nittrouer and Tarr, however, all formants consisted of harmonics (i.e., had periodic fine structure), regardless of whether those harmonics were similarly spaced or not. Consequently, periodicity itself could have been the basis of integration. The masking noise was aperiodic, so the strategy employed by children could be to integrate the parts of the sensory input consisting of periodic structure.

The current study was designed to test the hypothesis that spectral components are integrated in children's speech perception based on their simply being periodic by including a condition in which stimuli were identical in gross spectral shape to the constant-formants condition of Nittrouer and Tarr (2011), but the cosignal was comprised of noise. There is some evidence to suggest that children might be less attentive to noise components in speech signals. For example, in fricative perception, children weight fricative noises less strongly in their phonemic decisions than adults (e.g., Nittrouer, 1992, 2002; Nittrouer and Miller, 1997; Nittrouer and Studdert-Kennedy, 1987), but weight formant transitions more strongly. The interpretation of those findings has been that children are more attentive to time-varying changes in formant frequencies. However, those outcomes might instead suggest that children are especially attentive to periodic structure in the speech signal. Accordingly, children may group together spectral components that have periodic structure. The current study examined that possibility. If true, children should show diminished CMP when an aperiodic cosignal is used.

3. Hypothesis 3: Temporal synchrony

Darwin (1981) demonstrated that formants comprising a vowel are less likely to be grouped together into a unitary percept when they start at different times than when they have synchronous onsets. In agreement with that finding, Gordon (1997a, 2000) reported that CMP was hindered when the target and cosignal had asynchronous onsets or offsets, and that finding was observed for both speech and nonspeech stimuli. The interpretation applied to those results by Gordon was that one mechanism accounting for CMP is that the cosignal serves as a temporal marker that helps the listener locate the target in the noise, thus facilitating its recognition. This idea is described by Gordon as a corollary of another idea often invoked to explain comodulation masking release, known as "dip listening." According to that latter idea, flanking noise bands help mark the locations in which the masking noise over the signal is lowest, thus indicating when the listener should attend. Gordon coined the term "peak listening" to refer to the notion that in CMP the cosignal similarly marks the interval when the listener should be attending. The current study tested the temporal-synchrony hypothesis by using a cosignal with a flat spectral shape and onsets and offsets temporally synchronous with the target. Specifically, flat noise above 1 kHz was the cosignal. If indeed one mechanism underlying the CMP observed in past experiments was that having the cosignal start and stop at the same time as the target served the purpose of having a marker outside the critical band of the target signal, then CMP should be observed in this condition, thus supporting the temporal synchrony hypothesis. An implication would be that CMP does not necessarily arise due to integration across the spectrum. Rather, that high-frequency cosignal draws attention to the target. That means the target might evoke the vowel category label by itself. Work by Werner *et al.* (2009) showed that infants fail to attend to marked (cued) temporal intervals to the same extent as adults when asked to detect nonspeech tones, so it could have been predicted going into this experiment that adults might show the effect, but not children.

This stimulus condition was designed as it was (rather than using a cosignal that had an asynchronous onset, offset, or both with the F1 target) so that the pattern of results that would support this third hypothesis differed from those that would support the first two hypotheses. Support for each of the first two hypotheses depended on failing to evoke CMP with the new stimuli. If predictions were identical across all three hypotheses, and if failure to evoke CMP was observed for all three new sets of stimuli, the most that could be concluded is that CMP does not occur for any kind of stimulus other than the constant-formants stimuli. That outcome would restrict interpretations that could be made about what does explain CMP. In particular, this third condition was designed to create different predictions from those made for the shaped-noise condition. In this way, predictions are opposite across the two noise conditions.

4. An alternative explanation: Speech-like cosignals trigger integration

Although not a hypothesis that could be tested in the strictest sense, results across the various conditions in the current study were able to provide evidence regarding the suggestion from Nittrouer and Tarr (2011) that a schema-based principle might explain children's strong spectral integration for speech signals. In particular, three of the four conditions involving a cosignal imposed a formant-like shape on that cosignal; the fourth did not. If listeners' CMP effects—especially those of children—were related across those conditions, but not related to CMP in the fourth condition, it could be suggested that those listeners were influenced by the cosignal having a speech-like shape.

C. Summary

The current study tested three new hypotheses about potential sources of CMP for adults and children: unique spectral shape, periodicity, and temporal synchrony. To test these hypotheses, stimuli used in an earlier experiment were presented, both to replicate the earlier work and to serve as benchmarks against which to evaluate outcomes with the three new sets of stimuli. These previously used stimuli are described as *F1 only* and *constant formants*. The three new stimulus sets are described as (1) *shifted formants*, designed to provide two stimuli in which the gross spectral shape of F1 + cosignal stimuli were the same across F1 frequencies; (2) *shaped noise*, designed to provide stimuli in which the cosignal lacked periodicity, but preserved the gross spectral shape of the constant-formants stimuli; and (3) *flat noise*, designed to provide a condition in which the cosignal served only to mark the temporal interval in which the target could be found. Table I succinctly displays the three hypotheses tested by this work. In addition, this study evaluated the proposal that children's strong and apparently obligate spectral integration might be evoked by signals shaped as they would be in the course of speech production.

II. METHOD

A. Listeners

Sixty-two listeners were tested in this experiment: 21 adults between the ages of 18 and 39 yr; 20 children between 8 yr, 0 months and 8 yr, 11 months; and 21 children between 5 yr, 2 months and 5 yr, 11 months. All participants (or in the case of children, their parents on their behalf) reported having normal hearing, speech and language. None of the children had more than five episodes of otitis media before the age of 3 yr. All participants passed hearing screenings of the frequencies of 0.5, 1.0, 2.0, 4.0, and 6.0 kHz presented at 25 dB hearing level to each ear separately at the beginning of the first test.

B. Equipment and materials

Testing occurred in a soundproof booth, with the computer that controlled stimulus presentation and recorded responses in an adjacent room. A Welch Allen TM-262 audiometer and TDH-39 headphones were used for the hearing screening. Stimuli were presented from the computer using a Soundblaster digital-to-analog converter, a Samson C-que 8 headphone amplifier, and AKG-K141 headphones.

Two pictures on cardboard (6 in. × 6 in.) were used so that listeners could point to the picture representing their response choice after each stimulus presentation. One picture was of a dog biting a woman's leg (*bit*), and the other was of a man with playing cards in his hands and stacks of poker chips in front of him (*bet*).

C. Stimuli

Five sets of stimuli were created using a sampling rate of 10 kHz, with low-pass filtering below 5 kHz and 16-bit digitization. Each set consisted of a pair of synthetic vowels: /ɪ/ and /ɛ/. The F1 targets were the same for all sets of stimuli, and were derived from vowels created with the Sensimetrics "SENSYN" software, a version of the Klatt synthesizer. All stimuli were 60 ms long, which included 5-ms on and off ramps. The synthesized vowels from which the two F1 formants were derived consisted of three steady-state formants. F1 was 375 Hz for /ɪ/ and 625 Hz for /ɛ/. F2 and F3 were 2200 and 2900 Hz, respectively in both vowels. Formant bandwidths (at 3 dB below peak amplitude) were 50 Hz for F1, 110 Hz for F2, and 170 Hz for F3. Fundamental frequency (f0) was stable at 125 Hz.

In all five stimulus sets, F1 targets were embedded in masking noise that was 600-ms long. That noise was created in MATLAB and low-pass filtered below 1000 Hz with a transition band to 1250 Hz. In all conditions, stimuli started 420 ms after the onset of the masking noise.

1. Previously used conditions

Two conditions, F1 only and constant formants, were included in this experiment for comparison because they had been used in previous CMP experiments (Gordon 1997b; Nittrouer and Tarr, 2011).

a. F1 only. This condition was created by low-pass filtering the two synthetic vowels using a digital filter with attenuation starting at 1000 Hz and a transition band to 1250 Hz. This condition is required in order to compute CMP, which is obtained by subtracting labeling thresholds for the cosignal condition from thresholds for the F1-only condition. These F1-only stimuli were combined with a spectrally adjacent *cosignal* in all remaining conditions.

TABLE I. Three major hypotheses tested, claim of each hypothesis, the stimulus condition created to test the hypothesis, and where thresholds would fall relative to each of the benchmark conditions, if the hypothesis is supported.

Hypothesis to be tested	Claim of the hypothesis	Stimulus condition created to test the hypothesis	How thresholds for test stimuli should compare to F1-only stimuli	How thresholds for test stimuli should compare to constant-formants stimuli
Unique spectral shape	The overall shape of the spectrum provides information that distinguishes the stimuli from each other, facilitating CMP	Shifted formants	Same thresholds	Higher thresholds
Periodicity	Separate spectral components cohere because they each consist of periodic signals, distinguishing them from the noise masker and facilitating CMP	Shaped noise	Same thresholds	Higher thresholds
Temporal synchrony	A cosignal draws attention to the interval in which the target signal occurs, facilitating CMP	Flat noise	Lower thresholds	Same thresholds

b. Constant formants. This condition consisted of the F1-only stimuli combined with a synthetic cosignal, acoustically identical in both vowels. To create the cosignal, the original /ɛ/ stimulus was high-pass filtered with a low-frequency cutoff at 1250 Hz and a transition band down to 1000 Hz. This cosignal was combined with both F1-only stimuli to ensure it was acoustically identical across tokens. In vowel contrasts involving only a height difference, as used here, F1 is sufficient to cue the distinction. The F2/F3 cosignal in these stimuli was set to be 12 dB lower than the F1-only stimuli. The top panel of Fig. 1 shows smoothed spectra of the F1-only stimuli (below 1 kHz) and the constant-formants stimuli (across the spectrum). Incorporating this condition provided measures of both reliability and effect sizes. An estimate of the reliability of measurement was obtained by comparing results for this condition from Nittrouer and Tarr (2011) to results from the current study. And because these constant-formants stimuli are the standard cosignal condition, so to speak, effect sizes from other conditions can be compared to the magnitude of CMP obtained for these stimuli.

2. Novel conditions

Three new conditions were included to test the three experimental hypotheses. These involved cosignals with shifted formants, (speech) shaped noise, and flat noise.

a. Shifted formants. This condition used the same target and cosignal for the /ɛ/ stimulus as used in the constant-formants condition. However, a new cosignal was created for the /ɪ/ target stimulus. The new cosignal was synthesized with F2 and F3 peaks at 1950 and 2650 Hz, respectively. Therefore, the formant resonances in the /ɛ/ and /ɪ/ stimuli were separated by the same linear distance, making the spectral envelope the same for each vowel. This change in the cosignal had no perceived effect on the quality of the /ɪ/ vowel. Formant peaks of the new cosignal were again 12 dB below the F1 peak. The bottom panel of Fig. 1 shows smoothed spectra of the shifted-formants stimuli. This condition tested the hypothesis that listeners who rely on unique, overall spectral shapes for vowel labeling (and so for CMP) would have difficulty with stimuli in which that shape remains constant across stimuli.

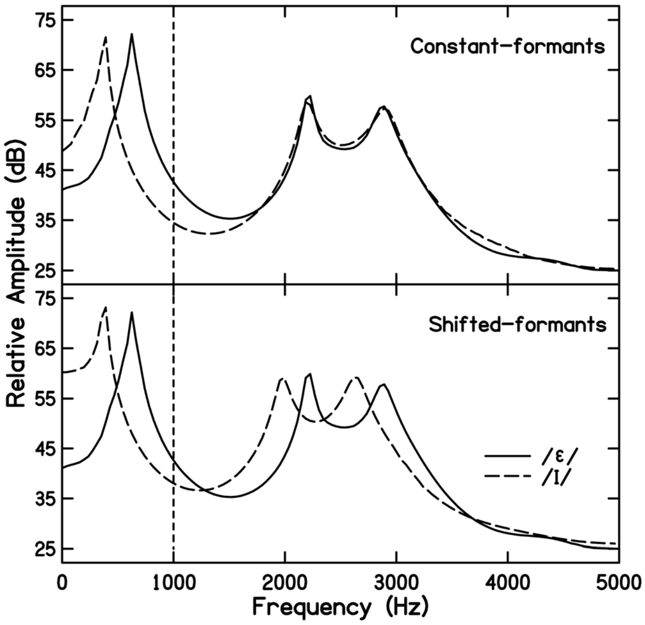

FIG. 1. In both panels, F1-only stimuli are shown as that portion of the spectrum below 1 kHz. In the top panel, full spectra are the stimuli in the constant-formants condition. In the bottom panel, full spectra are the stimuli in the shifted-formants condition.

b. Shaped noise. This condition consisted of the F1-only stimuli combined with a (speech) shaped-noise cosignal with formant resonances identical to the constant-formants condition. To generate that cosignal, spectrally flat noise was created using a random number generator in MATLAB. LPC source-filter separation was performed on the /ɛ/ constant-formants stimulus to extract the spectral envelope. This envelope was applied to the noise signal and high-pass filtered using the same digital filter described for the constant-formants condition. Thus the cosignal had the identical speech-like shape as the constant-formants stimuli, but without the harmonic source. The amplitude of the shaped-noise cosignal was adjusted so that the resonant peaks of F2 and F3 were 12 dB lower than the F1 resonant peak. This condition tested the hypothesis that listeners integrate components consisting of periodic sources.

c. Flat-noise. This condition consisted of the F1-only stimuli combined with a high-pass filtered noise cosignal. The noise was obtained by applying the filter function used to obtain the cosignal in the constant-formants condition to noise generated in MATLAB. The flat-noise cosignal was adjusted in amplitude to match the RMS amplitude of the shaped-noise cosignal. This condition tested the hypothesis that onset/offset synchrony between the low-frequency target and the cosignal explains CMP. If temporal synchrony is sufficient to evoke CMP, then listeners should continue to show masking protection when the cosignal is spectrally flat noise. If listeners show masking protection for the shaped-noise condition, but not the flat-noise condition, then it can be concluded that temporal synchrony is not sufficient to explain CMP, thus supporting the general position that the F1 target acquires its effect on labeling most strongly as part of a coherent phonetic percept, rather than by itself.

D. Procedures

The hearing screening was completed at the beginning of the session. Adults and 8-yr-olds completed all five stimulus conditions in a single, 45 min session. Five-year-olds completed the experiment in two, 45 min sessions, on different days. During the first session they completed two of the stimulus conditions and during the second session they completed the remaining three stimulus conditions.

The order of presentation of stimulus conditions varied across listeners. The starting condition was randomly selected from the four cosignal conditions (constant-formants, shifted-formants, shaped-noise and flat-noise), with the stipulation that each starting condition was evenly distributed across subjects in each age group. If a listener started with a cosignal condition involving synthetic speech (constant-formants or shifted-formants), the second condition was randomly selected from the noise cosignal conditions (shaped-noise or flat-noise). If the first condition involved a noise cosignal, the second condition was synthetic speech. The third condition was always the F1-only stimuli. The fourth condition matched the cosignal type (synthetic speech or noise) of the first condition, but was the condition that had not been tested. The fifth condition was the remaining untested stimulus condition.

All listeners completed general training at the beginning of the experiment, followed by four steps in each of the stimulus conditions: (1) condition-specific training, (2) a pretest, (3) actual testing, and (4) a posttest.

1. General training

The experimenter introduced each picture separately, and told the listener the name of the word associated with that picture. Listeners practiced pointing to the correct pictures and saying the correct words after they were spoken by the experimenter (five tokens of each word in random order). Having listeners both point to the picture and say the word ensured that they were correctly associating the word and the picture.

Next the 60-ms constant-formants stimuli were introduced, without noise. Twenty-five tokens of each stimulus were presented, in random order, with the only stipulation that no more than two tokens of one stimulus could be presented before the other stimulus was presented. Stimuli were presented at 74 dB sound pressure level (SPL). Listeners were instructed that they would be hearing "a little bit" of each word. They had to point to the correct picture and say the correct word that the little bit came from. Feedback was given.

2. Condition-specific training

This training consisted of 50 presentations of stimuli (25 tokens of each) to be used in the condition to follow. Stimuli were presented at 74 dB SPL without masking noise. Again, listeners said each word and pointed to the picture associated with it. Feedback was provided.

3. Pretest

Up to 50 stimuli were presented without noise or feedback in the pretest. As soon as the listener responded correctly to nine out of ten consecutive presentations, the pretest was stopped. If 50 stimuli were presented without the listener ever responding correctly to nine out of ten consecutive presentations, that listener was not tested in that particular condition.

4. Adaptive testing

An adaptive procedure (Levitt, 1971) was used to find the signal-to-noise ratio (SNR) at which each listener could provide the correct vowel label 79.4% of the time. The noise was held constant throughout testing at 62 dB SPL, and the level of the signal varied. The initial signal level was 74 dB SPL. After three consecutive correct responses, the level of the signal decreased by 8 dB. That progression, or "run," of decreasing signal level by 8 dB after three correct responses continued until the listener made one labeling error, at which time the level of the signal increased by 8 dB. That shift in direction of amplitude change is termed a "reversal." Signal amplitude continued to increase until the listener gave three correct responses, when another reversal occurred. During the first two runs (one with decreasing amplitude and one with increasing), signal level changed by 8 dB on each step. During the next two runs, signal level changed by 4 dB. Across the next and final twelve runs, level changed by 2 dB on each step. The mean signal level at the last eight reversals was used as the threshold. No feedback was provided and stimuli were presented in an order randomized by the software with no restrictions on how many times one vowel could be presented before the other was presented.

This test procedure matched what was used by Nittrouer and Tarr (2011), but is not as extensive as what was done by Gordon (1997a,b, 2000), who had only adults as listeners. In all of Gordon's work, more test runs were presented. The reason for modifications in procedures was that children do not tolerate testing near threshold for prolonged periods of time very well. To compensate for that fact, training was enhanced, compared to what Gordon had done. These modifications of enhanced training, but truncated adaptive testing are similar to those suggested by Aslin and Pisoni (1980) as ways to accommodate the special circumstances of working with children.

5. Posttest

After testing in each condition was completed, listeners heard ten stimuli at 74 dB SPL without noise and without feedback. They needed to respond correctly to nine of them. If they did not do so their data were not included in the analysis.

Listeners had to meet the pre and posttest inclusionary criteria for all conditions in order for their data to be included. This restriction ensured that the adaptive tracking procedure was not affected by listeners not reliably knowing the vowel labels.

III. RESULTS

One adult (5%) and seven 5-yr-olds (33%) failed to meet either the pre or posttest criterion described above. The adult failed the posttest for both the F1-only and constant-formants conditions. Five-year-olds who failed any part of testing on the first day were not tested on the second day, so it is not possible to get a complete account for these children of what they could or could not do. Six 5-yr olds failed at least one condition the first day, so none of them was tested in the F1-only condition. Four of those six children failed both conditions in which they were tested; the other two failed just one condition. Thus, ten conditions were failed in total across the six 5-yr-olds. Four of those failures were in the flat-noise condition, and the remaining six failures were evenly distributed across the constant-formants, shifted-formants, and shaped-noise conditions. One 5-yr-old passed all pre and posttests on the first day, but failed the posttest of

TABLE II. Means (and standard deviations) of thresholds (in dB SPL) obtained in each condition.

		Condition									
Age		F1-only		Constant-formants		Shifted-formants		Shaped-noise		Flat-noise	
	n	M	SD	M	SD	M	SD	M	SD	M	SD
Adults	20	61.4	1.5	58.1	0.8	56.3	2.0	58.9	2.0	59.1	2.3
8-yr-olds	20	65.6	4.2	59.2	1.8	59.1	2.4	61.7	3.2	64.9	4.9
5-yr-olds	14	67.7	4.0	60.5	2.3	60.1	4.4	61.7	5.0	65.4	5.6

the F1-only condition on the second day. After data for these participants were eliminated, data remained from 20 adults, 20 8-yr-olds, and 14 5-yr-olds.

Table II shows thresholds for all groups and conditions used in this experiment. Before any statistical analyses were performed, these results were examined to see if they met assumptions of normal distribution and homogeneity of variance. Generally speaking, the assumption of normal distribution of outcomes in each condition was met. As can be seen in Table II, children showed somewhat greater variance than adults, but it was not considered so severe of a difference that parametric tests could not be performed.

A. Comparison of current results to Nittrouer and Tarr (2011)

Mean thresholds (and SDs) from Nittrouer and Tarr (2011) for the F1-only condition were 61.2 (3.4), 65.0 (4.1), and 70.2 (3.8) for adults, 8-yr-olds, and 5-yr-olds, respectively. For the constant-formants condition, these values were 57.9 (1.4), 58.8 (1.1), and 61.1 (2.9), for the same groups in the same order. These group means differ by no more than 0.6 dB across studies, except for 5-yr-olds' thresholds for the F1-only condition. In the current experiment, this mean threshold is 2.5 dB lower than in the earlier experiment. For each age group, *t* tests were performed on thresholds from each condition used in both experiments (i.e., F1 only and constant formants). All were non-significant with $p > 0.10$, except for 5-yr-olds' thresholds for the F1-only condition, $t(35) = 2.21$, $p = 0.034$. This one significant outcome did not affect the current experiment because the 2.5 dB difference across experiments in this one threshold is modest compared to the size of the CMP effect itself for 5-yr-olds: 9.1 dB in Nittrouer and Tarr and 7.2 dB in this study. These comparisons for conditions common to both studies reveal good reliability for the test measures.

B. Age and condition effects

To examine overall age and condition effects, a two-way analysis of variance (ANOVA) was performed on the thresholds shown in Table II, with age as the between-subjects factor and condition (F1-only, constant-formants, shifted-formants, shaped-noise, and flat-noise) as the within-subjects factor. The main effect of age was significant, $F(2, 51) = 13.37$, $p < 0.001$, $\eta^2 = 0.344$. *Post hoc* comparisons showed that adults' thresholds were lower overall than those from both groups of children ($p < 0.001$), but thresholds were similar across children's groups ($p > 0.10$). The main effect of condition was also significant, $F(4, 204) = 78.67$, $p < 0.001$, $\eta^2 = 0.607$. In addition, the age × condition interaction was significant, $F(8, 204) = 4.52$, $p < 0.001$, $\eta^2 = 0.151$. This last outcome indicates variability in the pattern of performance across conditions for each age group. That variability was explored by examining outcomes across conditions for each age group separately, according to the predictions made in the Introduction.

C. Testing the hypotheses

To test the three hypotheses posed in the Introduction, one-way, repeated-measures ANOVAs with condition as the factor were performed on data for each age group separately. The *post hoc* comparisons derived from those ANOVAs served as the outcomes used to evaluate whether or not there was support for each of these hypotheses.

1. Hypothesis 1: Unique spectral shape

This hypothesis suggests that the unique spectral shapes of stimuli in the constant-formants condition enhance the distinctiveness between the two stimuli beyond what is found in the F1-only condition, and that enhanced distinctiveness could help CMP by facilitating vowel labeling. This suggestion offers one possible explanation for the CMP observed in earlier studies (Gordon, 1997b; Nittrouer and Tarr, 2011). Support for the position would be provided if the stimuli in the shifted-formants condition diminished or eliminated CMP. Thus, predictions were that thresholds in the shifted-formants condition would be the same as those for the F1-only condition (i.e., not significantly different) and would be higher than those for the constant-formants condition. To test that hypothesis, outcomes for the shifted-formants condition were compared to both those benchmark conditions. Table III shows those results for each group. In this and the two following tables, Bonferroni significance levels are shown, along with a statement of whether those outcomes support acceptance or rejection of the hypothesis. This information is provided for threshold comparisons with the F1-only condition on the left and the constant-formants condition on the right.

For children, outcomes clearly supported rejecting the hypothesis: thresholds for stimuli in the shifted-formants condition were lower than in the F1-only condition, thus showing CMP, and similar to those in the constant-formants condition. For adults, outcomes also supported rejecting the hypothesis, even though there was a significant difference between thresholds for the constant-formants and the

TABLE III. Results of comparisons between thresholds in the shifted-formants condition and each of the two benchmark conditions (F1-only and constant-formants) for each age group separately. Shown here are p values using Bonferroni adjustments for multiple comparisons. Also shown is whether each specific outcome supports acceptance or rejection of the unique spectral shape hypothesis.

	F1-only		Constant-formants	
Shifted-formants	Significance	Hypothesis	Significance	Hypothesis
Adults	<0.001	Reject	0.006	Reject[a]
8-year-olds	<0.001	Reject	NS	Reject
5-year-olds	<0.001	Reject	NS	Reject

[a]Although adults' thresholds for these two conditions are different, the hypothesis would only be supported if those thresholds were higher in the shifted-formants condition. Instead thresholds were lower in the shifted-formants condition, so the hypothesis is rejected.

TABLE IV. Results of comparisons between thresholds in the shaped-noise condition and each of the two benchmark conditions (F1-only and constant-formants) for each age group separately. Shown here are p values using Bonferroni adjustments for multiple comparisons. Also shown is whether each specific outcome supports acceptance or rejection of the periodicity hypothesis.

	F1-only		Constant-formants	
Shaped-noise	Significance	Hypothesis	Significance	Hypothesis
Adults	<0.001	Reject	NS	Reject
8-year-olds	<0.001	Reject	0.001	Support[a]
5-year-olds	0.004	Reject	NS	Reject

[a]Although thresholds for the shaped-noise stimuli were higher than for the constant-formants stimuli, CMP was nonetheless observed; thresholds were lower than for the F1-only stimuli. Consequently, the hypothesis that periodicity accounts entirely for CMP for 8-year-olds cannot be accepted.

shifted-formants conditions. Adults actually showed *lower* thresholds for the shifted-formants than for the constant-formants condition; accepting the hypothesis depended on finding *higher* thresholds, indicating diminished CMP. Apparently the shifted-formants stimuli were more distinctive than the constant-formants stimuli for adults. That outcome might reflect adults' abilities to attend to specific frequency regions within the stimuli: For the traditional, constant-formants stimuli, only one formant differed in frequency across vowels, but for the shifted-formants stimuli, all three formants differed in frequency. But regardless of the mechanism underlying adults' enhanced CMP, the hypothesis that CMP is facilitated by having spectra that differ in overall shape can be rejected for all groups because CMP was equivalent or enhanced when overall spectral shape was the same across vowels.

2. *Hypothesis 2: Periodicity*

This hypothesis suggests that the basis of children's CMP is that they integrate signal components that are periodic. To examine this hypothesis, a cosignal identical to that of the constant-formants cosignal was generated, but using aperiodic noise instead of periodic structure. Predictions were that thresholds for this shaped-noise cosignal would be the same as those for the F1-only condition and would be higher than those for the constant-formants condition. To test that hypothesis, thresholds were compared between the shaped-noise condition and each benchmark condition. Table IV shows those results for each group.

The hypothesis that separate spectral components cohere based on the fact that they each consist of periodic signals can clearly be rejected for adults and 5-yr-olds: For both groups, thresholds for the shaped-noise stimuli were lower than for the F1-only stimuli and equivalent to those of the constant-formants stimuli. Both of those outcomes support rejection of the hypothesis. For 8-yr-olds, the evidence is mixed. Thresholds were significantly lower in the shaped-noise condition than in the F1-only condition, supporting rejection of the hypothesis. However, 8-yr-olds had higher thresholds for the shaped-noise stimuli than for the constant-formants condition. On its own, that outcome could support the periodicity hypothesis because CMP is diminished when the cosignal in not periodic, but two factors militated against accepting it. First, 8-yr-olds showed some CMP for the shaped-noise stimuli (3.9 dB); the magnitude of that CMP was just not as great as the CMP measured for the constant-formants stimuli. Second, both younger and older listeners showed CMP comparable in magnitude for the shaped-noise and constant-formants conditions. These outcomes mean that if periodicity was the primary basis of CMP for 8-yr-olds, the developmental course would have to be U shaped: at younger and older ages it is not the primary basis. That scenario seems unlikely, especially in the context of there being some amount of CMP for 8-yr-olds.

3. *Hypothesis 3: Temporal synchrony*

This hypothesis suggests that CMP may arise from having a cosignal that marks the interval in which the target signal occurs (Gordon, 2000). That hypothesis was tested by designing a cosignal that could do no more than mark the temporal location, namely, rectangular, or flat, noise. Predictions were that if this hypothesis is valid then thresholds in the flat-noise condition should be lower than those for the F1-only condition and similar to those of the constant-formants condition. Results for the relevant comparisons are shown in Table V.

For the testing of this hypothesis, the evidence is clear for each age group. The hypothesis that CMP is facilitated by a cosignal that draws attention to the temporal interval in

TABLE V. Results of comparisons between thresholds in the flat-noise condition and each of the two benchmark conditions (F1-only and constant-formants) for each age group separately. Shown here are p values using Bonferroni adjustments for multiple comparisons. Also shown is whether each specific outcome supports acceptance or rejection of the temporal synchrony hypothesis.

	F1-only		Constant-formants	
Flat-noise	Significance	Hypothesis	Significance	Hypothesis
Adults	0.003	Support	NS	Support
8-year-olds	NS	Reject	<0.001	Reject
5-year-olds	NS	Reject	0.007	Reject

which the target signal occurs is supported for adult listeners, but rejected for both children's groups. Thus, the idea is supported that temporal synchrony contributes to CMP for adults, to at least some extent, as Gordon (2000) demonstrated. Of course, it is not the only mechanism that can be shown to explain the effect. For example, Nittrouer and Tarr (2011) observed that CMP for adults was eliminated when harmonicity was dissimilar across the target and cosignal. For adults at least, several mechanisms apparently underlie the effect.

D. An alternative explanation

Testing the three hypotheses explicitly proposed in this study involved comparing CMP for the three novel sets of stimuli created for the study against CMP for stimuli used by Gordon (1997b) and Nittrouer and Tarr (2011). Those comparisons provide evidence for the claims that children's CMP does not seem to be explained by (1) the unique spectral shape of the overall stimuli, (2) spectral components sharing the quality of periodicity, or (3) temporal synchrony of spectral components. Adults' outcomes, on the other hand, support the conclusion that the last of these principles can account for their CMP, to at least some extent.

An alternative to the hypotheses offered above and explicitly tested by the separate stimulus conditions is the suggestion that children's strong spectral coherence is explained by the fact that all components are heard as arising over the course of speech production, most likely because they have speech-like resonances. To evaluate the veracity of that account, it was necessary to look at responses across stimulus conditions that used cosignals. Three of these four conditions involved cosignals that had speech-like resonances; the fourth did not.

Table VI shows between-conditions Pearson product-moment correlation coefficients for the CMP effect, for each listener group separately. The prediction was that if listeners base their spectral coherence on signals having speech-like resonances, CMP effects would be correlated among the three stimulus conditions that preserved that structure, and uncorrelated for the fourth condition, which did not. Looking at the correlation coefficients in Table VI makes it clear that outcomes for the 5-yr-olds match these predictions: significant correlation coefficients are seen for the three conditions with speech-shaped spectral structure, shown in the first two columns, but not between any of those conditions and the flat-noise condition, shown in the last column. To a great extent, the same pattern is seen for correlation coefficients from 8-yr-olds. However, there is some relationship found between CMP for the shaped-noise and flat-noise conditions, reflecting their diminished CMP for the shaped-noise condition.

TABLE VI. Correlation coefficients between CMP effects in each condition with a cosignal.

		Shifted Formants	Shaped Noise	Flat Noise
Adults	Constant formants	0.58[a]	0.26	0.44[b]
	Shifted formants		0.50[b]	0.57[a]
	Shaped noise			0.77[a]
8-year-olds	Constant formants	0.87[a]	0.74[a]	0.39
	Shifted formants		0.68[a]	0.22
	Shaped noise			0.54[b]
5-year-olds	Constant formants	0.61[b]	0.58[b]	0.33
	Shifted formants		0.63[b]	0.37
	Shaped noise			0.24

[a] Significant at α of 0.01.
[b] Significant at α of 0.05.

For adults, correlation coefficients across speech-like stimulus conditions are slightly weaker than what is seen for children. However, CMP effects in all conditions incorporating speech-shaped spectra are related to CMP effects in the flat-noise condition. These results suggest that perhaps adults' CMP can be explained by both kinds of signal structure: presence of speech-like spectral prominences and temporal synchrony between target and cosignal. The slightly weaker relationships, compared to what is observed for children, may indicate that different adults depend on each principle to a slightly different extent.

E. Is it really coherence?

Finally, the question of whether CMP necessarily involves integration of spectral components was addressed. An overarching idea that motivated the study of CMP in children was that children seem obliged to integrate spectral components in speech perception, meaning that they appear unable to attend to selected frequency regions, even when it would be advantageous to do so (Nittrouer and Crowther, 2001). That suggestion for children's perception is complementary to Gordon's (1997b) suggestion that CMP demonstrates that a single spectral component has its strongest effect on vowel labeling not by itself, but through its contribution to a broader spectral pattern. But even though the CMP effect is presumed to reflect coherence of target and cosignal, it can be hard to evaluate whether or not it actually entails that coherence.

In Nittrouer and Tarr (2011), one kind of evidence used to support the claim that children's strong CMP illustrates that they are obliged to integrate across broad spectral sections in speech perception was the finding that their labeling thresholds were more similar to those of adults for the constant-formants stimuli than for the F1-only stimuli. In the current experiment, that trend was replicated: Comparing outcomes for adults and 5-yr-olds, it is found that Cohen's *d* was 1.39 for the constant-formants condition and 2.09 for the F1-only condition. These values mean that 5-yr-olds' thresholds were higher, relative to those of adults, for the F1-only condition than for the constant-formants condition. Consequently, one interpretation of the current results could be that children are more hindered than adults in their vowel labeling when broad spectral structure is unavailable.

That last interpretation would be strengthened if there were evidence that the individual listeners who showed the highest thresholds for the F1-only condition were the same listeners who had the strongest CMP in conditions involving cosignals with speech-shaped spectra. To evaluate that suggestion, Fig. 2 was created. This four-panel figure shows scatter plots of thresholds in the F1-only condition on each of the x axes, and the CMP effect for one of the cosignal conditions on the y axis. Going clockwise from the top left, conditions shown are constant formants, shifted formants, flat noise, and shaped noise. Thus, results for the least speech-like condition (flat noise) are shown in the lower right panel. For all three of the speech-like cosignals, a strong relationship between thresholds in the F1-only condition and the magnitude of CMP can be seen. In fact, the Pearson product-moment correlation coefficients between the two variables for these three conditions range from 0.91 for the constant formants to 0.57 for the shaped noise. All three of those correlation coefficients are statistically significant ($p < 0.01$). However, the Pearson product-moment correlation coefficient between thresholds for the F1-only condition and CMP effect in the flat-noise condition was only 0.12, which was not significant. Thus, the magnitude of the CMP effect was related to thresholds in the F1-only condition only when the cosignal used in the comparison condition was shaped as it would be in a speech signal. This outcome emphasizes that the listeners with the highest thresholds for the F1-only condition showed the greatest benefit of having a speech-like cosignal, but were not differentially influenced by having a cosignal that was not speech-like.

IV. DISCUSSION

This study was a follow up to an earlier one (Nittrouer and Tarr, 2011), which had shown that CMP is greater in magnitude for children (ages 5 and 8 yr) than for adults, and that the effect is not explained by the low and high formants sharing the same harmonic structure. Rather, it was suggested, children integrate spectral components that all seem to be part of a speech percept. If true, that account would mean that a schema-based principle might be responsible for the strong spectral integration observed in children's speech perception. For adults in that earlier experiment, CMP disappeared when harmonicity across signal components (target and cosignal) was dissimilar. This finding suggests that the primitive principle of harmonicity can account for spectral integration across formants for adults. Although likely not the only principle accounting for the effect, it is at least one that can be demonstrated to do so.

Of course, one problem in attributing children's strong integration of separate components to a schema-based principle, rather than a primitive principle, based on the findings of Nittrouer and Tarr (2011) is that only one primitive principle was tested in that earlier study, namely, harmonicity. The current experiment extended that work by examining three other potential sources of CMP for children and adults. All three possible accounts were related to properties of the signal: Two of those accounts clearly meet the definition of primitive principles of perceptual organization, according to auditory scene analysis: periodicity and temporal synchrony. Although not discussed as a primitive principle for spectral integration by Bregman (1990), the third hypothesis tested in this study nonetheless was based on signal structure: unique spectral shape. Primitive principles of auditory scene analysis are generally viewed as being present from birth and related to properties of the signal. Accordingly, these principles should have been sufficient to evoke CMP in the children serving as listeners.

Another challenge to using the data from the earlier study (Nittrouer and Tarr, 2011) to bolster the claim that a schema-based principle explains children's strong integration across the spectrum for speech was that the only support that could be offered came in the form of eliminating a primitive principle: Because CMP continued to be observed for children when the harmonicity principle was violated, the argument was made that a schema-based principle likely underlies their CMP. Looking across stimulus conditions in the current experiment provided opportunity to test the claim more rigorously by seeing if CMP was observed, both when

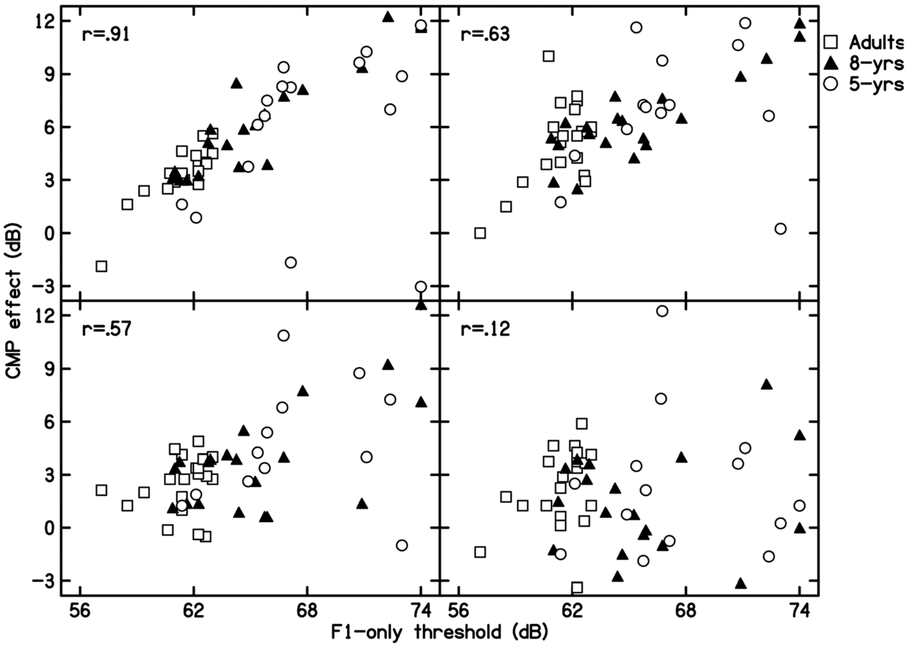

FIG. 2. Scatter plots of thresholds for the F1-only condition and CMP effects for each of the four cosignal conditions. Going clockwise from the top left, conditions shown are constant formants, shifted formants, flat noise, and shaped noise.

there was and when there was not opportunity for the proposed schema-based principle to be applied.

In sum, three hypotheses were constructed and tested regarding the possible mechanisms underlying CMP for adults and children. (1) The *unique spectral shape* hypothesis suggested that differences in overall shape of the spectrum for the fused signals might make the two stimuli more distinct, and so facilitate CMP because it derives from a labeling response. (2) The *periodicity* hypothesis suggested that those components of the sensory input that are periodic are segregated from the aperiodic components, and fused based on that periodicity, regardless of whether they share a common harmonic structure or not. (3) The *temporal synchrony* hypothesis suggested that because the cosignal has a synchronous onset and offset with the target, it marks where in the noise the listener should attend, thus facilitating CMP. To test these hypotheses, five sets of stimuli were constructed, four of which included cosignals. Three of those conditions with cosignals, but not the fourth, provided opportunity for listeners to apply a schema-based principle of integrating across the spectrum if the shape of that spectrum matches what might be generated over the course of speech production.

A. CMP in children

When it comes to children's responding, evidence was found in the current study to allow rejection of all three hypotheses explicitly tested with the novel cosignals. First, the magnitude of CMP was as great for children when the envelope of the cosignal was shifted to create stimuli with identical spectral shapes as when the gross spectral shapes of the two stimuli were different. Therefore, unique spectral shapes were not necessary for children to demonstrate CMP; the spectral envelopes created by target and cosignal just needed to elicit different vowel percepts. That finding suggests that children may not be as sensitive as adults to details of spectral shape. As long as the spectrum could reasonably have been created by a vocal tract, it is apparently sufficient to evoke CMP. Results from earlier studies on other speech phenomena support that suggestion. As mentioned in the Introduction, studies investigating fricative labeling have shown that children are sensitive neither to the precise frequencies of separate fricative poles (e.g., Nittrouer, 1992; Nittrouer and Studdert-Kennedy, 1987) nor to the overall shapes of fricative noises (Nittrouer and Miller, 1997).

Another major finding of the current study was that children continued to show substantial CMP when the cosignal consisted of noise instead of periodic structure, as long as that noise was shaped as it would be by a vocal-tract filter. For 5-yr-olds, thresholds in that condition were not statistically different from those found when both target and cosignal were periodic. For 8-yr-olds, there was a slight increase in thresholds for the aperiodic cosignal, compared to the periodic cosignal. Nonetheless, it was not a sufficient increase to alter the conclusion that periodicity fails to account for CMP in 8-yr-olds' responses.

Finally, CMP was not observed for children when flat noise served as the cosignal, so we could do no more than mark the temporal location of the target. Thus, the suggestion is offered that common fate between target and cosignal is not sufficient to elicit CMP in children's perception. From this study alone, however, it cannot be determined whether or not it is necessary. But in general, principles that could reasonably be assigned the designation of primitive, according to auditory scene analysis, were not found to explain any part of CMP for children.

In general, these outcomes for children support earlier suggestions (Nittrouer and Crowther, 2001; Nittrouer and Tarr, 2011) that children are what might be termed obligate integrators when it comes to speech signals, meaning that they so strongly integrate across the broad spectrum of speech that it is more difficult for them than for adults to judge the auditory qualities of specific properties of the signal. A challenge that can be offered to that suggestion, however, comes from findings that children are simply less sensitive to acoustic structure than are adults, as demonstrated by psychophysical measures obtained with nonspeech signals (e.g., Jensen and Neff, 1993). Of course, this situation reveals the classic chicken-and-egg dilemma: Are children less attentive to specific attributes of acoustic signals because they so strongly integrate across the spectrum, or do outcomes showing strong cross-frequency integration actually reveal poor auditory sensitivity? Reconciling those two possibilities is further complicated by the fact that studies using speech signals have an ecological advantage over studies with nonspeech signals, such as tones. In any event, this dilemma has never been thoroughly addressed, and in the end may not be critical to understanding children's perceptual organization with speech signals. Regardless of the reason, children appear to integrate strongly across the entire spectrum of speech, if that spectrum bears a resemblance to what would be created by a human vocal tract.

B. CMP in adults

For adults, slightly different patterns in outcomes were observed from those found for children. First, no support was obtained for the idea that unique spectral shapes of the combined target and cosignal facilitated labeling, and so accounted for CMP in earlier experiments. In fact, CMP was greater in magnitude for adults in the current experiment when the gross spectral shapes of the two stimuli were the same, only shifted in frequency. This finding suggests that adults are likely able to attend selectively to specific frequency regions within stimuli, because even though the gross spectral shape was the same for the two stimuli, all formants were different. A listening strategy in which the energy in discrete frequency regions is monitored could account for this outcome.

A second finding concerning adults in the current experiment was that they were not found to group signal components based on periodicity alone: CMP was similar in magnitude regardless of whether the cosignal was periodic or not. At first that finding might seem to conflict with the outcome of Nittrouer and Tarr (2011) showing that CMP was eliminated for adults when the target and cosignal had different harmonic structures. However, in that case, the two components might be perceived by adults as emanating from different speakers, based on those different harmonic structures. When the cosignal was speech-shaped noise instead, the target and cosignal combined might reasonably be expected to be perceived as emanating from a single speaker. In fact, speakers with breathy voices have vowel spectra that are noisy in the higher frequencies.

Still another finding was that CMP could be evoked from adults simply based on temporal synchrony between target and cosignal, at least if they did not consist of different harmonic structures, as was the case in Nittrouer and Tarr (2011). In that earlier experiment, disrupting harmonicity across target and cosignal was enough to eliminate CMP for adults. In this current experiment, however, the cosignal was noise, so the target and cosignal did not have different harmonic structures, strictly speaking, because only one of them had a harmonic structure at all. Adults have been shown to use temporal expectancy readily (Werner *et al.*, 2009), but so have infants: Listeners in both age groups in the Werner *et al.* study were sensitive to consistent timing of stimulus presentation after the onset of noise. Consequently, it might be suggested that adults were only making use of that principle of expectancy in this experiment, even without a cosignal, because the target appeared at a constant interval after noise onset in all trials. There is no way of knowing whether or not adults did apply that kind of expectancy, but presumably adults and children could have done so to a similar extent. However, the flat-noise cosignal in the current experiment served as an explicit cue to the presence of the target, and adults seem to have made use of that additional cue to a greater extent than children.

Finally, one interesting difference in results for adults and children seemed to be their flexibility in applying various mechanisms that might promote CMP. Based on the correlation coefficients shown in Table VI, it seems that the only mechanism accounting for CMP in children was that broad spectral shape of the signal components: that shape had to be speech-like. For adults, that was one factor, but here it was found that temporal synchrony also explained the effect to some extent. In Nittrouer and Tarr (2011), it was found that CMP is promoted by target and cosignal sharing a common harmonic structure. This age-related difference in the ability to use multiple mechanisms matches outcomes of earlier studies where adults were found to have more flexible listening strategies than children (e.g., Nittrouer *et al.*, 2000).

C. Conclusions

Outcomes of the current study for children and adults provide general support for the conclusions reached by Nittrouer and Tarr (2011): Children exhibit stronger fusion of spectral components in speech perception than do adults, and that effect cannot be explained by primitive principles of auditory scene analysis. Rather, it seems that children integrate these signal components when they are speech-like. That conclusion follows from the finding that the only condition across these two studies in which children failed to show CMP was the condition in which the cosignal lacked the shaping imposed by a vocal tract filter, namely, the flat-noise condition of the current experiment. The failure to find CMP in that condition for children is striking in light of the fact that the phenomenon had appeared fairly unshakeable in children's responses in all other conditions tested. The overall developmental course suggested by the findings of this study and Nittrouer and Tarr is that children integrate spectral components based on them all having speech-like spectral envelopes. As experience is gained with speech signals, children discover the details of those signals and learn to attend to those details. These details include properties such as consistent harmonic structure across the spectrum generated by a single speaker, but different harmonic structures for different speakers; paying attention to the details of the filter functions, which means listening selectively to narrow frequency regions; and the fact that all structure associated with a single utterance typically start and stop at the same time. Thus, children exhibit what appears to be a speech-specific schema, or strategy, failing to attend to conflicting information that fits with primitive principles of perceptual organization. It is only after children acquire additional listening experience that they learn to attend to acoustic details in the signal.

ACKNOWLEDGMENTS

This work was supported by Grant No. R01 DC000633 from the National Institutes of Health, National Institute on Deafness and Other Communication Disorders to S.N. The contribution of Katherine Kerns to data collection is gratefully acknowledged. We also thank Amanda Caldwell Tarr and Joanna Lowenstein for reading earlier drafts of this manuscript.

Aslin, R. N., and Pisoni, D. B. (**1980**). "Some developmental processes in speech perception," Child Phonology **2**, 67–96.

Bargones, J. Y., and Werner, L. A. (**1994**). "Adults listen selectively; Infants do not," Psychol. Sci. **5**, 170–174.

Best, C. T., Studdert-Kennedy, M., Manuel, S., and Rubin-Spitz, J. (**1989**). "Discovering phonetic coherence in acoustic patterns," Percept. Psychophys. **45**, 237–250.

Boysson-Bardies, B. de, Sagart, L., Halle, P., and Durand, C. (**1986**). "Acoustic investigations of cross-linguistic variability in babbling," in *Precursors of Early Speech*, edited by B. Lindblom and R. Zetterstrom (Stockton Press, New York), pp. 113–126.

Bregman, A. S. (**1990**). *Auditory Scene Analysis* (MIT Press, Cambridge, MA), pp. 1–790.

Carney, A. E., Widin, G. P., and Viemeister, N. F. (**1977**). "Noncategorical perception of stop consonants differing in VOT," J. Acoust. Soc. Am. **62**, 961–970.

Darwin, C. J. (**1981**). "Perceptual grouping of speech components differing in fundamental frequency and onset-time," Quart. J. Exp. Psychol. **33**, 185–207.

Gordon, P. C. (**1997a**). "Coherence masking protection in brief noise complexes: Effects of temporal alignment," J. Acoust. Soc. Am. **102**, 2276–2283.

Gordon, P. C. (**1997b**). "Coherence masking protection in speech sounds: The role of formant synchrony," Percept. Psychophys. **59**, 232–242.

Gordon, P. C. (**2000**). "Masking protection in the perception of auditory objects," Speech Commun. **30**, 197–206.

Hall, J. W., Buss, E., and Grose, J. H. (**2008**). "Spectral integration of speech bands in normal-hearing and hearing-impaired listeners," J. Acoust. Soc. Am. **124**, 1105–1115.

Jensen, J. K., and Neff, D. L. (**1993**). "Development of basic auditory discrimination in preschool children," Psychol. Sci. **4**, 104–107.

Levitt, H. (**1971**). "Transformed up-down methods in psychoacoustics," J. Acoust. Soc. Am. **49**, 467–477.

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

McMurray, B., Tanenhaus, M. K., and Aslin, R. N. (**2002**). "Gradient effects of within-category phonetic variation on lexical access," Cognition **86**, B33–B42.

Moore, D. R., Ferguson, M. A., Halliday, L. F., and Riley, A. (**2008**). "Frequency discrimination in children: Perception, learning and attention," Hear. Res. **238**, 147–154.

Nittrouer, S. (**1992**). "Age-related differences in perceptual effects of formant transitions within syllables and across syllable boundaries," J. Phonetics **20**, 351–382.

Nittrouer, S. (**2002**). "Learning to perceive speech: How fricative perception changes, and how it stays the same," J. Acoust. Soc. Am. **112**, 711–719.

Nittrouer, S., and Crowther, C. S. (**2001**). "Coherence in children's speech perception," J. Acoust. Soc. Am. **110**, 2129–2140.

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

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.

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., and Tarr, E. (**2011**). "Coherence masking protection for speech signals in children and adults," Atten. Percept. Psychophys. **73**, 2606–2623.

Nygaard, L. C. (**1993**). "Phonetic coherence in duplex perception: Effects of acoustic differences and lexical status," J. Exp. Psychol. Hum. Percept. Perform. **19**, 268–286.

Remez, R. E., Pardo, J. S., Piorkowski, R. L., and Rubin, P. E. (**2001**). "On the bistability of sine wave analogues of speech," Psychol. Sci. **12**, 24–29.

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.

Studdert-Kennedy, M. (**2000**). "Imitation and the emergence of segments," Phonetica **57**, 275–283.

Werner, L. A., Parrish, H. K., and Holmer, N. M. (**2009**). "Effects of temporal uncertainty and temporal expectancy on infants' auditory sensitivity," J. Acoust. Soc. Am. **125**, 1040–1049.